Digital Image Users and Reuse:

Enhancing practitioner discoverability of digital library reuse based on user file naming behavior.

Dissertation

Zur Erlangung des akademischen Grades

Doctor philosophiae (Dr. phil)

Im Fach Bibliotheks- und Informationswissenschaft

eingereicht an der

Philosophische Fakultät

Institut für Bibliotheks und Informationswissenschaft

Humboldt-Universität zu Berlin

von

Michele Reilly

Die Präsidentin der Humboldt-Universität zu Berlin

Prof. Dr.-Ing. habil. Dr. Sabine Kunst

Die Dekanin der Philosophischen Fakultät
Prof. Dr. Gabriele Metzler

Erstgutachter: Prof. Dr. Elke Greifeneder

Zweitgutachterin: Prof. Dr. Claudia Lux

Tag der Verteidigung: 06.10.2021

Selbstständigkeitserklärung

Hiermit erkläre ich, Michele Reilly, Matrikel-Nr: 587125, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Dissertation wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt oder veröffentlicht.

Fayetteville, Arkansas, USA, den 2021-10-11

Abstract

This dissertation explores devices practitioners utilize to discover the reuse of digital library materials. The author performs two verification studies investigating two previously employed strategies that many practitioners use to identify digital object reuse, specifically Google Images reverse image lookup (RIL) and embedded metadata. The dissertation describes these strategy limitations and offers a new, unique approach for tracking reuse by employing the author's search approach based on user file naming behavior. The author asks the following questions:

1. Can practitioners use Google Images and embedded metadata as practical tools to track digital object reuse?
2. Are there patterns in descriptive and administrative embedded metadata from reused digital images that practitioners can use to inform reuse discovery?
3. If yes to question two, what is a strategy that practitioners can implement to discover and trace the reuse of digital objects more effectively?

While exploring the utility and limitations of Google Images and embedded metadata, the author observes and documents a pattern of user file naming behavior that exhibits promise for improving practitioner's discoverability of reuse. Finally, the author conducts a file naming assessment investigation, an original study, to examine this pattern of user file naming behavior and the impact of file naming on search engine optimization.

The author derives several significant findings while completing this study. First, the author establishes that Google Images is no longer a viable tool to discover reuse by the general public or other users except for industry users because of its algorithm change. Second, embedded metadata is not a reliable assessment tool because of the non-persistent nature of embedded metadata. Third, the author finds that many users generate their own file names, almost exclusively human-readable when saving and sharing digital images. Fourth, the author argues that when practitioners model search

terms after the "aggregated file names" (based on the user file naming behavior), they increase their discovery of reused digital objects.

Zusammenfassung

In dieser Dissertation werden Tools untersucht, mit denen Berufsprraktiker die Wiederverwendung digitaler Bibliotheksmaterialien entdecken können. Der Forscher führt zwei Verifizierungsstudien durch, in denen zwei beliebte Tools untersucht werden, mit denen viele Praktiker die Wiederverwendung digitaler Objekte identifizieren, insbesondere die RIL (Reverse Image Lookup) von Google Images und eingebettete Metadaten. Die Dissertation beschreibt Tool-Einschränkungen und bietet einen neuen, einzigartigen Ansatz zur Verfolgung der Wiederverwendung, indem die Suchstrategie verwendet wird, die auf dem Benennungsverhalten von Benutzerdateien basiert. Die Autorin oder die Forscherin Forscher stellt folgende Fragen:

1. Können Praktiker die Google Bildsuche und eingebettete Metadaten als praktische Tools verwenden, um die Wiederverwendung digitaler Objekte zu verfolgen?
2. Gibt es Muster in beschreibenden und administrativ eingebetteten Metadaten aus wiederverwendeten digitalen Bildern, anhand derer Praktiker die Entdeckung der Wiederverwendung informieren können?
3. Wenn ja, um Frage zwei zu beantworten, welche Strategie können Praktiker implementieren, um die Wiederverwendung digitaler Objekte effektiver zu entdecken und zu verfolgen?

Bei der Untersuchung des Nutzens und der Einschränkungen der Google Bildsuche und der eingebetteten Metadaten beobachtete und dokumentierte der Forscher ein Muster des Benennungsverhaltens von Benutzerdateien, das vielversprechend war, um die Auffindbarkeit der Wiederverwendung durch den Praktiker zu verbessern. Der Forscher führte eine Untersuchung zur Bewertung der Dateinamen durch, eine Originalstudie, um dieses Muster des Benennungsverhaltens von Benutzerdateien und die Auswirkungen der Benennung von Dateien auf die Suchmaschinenoptimierung zu untersuchen.

Der Forscher entwickelte während des Abschlusses dieser Studie mehrere signifikante Ergebnisse. Zunächst stellte der Forscher fest, dass die Google Bildsuche kein praktikables Werkzeug

ist, um die Wiederverwendung zu entdecken, da sich der Algorithmus geändert hat. In ähnlicher Weise sind eingebettete Metadaten kein zuverlässiges Bewertungsinstrument, da eingebettete Metadaten nicht persistent sind. Drittens stellte der Forscher fest, dass viele Benutzer ihre eigenen Dateinamen generieren, die beim Speichern und Teilen digitaler Bilder fast ausschließlich für Menschen lesbar sind. Viertens argumentiert der Forscher, dass Praktiker, wenn sie Suchbegriffe nach den „aggregierten Dateinamen“ modellieren (basierend auf dem Benennungsverhalten der Benutzerdateien), die Entdeckung wiederverwendeter digitaler Objekte erhöhen.

Acknowledgments

Many people have helped me on this journey. I want to take a moment to thank a few of them.

I wish to thank my first reader Professor Elke Greifeneder. From the very beginning, she been generously there to advise and aid me in this quest. From meeting her in Croatia to her helping me register, and to the very end, I could count on her help to attain my goal.

I want to thank my second reader, Professor Dr. Claudia Lux. I very much her willingness to take on the task of the second reader. Her sage comments, thoughtful suggestions, and recommendations were greatly appreciated.

To my past two supervisors, Carolyn Allen, former Dean of Libraries for the University of Arkansas Libraries, and Patricia Cutright, Dean of Libraries Emeritus, Retired, Central Washington University Libraries, I could not have done it without your unflagging support and consistent guidance.

To my colleague Santi Thompson, I'd like to offer the sincerest thanks. You put up with me stalling other research projects, throwing out ridiculous ideas, and reading countless pages of my meandering writing. You've been the tugboat that led my ship going into safe harbor.

To my friends, Jeff Banks, Jerome Crowder, Nancy Beck-Young, Becky Severin, and Ronda Boyd, for your steadfast support, previewing chapters, letting me float ideas and concepts, and all-around propping me up during all this time. Thank you!

To my sons Benjamin and Noah Reilly, your encouragement has meant the world to me. Making you proud is all I really could ask for as a mom. Though a grandchild or two wouldn't be scoffed at.

To my wonderful husband, Jon Reilly, your understanding, support, and love have gotten me through this whole process. I couldn't have done it if you hadn't been there for me. So now, maybe we can have some quiet get-a-ways without my laptop or thick folder full of articles to read.

Dedication

For Jon, who has inspired, loved, and supported me throughout this process. Who has been my rock and who I am sure will be glad to have his wife back.

For my sons, Benjamin and Noah, who have been my biggest cheerleaders.

For my mother, Lois, whose love has supported me all my life.

For my friend and colleague, Santi Thompson, I can't wait to get back to just doing research with you.

Table of Contents

List of Figures

List of Tables

Glossary of Abbreviations and Terms

BnF	Bibliotheque Nationale de France
Browsers	A software application used to access the World Wide Web.
BMP	Bitmap image file
CBIR	Content-Based Image Retrieval
CERN	European Organization of Nuclear Research
CHI	Cultural heritage institutions
CKRN	Canadian Research Knowledge Network
COBOL	Common Business Operating Language
CRL	Council of Research Libraries
DDB	Deutsche Digitale Bibliothek
Digital Library	The collection of digitized and/or born-digital items stored in a content management system to display, preserve, store, and manage.
DL	Digital Library
DLF AIG	Digital Library Federation Assessment Interest Group
DPLA	Digital Public Library of America
Exif	Exchangeable image file format
Floppy disk	Thin, flexible computer disk storage
FORTRAN	Formula Translation - computer programming language
GIF	Graphics Interchange Format
GLAM	Galleries, Libraries, Archives, and Museums
GUI	Graphical user interface
IBM	International Business Machines Corporation
IPTC	International Press Telecommunications Council
ISO	International Organization for Standardization
JFIF	File Interchange Format
Jpeg/Jpg	Joint Photographic Expert Group
LAM	Libraries, archives, and museums
LOC or LC	Library of Congress. United States

LOCDC	Library of Congress's Digital Collections
MARC	Machine-Readable Cataloging
MARC21	Machine-Readable Cataloging for the 21st century
MDC	Metadata for Digital Content
MESL	Museum Educational Site Licensing
MfDG	German Historical Museum
MODS	Metadata Object Description Schema
NCSA	National Center for Supercomputing Applications
NDLP	National Digital Library Project. United States Library of Congress
NEH	United States National Endowment for the Humanities
NSF	United States Government National Science Foundation
NYPL	New York Public Library
NYPLDC	New York Public Library Digital Collections
PDF	Portable Document Format
PNG	Portable Network Graphics
Practitioner	Digital image information professionals who provide and manage digital libraries.
Reuse	The use of a digital image in a setting other than its original intent.
RIL	Reverse Image Lookup
ROI	Return on investment
SMK	Statens Museum for Kunst
TCP/IP	Transmission Control Protocol (TCP). Internet Protocol (IP)
TIFF	Tagged Image File Format
UNIVAC	Universal Automatic Computer
URL	Uniform Resource Locator
Users	Those who search, discover, and/or download images from digital libraries.
Web-blogging	Websites authored by either a single or multiple authors. Usually contains discussion or informational content.
WebP	Image format using lossy and lossless compression
Wifi	Wireless network protocols

Wikis	A wiki is a website that allows collaborative user editing.
XML	Extensible Markup Language
XMP	Extensible Metadata Platform

Chapter 1: Introduction

The theory, investment, and implementation of cultural heritage digital libraries and mass digitization began over 25 years ago. Galleries, libraries, archives, museums (GLAM), and other cultural heritage institutions have envisioned digital libraries as a way to highlight and share their unique and rare analog collections. In addition, they have embraced their role as digital civic and cultural innovators, participators, and creators. However, this work has not been undertaken systematically or consistently, “The legacy of 30 years of investment in cultural heritage digitisation is a patchwork of small to large scale content, held in different locations, formats and under different reuse licenses, with different institutional approaches to risk, public engagement and entrepreneurship.” (Terras, Coleman, Drost, Elsden, Helgason, Lechelt, Osborne, Panneels, Pegado, Schafer, Smyth, Thornton, Speed, 2021, p. 11).

As this patchwork of digital libraries has evolved, digital library practitioners from the GLAM industries have attempted to assess this growing medium. Kelly (2014) states, "digital library assessment allows libraries to create effective and sustainable evaluation models based on the successes and shortcomings of previously completed projects." (p. 384). The multiple types and number of assessment projects demonstrate the difficulty that GLAM institutions experience when evaluating and assigning value to digital libraries, collections, and images. Some of these projects include testing user interfaces, platform functionality and navigation, practitioner interface functionality, and digital image user behavior. In addition, the physical distance of users makes creating evaluation criteria and instruments challenging.

This dissertation focuses on the search strategies practitioners should employ to overcome some of the assessment challenges and offers a novel approach to enhance the search results of Google Images to discover digital image reuse. One type of evaluation and the starting point for the author's research, is who uses digital images and for what purposes do they use them, sometimes called digital image reuse. (Kelly, 2015; Chung & Yoon, 2011; McCay-Peet & Toms, 2009; Reilly & Thompson, 2016). Professional literature documents the difficulties practitioners face when

attempting to do this type of reuse assessment (O'Gara, Woolcott, Kelly, Muglia, Stein, & Thompson, 2018; Thompson & Reilly, 2017).

Section 1.1 Research questions

This dissertation contributes to the critical discourse on digital image reuse assessment. The author challenges the existing methods that practitioners employ to increase digital image user reuse and engagement. The author uncovers a unique approach for tracking reuse by employing a search strategy based on a newly identified user file naming behavior. In the investigation, the author verifies previous digital image reuse assessment methods, specifically Google Images reverse image lookup (RIL) and embedded metadata. The author asks the following questions:

1. Can practitioners use Google Images and embedded metadata as practical tools to track digital object reuse (as detailed in chapters four and five of this dissertation)?
2. Are there patterns in descriptive and administrative embedded metadata from reused digital images that practitioners can use to inform reuse discovery strategy (as detailed in chapters four and five)?
3. If yes to question two, what is a strategy that practitioners can implement to discover and trace the reuse of digital objects more effectively? (as detailed in chapter six)

This dissertation emphasizes the importance of implementing descriptive, user-friendly file naming practices. In addition, it stresses the connection between user file naming behavior and discoverability of reuse; a relationship largely overlooked in the existing literature.

Section 1.2 Definitions

Throughout this dissertation, the author employs the terms "users," "practitioner," "reuse," "verification study," "digital library," and others. To aid in interpreting these terms within the dissertation, the author defines them as,

"Aggregated file name" The author defines an aggregated file name as a text string based on the most frequent descriptive terms found in user-generated file names. When there are no frequently repetitive descriptive terms, the author uses the original title, creator, and/or date to derive the aggregated file name. An example of an aggregated file name is: Self-portrait-Vincent-van-Gogh-1887

"Digital Library" The author defines a digital library as a suite of online services that collects, displays, manages, and preserves collections of digitized or born-digital materials in any format from a cultural institution, library, archive, or gallery. Not included within the term is digital content purchased from publishers (Matusiak, 2012; Xie, Babu, Joo & Fuller, 2015; Burns, Sundt, Pumphrey, and Thoms, 2019; and Hebron & Mowry, 2021).

"Digital Image" The author defines a digital image as a digital surrogate of an analog work such as a photograph of a painting or sculpture.

"GLAM industries" The author includes galleries, libraries, archives, and museums as types of GLAM industries.

"Original file name" The author defines the original file name as the name generated during the download of an object from its original digital library source.

"Practitioner" The author defines practitioner as the group of information professionals who make digital images accessible to users and manage digital libraries.

"Private industries" The author groups any for-profit organization or company, except for galleries, libraries, archives, and museums, as types of private industries.

"Reuse" The author adopts the definition of the reuse of a digital image as (a) using, transmitting or sharing a digital image in a new setting that is not its original platform, purpose, or context, or (b) repurposing, or transforming the image into a new digital file (Thompson & Reilly, 2107; Shiri, Kelly, Kenfield, Masood, Muglia, Thompson & Woolcott, 2020; Thompson, Woolcott, Muglia, O'Gara, Kenfield & Kelly, 2019). In this definition, "Reuse" cannot occur by the originating digital library.

"Users" The author defines users as those who search, discover, and/or download images from digital libraries to fulfill various objectives, such as personal research, entertainment pursuits, and work-related activities.

"Verification study" The author adopts the United States Code of Federal Regulations Title 21's definition of a verification study as: "confirmation by examination and provision of objective evidence that specified requirements have been fulfilled."

Section 1.3 Chapters overview

Chapter two discusses the development of the technology needed to create digital libraries and the advancement of digital libraries. It examines scholarship chronologically, starting from 1945 to the present time. This chapter explores the terminology "digital library" from the GLAM scholarship, the GLAM institution websites, and user perceptions. Finally, it provides an overview of selected global digital library aggregators.

The author provides a digital library assessment literature review in chapter three. The assessment review is structured chronologically on general assessments and breaks out the digital library reuse assessment into more focused sections.

Chapter four explores Content-Based Image Retrieval (CBIR) and Reverse Image Lookup (RIL). It defines and explains the technology, identifies private industries using CBIR and RIL, and documents a Google Images verification study.

Chapter five focuses on definitions, specifications, and uses of embedded metadata, discusses an application tool used to edit, analyze, or extract metadata, relates industrial, GLAM, and website use of embedded metadata, and includes results from an embedded metadata verification study.

The author employs the same methodologies used in previous RIL and embedded metadata studies for chapters four and five. In addition, the author gives a detailed explanation of these methods in each chapter. Chapters four and five likewise articulate the challenges of using these platforms as reuse assessment techniques. While exploring the utility and limitations of Google Images and embedded metadata as a means to discover reuse, the author observes and documents a pattern of user file naming behavior that exhibits promise for improving practitioner's discoverability of reuse.

Chapter six examines the pattern of user file naming behavior and the impact of file naming on search engine optimization discovered in the chapter five embedded metadata verification study. Finally, the author conducts a file naming investigation study that compares Google Images search results using two text strings: the original file name and an "aggregated file name."

Chapter seven communicates the total research conclusions. It articulates the limitations of the Google Images verification study, embedded metadata verification study, and the file naming investigation study. It concludes with the following implication for digital library practitioners: employing a search strategy based on user file naming behavior to enhance digital object discovery reuse**.**

Section 1.4 Significant findings

Chapter four CBIR/RIL verification study results establish that, contrary to past reverse image lookup reuse studies, Google Images is less viable as a means to discover reuse than it once was just a few years ago. This conclusion is due to a change in the Google Images search algorithm.

Chapter five embedded metadata verification study demonstrates that embedded metadata is not a reliable assessment tactic because of the non-persistent nature of embedded metadata. This result aligns with previous studies but yields an outcome that significantly impacted this dissertation's research findings.

Chapter six reveals that many users generate individual file names when downloading, uploading, and sharing digital images on the web. In addition, the chapter discovers that user file names are almost exclusively human-readable. This discovery is in contrast to the practices of GLAM institutions studied in this dissertation, which prioritize administrative file naming customs. It also notes that the human-readable user-generated file names are more discoverable in Google Images search results than other approaches, such as using the digital library's original file name or using the image as the RIL search query. Finally, the chapter addresses possible reasons for this pattern.

In chapter seven, the author argues that when practitioners model search terms after the "aggregated file names" (based on the user file naming behavior), they increase their discovery of reused digital objects using Google Images significantly.

Chapter 2: Development of digital technology and digital libraries

This chapter discusses the advent and advancement of technology, principally computers, and the World Wide Web, which made digital libraries, digitization, and online discovery possible. Computer technology has enabled users to discover, reuse, and engage with place-based cultural heritage materials in new and exciting ways. It has opened new areas of research for practitioners, such as user experience and reuse. During the COVID19 pandemic, computer technology has aided GLAM institutions in engaging and growing their user base in ways never done previously.

Computer technologies evolved from simple pieces of punched wood to circuits, from something found in factories, laboratories, and major businesses to ubiquitous devices found in most modern classrooms and homes. The increasing momentum of computers, imagining, and web technologies have dramatically extended the traditional GLAM institutions into the digital age by providing more and better service, fluid access to resources, and the ability to target those resources to their community (Lagoze, Krafft, Payette, & Jesuroga, 2005; Borgman, 2000; Borgman, 2003b; Borgman, 1999; Surowiecki, 2004). The participatory nature of the web alters the ways in which users find, download, and share with others those things that are of interest to them, creating a "digital culture." This emerging digital culture is discussed by Giannini & Bowen (2019) when they wrote

> "as the space between digital and physical, real and virtual blurs, recasting art, cultural, social life and human behavior into new digital forms, places and spaces, digital life and culture are merging and redefining the way we live, while more and more we think and see digitally." (p. Preface).

Understanding the development of digital technology and its impact on art, social life, and human behavior will help practitioners recognize how users have changed digital culture through their increased adoption, reuse, and customization of digitized images.

Section 2.1 Development of digital technology

Unlike some major innovations such as the wheel, there have been documented records of computer technology's development, beginning with the "first attempts to develop mechanical calculators took place in Europe in the 17th century." (Berg, Bommes, Hardle & Petukhina, 2017, p. 5). Joseph Marie Jacquard, in 1801, created a punched wooden loom design card, which became a precursor to the punch-card drive computer systems developed nearly one hundred years later (Berg et al., 2017). The 1800s saw exponential growth in the development of calculation machines, including in 1822, when Charles Babbage, an English mathematician, promoted a machine that would calculate numbers (Berg et al., 2017). A successful punch-card calculation machine was developed and used by Herman Hollerith to calculate the 1890 United States Census. This machine calculation reduced the time from more than seven years to two and half years (Enzo, 2020). "In 1896 Hollerith founded the *Tabulating Machine Company,* today known as IBM." (Berg et al., 2017, p. 161). This concept of number calculation would continue until after the turn of the twentieth century.

Envisioning the theory of machine memory would not happen until the 1930s. Belgian Paul Otlet presented his ambitious Mundaneum project to create a giant organized card catalog of what he claimed was all the world's knowledge. Considered the precursor to the modern-day "search engine, the Mundaneum was still a paper-based system (Computer History Museum, 2020a). In the same decade, Alan Turning presents the idea of a machine that will calculate anything that can be computed, giving birth to the essential hypothesis of the modern computer (Enzo, 2020). Computers were either a theory or a mechanical punch system with gears, shafts, and cams at this stage. An exception would be in 1937 when J. V. Atanasoff, from Iowa State University in the United States, attempted to build a computer without mechanical parts (Enzo, 2020). It would take four more years before this goal would be realized by German engineer Konrad Zuse when he generated the first computer using relays (Zimmermann, 2017).

It is this relay-driven computational equipment that spurs the next wave of innovation from 1943 to 1956. Vacuum tubes, relays, switches, and transistors advanced the computer into a commercially available product and established the need for computer languages. John Mauchly and J. Presper Eckert built a calculating computer using over 15,000 vacuum tubes in 1943-44. In 1946, they constructed the ENIAC (Electronic Numerical Integrator and Computer) at the University of Pennsylvania, the first commercial computer for the United States Army (Berg et al., 2017; Ceruzzi & Paul, 2003). One year later, Bell Labs engineers William Shockley, John Bardeen, and Walter Brattain invented the transistor to replace vacuum tubes. That same year, Manchester University engineers Frederic Williams and Tom Kilburn developed a high-speed electronic memory tube. Punch cards still directed the work of computers until 1953. That year, Dr. Grace Hopper, while working for Eckert-Mauchly Computer Corporation, authored the first computer program, Common Business Operating Language (COBOL) (Howell, 2016). A team of computer scientists from the University of Michigan developed the programming language *Formula Translation* or FORTRAN in the late 1950s (Berg et al., 2017). With the invention of the keyboard in 1956, the transition from punch cards to direct programming was complete (Berg et al., 2017; Zimmermann, 2017).

Between 1957 to 1969, many developments in computer hardware, software, and peripherals (external devices connected to computers that provide input or output such as printers, scanners, and others) transpired. Some notable technology advances include:

- the work of Robert Kirsch to design a drum scanner in 1957 (Computer History Museum, 2020b).
- The first personal computer was developed though the term “personal computer” meant that it was a single-person computer, “in contrast to a mainframe” computer (Berg et al., 2017, pp. 51-52).
- Ivan Sutherland’s computer drawing system called Sketchpad in 1963 (Myers, 2020).
- the invention of the computer chips in 1958 by Jack Kilby and Robert Noyce (Enzo, 2020).

- ground-breaking concept of the hyperlink and hypertext (Myers, 2020).
- the semiconductor, the device that replaced the vacuum tube and increased the computing power by overwhelming amounts, was invented (Berg et al., 2017).

From 1970 to 1979, computers became more accessible and networked. The advent of home personal computers (PCs) appeared in the late 1970s, as did the first commercially available microprocessor (Berg et al., 2017; Ceruzzi & Paul, 2003). The initial PCs usually came as kits to be assembled by buyers. They came with a keyboard, a monitor, and the ability to be connected to a television if preferred (Berg et al., 2017). At the same time, computer companies spring to life. IBM launched the first desktop computer, and the founding of Microsoft and Apple computer companies occurred (Berg et al. 2017; Reuters, 2009). A few other significant developments include:

- the "floppy disk" invented by IBM engineers, which according to Berg et al. (2017), "was soon adapted as a general storage medium for early personal computers." (p. 165).
- The PET (Personal Electronic Transactor), the first fully assembled computer, ushered in the era of 8 bit systems with preinstalled programming languages and data that was read and written on a data cassette (Berg et al., 2017).
- Ethernet network technologies used to connect multiple computers (Enzo, 2020).
- TCP/IP protocol which makes the internet possible (Computer History Museum, 2020c).

From 1980 to 1990, personal computers became more popular, with IBM and Compaq companies leading the way (Enzo, 2020; Reuters, 2009). IBM launched its first personal computer while Compaq created and sold the first foldable laptop. The graphical user interface (GUI) and the

computer mouse were introduced in 1983, replacing the command line as the only way to interact with the machine (Berg et al., 2017).

These years would be especially noteworthy for the conception and establishment of the World Wide Web. Tim Berners-Lee, a European Organization of Nuclear Research (CERN) scientist, instigated the World Wide Web in 1989. By 1990, he had created the first Web browser, NEXUS. Additionally, he wrote the Hypertext Markup language, the language that makes the internet operate (Myers, 2020.) The first search engine would be developed that same year by McGill University student Alan Emtage. "You got mail," an email service, began in 1989 by America Online, Inc (AOL) (Pew Research Center, 2014).

From newly discovered web cameras to the jpeg image format to the World Wide Web coming into the public domain, the short period from 1991 to 1993 saw considerable personal computing changes. In 1991, researchers from the University of Cambridge, England, set up the first webcam in front of the only coffeemaker in the building. An accompanying computer program displayed an icon on the researchers' monitors, letting them know when a fresh pot of coffee was finished brewing (Stafford-Fraser, 1995), proving that desperation is the mother of invention. By 1992, the Joint Photographic Expert Group set the jpeg (.jpg) photo format standards (Computer History Museum, 2020d). A year later, CERN gave the World Wide Web technology to the public domain, and the National Center for Supercomputing Applications (NCSA) released the first general population web browser, Mosaic 1.0 (Myers, 2020; Pew Research Center, 2014).

Between 1994 to 2001, the world witnessed a tremendous spike in the number of global internet users and consequently the number of consumer websites. In 1995, the Dublin Core metadata standard, descriptive language for a digital image (Besser, 2016), made searching across multiple collections possible. During this period, scanners became inexpensive and commonplace, making it easier to reformat analog images and documents to digital. Simultaneously, consumer websites such as Amazon, Craigslist, and Match.com (Pew Research Center, 2014) sprung up, and Apple produced a Wi-Fi router and built-in Wi-Fi capability into their Mac computers (Computer History Museum, 2020e).

The time from 2001 to 2020 will stand as the period of advanced social media, web browser, and computer transformation. Marshall McLuhan in 1962 promoted the hypothesis of a global village. His theory suggested that as media, technology, and communications became commonplace and interconnected, the world's social fabric would be irrevocably changed (Bowen & Giannini, 2019). When Facebook, Pinterest, Twitter, YouTube, and Instagram, to name a few social media sites, came into existence, they became examples of the interconnectedness of the world's social fabric or McLuhan's global village. They hasten the recognition of cultural identity, greater awareness of cultural heritage, and the arts (Bowen & Giannini, 2019). These sites, along with web-blogging and wikis, give users the ability to develop, edit, upload, and share content. Bowen & Giannini, (2019) state, "Indeed, IT research has become the engine of change positioned at the heart of innovation and creativity while empowering interaction of individuals and institutions across the globe." (p. 572). Home video gaming has become a big business that captures a generation of young adults. Browsers have become more sophisticated with the addition of Mozilla's Firefox and Google's Chrome (Zimmermann, 2017). Computers transformed to smaller and lighter, as was evidenced when in 2008 Steve Jobs "unveiled the MacBook Air … by pulling it out of an envelope." (Berg et al., 2017, p. 110).

Similarly, computing speeds, storage, and computer processing advance to the molecular level, anticipating a possible future towards human augmentation. "In parallel to the advances in hard drive technology, better and faster internet connections have led to cloud storage services becoming more popular." (Berg et al., 2017, p. 173). The World Wide Web has made the world smaller as the number of global internet users has grown from 16 million in 1995 to over 4.574 billion in January of 2020 (Internet World Stats, 2020).

The exponential growth of online digitized images is fulfilling McLuhan's global village theory. "Acting as a catalyst for growing participation online, the Web turned the digital tide toward visual interface, smartphones and personal digital devices that sparked the rapid rise of digital culture." (Giannini & Bowen, 2019, p. ix). The GLAM community has increased its digital output and established policies and protocols that aid in the categorization, description, and accessibility of

digital library materials. They bring the world's diverse and unique cultural heritage to users through their interconnectedness of digital collections.

Section 2.2 Development of digital libraries

The concept of digital libraries began seventy-five years ago. Originating with Vannevar Bush (1945), research over the next fifty years focused broadly on the strategies for storing the explosion of scientific research publications produced post-World War Two in digital format. Scholars concentrated on defining what a digital library is and identifying the technological solutions to make it a reality in this era. By the mid-1990s, academic and cultural heritage organizations entered into the digital library scholarly conversation. Scholarly research reveals how these groups sought solutions for making research documents, articles, university records, images, and multimedia collections more accessible using the evolving technology and the emergence of the World Wide Web. Scholars such as Jeng (2005) discussed the need to make information "organized, well-managed, and supports the creation, use, and searching of digital objects. Digital library should be looked at as a tool that supports a user's information task. Users are looking for an information system that is easy and intuitive to use." (p. 111).

According to Abby Smith (1999), one such example of digitized special collections material use in the classroom. She writes,

> "Digital technology can also make available powerful teaching materials for students who would not otherwise have access to them. Among the most valuable types of materials to digitize from a classroom perspective are those from the special collections of research institutions, including rare books, manuscripts, musical scores and performances, photographs and graphic materials, and moving images." (para. 24).

As the scanning and imaging technology advanced, it was now possible to digitize the growing cache of rare and unique GLAM materials and institutional products. After 2005, scholarship turned its attention to the users of digital libraries. Studies addressed user engagement with existing interfaces, other user interaction types, and users' roles in creating content. Additionally, the scholarship confronted the storage space and preservation needs of digital objects.

The development of digital libraries suggests that they were designed originally for scholarly and research pursuits. Only when the GLAM industry saw the value in presenting unique content via the World Wide Web to those outside the academy and research areas that digital library use exploded. This explosion of use led practitioners to seek out who these new users are, what they are looking for, how they find content, and how they reuse the digital content. Answering these questions can guide digital collection development, understand user behavior, deliver more highly valued content, supporting preservation, supports research, and provide greater accessibility. This dissertation contributes to this ongoing conversation by examining approaches to reuse and articulating an alternative method for practitioners to discover digital library reuse.

Section 2.2.1 Early scholarship: defining digital library (1945-1995)

Perhaps the first person to write on what scholars describe as the digital library was Bush (1945), who conceptualized the memex machine in 1945. According to Bush (1945), this hypothetical technology would offer the ability to "store[s] all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory." (p. 99). In Bush's mind, the memex machine offered new ways to store and make accessible human knowledge at rates that far exceeded the human mind. He wrote, "Wholly new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified." (p. 101).

Building on Bush's theorized memex machine, other professionals addressed the storage and retrieval of scholarly content. According to Licklider and Clapp (1965), the Ford Foundation

established the Council of Research Libraries (CRL) in 1956 to address the burgeoning problem of scientific publications and output and discover ways in which technology can play a part. CRL commissioned a group of engineers and psychologists to investigate the qualities of future libraries, published in the report *Libraries of the Future*. It focused on two areas: "man's interaction with the body of recorded knowledge" (p. 2) and "explorations of the use of computers in information storage, organization, and retrieval." (p. 2).

Libraries of the Future also highlighted additional studies that addressed human interactions with the hypothetical digital library, future digital libraries' storage capabilities, and the functional requirements needed to engage users. Beyond engagement, the *Libraries of the Future* report highlighted ways the computer could store, retrieve, and understand commands using natural language syntax and "precognitive functions" (p. 131) to access content held in its memory. Finally, the report also theorized the 13 functional requirements of a digital library. The requirements were:

1. "Present for examination a document specified by any sufficiently prescriptive segment of its bibliographic citation.
2. Turn pages, forward or backward, in response to the pressing of a key
3. Permit designation of a passage (segment of text) by pointing to the beginning and then the end with a light pen.
4. Accept labels from the typewriter and associate them with passages of text.
5. Record as a note, and preserve for later inspection, any designated passage.
6. Append bibliographic citations to extracted passages.
7. Accept retrieval prescriptions from the typewriter.
8. Accept from the typewriter coded versions of specifications of such operating characteristics as, "Consider a neighborhood to be five consecutive lines of text," or "Consider a search to be satisfied when any two of the three elements of the search have been satisfied."
9. Carry out retrieval searches and display passages in which the retrieval prescriptions are satisfied.

10. Compose graphs from tabulated data and present the graphs, against labeled coordinate grids, on the oscilloscope screen.
11. Set two graphs side by side to facilitate comparison.
12. Expand or compress the scales of graphs, under control from the light pen.
13. Change the number of grid lines or the calibration numbers associated with the lines, or both together, and recalculate and redisplay the calibration numbers when grid lines are added or deleted." (Licklider and Clapp, 1965, pp. 177-178).

It was not until the mid-1990s that the technology caught up to Licklider's and Clapp's (1965) vision of a digital library.

One of the first examples of an academic unit's association with the digital library came in 1986 when the University of California Berkeley's office of Information and Systems and Technology developed the ImageQuery software. ImageQuery was a networked compliant, graphical user interface system that allowed a user to search the image database using a mouse to point and click and/or Boolean operators, sort, browse, and view images as thumbnails to interact with images by annotating and associating with geographical locations. The program was the "first deployed multi-user networked digital image database system" (Besser, 2016, p. 95) using "high-quality digital images from its Art Museum, Architecture Slide Library and Geography Department." (Besser, 2016, p. 95). As Van House, Butler, and Schiff (1996a) noted, the goal of the project was:

> "to develop a massive, distributed, electronic, work-centered library of environmental information containing text, images, maps, numeric datasets, and hypertextual multimedia composite documents to support actual environmental planning decisions by means of a coherent, content-based view of a diverse distributed collection which will scale to very large collections and large numbers of clients and servers, and improved data acquisition technology." (p 1).

The ImageQuery software was showcased at the American Association of Museums and the American Library Association conferences in June 1987, marking the introduction of the networked digital image database system to GLAM communities (Besser, 2016).

In 1988, the term "digital library" may have first appeared in GLAM scholarly publications in Kahn and Cerf's (1988) report to the Corporation for National Research Initiatives (Trivedi, 2010; Kahn & Cerf, 1988). In the report, Kahn and Cerf (1988) wrote that the digital library should contain the "conventional archive of current or historically important information and knowledge, along with ephemeral material such as drafts, notes, memoranda, and files of ongoing activity." (p. 3) As Trivedi (2010) further defined a digital library as "a library in which collections are stored in digital formats (as opposed to print, microform, or other media) and accessible by computers. The content may be stored locally, or accessed remotely.".(p. 1) For the GLAM community, the scope of content expanded the current understanding of a digital library from the theoretical view of a technological memory device focused primarily on scientific publications, as espoused by Bush (1945) and Licklider and Clapp (1965) to include the historical-cultural heritage archive. Subsequent reports from the GLAM community continued to differentiate the term digital library even further. The term "digital library" will be discussed later in the chapter.

The Digital Library Initiative report, issued jointly with three United States Government agencies; Defense Advanced Research Projects Agency, the National Science Foundation, and the National Aeronautics and Space Administration, expressed the digital library as an "electronic library." (Borgman, 1999, p. 233). Burns et al., (2019) explained the "electronic library" project as one "designed to spur innovation in large-scale information storage and retrieval" (para. 3) by offering "grants to projects in computer and information science." (para. 3) This sentiment was echoed by Pomerantz, Choemprayong, and Eakin (2008), who wrote that the focus was on building technical infrastructure. Burns et al. (2019) note that:

> "Later, after the program was extended to include the National Endowment for the Humanities, the Library of Congress, and the National Library of Medicine, more emphasis was placed on projects focused in the arts and humanities. These disparate communities—

> computer and information science, library science, arts, and humanities—each provided their own specialized definition of the term." (para. 3).

Borgman (1999) went on to write "that the users of a national electronic library would include students, teachers/professors, researchers/scholars, librarians, authors, publishers, information providers and practitioners. Contributors of information resources would include publishers, universities, professional societies, libraries, authors, editors and compilers." (p. 233). Borgman would summarize Edward Fox's *Source Book on Digital Libraries*, chapter 1, section B: workshop report (Fox, 1993), in defining a National Electronic Library as "(1) a service; (2) an architecture; (3) a set of information resources, databases of text, numbers, graphics, sound, video, etc. and (4) a set of tools and capabilities to locate, retrieve and utilize the information resources available." (Borgman, 1999, p. 233)

Section 2.2.2 Academic digital libraries and cultural heritage organizations 1995-2005)

This next era ushered in the explosion of digital libraries in GLAM institutions. In this period, the GLAM industry becomes one of many "trailblazers" forecasted by Bush (1945). He noted:

> "There is a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record. The inheritance from the master becomes, not only his additions to the world's record, but for his disciples the entire scaffolding by which they were erected." (p. 45).

This phase witnessed the expanding scope of digital library content, web interfaces, and content management systems when the world wide web, web technology, and computers became household utilities. As a result, scholarship focused on at least three themes: defining the term "digital library" in a GLAM context, emphasizing the critical role GLAM communities play in digital libraries, and exploring how GLAM institutions implemented digital libraries.

While GLAM institutions relied more on technology to provide access to cultural heritage materials and library print and electronic collections, definitions of what a digital library is and what to call it proliferated in the GLAM literature. Multiple professionals deemed a digital library as an extension of a traditional library (Cleveland, 1998; Marchionini, Plaisant, & Komlodi, 1998; Waters, 1998; Schwartz, 2000). For example, Cleveland (1998) described a digital library as having "the same purposes, functions, and goals as traditional libraries collection development and management, subject analysis, index creation, provision of access, reference work, and preservation." (p. 2) At the same, professionals were also debating the name of this construct. Cleveland (1998) recounts "phrases like 'virtual library,' 'electronic library,' 'library without walls' and, most recently, 'digital library,' all have been used interchangeably to describe this broad concept." (p. 1) While not agreed upon by the GLAM industry, the term' digital library' has become synonymous with online image collections.

Contributions from the GLAM industry also included the concept of community in the definitions and scope of a digital library. The view of a digital library community initially was narrow in its focus. It mainly encompassed researchers who were generating scholarly output. When acknowledging the benefits of digital content and the process of further digitization, Praveena (2019) notes that it will "lead to new knowledge by enabling scholarly use that was not possible with print collections." (p. 626) He continues to say that within this community, "wider dissemination of unique collections will encourage scholarly use." (p. 626) Howard Besser (2004) expands on the idea of academic use. He states,

> "digital libraries will be critical to future humanities scholarship. Not only will they provide access to a host of source materials that humanists need in order to do their work, but these libraries will also enable new forms of research that were difficult or impossible to undertake before." (p. 557).

GLAM literature broadened the idea of community in several ways. As Punzalan, Marsh, and Cools (2017) state, "digital access is seen as a force for good and a means for LAMs to creatively reach wider audiences and communities." (p. 62) Similar literature acknowledged that a digital library

has multiple, distinct user communities, expanding the conversation beyond just researchers (Shiri, 2003). The GLAM literature also recognized the role practitioners play in administering and curating digital libraries' content, giving standing to an emerging community of practice (Shiri, 2003). According to Griffin (1998), participants at the 1997 Santa Fe Planning Workshop on Distributed Knowledge Work Environments sponsored by the National Science Foundation expanded the notion of what a digital library is, noting, digital libraries cultivate a community that brings together "collections, services, and people in support of the full life cycle of creation, dissemination, use, and preservation of data, information, and knowledge." (para. 17).

During this era, GLAM institutions, particularly national libraries and large academic libraries, also developed and implemented digital library platforms to expand access to rare and unique digitized collections. One of the first was the National Digital Library Project (NDLP) out of the United States Library of Congress (LC). Started as a pilot project from 1990-1994, the NDLP project team experimented with digitization technology, platform design, and audience identification to highlight LC's world-class collections. NDLP development led to the eventual formation of LC's American Memory Project, the flagship of its historical digital library efforts, launched in October 1994 (American Memory, Library of Congress, 2020; Marchionini et al., 1998). In the following year, LC collaborated with over a dozen museum and academic institutions, including but not limited to the Fowler Museum of Cultural History at the University of California, Los Angeles, The Museum of Fine Arts, Houston, The National Gallery of Art, and the Harvard University Art Museums, Cornell University, Columbia University, and the University of Illinois Urbana-Champaign, to launch the Museum Educational Site Licensing (MESL) project. MESL developed a digital library infrastructure for digitized museum content to be delivered to university networks (Besser, 2016). The Gateway and Bridge to Europe's National Libraries (GABRIEL), founded in 1997 by the "national libraries of the United Kingdom, the Netherlands, Finland, and Germany" (Cousins, 2017, p. 262), was the "gateway to the bibliographic holdings and treasures of 41 national libraries, representing the 39 member states of the Council of Europe." (Hillson, 2002, Para. 1). Europeana credits GABRIEL as its foundation (Jefcoate, 1996). In the same year, the Bibliothèque Nationale de France launched Gallica to digitize

collections representing France's national heritage accessible online (Bibliothèque Nationale de France, 2020). Responding to the proliferation of digital library platforms throughout the world and their disparate systems and terminology, the European Commission designed and co-funded the DELOS Digital Library Reference Model. This model was established to develop digital library structures and standard practices (Candela, Castelli, Ferro, Ioannidis, Koutrika, Meghini, Pagano, Ross, Soergel, Agosti, Dobreva, Katifori & Schuldt, 2007). Section 2.5 Overview of selected global digital libraries provides additional information about several of these projects.

Section 2.2.3 The user experience in digital libraries (2005 - Present)

In the preceding years, digital libraries evolved and matured. Their work primarily centered on workflows, interfaces, and to a small degree, the interplay between the system interfaces and the user. From around 2005 to the present day, the user and their engagement with digital libraries became a significant study area. While still focusing primarily on the user interface, how user's experience and engage with digital libraries and use the images they discover is becoming more sophisticated and advanced. The topic of digital library assessment development will be discussed later in chapter three.

In 2000, Saracevic (2000) asked, "How well does a digital library support the institutional or organizational mission and objectives?" (p. 363). Digital libraries were prepared to address this question by aligning their mission to the founding institution's mission more closely but focusing on creating an understanding of and providing access to primary cultural sources (Trivedi, 2010). Not bound by format, digital libraries expand the barriers placed on physical library resources. They present the historical and cultural assets of an institution to a worldwide audience. In addition, the World Digital Library (n.d.,a) states, they make "available on the Internet, free of charge and in multilingual format, significant primary materials from all countries and cultures." (para. 2). Recurring themes in the mission statements of digital libraries, is the conviction that making GLAM riches available for free, over the internet and in various formats creates a philosophy of international

peace and cultural understanding, enable life-long learning, and deliver resources to students, scholars, researchers, and the general public.

This era ushered in issues and questions that digital library practitioners still grapple with today, including the inevitable storage concerns for the preservation of the growing digital assets, the creation of multiple metadata schema, the obstacles resolving interoperability using the Open Archives Initiative Protocol for Metadata Harvesting, the integration of social media and user-generated content, and the evolution of consortial infrastructure (Brogan, 2006; Besser, 2016; Burns et al., 2019). These concerns have invoked conversation, workflows, policy, process, and technology innovation.

Section 2.3 Interchangeable digital library terminology

How institutions label collections of digitized images has varied over time. Initially, many institutions labeled collections of digitized materials digital libraries. Burns, Sundt, Pumphrey, and Thoms (2019) postulated that the widespread acceptance of the term digital library could result from adopting the Digital Library Initiative program grant name, which started in the 1990s. However, as discussed in section 2.2.2, the term "digital library" has been debated for many years. The following sections will break down the issue of how this phrase differs depending on the audience.

Section 2.3.1: Terminology used in GLAM literature

Within the GLAM literature, librarians, archivists, and other cultural heritage professionals are prone to using the terms *digital library, digital image collections, digital archives, image repositories, and image libraries* interchangeably (Green & Lampron, 2017; Munster, Kamposiori, Friedrichs, & Krober, 2018; Coburn, 2020; Feliciati. 2020). Some scholars and practitioners use the label *digital library* to describe the technological infrastructure or a set of services where the digitized content

resides (Xie, Joo, & Matusiak, 2021; Sharma & Chauhan, 2019; Liang & Chen, 2018; Kalisdha, & Suresh, 2017). They describe a digital library as the place where digital content such as text, still images, audio, and video are collected, organized, stored, retrieved, searched, and disseminated (Xie, Joo, & Matusiak, 2021; Sharma & Chauhan, 2019; Liang & Chen, 2018; Kalisdha, & Suresh, 2017). According to Agosti, Di Nunzio, Ferro, Maistro, Marchesin, Orio, Ponchia, and Silvello (2018), "digital libraries are heterogeneous systems with functionalities that range from data representation to data exchange and data management." (p. 30).

Section 2.3.2: Terminology used by GLAM institutions

Other practitioners have begun to study the prominence of terms used by institutions to describe locations for digitized images. Burns et al. (2019) noted that the terms *Digital Collections* and *Digital Library* were the two most common terms used by Association of Research Libraries (ARL) members in the United States. Stating, "Although the term digital collections appears to be increasingly popular, the use of digital library, digital archives, and similar variations continue to persist and are still quite popular." (p. 2).

Section 2.3.3 End-user perception of digital library labels

Going beyond the study of what terminology is used by GLAM institutions, Burns et al. (2019) explored "how users perceive different labels and associate them with different types of digital materials." (p. 2) They found that particular audiences, including undergraduate students, local community members, academic library staff, and public library staff, had differing expectations concerning what content would be associated with each term. When surveyed, most user types, particularly the Community Members and the Library's Staff groups, were more likely to choose the term titled *Digital History Collections* to find historical photographs. For the term *Digital Library,* respondents were more likely to associate that term with e-books. *Digital Archives* was associated with yearbooks and newspapers, and *Digital Collections* was associated with digital oral histories.

Burns et al.'s (2019) results illustrate the need for librarians and others working with online cultural history content to clarify the labels used to better reflect and clarify the materials within the online environment.

Section 2.3.4 Terminology used by GLAM institutions within this dissertation

For the digital libraries/collections referenced in this dissertation, the predominant term is Collections with variants for digital collections, image collection, online collections, and online collections database. Figure 1 illustrates the varying terminology used by GLAM institutions researched in this dissertation.

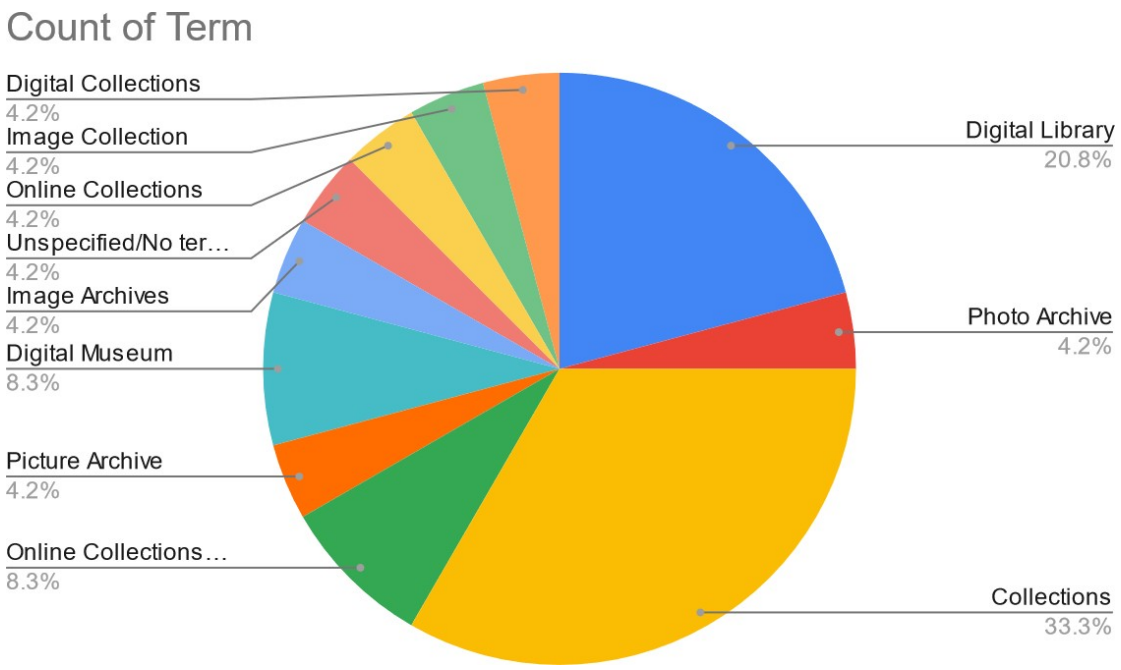

Figure 1 - Terminology used by GLAM institutions researched in this dissertation

As evidenced in the literature and current practice, the label digital library has evolved as a term. More GLAM institutions are using the term digital collections or collections to encompass their digital assets. The digital collections frequently exist hierarchically within a cultural heritage

organization's home page. Digital image and multimedia materials are gathered into a "collections or digital collections" page, which is a child of the institution's main page. An example (figure 2) of the hierarchical web presence of digital collection from the Slovak National Gallery shows the Online page as a second child page under the Collection page from the gallery's home page.

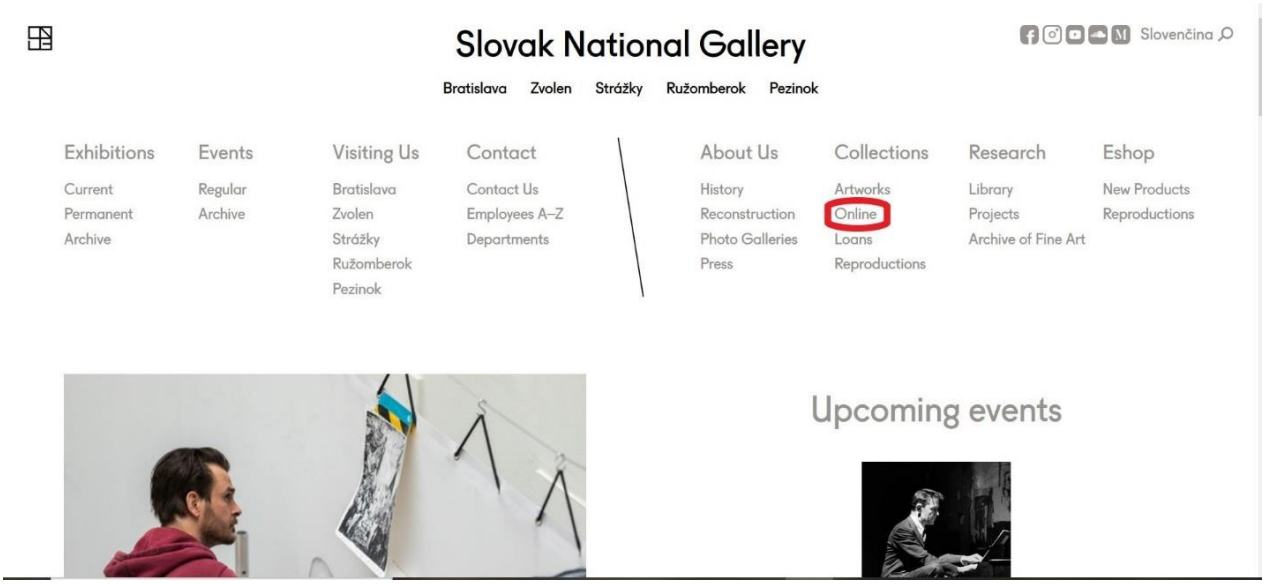

Figure 2 - Example of "Online" page as a second child page

For the purposes of this research, the term *digital library* is an umbrella term for any digital collection within a cultural heritage institution or digital library aggregator's web presence. The ultimate purpose of this dissertation is not to define what a digital library is or is not. It also does not address the characteristics that define a digital library. Instead, it adds to the critical discourse on reuse assessment, offers ways for practitioners to discover reused digital images on the web, and discusses methods that practitioners can use to increase digital image user reuse and engagement.

Section 2.4 Digital images in digital libraries

This dissertation focuses on digital images within a digital library. As defined in section 1.2, a digital image is the digital surrogate of an analog work, such as a photograph of a painting or sculpture.

Section 2.5 Overview of selected global digital libraries

This section provides an overview of a select number of digital libraries and digital library aggregators making an impact by bringing cultures together by digitizing and displaying their unique collections and/or aggregating and exhibiting content from other digital libraries. In addition, it serves to provide a knowledge base of digital library development through a global perspective.

The author defines a digital library aggregator as one that accumulates or harvests another digital library's materials, metadata, or hyperlinks. The growing number of disparate digital libraries worldwide has necessitated digital library aggregators and large-scale digital library platforms to facilitate greater access to GLAM collections. As Matusiak (2017) states,

> "Large-scale digital libraries (DLs) represent the next step in digital library development by providing a single access point to and the ability to search across a multitude of scientific and cultural heritage collections. Large-scale distributed systems, such as the Digital Public Library of America (DPLA) or Europeana, gather metadata from individual DLs or other aggregators and offer a central portal for searching and linking to digital objects." (p. 157).

For each selected digital library, the author recorded facts and figures when this data was publicly available. This information included: digital library's goal and mission, date of founding and its brief development over time, governing body, number of institutions that contribute to aggregators, number of items contained in the digital library, types of services and programming offered, financial information, and usage statistics. The author focused on this data to provide background information on how selected global digital libraries evolved. The author did not select images from all of these digital libraries as part of this dissertation because some did not meet the image selection criteria discussed later in chapters four, five, and six.

Section 2.5.1 Europeana

Europeana is a digital portal that partners with thousands of museums, archives, cultural heritage organizations, and libraries to share their unique and engaging materials for enjoyment, research, and education. Europeana Collections states that they display online "millions of cultural heritage items from around 4,000 institutions across Europe" (Europeana, n.d., About page). According to their mission statement, "We work with them through a number of regional, national, domain and thematic aggregators, communicating through a dedicated Aggregator Forum." (Europeana. n.d.,b, Mission page). Their mission is to "transform the world with culture. We build on Europe's rich cultural heritage and make it easier for people to use for work, learning or pleasure. Our work contributes to an open, knowledgeable and creative society." (Europeana, n.d.,a, About page). Europeana has three overarching goals. The first goal is to make it easier for users to discover materials from Europeana GLAM partners. The second goal is to extend the reach of these partner materials to markets and audiences who may not be able to access them easily. The third goal is to get the audience excited about cultural heritage materials by offering over 50 million and growing digitized works such as books, documents, art, photographs, and music. European also designed and hosts novel interactive projects and programs such as "GIF IT UP," (GIF IT UP, n.d.,a) a contest that invites users to craft a GIF using Europeana, Digital Public Library of America, Digital NZ (New Zealand), and Trove (digital library of Australia) openly licensed content. Financed by the European Union's Connecting Europe Facility and European Union Member States, the prototype went live in late 2008. At its launch, the world received access to 4.5 million digital objects.

Section 2.5.2 The Deutsche Digitale Bibliothek

The Deutsche Digitale Bibliothek (DDB) (n.d.a), mission is to promote unlimited access by providing a single website for users to discover and use the wealth of Germany's cultural heritage and scientific discoveries (Arora, 2018) through its web portal, as well as through the Europeana website (German Digital Library, n.d.). The DDB acts primarily as an aggregator, making content available from 23 institutions, including libraries, archives, museums, monuments offices, media libraries, universities, and other research organizations. According to the DDB, the platform and service "enables and encourages them [scientific and cultural partners] to network, cooperate and to develop and use together services and innovative tools. These make possible, in particular, new and more effective forms of presenting, managing and processing digitised contents." (DDB, n.d.b, para. 10) Per the DDB page portal, as of March 21, 2020, there are "32,827,731 objects, including 10,524,237 with digitised media." (DDB, n.d., para. 1). Founded in 2007 with a contribution of eight million euros from the Federal Government of Germany and launched online in 2012, the DDB started with an operational budget of 2.6 million euros for its first five years. With the start-up money, the DDB built its infrastructure and partnerships with state and local authorities (Arora, 2018). The German Federal Government and Länder. (DDB, n.d.) provide continued funding.

Section 2.5.3 Australia: Trove Discovery Services

Launched as a metadata aggregator, the Trove Discovery platform is Australia's national digital image and scholarship repository with contributions from GLAM and research institutions of various types and sizes. The National Library, Australia's State and Territory libraries, and hundreds of cultural and research institutions around Australia collaborate "to create a legacy of Australia's knowledge for now and into the future." (National Library of Australia, 2020, para. 3). From its inception in August 2008, Trove focused on collating disparate digital items across Australia, including "the Register of Australian Archives and Manuscripts, Picture Australia, Libraries Australia, Music Australia, Australia Dancing, PANDORA web archive, ARROW Discovery Service

and the Australian Newspapers Beta service." (National Library of Australia, 2020, para. 2). To date, Trove provides access to approximately 6.5 billion items, including scholarly journals and articles, books, maps, diaries, digitized newspapers, audio/video recordings, and images, to name a few.

Zone	Work Count
Journals, articles and research	12,111,614
Books	20,432,221
Maps	445,215
Diaries, letters, archives	744,075
Government Gazettes	3,535,033
Lists	132,584
Music, sound and video	2,787,977
Digitised Newspapers and more	226,952,521
People and organisation	1,213,992
Pictures, photos, objects	4,520,175
Australian web archive	6,217,613,420
Total	6,490,488,827

Table 1 - Trove Discovery Services Item Count. (National Library of Australia, 2020)

Section 2.5.4 Canadiana

Canadiana (Canadiana webpage, n.d.) is a consortium and national digital library aggregator that consists of 76-member academic libraries, two national libraries, and the largest public library system in Canada. The Canadian Research Knowledge Network (CRKN) manages the platform. Focused on expanding access to Canada's rich cultural heritage, CRKN's mission states that they "advances [sic] interconnected, sustainable access to the world's research and to Canada's documentary heritage content." (Canadian Research Knowledge Network, 2020a, para. 1). CKRN merged with Canadiana in 2018. Canadiana, rooted in the foundation of in three major organizations: the Canadian Institute for Historical Microreproductions (founded in 1978), the Canadian Initiative on Digital Libraries

(founded in 1997), and its successor, AlouetteCanada (founded in 2007) (Canadian Research Knowledge Network, 2020b). The three flagship collections that comprise the Canadiana online collections are; Early Canadiana Online, Heritage, and Canadiana Online. Combined, the total number of monographs, periodicals, newspapers, images, ephemera, and more is over 60 million pages (Canadian Research Knowledge Network, 2018).

Section 2.5.5 Library of Congress Digital Collections

The central research division of the U.S. Congress and the home of the U.S. Copyright Office, the Library of Congress, is considered one of the world's largest libraries and is home to the Library of Congress's Digital Collections (LOCDC) (LOC, 2018). Its collection consists of "millions of books, recordings, photographs, newspapers, maps and manuscripts." (LOC, 2018, about page). The LOCDC makes available digital surrogates that document the American experience, creativity, cultural and historical heritage, and knowledge progression throughout the world. Viewers can access maps, photographs, letters and documents, newspapers, and audio/visual recordings from the digital collection's portal. The LOCDC was founded originally as the American Memory Project pilot, which ran from 1990 through 1994 (American Memory, Library of Congress, n.d.). From 1994 until the present day, the Library of Congress has quietly migrated the American Memory Project to its current platform and title. Its broad and comprehensive subject areas consist of American history, government, law and politics, performing arts, world cultures and history. Additionally, the LOCDC aggregates millions of pages of materials from thousands of GLAM institutions around the United States of America (The Library of Congress, n.d.).

Section 2.5.6 Gallica

Launched in 1997, Gallica is the digital library of France. Its "mission is to preserve French heritage and to serve as a national encyclopedia of France." (Borbely, 2013, para. 1). Gallica partners with over 270 institutions, including libraries, museums, government, higher education, and archives.

Gallica aggregates and links to the partner's digital library or displays the digital content for partners lacking a digital library platform (Bibliothèque Nationale de France, n.d.). The Gallica website has gone through several iterations since its inception, the last in 2015 (Bibliothèque Nationale de France, n.d.). Funded by the Bibliotheque Nationale de France (National Library of France), it supports the digitization of France's cultural heritage materials, including tens of thousands of documents, images, and sound recordings for free and open access (Bibliothèque Nationale de France, n.d.).

Section 2.5.7 National Digital Library of India

The Indian Institute of Technology Kharagpur developed the National Digital Library of India. It operates under the auspices of the "Ministry of Human Resource Development (MHRD) through its National Mission on Education through Information and Communication Technology (NMEICT)." (National Digital Library of India, 2020, popup screen). It is intended as a multi-lingual, multi-user, and multi-disciplinary portal for Indian digital resources, it includes 47,944,566 items (National Digital Library of India, homepage search box, 2020).

Section 2.5.8 Japan Search

The mission of Japan Search is to preserve, transmit, and promote the use of the digital cultural, tourism, educational, and research archives and resources of Japan. It aspires to communicate new values and promote innovation (Japan Search, 2021a, about page). Operating under the auspices of the National Diet Library of Japan and "in cooperation with a variety of organizations in Japan under policies established by the Working Group Steering Committee of the Digital Archives Japan Promotion Committee." (Japan Search, 2021b, home page). Japan Search is a metadata aggregator of many of Japan's digital resources. Japan Search resources are categorized by "Education or commercial use," "Searchable contents," and "Online access."

Section 2.6 Conclusion

As these sections have demonstrated, there have been significant developments in computer technology and digital library platforms. This dissertation would not be possible without the foundational theories, concepts, and accomplishments of these pioneering technologies, technologists, and theorists. The research presented in this dissertation uses internet-based systems and protocols to understand user file naming behavior and reuses of digital library content not addressed in previous studies.

Chapter 3: Digital Library Assessment

This chapter provides a brief review of the changing scholarship trends surrounding the evaluation of digital libraries and digital image reuse.

Virtually, as soon as digital libraries came into existence, assessment of them commenced. According to Xie, Joo, and Matusiak (2021),

> "Digital library evaluation is crucial to digital library development and enhancement. Without evaluation, the success of a digital library cannot be valued, the problems of a digital library cannot be identified, and, most importantly, the life cycle of digital library development cannot be sustained." (p. 131).

The author accumulated a representative set of relevant articles from various sources, including Library Literature & Information Science Full Text (H.W. Wilson), Library, Information Science & Technology Abstracts, and Information Science & Technology Abstracts, to retrieve the largest and most diverse articles on the topics. Additional relevant articles were located from journals library journals such as D-Lib Magazine, Library Review, and Journal of Web Librarianship. Search terms included "digital library assessment," "Digital library evaluation," and "digital library use" "design and evaluation." Author keywords comprised many of the search terms. Another article gathering strategy was investigating already selected article bibliographies and using the "cited by" function on Google Scholar for articles already selected.

This chapter is divided into decades and reflects the maturity of digital library assessment over time. It shows how assessment approaches became more diverse and sophisticated, shifting from investigations focusing on interface design, website functionality, and user group identification in the 1990s to user studies and the standardization of metrics by the 2010s. During this time, assessment has gone beyond traditional approaches such as focus groups to adopting technology to obtain user needs information, usability, reuse, and content collection.

Section 3.1 Digital library assessments - 1990 to 1999

Scholars questioned whether traditional metrics were complex enough to measure this new media during this early decade of digital library growth. Instead, they focused on interface design, website functionality, research criteria, and determining user groups. As a result, very little actual assessment took place as the digital library community grappled with building technology infrastructure and provisioning digital libraries.

Discussions on the topic of digital library assessment started with identifying the critical components of an evaluation. Van House, Butler, Ogle, and Schiff (1996b) state, "Digital libraries [DL] can be described and evaluated on three key components: contents, functionality, and interface." (para. 4) Whereas Marchionini, Plaisant, and Komlodi (1998) designated four categories of digital library research/assessment: content, services, technology, and culture. Despite the taxonomy differences, the first three components from each are the same. The content category referred to items associated with the digital library, such as the format types, metadata scheme, or collection scope. The functionality and/or service category signified the search, browsing, and filtering operations. Interface and/or technology indicate the user platform and interoperability. Marchionini et al. (1998) define the term culture as privacy, rights management, and ensuring data quality. However, some minor crossovers in these broad categories are related to the user interface and search capabilities.

There was a consensus that user assessment is an iterative process. Van House et al. (1996b) contended that digital library administrators should assess user needs periodically. They assert that administrators start iterative assessment with the front-end user interfaces and that additional future assessments should continue to understand the use of digital materials (Van House et al., 1996b).

The literature provides a limited number of examples of user studies assessments conducted during this period. Early evaluations, employing focus groups and survey data, concentrated on developing interfaces and query path search behavior to meet user needs. The Human-Computer Interaction Lab at The Library of Congress (LC) completed one such user interface assessment (Plaisant, Marchionini, Bruns, Komlodi, & Campbell, 1997). The goal of the project was divided into

three iterations, "to establish a user-centered design team for the NDLP, to create interface prototypes that serve a wide range of users, and to develop a variety of tools and widgets that LC may incorporate into future implementations." (Plaisant et al., 1997, p. 518). The researchers surveyed K-12 teachers, school library media specialists, parents, and daycare center workers as part of the user interface development (Marchionini & Plaisant, 1996). Beyond interface design, the Lab also tested elements such as search, browse, metadata display, user roles, and filter criteria (Plaisant et al., 1997).

Section 3.2 Digital library assessments - 2000 to 2010

Digital libraries matured between the years 2000 to 2010. Daniel Greenstein (2002) states, "most [sic; digital libraries] are still at a stage where limited experimentation is more important than well-informed strategic planning." (p. v). While the evaluation of digital libraries was not well developed at this time, Barton (2004) expresses the importance of performance assessment, stating that it is "a means of demonstrating the value of digital library services and their contribution to institutional goals, thereby securing resources for the future." (p. 141). This period in the digital library assessment growth concentrates on evaluation strategies, frameworks, approaches, gaps in assessment practice, and evolving trends. Digital library practitioners pay little attention to the content users were engaging in. Professionals debate the merits of using traditional library assessment metrics or creating new ones out of whole cloth.

Researchers in the 2000s identified multiple imbalances with digital library assessment. Greenstein (2002), Bertot (2004), and Barton (2004) all observe that digital library assessment is new and perplexing to professionals. The prospect of evaluating digital libraries and their users is challenging for practitioners. Reeves, Buhr, and Barker (2005) assert that evaluation requires a sophisticated approach using the "triangulation" method of "multiple models, procedures, and tools." (p. 420). Greenstein (2002) goes on to state the librarians have insufficient data on "how library users behave in a network environment" (p. v), "which is in contrast with the traditional assessment of library collections and services." (p. v). Saracevic (2004) notes a glaring difference when comparing

the number of evaluations of institutional repositories to digital libraries. Saracevic (2004) additionally states, "evaluation [of digital libraries] seems to be an exception rather than a rule." (p. 10) He divides digital library evaluations into two categories:

1. "Meta or "about" literature: works that suggest evaluation concepts, models, approaches, methodologies or discuss evaluation, but do not contain data
2. object or "on" literature: works that report on actual evaluation and contain data; even data reporting is of two kinds: hard data or soft (sort-of) data (impressions)." (Saracevic, 2004, p. 9).

Scholars made use of many different methods to assess digital library user behavior. Illustrated by Jela Steinerova (2007), who used the term 'relevance' to evaluate the reuse of digital library materials, opposed to the traditional definition of "an evaluation of services, system performance, or the success of the digital library." (p. 38). Steinerova (2007) stated, "One way to improve the success of digital libraries could be a phenomenological inquiry into the experience of users when judging relevance." (p. 38).

Whereas Steinerova (2007) concentrated on traditional methods such as focus groups to study user behavior, Sfakakis and Kapidakis (2002) used transaction logs to understand how users interact with the digital library platform. However, a limitation of this method was that it did not indicate the user's purpose for using the digital library. In later chapters, this dissertation will discuss how digital library practitioners use web and software resources to determine how users interact with digital library platforms and digital library images and for what purposes.

Many researchers developed strategies or key evaluation criteria around the assessment of digital libraries to understand user behavior or engagement with content. Assessment toolkits consisted of usage metrics, user identification, technology services, and collection content (Fuhr, Hansen, Mabe, Micsik & Sølvberg, 2001; Barton, 2004; Reeves et al., 2005; Marchionini, 2000; Long, 2002; Borgman, 2003a & b; Jeng, 2005; Buchanan & Salako, 2009).

Meta-analysis showed up frequently in the literature during this period. Meta-analysis is defined as the analysis of combining the results of multiple studies of different types of digital library assessments. A prime example of a meta-analysis is the research of Thong, Hong, and Tam (2002). They asserted that "the existing digital library research has focused on the technical development of the system, such as information storage, information retrieval, and system integration. Most of the prior research are in-depth case studies of an individual digital library." (p. 216).

To improve evaluation methods, Thong et al. (2002) emphasize the importance of establishing "the critical factors that affect the intention to use digital libraries, examine the relative impact of each individual factor, and understand the mechanism through which they act." (p. 216). Tsakonas and Papatheodorou (2006) did just this when they identified five assessment criteria: relevance, format, reliability, level, and timeliness. (p.403). Establishing criteria and understanding mechanisms was the optimal way to increase digital library interface functionality and collection development.

As digital library evaluators develop assessment criteria, a limited number of documented user studies around system design appear in the literature. For example, Somerville and Brar (2009) and Toms, Dufour, and Hesemeier (2004) discuss how the assessment teams recruited users to evaluate their current interface and/or website and test future prototypes. An anomaly to user studies was a study completed by Bui and Park (2006), who focused on digital library assessment through evaluating metadata quality.

Section 3.3 Digital library assessments - after 2010

The period after 2010 has seen a shift away from usage assessment and user interface upgrade assessments to focus more closely on the user experience. User needs, satisfaction, discovery, accessibility, and interaction with the growing cache of digital image collections have led to various studies (Felicitate, 2020; Felicitate, 2020; Glowacka-Musial, 2020; Dang, 2020; Wagner, 2020; Münster, Kamposiori, Friedrichs & Kröber, 2018). Münster et al. (2018) state,

> "despite the wide availability of digital image collections today, previous research has shown that scholars regularly face problems with accessing and interacting with them. These problems relate to issues such as discoverability, copyright, width and depth of material, or low image quality and can significantly limit the image collections' potential utility." (p. 367).

This dissertation presents a novel approach to rectifying user and reuse discoverability by taking advantage of newly discovered user file naming behavior not addressed in the scholarly literature.

One area of scholarship concentrates on creating standardized methods for assessment. The Digital Library Federation Assessment Interest Group (DLF AIG) attempts to tackle standardized methods by developing a set of best practices and guidelines for assessing user and usability studies, return on investment, the reuse of digital library content, and data collection (Chapman, DeRidder, Hurst, Kelly, Kyrillidou, Muglia, O'Gara, Stein, Thompson, Trent, Woolcott, & Zhang, 2015; Bragg, DeRidder, Johnston, Junus, Kyrillidou, Chapman, & Stedfeld, 2016; Chapman, DeRidder, & Thompson, 2015).

Several scholars are endeavoring to identify who are the users of digital libraries, for what purposes do they use digital libraries and images and is it possible to equate usage to return on investment (ROI) (Konkiel, Dalmau, & Scherer, 2015). In striving to provide cultural heritage content to users, the actual production and display costs are often overlooked, both in reality and scholarly literature. Chapman et al. (2015) states,

> "very little published research exists on the topic of ROI for digital libraries. This is surprising given the significant interest in the literature of higher education in general over the past decade, as well as in academic library literature within the past five years. The most common theme in the existing literature on ROI for digital libraries is analysis of "cost" in the form of both time and money." (p. 25).

They discuss the importance of digital libraries undergoing ROI assessments to convey the value of digital libraries to the institution.

In 2014, Kelly (2014) reviewed the digital library literature from 2004-2014. She identified eight scholarship themes, Altmetrics (Konkiel et al., 2015; Kelly 2017), Google analytics (Bragg, DeRidder, Johnston, Junus, Kyrillidou, Chapman & Stedfeld, 2015), Google alerts (Kelly, 2018b), Reverse Image Lookup (RIL) (Reilly & Thompson, 2017; Kelly, 2015), Use logs (Reilly & Thompson, 2014), embedded metadata (Thompson & Reilly, 2018), usability study (Balog, 2011; Matusiak, 2012; Jabeen, Qinjian, Yihan, Jabeen & Imran, 2017; Hu, 2018; Campbell, 2018), and "content usefulness." (Lamont, 2014). Kelly concluded her article by acknowledging the growing interest in these new and emerging areas of study as an augmentation to past usability and web statistic digital library assessments. The author agrees with Kelly that the "majority of scholarship about digital library assessment" (p. 384) involves usability and web statistics.

A developing area of collection assessment and research is the concept of "Collections as Data." According to Padilla (2017), "collections as data entails thinking about ways to increase meaning making capacity by making collections more amenable to use across an expanded set of methods and tools, typically but not exclusively computational in nature." (p. 2.). This area of research theorizes that digital collections have computational research value beyond the traditional use of viewing or reusing the object and should become a core activity of the cultural heritage sector (Wittmann, Neatrour, Cummings, & Myntti, 2019; Padilla, Allen, Frost, Potin, Russey Roke & Varner, 2019; Ames, 2021). According to Candela, Saez, Escobar & Marco-Such (2020), "making digital collections available as data and ready for computational analysis have an impact on user

engagement since collections are more accessible and interoperable, using open licences and reactivating legacy material held by GLAM institutions." (p. 1).

Another area of digital collections research is Linked Open Data (LOD). Researchers contend that LOD will make digital collections and images more accessible, interoperable and create relationships between digital collections by enhancing descriptive metadata vocabularies (Candela, Saez, Escobar & Marco-Such, 2020; Marcondes, 2020).

A novel research topic uses visualization to assess digital library collections, digital library users, and/or digital image reuses that create a new product. These new products include overlaying maps to detail the historical movement of people or telling a personal or historical story graphically. Glowacka-Musial (2020) describes visualization as "graphic representations of data objects and their relationships. It is commonly viewed as an essential part of textual data analysis that helps to make quantitative information legible and easy to comprehend." (p. 5). Additionally, curators are using this tool to determine the structure and organization of their collections, the "content and provenance, relationships among the collection's items, the scope and size of the collection, and the number of files and their formats, as well as text patterns in documents and visual patterns among images." (Glowacka-Musial. 2020, p. 6).

Social media assessment and research are underway that involve discovering insights into user interests and engagement (Feliciati, 2020). Examples include the act of curation employing social media platforms such as Instagram and Pinterest (Ullrich & Geis, 2021; Thompson & Reilly, 2019), adding the use of Twitter and Instagram in the art history classroom, (Jimerson & Leigh, 2020) and gamifying cultural heritage venues to increase topic knowledge and engagement of museum patrons (Kontiza, Liapis, & Jones, 2020).

Other evaluations and research studies center around the impact of digital culture. For example, creating digital libraries, collections, and images affects not only the development and construction of the digital libraries, collections, and images but affects the users, user satisfaction, user research productivity, and user teaching and learning (Feliciati, 2020; Green & Lampron, 2017; Dang, 2020; Xie, Joo & Matusiak, 2021).

The growth of GLAM Labs is a direct subset of the open access movement. The emerging GLAM Labs facilitate the creation of unique and creative reuse and experimentation with cultural heritage content (Valeonti, Terras & Hudson-Smith, 2019; Glowacka-Musial, 2020; Candela, Sáez, Escobar & Marco-Such, 2020; Mahey, Al-Abdulla, Ames, Bray, Candela, Chamber, Derven, Dobreva-McPherson, Gasser, Karner, Kokegei, Laursen, Potter, Straube, Wagner & Wilms, 2019). Mahey et al. (2019) state

> "GLAM Labs come in a variety of shapes and sizes. They use experimental methods to make cultural heritage collections available in innovative, engaging and unexpected ways. Operating at the intersection of digital cultural heritage, innovation, technology and creativity, they provide significant benefit for organisations, users, society and culture." (p. 33).

Section 3.4 Digital image reuse

There are various types of digital reuse, but two of the most prominent examples are digital image reuse and research data reuse. The author's research focuses only on digital image reuse. As defined in Section 1.2, the author utilizes the following definition for reuse: the act of (a) using, transmitting, or sharing a digital image in a new setting or (b) repurposing or transforming the image into a new object that is not its original platform, purpose, or context. Kenfield, Kelly, Muglia, O'Gara, Thompson & Woolcott (2019) state that "reuse shows engagement with collections and impact of digital repository resources in a more meaningful fashion" (p. 51).

Baldacci (2019) uses the term 'recirculation' to describe the flow of digital images from one user to another. She states that recirculation is "a process through which both visual and cultural imagery are put in motion over and over again in the current information age, and in the context of post-Internet art in particular." (p. 25). Closely related or a subset of reuse is 'remix.' Waysdorf (2021) describes remix as "the use of preexisting material to make something new." (p. 1) and goes on to say, "in its defining form, remix is about reuse: about doing something with existing material." (p.

11). Remix culture represents products such as memes, videos made with existing digital images and multimedia, and other resources found online. Some excellent examples of projects that inspire and promote the reuse of digital images are,

- Europeana's GIF IT UP (GIF IT UP, n.d.,b, Winners 2020) competition, which awards creative "gif-making" of digital images from Europeana, the Digital Public Library of America, Digital NZ, and Trove.
- The British Library Labs, which annually awards exceptional digital reuse projects.
- The Library of Congress Innovator in Residence program which annually awards a creative individual to use the library's digital collections to create novel and innovative tools, artwork, or applications.

Why would GLAM institutions want to promote the reuse of their digital content? Several authors agree that the reuse of digital images adds value perceived or otherwise to the GLAM institution by driving foot traffic to the brick-and-mortar building, increasing knowledge and culture, heightening community development and engagement, affirming civic aspirations and inspiration, aiding in public education, and fostering public and social good (Terras et al., 2021; Liew, Goulding & Nichol. 2020: Thompson and Reilly, 2019; Konkiel et al., 2015; O'Neill, 2017). Konkiel et al. (2015) write,

> "Digital special collections in particular can have value to the public, beyond their use for research and scholarship. Many collections are reused by the casual reader in ways that can leave traces of impact like unexpected references to the source collections on the Web in the form of memes, "fan" websites, and other "pop culture" formats." (p. 4).

Liew et al. (2020), reaffirms the concept of social good by stating, "Digital technologies offer potential solutions to issues around facilitating participation and barriers to community engagement in the cultural heritage sector. Digital tools can provide innovative approaches for collecting, curating, sharing and visualising cultural memories." (Liew et al., 2020, p. 5).

Aleksandra Strzelichowska (2020) from the Europeana Foundation outlines the reasons best in her blog post: *Why should you open up your digital collections for reuse - explained in GIFs.*

1. "To give content another life.
2. To allow content to be used in education.
3. To remain relevant.
4. To gain visibility.
5. To introduce people around the world to your content.
6. To be able to share a bigger part of your collection.
7. Because you can start small." (para. 3-8).

User engagement in and the reuse of digital images motivate the stimulation of ideas, encourage self-expression, inspire art and culture, and create a mash-up/maker movement. Orlandi (2020) writes

> "If the Museum releases images and contents with free licenses, it authorizes, allows and hopes that people rework them, make them their own, share them and therefore participate in new narratives, and create new points of view on what we value as heritage. And heritage becomes alive when it is reinterpreted right here, right now, and, sure is, the museum can be stimulated by ideas coming from the public." (p. 62).

Understanding a user's purpose for utilizing digital images, how they search, how they engage with digital images, and what terminology/folksonomy they use to search can impact how digital libraries make existing content more user-friendly and discoverable, retain and excite current users, and connect with new users. As Terras et al. asserts, "Most transparently, there are opportunities to render collections in new ways, providing novel means for audiences to experience and engage with them, cocreating new meaning." (Terras et al., 2021, p. 6).

In addition, these user-friendly collections garner collaborations with other GLAM institutions, encourage research scholarship use possibilities, fuel attention from scholars and the general public, stimulate creative innovation and economic growth, and promote equitable opportunities for users to engage with collections. “The most successful digital initiatives are often those that are developed in partnership. Specialist curators have extraordinarily rich knowledge of collections and are in close contact with users and researchers.” (Siefring, 2019, p. 25).

Section 3.5 Digital image reuse assessment

A distinct type of user assessment is digital image reuse. This dissertation focuses on this type of assessment using technology to evaluate reuse, unlike some reuse studies that use more traditional data collection methods such as focus groups, user surveys, and download statistics. Since 1999, several user studies have focused on who is using digitized materials and for what purposes. The following is an overview of a limited number of these studies, divided into two groups; studies focused on general audiences and studies focused on scholarly researchers.

The reuse of digital image library materials is a developing area of research among scholars and practitioners. Several directions of research have emerged. Smith (1999) discusses the need for more robust user studies. She writes, “We need more user studies before we can assert confidently what may seem self-evident to us now: that adding digitized special collections to the mass of information available on the Internet is in the public interest and enhances education.” (para. 24).

Case studies focused on general audiences often addressed various methods for collecting user populations’ data and DL consumer reuse patterns. For example, Reilly and Thompson (2014) used log data from a digital cart service (a predecessor to a “download” feature prominent in digital library interfaces today) within a digital image library to assess the “ultimate use” (p. 196), or reuse, of digital objects. Reilly and Thompson (2014) asked three questions: “why images were used, what products were created from the images, and what implications this has on digital library management.” (p. 196). They concluded that the most extensive user type was in the “Visitor”

category. These users were using the digital library collections for their own "Personal" agenda. Reilly and Thompson believed that reuse assessment provides practitioners with better ways to develop collection management priorities.

Furthermore, practitioners can create promotional material for those collections. Using a different data collection method, Chung and Yoon (2011) analyzed patron search requests from the Yahoo! Answers.com portal and created a set of codes to categorize image reuse. They determined that "illustrative uses" (p. 163), and "generation of idea uses" (p. 163) were the majority of uses. In another article, Kelly (2018) evaluated Wikipedia citations to identify instances of cultural heritage institution digital image collection reuse within the Wikipedia community. She contends that Wikipedia use patterns can help practitioners discover what the "community finds most relevant and derive recommendations for collection development and digitization priorities." (p. 102).

Punzalan, Marsh, & Cools (2017) devised a framework for "documenting, demonstrating, and assessing the impact of digitized ethnographic collections" (p. 82). This framework identified six parameters, including "knowledge, professional discourse, attitudes, institutional capacity, policy, and relationships," detailing what the authors contend are meaningful impacts for ethical collection development of digitized material from indigenous repositories (p. 64). Incorporating this framework in digital collection selection and access decisions, according to the authors, will mitigate adverse outcomes or unintended harm to indigenous peoples (Punzalan et al., 2017). This argument echoes an observation made by Reilly and Thompson (2014) and Harris and Hepburn (2013), who all advocate for increased user participation in collection development decision making. Like Punzalen et al. (2017), Kelly (2018a) developed a framework intended to assist digital library practitioners in making collection development decisions and setting digitization priorities. Kelly (2018a) asserts that this framework should be used "in conjunction with other assessment methods to inform CHI [cultural heritage institutions] how their resources are being utilized online." (p. 102).

Case studies on reuse assessment have also explored the ways scholars and researchers use digital images. Terras (2015) explores the impact of open licensing on the reuse of images by scholars in the arts and humanities. Terras (2015) notes that despite the proliferation of digital image content

online, the move towards licenses that promote open sharing and reuse has been slow to evolve. Terras (2015) outlines future areas of focus. These include the need for practitioners to pay more attention to the study of open cultural heritage image data, advocate for adopting licenses that favor open access, and develop search technologies that make open access content more discoverable.

McCay-Peet and Toms (2009) interviewed historians and journalists to understand how these two audiences reuse digital image content. The authors found that historians and journalists used digital images for "illustrative purposes" (p. 2427), often relying on the images to bolster the historical arguments or storylines presented. Conversely, they observed fewer instances of image reuse for "information purposes" (p. 2427). McCay-Peet and Toms (2009) also found that the discoverability of images depended more on "conceptual attributes" such as the subject (person, event, or location) of the image (p. 2427) and suggested that practitioners incorporate some of these attributes within an object's metadata.

Harris and Hepburn (2013) also analyzed digital image reuse practices among historians by conducting a content analysis of scholarly articles in history journals. Given the vast amounts of digital images now available online, the authors predicted that they would find an uptick in digital image reuses in recent scholarly publications. However, after conducting their analysis, they discovered that recently published articles were not increasingly incorporating digital images, leaving them to conclude that "historians are not finding images suitable to their research" (p. 278).

Beaudoin (2014) performed another discrete research project that included image reuses of archaeologists, architects, art historians, and artists. By collecting data through interviews, she identified multiple types of reuses and created a framework to analyze reuse categories. Beaudoin (2014) noted that the most prominent form of reuse was the "development of knowledge." (p. 131). She notes other reuse types included: "developing creative works," "critical thinking development," "translating verbal information," "engaging students," "creating emotion," and "marketing" (p. 131).

Reilly and Thompson (2014) merged Beaudoin's (2014) work with the research that Chung and Yoon (2011) conducted to develop a more comprehensive protocol for studying users and reuses

of digital images. This approach "expanded the user base to include general visitors, scholars, researchers, students, and university staff" (p. 201).

Section 3.6 Conclusion

As discussed in this chronological literature review, digital library assessment has evolved since digital library inception. Early work in digital library development generated questions about how this new information technology could be assessed. Scholars considered whether brick and mortar library assessment criteria and methods were applicable or whether new techniques would need to be employed. Very little actual assessment happened during the early years of digital library development as interface design, website functionality, and collection building took precedence over assessment. By the mid-digital library development period, assessment criteria were created and used on a limited number of user assessments. Researchers have refined the types of assessment, continuing to do platform and user studies while also exploring new areas such as image reuse. Digital library reuse is an emerging research topic that uses traditional methods such as focus groups and surveys as well as methods that use technology to determine reuse, such as RIL and website searching. Assessments are less about how the digital library platform or interfaces work and more about considering who is using the digital materials displayed. This dissertation builds on the results of previous reuse assessment studies.

The first dissertation research question intends to verify whether RIL and embedded metadata continue to have some assessment utility for practitioners. The second research question, which uses a methodology from an earlier study completed by the author (Thompson and Reilly, 2018), identifies user file naming behavior patterns that could be helpful in the discovery of reused digital images. These user file naming behaviors identified by the author are new contributions to the reuse assessment literature. Finally, the third research question diverges substantially from previous reuse research. It focuses not on popular research tools to discover reuse but, instead, it addresses user behavior file naming as a reuse search strategy for practitioners. This research contributes a unique approach to reuse assessment because it emphasizes a search engine agnostic strategy to collect reuse data, which the author argues is more effective than other methods studied previously.

Chapter 4: Content-Based Image Retrieval and Reverse Image Lookup

This chapter will discuss the use of Content-Based Image Retrieval (CBIR) and Reverse Image Lookup (RIL) to track the reuses of digital images. It will define CBIR and RIL, how the tools operate, and the types of applications available online. It will also provide an overview of the industries that take advantage of these tools and outline studies that have used these tools in a digital library reuse assessment. Additionally, this chapter will perform a verification study to evaluate the effectiveness of using Google Images to discover reuse.

Section 4.1 CBIR/ RIL definitions and types

CBIR and RIL are image-based search applications that offer users the ability to query exact or similar images based on image attributes.

Section 4.1.1 Content-based image retrieval - CBIR

Van House, Butler, Ogle, and Schiff (1996b), Eakins and Graham (1999), Marques (2016), Beaudoin (2016), Thompson and Reilly (2017), Beskow & Carley (2020), and others have provided broad summaries and histories of CBIR technology. CBIR, coined at a 1992 National Science Foundation workshop, relies upon underlying search algorithms (Shapiro & Stockman, 2001) to provide users with various ways to search for images, including pre-existing or user sketched images, semantic retrieval, and relevance feedback. Beaudoin (2016) states, "unlike systems that rely on users' text-based queries, CBIR systems allow users to query image collections by purely formal visual characteristics." (p. 350). These visual characteristics include color, shape, and texture within an image. CBIR retrieves exact or similar images (Eakins & Graham, 1999). Sharma and Batra (2014) identified another benefit to using a CBIR search query. They write,

> "Having humans manually annotate images by entering keywords or metadata in an astronomically immense database can be time consuming and may not capture the keywords desired to describe the image. The evaluation of the efficacy of keyword image search is subjective and has not been well-defined." (para. 2).

As Marques (2016) notes, CBIR "are visual-based solutions whose main characteristic is the ability to answer the question "What is this?" (or, more realistically, "Can you find more images/videos that look like this?") using only an image as input to the query" (p. 8). CBIR search tools answer these questions through a series of functions. After the user drags and drops or uploads an image into the search box, the search algorithm parses the image attributes. It compares the results with images in the database (usually large, unstructured, and distributed) or on the internet. CBIR extracts pixel data to encode the visual content also referred to as the feature vector. CBIR feature vector extraction is an offline index query using the Query By Example (QBE) technique (Marques, 2016). Using relevance feedback (Datta, Joshi, Li, & Wang, 2008), CBIR (and RIL) uses machine learning to refine search results based on an evolving understanding of user intent (Cardoso, Muller, Alexandre, Neves, Trevisani & Giraldi, 2013).

One early example of a database that integrated CBIR functionality was Cyprus, a University of California Berkeley digital libraries research project that experimented with developing more sophisticated, efficient image searching methods. According to Van House et al. (1996b), Cyprus "adds the capability of searching for color "blobs"--clusters of color within an image that are likely to correspond to objects such as yellow flowers, a red car, or an orange fish" (para. 9). Projects like Cyprus were the beginning of designing for RIL functionality.

Section 4.1.2 Reverse image lookup - RIL

Linder, Webb, and Kerne (2013), Terras and Kirton (2013), Chutel and Sakhare (2014), Thompson and Reilly (2017), and others have also described how RIL applications are a refinement of earlier CBIR technology, particularly in the design of the user interface. CBIR searches across discrete image databases, whereas RIL identifies instances of images posted on the web (Terras and Kirton, 2013; Chutel & Sakhare, 2014; Thompson & Reilly, 2017). As Chutel and Sakhare (2014) note,

> "These engines are to be provided with a query image rather than keywords. The searching process of these search engines comprises of first indexing the images available over the web and then performing matching of the query image with the images in the web to retrieve the desired images." (p. 1430).

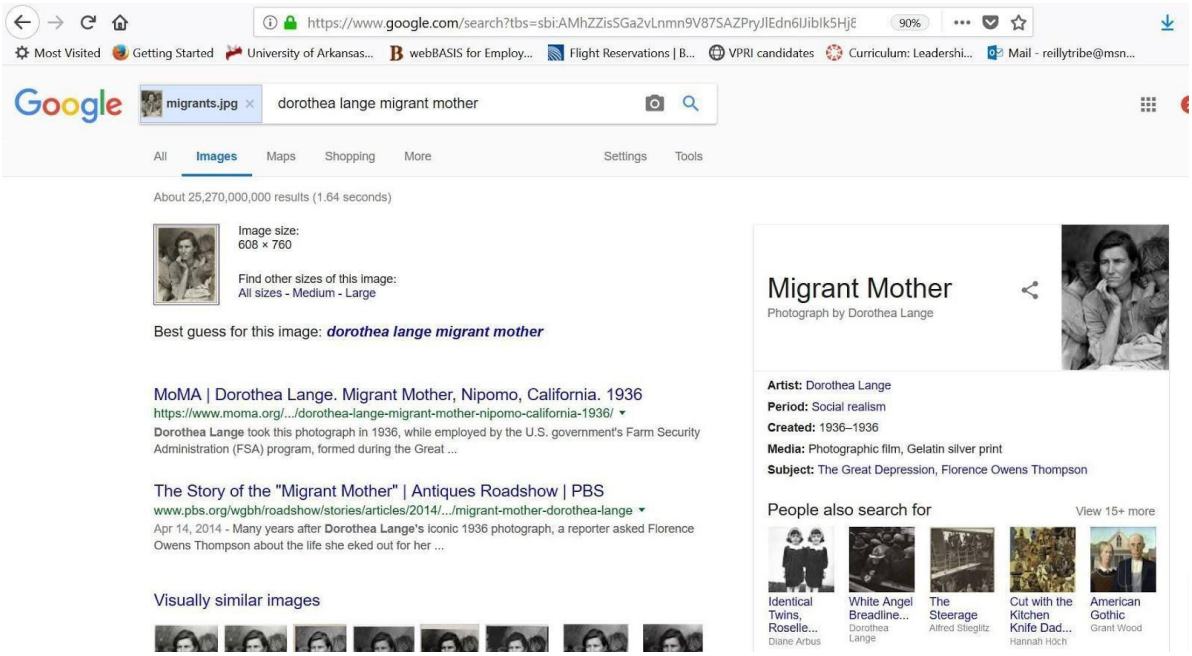

Figure 3 - Google Images search results

Section 4.1.3 Types of CBIR/RIL applications

CBIR/RIL online applications have expanded over time. There are now many CBIR/RIL applications, including Google Images, Tineye, Bing Image Trending, RevIMG, and Pinterest Visual Search Tool (Ajinkye, 2019; Gaikar, 2012).

Section 4.2 Private industries that use CBIR or RIL applications

Since its inception, various private industries have applied CBIR/RIL technologies to solve "real world" problems. Beaudoin (2016) describes how the building construction and design industry has used CBIR/RIL to research floor coverings and retrieve images of construction materials for users. She writes, [images] "illustrated how aspects such as materials, manufacturer, tuft sizes, pattern, and color could be weighted and searched upon by users." (p. 351). In criminology, this technology aids fingerprint identification, facial recognition, and "forensic inquiries" (da Silva Torres & Falcao, 2006; Katira, Vora, Wali, & Medhekar, 2015; Lesch, 2017). Patterson (2016) observes how CBIR/RIL can be used to track intellectual property rights on orphan works. Katira, Vora, Wali, and Medhekar (2015) discuss the use of CBIR/RIL in determining trademark image registration and property ownership. Numerous authors, including Lesch (2017), Faruque, Antani, Long, Kim, and Thoma (2016), Junior, Oliveira, and Azevedo-Marques (2017), Oliveira and Kaster (2017), and Petrov (2015), consider the role that CBIR/RIL plays in making patient diagnoses. As one example, Oliveira and Kaster (2017) state, "this operation enables physicians to use knowledge from previous diagnoses in new cases based on imaging exams or to search for correlations among images based on visual findings in exams." (p. 1). Beyond diagnosis, CBIR/RIL has other uses such as "treatment planning," "assessing response to treatment," and "teaching and research" (Oliveira & Kaster, 2017, p. 2). Lesch (2017) notes how military forces use CBIR/RIL as landscape recognition and target identification. In another example, Katira et al. (2015) document how various military recognizance efforts use this technology, including "recognition of enemy aircraft from radar screens" and the "identification of targets from satellite photographs." (p. 9811).

Section 4.3 CBIR/RIL verification study

The CBIR/RIL verification study aims to confirm whether CBIR/RIL, explicitly Google Images, is still a viable tool that practitioners can use to understand who uses GLAM digitized images and for what purposes. The author used previously published analyses to recreate and replicate a study that revealed the possible benefits and limitations of CBIR/RIL as an assessment tool. These results contribute to the ongoing CBIR/RIL assessment conversation. Complete data sets are listed in the appendix #2.

Section 4.3.1 CBIR/RIL previous studies

The author relied on several previous studies to design the verification study. A substantial amount of research has been conducted on CBIR/RIL in the GLAM profession. For example, Nieuwenhuysen (2018), Reilly and Thompson (2017), Thompson and Reilly (2017), Wan and Liu (2008), Kousha, Thelwall, and Rezaie (2010), Terras and Kirton (2013), Kelly (2015), and Beaudoin (2016) explore how practitioners use CBIR/RIL to identify users and reuses of cultural heritage materials. These studies also examine images used for commercial purposes, copyright violations, research, artistic expression, education, and instruction, among others.

Several of these studies, including Chung and Yoon (2011), Reilly and Thompson (2016), and Górny and Mazurek (2012), developed categories and definitions of the user (Table 2) and reuse types (Table 3). Therefore, the author adopted Reilly and Thompson's categories and definitions for use in the verification study.

User Type	Definition
Cultural Heritage	Includes museums, archives, and historical societies
Educator, Higher Education	An institution or educator from an academic or higher education above twelfth grade
Educator, K-12	An institution or educator from kindergarten to 12 grade.
Industry	A vendor or entity that sells a product
Non-Profit	An entity that is not directly an educational institution, i.e., reference materials, .org
Personal	An individual using the image for personal reasons.

Table 2 - Categories and definitions of the user according to Reilly and Thompson (2016)

Reuse Type	Definition
Social Media	Image used for entertainment purposes in a social media platform, Facebook, Twitter, Pinterest, Flickr.
Popular Culture Research	The image was used for personal research purposes or personal fulfillment, i.e., a personal blog post but not a publication.
Commerce	Image used to sell a product
Exhibit	Image used for displaying items in an exhibition, including image galleries.
Instruction	Image used for teaching purposes
Scholarly Research	Image used for academic research purposes
Popular Culture Publication	Image used for online, non-scholarly publication
Scholarly Publication	Image used for online, scholarly publication, i.e., journal article, scholarly blog
Promotion and Marketing	Image used for advertising
Artwork	Image for decorative purposes
Other	Images that did not fit into any other category or use was not immediately apparent.

Table 3 - Type and Definition of Reuses according to Reilly and Thompson (2016)

Section 4.3.2 Research Design for CBIR/RIL verification study

Section 4.3.2.1 CBIR/RIL verification study evaluation criteria

To be a successful assessment tool for evaluating users and reuses, Google Images needs to:

1. Show a diversity of user types.
2. Show a diversity of reuse types.

The author maintains that diversity of user and reuse types are necessary because practitioners need tools capable of tracking the variety of audiences that consume digital library content. The author focused on Google Images because it is the prominent CBIR/RIL tool evaluated in the GLAM literature and its popularity with users. For example, from a 2010 Google event, Google Images experienced one billion page views per day and indexes over ten billion images. (Siegler, 2010)

Section 4.3.2.2 CBIR/RIL verification study digital library selection

The author chose images from three significant GLAM digital libraries for use in chapters four and five because all three offered a large number of images in the public domain and contained digital images with the most comprehensive metadata. The author provides additional details on these requirements in Section 4.3.2.3 Verification Study Image Selection Criteria.

One digital library is the Rijksmuseum (Rijksmuseum, n.d), the largest art museum in the Netherlands. Founded in 1798, it moved to its current location in 1808 and received a 375 million euros remodel of its space in 2013. It displays over 8000 objects spread over 80 rooms, and its total collection exceeds 1,000,000 items. The paintings, artifacts, and historical objects tell the story of 800 years of Dutch history. Over 2 million visitors annually marvel at the masterpieces by Rembrandt, Vermeer, and Hals that are on exhibit.

Another digital library is the Library of Congress's Digital Collections. For detailed information about LOCDC, see section 2.5.5.

Finally, the author used images from the New York Public Library (NYPL, 2020). NYPL, located in New York City, New York, is the largest public library in the United States based on the collection's size, as evidenced by the 2016 Institute of Museum and Library Services public libraries survey (Appendix #1). NYPL Digital Collections (NYPLDC, 2020), launched in 2005, is a growing collection of 895,780 prints, maps, manuscripts, photographs, and more (Hadro, 2015). The NYPLDC contains a tiny fraction of the physical library's overall holdings. The library claims that its digital collections are representative of the enormous diversity of its physical collections.

Section 4.3.2.3 CBIR/RIL verification study image selection criteria

The author developed selection criteria based on two conditions: distinctive image characteristics and image popularity.

CBIR/RIL, specifically Google Images, uses an image's attributes to conduct a search. Attribute-based searching makes it essential that the image has distinctive image characteristics and the prime motivation of image selection. According to Beaudoin (2016) and Reilly and Thompson (2017), CBIR/RIL searches cached images based on image attributes such as forms, blocks of color or shapes, and sketched or digitized images instead of traditional text-based searching. Reilly and Thompson (2016) noted that "one advantage of RIL is that users must not [need] to know keywords or phrases to describe the image being searched." (p. 57).

A second criterion is the popularity of an image. Popularly shared images are more likely to be found by the CBIR/RIL tools because integration into the collective database happens when individuals share images across online platforms. The author made efforts to obtain the highest number of results by choosing images that would be the most recognizable. Two examples are an image of a Jackie Robinson baseball card (Figure 4) and a digital surrogate of a painting of Napoleon

(Figure 5). This approach supports Patterson's (2016) assertion that "image searches work much less well on anything from a more obscure source or which has been cropped too much from its original state." (p. 9).

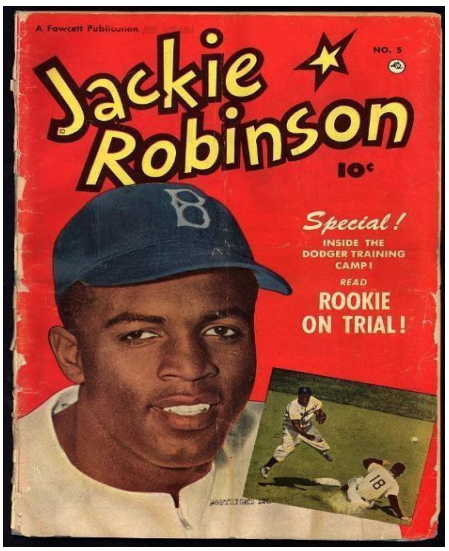

Figure 4 - Jackie Robinson baseball card[1]

Figure 5 - Painting of Napoleon[2]

[1] Jackie Robinson baseball card image. Accessed from https://www.loc.gov/resource/ppmsc.00133/

[2] Painting of Napoleon Accessed from https://www.europeana.eu/portal/en/record/90402/SK_C_1120.html?q=Napoleon+Bonaparte

The author collected images and associated metadata for this verification study and the verification study in chapter five at the same time. These data included the images themselves, a list of Google Images search result URLs, and a list of embedded metadata values per image. The author kept the number of images selected and the number of URLs collected for the verification studies intentionally small and United States GLAM institutions centric for two reasons. First, the author only uses images to verify the validity of previous U.S. studies' conclusions. Second, as stated by Kirton and Terras (2013); Kelly (2015); Kousha et al., (2010); Reilly and Thompson (2017); Terras and Kirton (2013), search results after initial 15 hits are often duplicates, broken links, or contain malicious content. After some trial and error, the author decided that ten images over three different digital libraries were optimal because they fit the highest number of criteria listed above. The ten images produce 100 total verification search results. See table 4 for the images selected and the digital library where they reside.

Image Title	Creator	URL	Digital Library/Collection
The City of New York		http://hdl.loc.gov/loc.gmd/g3804n.pm005950	Library of Congress Geography and Map Division Washington, D.C. 20540-4650 USA
Five generations on Smith's Plantation, Beaufort, South Carolina	O'Sullivan, Timothy H., 1840-1882, photographer	http://hdl.loc.gov/loc.pnp/ppmsc.00057	Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Destitute pea pickers in California. Mother of seven children. Age thirty-two. Nipomo, California	United States. Farm Security Administration (Sponsor) Lange, Dorothea (Photographer)	https://digitalcollections.nypl.org/items/639b2760-2289-0132-a9cd-58d385a7bbd0	New York Public Library, Digital Collections / "Farm Security Administration Photographs"
John Howell, an Indianapolis newsboy, makes $.75 some days. Begins at 6 a.m., Sundays. (Lives at 215 W. Michigan St.) Location: Indianapolis, Indiana.	Hine, Lewis Wickes, 1874-1940, photographer	http://hdl.loc.gov/loc.pnp/nclc.03225	Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Abraham Lincoln, 1809-1865.		https://digitalcollections.nypl.org/items/510d47e2-c226-a3d9-e040-e00a18064a99	New York Public Library, Digital Collections / Print Collection Portrait File
Front cover of Jackie Robinson comic book		https://www.loc.gov/resource/ppmsc.00133/	New York Public Library, Digital Collections / Print Collection Portrait File
Moulin-Rouge (La Goulue)	Toulouse-Lautrec, Henri de	https://www.rijksmuseum.nl/nl/collectie/RP-P-1954-57	Rijksmuseum - Europeana
The Windmill	Maris, Jacob	https://www.rijksmuseum.nl/nl/collectie/SK-A-2986	Rijksmuseum - Europeana
Are you helping? with salvage	Federal Art Project, sponsor	http:////hdl.loc.gov/loc.pnp/cph.3b49079	Library of Congress Prints and Photographs Division Washington, D.C. 20540 USA
Portrait of Emperor Napoleon I	Gérard, François-Pascal Simon baron	https://www.rijksmuseum.nl/nl/collectie/SK-C-1120	Rijksmuseum - Europeana

Table 4 - Images selected for CBIR/RIL verification study

Section 4.3.2.4 CBIR/RIL verification study process

The RIL verification study process consists of steps resulting in coded data that documented user and reuse types. Figure 6 illustrates the CBIR/RIL verification process. For complete data, refer to the spreadsheet in appendix #2.

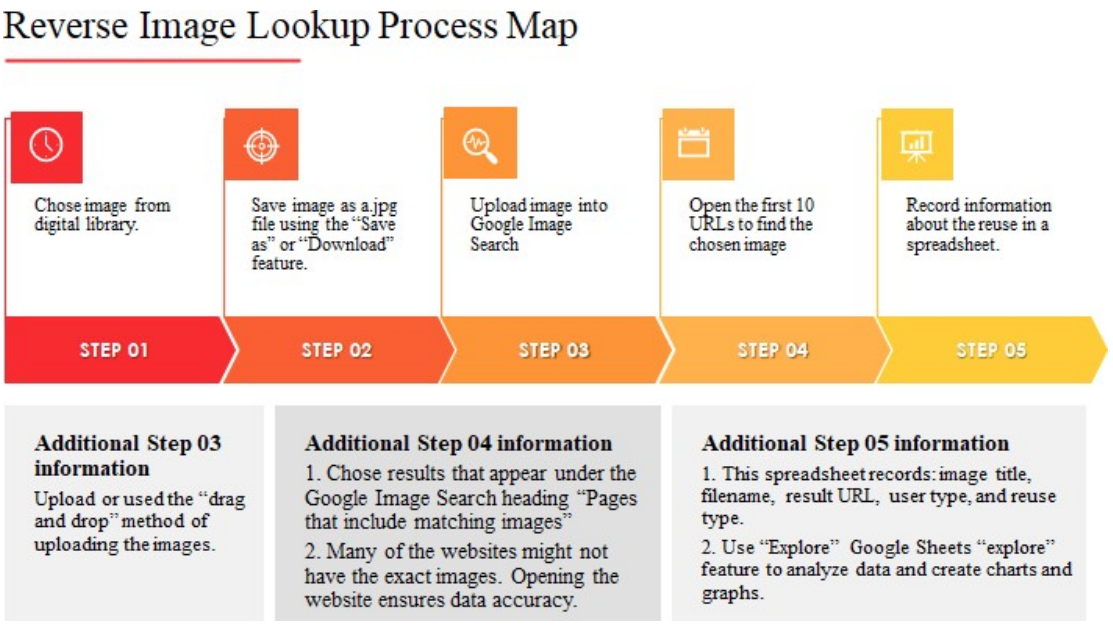

Figure 6 - CBIR/RIL verification process map

The author uploaded or used the "drag and drop" method of uploading the images defined in chapter 2 into the Google Images search engine to begin the verification study. Figure 7 is an example of a Google Images search result.

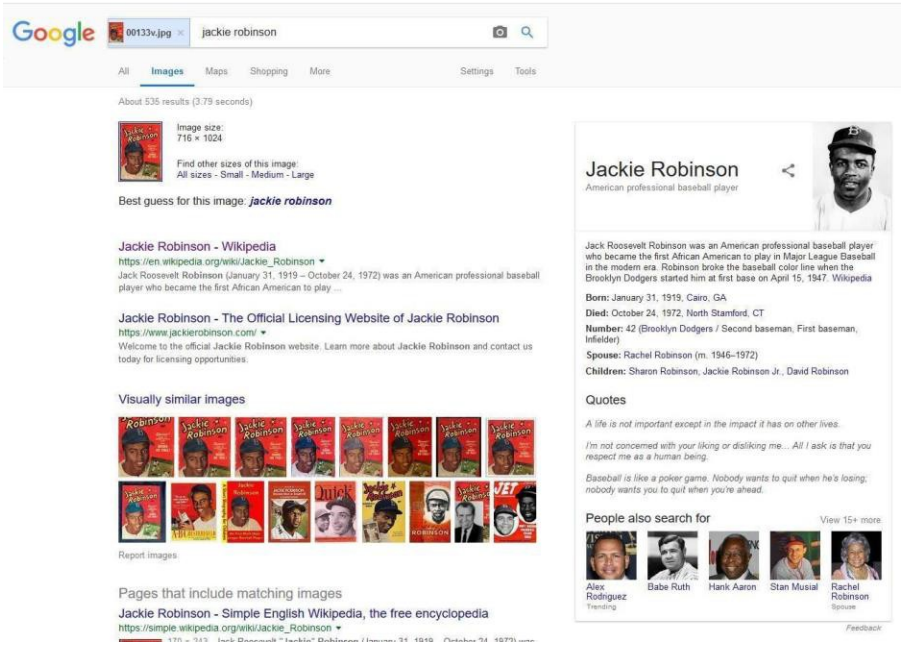

Figure 7 - Example of a Google Images search result

For each query, the author documented the first ten image results that appeared under the Google Images search results heading "Pages that include matching images" into a spreadsheet for coding and analysis. This spreadsheet records: image title, file name, result URL, user type, and reuse type. For each query, the author selected only one user type and reuse type. While compiling the results, the author established three exclusion criteria:

- Exclude any search result hidden behind a firewall that requires a username and password. Figure 8 is an example of a search result hidden behind a firewall.

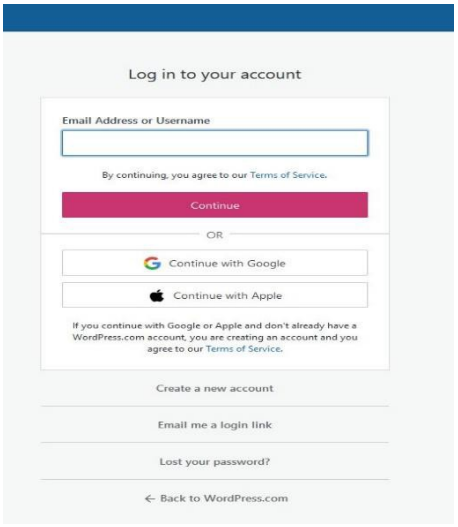

Figure 8 - Example of a search result hidden behind a firewall

- Exclude any search result locked behind pay-only access. Figure 9 demonstrates a pay-only access search result.

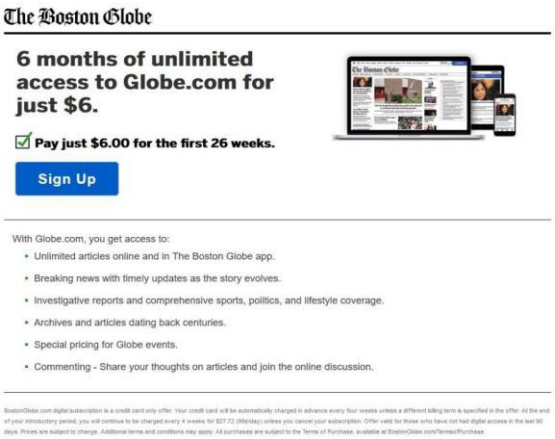

Figure 9 - example of a pay-only access search result

- Exclude any pages from the original content creator. The researcher does not consider the content creator webpages to be examples of reuse. See the definition of reuse in section 1.2. Figure 10 below is an instance of a content creator website.

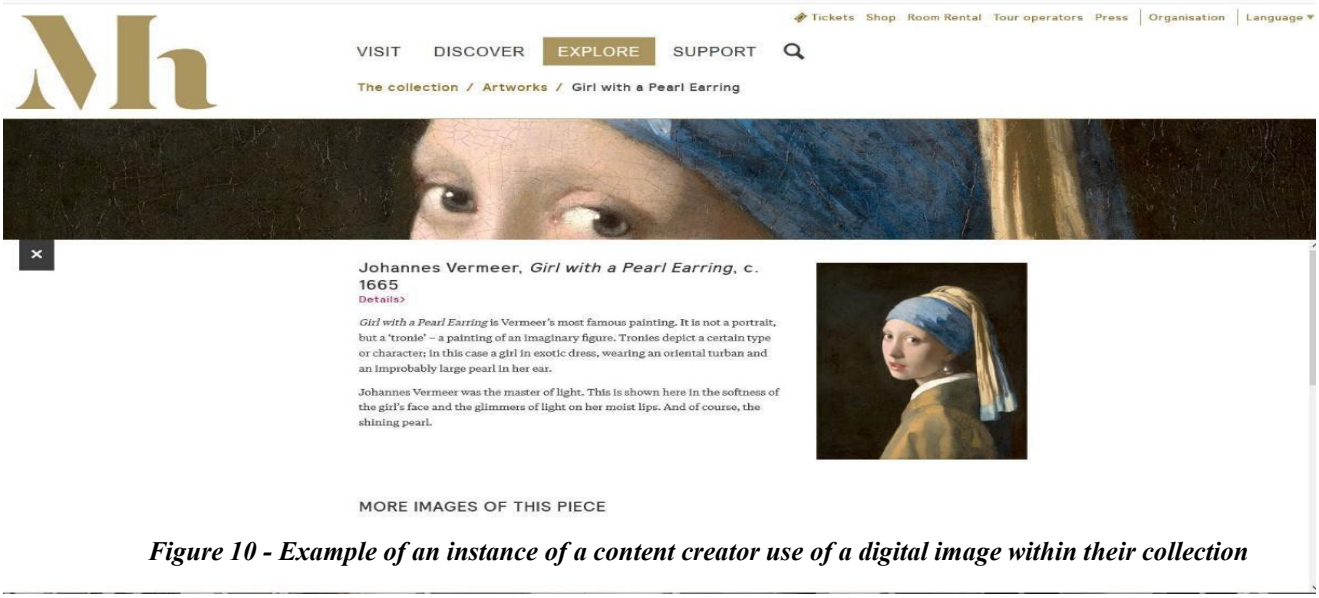

Figure 10 - Example of an instance of a content creator use of a digital image within their collection

Section 4.3.3 CBIR/RIL verification study User and Reuse type results

Section 4.3.3.1 User type results

After compiling the results for each image, the author analyzed the data using descriptive statistics. First, the author categorized individuals and groups who were reusing digital images on the web. These categories include "personal," "industry," "non-profit," and others described in Table 2: Type and Definition of Users according to Reilly and Thompson (2016). Figure 11 below shows the results for the User type results.

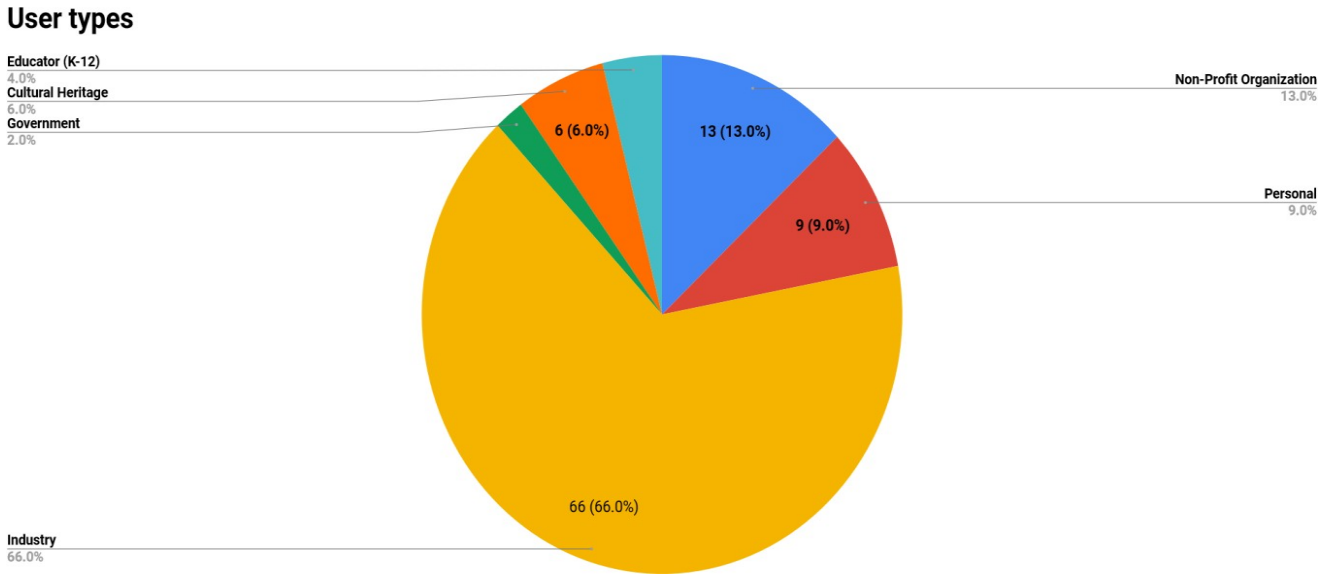

Figure 11 - Results of User types from CBIR/RIL verification study

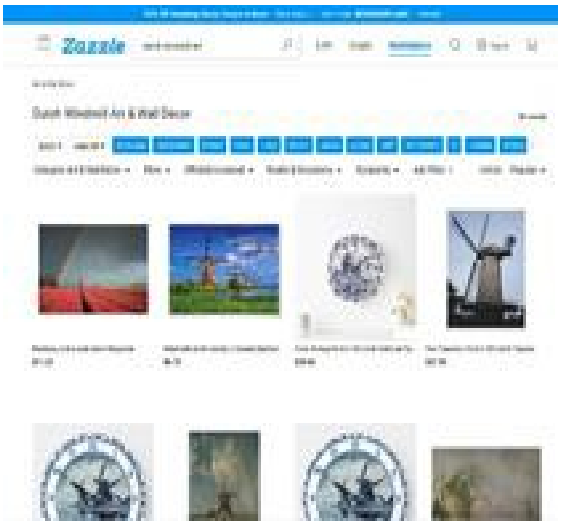

Figure 12 - Example of an Industry user search result

The majority of those reusing images over the web were in the "industry" category, with 66 URL results out of 100 total results. Figure 12 is an example of an industry user because the user type is selling a product. https://www.zazzle.com/

All other user types had significantly fewer results than the "industry" user type.

Figure 13 - Example of a non-profit organization user search result

The "non-profit organization" category resulted in 13 URLs. Figure 13 is an example of a Non-Profit Organization User because all Wikimedia/Wikipedia webpages are not-for-profit corporations.

https://commons.wikimedia.org/wiki/Main_Page

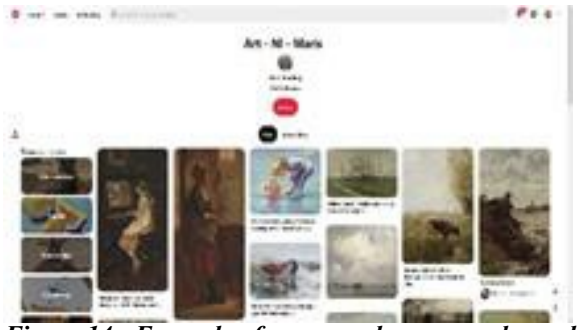

Figure 14 - Example of a personal user search result

The "personal" category had nine results. Figure 14 is an example of a "personal" user type because this is an individual user's pinboard on Pinterest.

https://ro.pinterest.com/marcbudding/art-nl-maris/

The "cultural heritage" category had six results. The "educator (K-12)" (kindergarten through 12th-grade level) had four results. The "government" category had two results. Finally, the "educator (higher education)" (higher education; college, university, masters, and doctoral institutions) category had zero results (and are not displayed in figure 15 due to zero results). Thus, this verification study's user type results yielded significantly different outcomes from other previous CBIR/RIL literature. The author will provide a possible explanation later in the chapter.

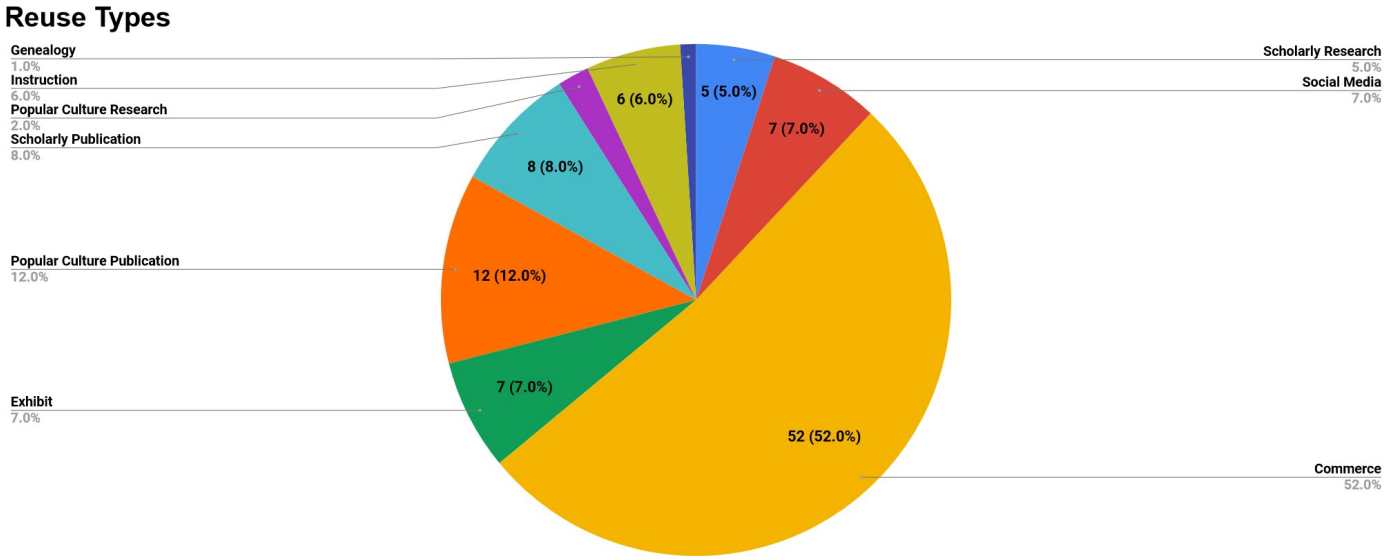

Figure 15 - Results of Reuse types from CBIR/RIL verification study

Section 4.3.3.2 Reuse type results

The author also categorized examples of digital images reuses on web pages. These categories included "commerce, "social media," "scholarly publications," and others described in Table 3: Type and Definition of Reuses according to Reilly and Thompson (2016) above.

The figure above shows the results for the reuse types that had at least one data point. The remaining reuse types results are listed below. Those types not displayed in figure 13 were due to zero results. They include "artwork," "video production," "promotion and marketing," "presentation," and

"other." The reuse type results from this verification study yielded significantly different outcomes from other previous RIL literature. The author will provide a possible explanation later in the chapter.

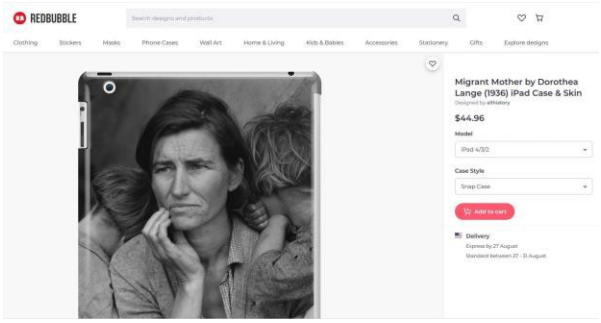

Figure 16 - Example of commerce reuse type search result

The majority of reuse types were for the "commerce" reuse type, with 52 results. Figure 16 is an example of a commerce reuse type. Image found at https://www.redbubble.com/i/ipad-case/Migrant-Mother-1936-by-PromoteProgress/34851341.MNKGF

Figure 17 - Example of popular culture publication type search result

Just as with the user types, all other reuse categories produced considerably fewer results. The "popular culture publication" category had 12 results. Figure 17 is an example of a popular culture publication reuse type search result. Image found at https://comicbookplus.com/?dlid=58299

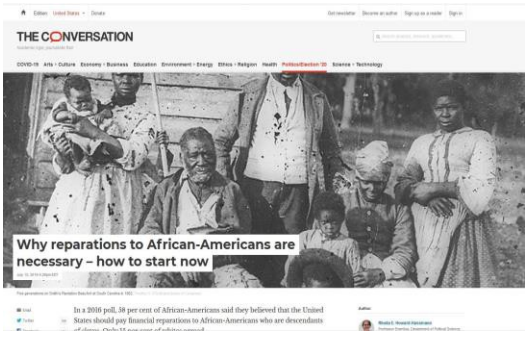

Figure 18 - Example of scholarly publication type search result

The "scholarly publication" category had eight results. Figure 18 is an example of a scholarly publication reuse type search result. Image found at https://theconversation.com/why-reparations-to-african-americans-are-necessary-how-to-start-now-119581

Section 4.4 CBIR/RIL verification study discussion

The author established two criteria to determine if the Google Images search engine was still viable for discovering digital library content users and assessing the reuse. The Google Images search should:

1. Show a diversity of user types
2. Show a diversity of reuse types

After analyzing the results, the author found that Google Images search technology, in its current incarnation, is particularly well suited for discovering "industry" user type and commerce reuse type. In part, the algorithm used by Google Images is responsible for these results because Google Images search algorithm has evolved based on user demand and, more recently, litigation against the company. In 2016, Getty Images filed a "competition law complaint" in the European Commission against Google. Getty claimed that Google Images was "creating captivating galleries of high-resolution, copyrighted content. Because image consumption is immediate, once an image is displayed in high-resolution, large format, there is little impetus to view the image on the original source site" (gettyimages, Company News, 2016, para. 3). This complaint stressed that in 2013 Google had changed Google Images search engine functions, limiting the searcher's ability to view a source file from its location. According to GettyImages (2016), "These changes have allowed Google to reinforce its role as the internet's dominant search engine, maintaining monopoly over site traffic, engagement data and advertising spend. This has also promoted piracy, resulting in widespread copyright infringement, turning users into accidental pirates." (para. 3). While these changes to Google's algorithm did not disable RIL capabilities, the author found that the latest updates to Google Images search's algorithm prioritized the "industry" user type and "commerce" reuse type websites over many other types of content, such as "personal" user type and "social media" reuse type sites.

These results directly contradict the observations of previous CBIR/RIL studies using Google Images search. Reilly and Thompson (2016) found that the "personal" user type and "social media" reuse type were the most prominent. The "personal" user type, at 51% of the total results in the 2016

study, fell dramatically in the verification study, comprising nine percent of the total. However, the "industry" user type, at almost 26% of the total results in their 2016 study, more than doubled in the verification study after the change in Google's algorithm. The trend in reuse type results aligns with the shift in prominent user type. In their 2016 study, Reilly and Thompson found that the "social media" reuse type had the highest results at 24%, while the "commerce" reuse type accounted for nearly 14% of the results. Furthermore, the author found that the social media reuse type fell to 6% in the verification study, while the "commerce" reuse type increased to 52% of total results.

Other previous RIL studies also indicate a shift in the Google Images search results. For example, Kelly (2015) found that university scholars and various news media outlets were the most popular users of images from the Loyola University New Orleans Digital Library. She also documented that the prominent reuses of images included instruction, news websites, and historical scholarly articles. For this verification study, the author maps these users as "educator (higher education)" and "industry" user types and reuse types as "instruction," "popular culture publication," and "scholarly publication" reuse types. Except for the "industry," all other users and reuse types results are not supported by the verification study.

Kousha, Thelwall, and Rezaie (2010) focused only on reuse, wrote that "informal scholarly communication," "background and layout," and "navigational" were the top three reuse types in their study. They defined "informal scholarly communication" as "education-related documents, scientific news stories, and messages posted to discussion boards and forums" (p. 1738). They described "background and layout" as "websites offering free downloads of templates and backgrounds" and "some from websites selling high-quality posters," and "user profile images in social networking websites (e.g. facebook)." (p. 1739). Finally, they classified "navigational" as "subject-specific online galleries (e.g., www.space.com) and general galleries (e.g. flickr.com)" that offer "no explicit evidence of direct or even indirect scholarly use, although they seem to perform a useful function for scholars and the interested public." (p.1739). For the verification study in this dissertation, the author maps these reuse examples to a variety of reuse types, including: "scholarly publication," "popular culture publication," "social media," "commerce," and "exhibit." While some Google Images search

queries returned numerous web pages, the author found that the changes to the Google Images search functionality diminished the variety of other user and reuse types.

Prior GLAM RIL studies, including Kousha et al. (2010), Kelly (2015), and Reilly and Thompson (2016), revealed that RIL technology was identifying a small but diverse set of users and reuses. However, based on the change in the algorithm used by Google Images and the results of this verification study, the author contends that the limitations of Google Images nearly prevent GLAM professionals from determining who is reusing digital library images. Therefore, the author found that Google Images is most appropriate for GLAM practitioners when trying to identify commercial reuses of images.

Chapter 5: Embedded Metadata

This chapter will discuss embedded metadata and use a desktop application to determine if embedded metadata is an alternative or additional way to track digital library images' reuse. It defines embedded metadata and provides examples of several embedded metadata specifications, schema, and tags. The chapter also explores the GLAM and private industries' use of embedded metadata. Finally, it shares the chapter's verification study, including the methodology and results, to evaluate the effectiveness of embedded metadata as a method to trace reuse.

The GLAM industry's accepted definition of metadata is data or information that describes other data or information. More simply put, metadata is data about data. Metadata is the underpinnings that allow software and web pages to render images and other types of online content. As Jenn Riley (2017) writes, "The core features of most software packages we use every day are metadata-driven." (p. 2). There are various types of metadata used by digital library practitioners to describe and administer digital content, including descriptive metadata, structural metadata, and administrative metadata.

Information professionals categorize and define these types of metadata in various ways. For example, Anne Gilliland (2016) defines descriptive metadata as data "used to identify, authenticate, and describe collections and related trusted information resources." (Table 2. *Different Categories of Metadata and Their Functions*, para 20). Examples include information found in catalog records or archival finding aids such as object title, creator, and description. Riley (2017) defines structural metadata as information that indicates "relationships of parts of resources to one another" (p. 6). Examples include data element names and an object's components and the relationship among those components, such as chapters in a digitized book. Riley (2017) also defines administrative metadata as data that supports the "long-term management of files" and places information used "for decoding and rendering files" and for "intellectual property rights attached to content" within this metadata type (pp. 6-7). Examples include file object type, size, and creation date. These types and examples are

illustrative and not intended to be a comprehensive commentary on the history, creation, and implementation of metadata.

Section 5.1 Embedded metadata definition, specifications, and applications

Section 5.1.1 Embedded Metadata Definition

Embedded metadata is information about an object that is "attached" to the object itself. As the Federal Agencies Digital Guidelines Initiative (FADGI)(2018) states, "embedded metadata is a component of a digital file that exists alongside the content (usually binary data) within the file, making the digital file self-describing" (para. 1). Riley (2017) states that "virtually all file format specifications include a metadata area" which can include "technical metadata about the file, used for decoding and rendering" and can "provide for the recording of descriptive metadata." (p.15). Scholars note that the creation and adoption of the Tagged Image File Format (TIFF) in the mid-1980s gave rise to the use of embedded metadata because the TIFF format automatically recorded information from the capture devices. Embedding the metadata decreases the opportunity of the object file and metadata file from being separated (Reser & Bauman, 2009; Lilly, 1995; Riley, 2017). Current embedded metadata typically contains administrative and technical information automatically generated by the capture device and photo editing software (Riggs, Douglas & Gagneja, 2018). As Riley (2017) writes, "there is also a history of software using system-level information to add extra administrative information to digital files, such as date created or the ID of the user logged into the system at the time the file was generated" (p. 37).

Section 5.1.2 Embedded metadata specifications, schema, and tags

There are varying types of embedded metadata specifications, schema, and tags, often generated by the capture devices and editing software used to process, view, and manipulate images.

The Exchangeable Information File Format (Exif) "is a standard used by camera manufacturers to store camera-created information in the file... [it] describes the characteristics of the image data itself so programs can know how to open the file properly" (Krough, 2018, para. 7). Riley (2017) writes that Exif "is not a file format, but rather a tag structure for embedded metadata within digital image files." (p. 26). According to Reser and Beauman (2009), Exif "has become a default standard and is found on just about every imaging device." (para. 9). It documents camera settings such as aperture and shutter speed, date and time taken, ISO setting, lens information, and information like the white balance selected for the image (Krough, 2018; Riley, 2017; Reser & Beauman, 2009).

Field	Value
ExifTool Version Number	10.64
---- File ---	
File Name	unnamed.jpg
Directory	.
File Size	1090 kB
File Modification Date/Time	2018:08:18 11:45:16-05:00
File Access Date/Time	2018:08:18 11:45:16-05:00
File Creation Date/Time	2018:08:18 11:45:16-05:00
File Permissions	rw-rw-rw-
File Type	JPEG
File Type Extension	jpg
MIME Type	image/jpeg

Table 5 - Example Exif data

Tools for extracting embedded metadata use composite tags when specifications, schema, and tags do not have fields to store extracted information. As Harvey (2016) writes, "the values of the composite tags are **Derived From** the values of other tags." (para. 1). Composite tag examples include "Camera: Lens," "Aperture," "Field of View," "35mm Focal Length," and "Shutter Speed" (Harvey. 2016; Krough, 2015).

Field	Value
Image Size	1875x2500
Megapixels	4.7
Y Resolution	300
Planar Configuration	
Resolution Unit	inches
Software	Adobe Photoshop CS3 Windows
Modify Date	2010:01:05 20:42:16
Artist	
Exif Image Width	2165
Exif Image Height	3000

Table 6 - Example of Composite data

The International Press Telecommunications Council (IPTC) (2021a) metadata consists of IPTC Core and IPTC Extension properties, which are:

> "Comprehensive sets of fields that allow users to add precise and reliable data about people, locations, and products shown in an image. It also supports dates, names and identifiers regarding the creation of the photo, and a flexible way to express rights information." (para. 4).

Martijn Kleppe (2015) states that IPTC enables the efficient management and sharing of files. Kleppe (2015) writes, "the exchange of files between journalists, but also academics, digital libraries and

cultural heritage institutions can use IPTC to include information about their objects in the digital files" (p. 1).

Field	Value
Coded Character Set	
Coded Character Set	
Application Record Version	0
Caption-Abstract	opname/scan 2002
Category	gc
Keywords	voorkeursbeeld
Copyright Notice	Rijksmuseum Amsterdam.P.O. Box 74888 .1070 DN Amsterdam.The Netherlands.+31 206747000

Table 7 - Example IPTC data

The International Color Consortium (ICC) (n.d.) profiles describe the color transmission and reproduction from one device to another. An example would be a photo initially captured by a camera or phone and transmitted and rendered on a computer monitor. (FAQ, What is an ICC profile.).

ICC_Profile	
Profile CMM Type	ADBE
Profile Version	2.4.0
Profile Class	Display Device Profile
Color Space Data	RGB
Profile Connection Space	XYZ
Profile Date Time	2007:03:02 10:07:41
Profile File Signature	acsp
Primary Platform	Unknown ()
CMM Flags	Not Embedded, Independent

Table 8 - Example ICC data

The Joint Photographic Experts Group (JPEG) File Interchange Format (JFIF) is a "light-weight file format" (Hamilton. 1992. p. 1) that allows the transfer of compressed images from one platform or device to another. JFIF data documents a file's version and resolution information.

JFIF	
JFIF Version	1.01
Resolution Unit	inches
X Resolution	300
Y Resolution	300

Table 9 - Example JFIF data

Adobe's Extensible Metadata Platform (XMP) is an open-source, editable, "file labeling technology" (Adobe, 2020, para. 2) that content creators can use to embed metadata into the file during creation (Christensen & Dunlop, 2011; Chastain, 2018). Example XMP data includes image and resolution properties and other descriptive information.

Field	Value
XMP Toolkit	Adobe XMP Core 4.2.2-c063 53.352624, 2008/07/30-18:05:41
Image Width	3395
Image Height	4463
Compression	Uncompressed
Photometric Interpretation	RGB
Samples Per Pixel	3
Planar Configuration	Chunky
X Resolution	350
Y Resolution	350
Resolution Unit	inches
Orientation	Horizontal (normal)

Table 10 - Example XMP data

Section 5.1.3 Embedded metadata application: ExifTool GUI

To extract, edit, or analyze embedded metadata, practitioners need specialized software. One of the most popular platforms is ExifTool (https://exiftool.org/install.html). In its command-line form, the tool "can be used to view and/or edit not only all tags for all known image formats but also for a wide variety of other file types including executables" (Harran, Farrelly & Curran, 2018, para. 24). While not as powerful as the command line version, the tool also has a graphical user interface (GUI) desktop application, which provides a more intuitive, user-friendly interface to read and modify embedded metadata (Toevs, 2015).

Beyond ExifTool, other software exists to view and manipulate embedded metadata with limited features and functionality, including Exif Pilot[3] and Photome[4].

Section 5.2 GLAM and private industries that use embedded metadata

Some cultural heritage sectors, like archives, have embraced embedded metadata as part of the born-digital collecting process. As Creighton Barrett (2017) writes, "archives are adopting digital forensics techniques to support acquisition, accessioning, appraisal, preservation, and access" (p. 1). An emerging area for the GLAM profession is utilizing embedded metadata to trace the reuse of images over social media (Thompson & Reilly, 2018). Additionally, information professionals and other private industries look to embedded metadata to verify the credibility or authenticity of images and multimedia (Boididou, Papadopoulos, Apostolidis & Kompatsiaris, 2017; Harran, Farrelly & Curran, 2018). Beyond the cultural heritage sector, other private industries have also deployed embedded metadata. For example, the IPTC (2021b) embedded metadata manifesto lists numerous use cases,

[3] available from www.colorpilot.com/exif.html
[4] available from www.photome.de/

such as audio recording software, university slide archives, and photo publishing websites. Because of its ubiquitous nature, embedded metadata even plays a role in everyday multimedia users' lives. As Reser and Bauman (2009) note:

> "One of the most prevalent examples is the metadata embedded in audio files. This data allows you to view your music files in a program like iTunes by artist, album, and song title, and easily transfer them together with all of their associated metadata to your iPod where you are able to access your files just as you do on your computer." (para. 1).

Embedded metadata is not just for keeping track of digital file details. For example, web pages rely on embedded metadata to link from one page to another, record user behavior, and index search results. Riley (2017) describes Google's Knowledge Graph project as an example of a web page relying on embedded metadata. She writes that:

> "The Knowledge Graph and other structured metadata stored by Google are used to enhance search results and provide other value-added features such as sports scores, integration of search results with maps, and the knowledge cards that appear on the search results screen providing details on notable people and places." (p. 2).

Section 5.3 Embedded metadata verification study

The goal of the embedded metadata verification study utilizing ExifTool is to confirm whether embedded metadata can discover both the reuse of digital images and the user behavior that demonstrates reuse. The author modeled the verification study from a previously published paper by Thompson and Reilly (2018). The recreation and replication conducted revealed the limits of embedded metadata as an assessment tool for reuse. These results contribute to the ongoing, embedded metadata assessment conversation. Complete data sets are list in the appendix #3.

Section 5.3.1 Embedded metadata previous studies

There is limited research on using tools to identify and examine embedded metadata found in online objects (Toevs, 2015; Thompson & Reilly, 2018). Toevs (2015) investigated "how the Exif toolset can be integrated into systems for scraping websites and reporting on specific tags that are contained within the images extracted from the website." (p. 26). Other studies focus on the use of embedded metadata in cultural heritage objects. Saleh (2018) sought to discover if GLAM institutions were embedding metadata into digital objects and whether embedded metadata could link users back to the object's originating institution. He found that most images surveyed did not contain embedded metadata elements and a minimal number of images contained ownership and rights metadata that would link back to the host institution. Thompson and Reilly (2018) also conducted a similar study, focused on "how or if embedded metadata followed the digital object as it was shared on social media platforms" in order to "clarify which embedded metadata fields, if any, migrated with the object as it was shared across social media." (p. 223). They found results comparable to Salah and documented that nearly all social media platforms stripped most existing embedded metadata.

Section 5.3.2. Research design for embedded metadata verification study

Section 5.3.2.1 Embedded metadata verification study evaluation criteria

The aim of the embedded metadata for reused online images verification study is to demonstrate that:

1. provenance fields can link back to the original object
2. other specific embedded metadata elements can track reuse

Section 5.3.2.2 Embedded metadata verification study digital libraries and image selection

The author used the same digital libraries and images generated for the CBIR/RIL verification study in chapter three. The author developed selection criteria based on embedded metadata robustness.

Virtually all capture devices, such as cameras, scanners, generate embedded technical metadata at the point of inception. The devices parse these metadata into approximately nine types: File, Exif, IPTC, XMP, ICC Profile, APP14, Composite, JFIF, and MakerNote. The author manually checked each image upon selection to ensure that embedded technical metadata was present.

The three digital libraries used in this verification study, Rijksmuseum (Rijksmuseum, n.d), New York Public Library (New York Public Library, 2020), and the Library of Congress (Library of Congress, n.d.), stood out as having the most thorough and inclusive metadata. Each uses a metadata schema to ensure discoverability and reuse. The Rijksmuseum provides "digital images of objects from the collection," "descriptive object information (metadata)," and "bibliographic data ...without restrictions on reuse." (Rijksmuseum. (n.d.). Open Data Policy page, para. 1). The New York Public Library uses Metadata Object Description Schema (MODS). Developed in 2002 by The Library of Congress' Network Development and MARC Standards Office, with interested experts, an XML schema was created to carry forward some selected MARC21 records onto digital objects.

> "MODS is intended to complement other metadata formats. For some applications, particularly those that have used MARC records, there will be advantages over other metadata schemes. Some advantages are: the element set is richer than Dublin Core, the element set is more compatible with library data, the schema is more end user oriented than the full MARCXML schema, and the element set is simpler than the full MARC format." (LC, 2016, para 3).

The Library of Congress uses "Metadata for Digital Content" (MDC), a schema developed by "an institution-wide working group." (LC, 2011, para. 3).

Two additional criteria, whether an image was in the public domain and was downloadable, affected the author's ability to access the file. The author chose public domain images because they could be downloaded using the "download" feature offered by the digital library or the "save as" function built into the user's preferred operating system. These two approaches retain the original embedded metadata. The author was careful not to employ photo editing software such as Photoshop

or Microsoft Image Viewer for this process as photo editing software will overwrite or alter some embedded metadata values. According to Thompson and Reilly (2018) and Smith, Saunders, and Kejser (2014), third-party image editing or viewing software will automatically overwrite the embedded technical metadata upon upload.

The author used the ExifTool GUI at the original image capture to extract the original embedded metadata. This process created the baseline metadata that the author used to compare all subsequent metadata extractions. As with all data collected, the author recorded metadata on a spreadsheet.

Section 5.3.2.3 Embedded metadata verification study process

For each URL, the author performed the following functions: identifying the corresponding image on each of the ten result pages and saving each image to a computer. During this process, the author used the "Save as" feature if available since importing or downloading directly into image editing software such as Adobe Photoshop alters some embedded metadata values. First, the author downloaded images to a folder directory. When downloading images, there were several instances when it was necessary to create an additional folder within the directory to avoid duplicate file naming and overwriting previously saved images. Next, the author employed the ExifTool GUI on each original image to document baseline values and recorded the output to a spreadsheet. Finally, the author repeated this process for the ten instances of reuse across all ten images.

To analyze the data, the author focused on the results related to provenance (file name, copyright notice, and web statement), rights management (copyright notice and web statement), and human editable metadata (keywords) fields. The author chose these fields because they are populated using natural language values that creators and other users could edit. Chastain (2018) notes that,

> "There are other types of metadata that allow users to add their own descriptive information within a digital photo or image file. This metadata might include the characteristics of the

photo, copyright information, a caption, credits, keywords, creation date and location, source information, or special instructions." (para. 3).

These fields contrast to the majority of the embedded metadata fields, which contain machine-generated values or file format information.

Section 5.3.3 Embedded metadata verification study results

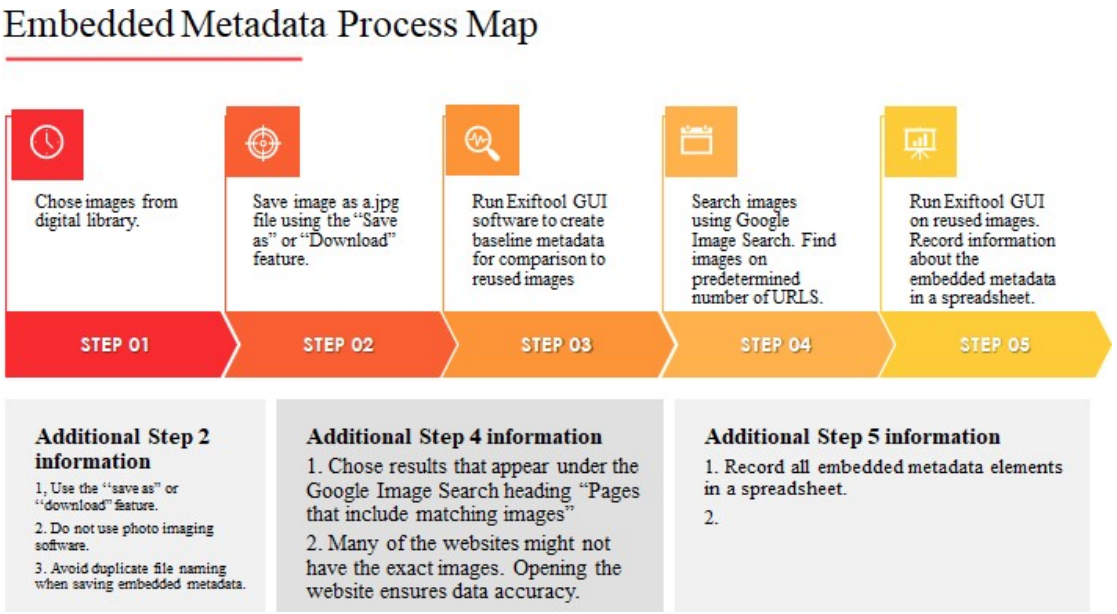

Figure 19 - Embedded Metadata verification study process map

The analysis results exhibited that those websites reusing digital images were more likely to retain human-readable elements of the file name. Users also tended to incorporate a combination of the title, creator, and/or date into their new file names. The data relating to copyright was not as promising, as few users supplied rights information or retained rights information from their source. Likewise, the finding revealed that the web statement and keyword fields were also not retained from the source or supplied when reused. Table 11 provides results for how frequently reused files retained the file name, copyright notice, web statement, or keyword from the original file. The table indicates

the number of times (out of a possible ten) that each reused image's embedded metadata matched original file values.

Image file name	File name	Copyright Notice	Web Statement	Keyword
Unnamed	Zero out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value
Windmill	Zero out of ten used the original file name	One out of ten transferred the copyright statement from the original object	One out of ten transferred the web statement from the original object	One out of ten transferred the keywords from the original object
90402	Zero out of ten used the original file name	Zero out of ten transferred the copyright statement from the original object. Note: there was only one result that did not match the original object	Zero out of ten transferred the web statement from the original object. Note: there was only one result that did not match the original object	Zero out of ten transferred the keywords from the original object
00133v	Zero out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value
The_City_Of_New_York	Zero out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value

Image file name	File name	Copyright Notice	Web Statement	Keyword
00057_150px	Zero out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value
03225_150px	Zero out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value
3b49079r	One out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value
Nypl.digitalcollections.639b2760-2289-0132-a9cd-58d385a7bbd0.001.w	Zero out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value
Nypl.digitalcollections.510d47e2-c226-a3d9-e040-e00a18064a99.001.w	Zero out of ten used the original file name	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value	Not applicable because the original image did not contain this value

Table 11 - Embedded metadata verification study analysis

Section 5.4 Discussion

To be successful, the assessment method for tracking embedded metadata for reused online images, embedded metadata values need to demonstrate that:

1. There are specific embedded metadata provenance fields that can link back to the original object
2. Other embedded metadata elements can track reuse

Given the lack of traceable results across the various reuses, embedded metadata does not appear as a viable method for tracking reuse in its current form. These results resemble previous studies, particularly the analysis of images uploaded to social media. Many social media sites such as Twitter, Facebook, and Pinterest "strip metadata from the images that are uploaded to their websites." (Toevs, 2015, p. 27). Additionally, social media sites may modify embedded metadata upon upload or download (Bushey, 2015; Corrado & Jaffe, 2017; Thompson & Reilly, 2018; Saleh, 2018). The results of the IPTC Photo Metadata Working Group's social media photo metadata test (2019) outlines embedded metadata removal practices by various social media sites. Several authors write that social media sites remove this information to prevent cyberbullying, cyberstalking, and other malicious acts made possible by using geographic information found in embedded metadata (Toevs, 2015; Riggs, Douglas, & Gagneja, 2018; Fazal, Nguyen, & Fränti, 2019; Riley, 2017).

Users have the option of keeping the original file name or renaming it with a value of their choice when downloading and uploading to websites. However, when analyzing embedded metadata from images downloaded from all sites, including social media, the author observed an interesting pattern in users' file naming behavior. File names from these shared files typically included some combination of the work's title, creator, or date. Table 12 demonstrates the pattern where the user chooses to apply some variation of the image's title, creator, or date in their file naming scheme.

Image title	**Number of User file name that uses some variation of the image's title, creator, or date (# of file names out of ten results)**
"Moulin-Rouge (La Goulue)"	Seven out of ten
"The Windmill"	Six out of ten
"Portrait of Emperor Napoleon I"	Two out of ten
"Front Cover of Jackie Robinson Comic Book"	Eight out of ten
"The city of New York."	Six out of ten
"Five generations on Smith's Plantation, Beaufort, South Carolina."	Five out of ten
"John Howell, an Indianapolis newsboy, makes $.75 some days. Begins at 6 a.m., Sundays. (Lives at 215 W. Michigan St.) Location: Indianapolis, Indiana."	Five out of ten
"Are you helping? with salvage"	Eight out of ten
"Destitute pea pickers in California. Mother of seven children. Age thirty-two. Nipomo, California"	Five out of ten
"Abraham Lincoln, 1809-1865."	Nine out of ten

Table 12 – File name that uses some variation of the image's title, creator, or date.

The author contends that these renaming patterns offer practitioners the prospect of tracking reuse discoverability and insights into user behavior. Even with the limitation of social media file naming alteration, the file naming pattern results were significant enough to warrant additional study. The following chapter will test and verify this discovery.

Chapter 6: File Naming

Chapter six makes the connections between digital image reuse, descriptive human-readable file naming, practitioner file naming customs, and reuse discovery. The author's analysis of the relationship between digital image reuse and file naming is significant because no current literature addresses this correlation.

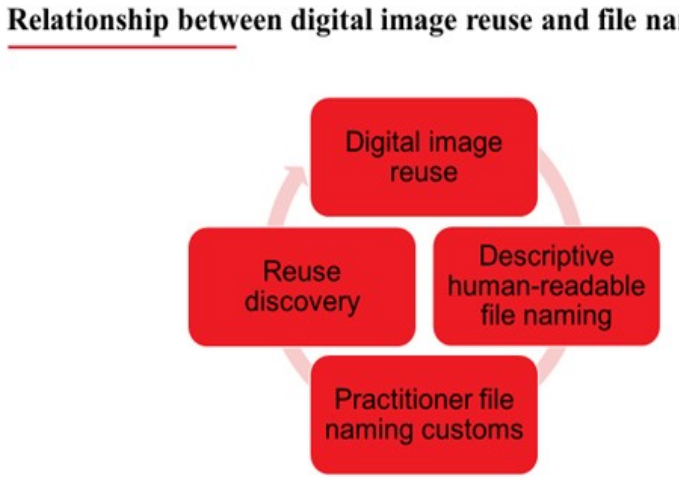

Figure 20 - Relationship between digital image reuse and file naming

The author explores file naming practices and investigates the impact that user renamed files have on reuse discoverability to draw these connections. To do this, the author developed a file naming assessment investigation that posed the following questions:

1. What is the user-generated file naming behavior? For example, does it match the combination of title, creator, or date based on the embedded metadata verification study outcome in chapter five?

2. Does user-generated file naming behavior yield more extensive search results than the original file name?

Chapter six provides an overview of documented file naming practices of both GLAM institutions and digital image users. It also outlines the role of human-readable file naming, containing a combination of the title, creator, and/or date, in producing better reuse search results. Additionally, the chapter addresses the role that file naming plays in search engine optimization. Finally, it discusses file naming guideline implications for practitioners.

Section 6.1 File naming conventions and practices

People have been file naming just as long as they have been organizing information. According to Parker-Wood, Long, Miller, Rigaux, and Isaacson (2014), file names perform two functions: "First, they serve to uniquely identify a file over time. Second, they serve to assist our memory, describing the contents of a file, and helping us to find it or recognize it when we look at it later" (para. 5). As witnessed in chapter five, many users employ descriptive file names that include the title, creator, and/or date as a memory or finding device. They are giving these images user-friendly, meaningful, and consistent names.

User-friendly and meaningful file naming requires the creator to represent files using the specific vernacular of the user group in which the files are relevant. Soules and Ganger (2005) eloquently express this concept when they wrote, "context is one of the key ways that users remember and locate data" (p. 120). Furthermore, Nardi, Anderson, and Erickson (1995) articulate the need for user-friendly meaningful file naming. They state:

> "Two things about naming are important. The first is name recognition. Although the file name might not be recalled verbatim, when searching for the file the name would be familiar when seen. The names of files are mini-indexes into the file. The other advantage of descriptive names is that when using Find, users could often remember at least part of the file name. Having part of the name enabled users to either hit the right file or find a file similar to or nearby the file they were trying to find."(para. 26).

Section 6.1.1 Descriptive file naming

Rogers (2014) states that in the cultural heritage sector, "there are generally two schools of thought on file naming; opaque and descriptive." (p.1). One file naming convention informs practitioners that they should include all of the "necessary" descriptive information regardless of where it is stored, such as "significant keywords that will be of use in future searches." (Council of Europe. Directorate General of Administration, 2017, p 2). An example of a descriptive file name would be 'wwI_poster_owens_0001.tif' because a person could use parts of the file name to search and find the file. This example aligns with another practice, the avoidance of "using initials, abbreviations and codes that are not commonly understood" (University of Edinburgh's Records Management division 2018, para. 2), which can obscure the meaning of a file name and make it more difficult to find when searching.

Section 6.1.2 Format file naming

Guidelines also advise practitioners on the file name's structure, format, and composition. Some call attention to the file name's length, stressing that shorter file names create long URL paths. The University of Edinburgh's (2018) Records Management division's website states, "File names should be kept as short as possible whilst also being meaningful." (para. 1). Several guidelines suggest that file names should be approximately 25 characters or fewer **(**North Carolina Department of Cultural Resources, 2008; Rehberger & Coates, 2019; Lloyd, 2014; Murray, 2017; CESSDA Training Team 2017 - 2020; Princeton University Library 2019; U.S. National Archives and Records Administration 2004). The University of Edinburgh's (2018) Records Management division further explains the importance of short file names by stating, "long file names mean long file paths and long URLs which increase the likelihood of error, are more difficult to remember and recognise, and are more difficult to transmit in emails as they often 'break'" (para. 1). Keeping this statement in mind, a less than optimal file name would be metapth743602_xl_TMFM_01-010_01.tif (this is an actual file name used

in the file naming assessment investigation later in the chapter) because it is too long and is not descriptive of the file contents.

Other guidelines recommendations include:

- Avoid "weird" symbols or special charactures that computers use for other things (including: "/ \ : * ? " < > [ ] & $)
- Ensure the first character of the file name should be an ASCII letter ('a' through 'z')
- Avoid spaces and use underscores (oral_history) or camel case (OralHistory) initial capital letters of words with no spaces
- It is preferable to use all lower case unless using camel case.

(North Carolina Department of Cultural Resources, 2008; Rehberger & Coates, 2019; Lloyd, 2014; Murray, 2017; CESSDA Training Team 2017 - 2020; Princeton University Library 2019; U.S. National Archives and Records Administration 2004; The World Digital Library, n.d.,b).

Section 6.1.3 Consistent file naming

Guidelines also stress the need for consistency in file naming. As the guidelines of the North Carolina Department of Cultural Resources (2008) maintain, "good file names are essential to accessibility" (p. 1). This consistency is critical for several reasons. Multiple people will be accessing files from multiple locations. Using consistent language, schema, and version control are essential (North Carolina Department of Cultural Resources, 2008; York University, n.d.). Additionally, Rogers (2014) writes that meaningful files names form a "safety net" that can provide "an easy way to ensure that digital images are accounted for within an institution's digital collection." (p.1).

Section 6.1.4 File naming behavior

Researchers have studied file naming behaviors among various user groups. Henderson (2005) observed several file naming activities among knowledge workers. She noted that "folder and file naming is a personal and idiosyncratic affair, and some names may only have meaning to their creator in a particular context" (p. 77). Dinneen and Julien (2019) summarized that users employ creativity in file naming and developed identifiable patterns. These patterns, based on the context in which they represent project titles, creation date, and special characters used to sort files in a particular sequence. Examples of file naming by context, exemplified when oral history files are named using a combination of the following (a) the interviewees, (b) the interviewer, and (c) the interview date (Barreau, 1995). According to Barreau (1995), "A typical naming convention used was to attach meaningful suffixes or file extensions to make filenames more distinguishable" (p. 335). Henderson (2005) notes that folder and file names most frequently "represent the genre, task, topic or time dimension of the documents they contained" (p. 75). Hicks, Dong, Palmer, and McAlpine (2008) examined the file naming strategies of engineers. They established that "75% of engineers use the title of the document, 60% also use the purpose or function …, 50% the project title…, and 45% the date" (p 16). They also noticed that engineers added a "memorable nickname to make the content easily recognizable" (p. 16). Barreau (1995) also emphasizes the importance of users generating a meaningful file name in a different investigation. According to one of her subjects, "he tried to create meaningful labels in naming files so that they would trigger his memory when he went to look for something" (p. 335).

Section 6.1.5 Consequences of non-meaningful file names

There are several consequences of not producing meaningful file names. Crowder, Marion, and Reilly (2015) note how relying on machine-generated file names often leaves users with "default names that can be uninformative and confusing (e.g., _IMG1163)" (p. 5). Adding to the chaos, they state "different types of files — from different machines, by different manufacturers — use different default naming conventions. (p. 5). As a result, "it is particularly valuable to assign names that are useful, i.e. that tell something about the data" (p. 5) and, according to the North Carolina Department of Archives and History (2008), meaningful descriptive file names promote "accessibility and ease of identification" (para. 3).

Despite these consequences, some have argued for less meaningful file name conventions. For example, according to Lacinak (2015):

> "File names are not the place for descriptive and structural information. That is what databases are for... Trying to cram excessive descriptive information into a file name creates unwieldy names and is often futile because conditions or conventions change and new scenarios arise over time." (p. 138).

The United States National Archives and Records Administration (2004) documentation also suggests that the scope of a digital project can have implications for file naming approaches:

> "Meaningful file names contain metadata that is self-referencing; non-descriptive file names are associated with metadata stored elsewhere that serves to identify the file. In general, smaller-scale projects may design descriptive file names that facilitate browsing and retrieval; large-scale projects may use machine-generated names and rely on the database for sophisticated searching and retrieval of associated metadata." (p. 60).

In contrast to Lacinak's (2015) argument, this dissertation is one example of how descriptive file names increase a practitioner's ability to discover reused digital images in websites, particularly blog posts and Wikipedia entries.

Section 6.1.6 Results of good file naming: optimized search results

There are numerous benefits to exposing descriptive file names to search engines. According to Hayes (2018), "search engines not only crawl the text on your webpage, they also crawl your image file names" (para. 6). The more descriptive information a digital library practitioner includes in the file name, search engines such as Google can index and move it higher in the search rankings. As Carr (2015) states:

> "when Google knows more about the content of your image, it can include it in Image Search, and sometimes the top images even show right at the head of a regular search page. Not only does it help people find your photos directly, but it also helps Google understand the content of the page that you posted the photo on, helping that page show up higher in regular search results for the subject." (para. 2).

Section 6.2 File naming assessment investigation

The author devised an investigation to explore:

1. What is the user-generated file naming behavior? For example, does it match the combination of title, creator, or date based on the embedded metadata verification study outcome in chapter five?

2. Does user-generated file naming behavior yield more extensive search results than the original file name?

The author designed the investigation methodology to address each question in a specific sequence. The author believes that responses to these questions can impact digital library practitioners' file naming conventions and assist in the discoverability of reused digital images. Complete data sets are listed in the appendix numbers 4-8.

Section 6.2.1 Research design for file naming assessment investigation

To answer question one, the author performed the following processes:

- select images,
- generated and documented data (file name, title, creator, digital library, and date of creation) about each image (for both original and user-generated file names),
- normalized data in languages other than English,
- compared user-generated file names to the original title, creator, and date metadata
- recorded the comparison data in a spreadsheet

To answer question two, the author used the same comparison data collected to answer question one to generate an aggregated file name. The author defines an aggregated file name as a text string based on the most frequent descriptive terms found in user-generated file names. When there were no frequently repetitive descriptive terms, the author used the original title, creator, and/or date to derive the aggregated file name. If the user file names were general terms such as "A Funeral," the

author added the creator's name to distinguish it from generic photos of funerals. Table 13 illustrates how an aggregated file name was generated based on the user file name.

User file name	**User file name**	**User file name**	**User file name**	**User file name**	**Aggregated file names - based on # of occurrences of user file names**
heron-in-snow-by-ohara-koson-b_90441.1506402920.1280.1280.jpg	Shoson_Ohara-No_Series-Unknown_bird-00034696-110802-F12.jpg	heron_in_snow_fine_art_by_ohara_koson_card-r56e0af21b9ff4a2ea24cf2cb0065e42d_em0cq_246.jpg	K2742D-324x486.jpg	heron-in-snow-ohara-koson.jpg	heron-in-snow-ohara-koson.jpg

Table 13 - Example of Aggregated File Name Generation

The author generated data on each image by performing two separate searches. First, using Google Images to search, the author placed the original file name as a text string into the search box and recorded the number of results that matched the original image. Then the author performed the exact search using the aggregated file name as a text string and recorded the number of results that matched the original image. Finally, the author compared both searches' results to determine which search approach yielded a higher number of matches to original images.

Section 6.2.2 File naming assessment investigation digital library selection

The author generated an image dataset derived from eighteen digital libraries for chapter six of the dissertation. The author selected images that Europeana featured on its Twitter feed and in their blogs, those individual digital libraries marketed on their homepages, highlights pages or blog pages, other social media, or advertising websites.

This section will discuss these digital libraries in more detail. The author collected publicly available data about each digital library encompassing:

- The mission of the institution
- The number of items held on the premises or displayed online
- The targeted audience
- The number of users per year
- The financial and/or governance information

This information was mined almost exclusively from the institution's websites and to a far lesser degree from journal articles or encyclopedias such as Britannica or New World Encyclopedias. A limiting factor to each institution's details includes public information accessibility and language translation.

Section 6.2.3 File naming assessment investigation digital libraries

Section 6.2.3.1 La Bibliothèque Nationale de France

The Bibliotheque Nationale de France (BnF) (Bibliotheque Nationale de France, 2020), located in Paris, is the National Library of France and is the national repository of all materials published and produced in France. The BnF has a long and storied history beginning in 1368 with Charles V to its current enterprise. The current library states that its mission is to collect, preserve, and make available French materials. The BnF makes over 5 million documents, images, and artifacts through the Gallica digital library portal from its historical and current collections available online. The Gallica online portal received over 15 million visitors in 2019 (Bibliotheque Nationale de France, 2019b). Sponsored

and supervised by the French Ministry of Culture, the BnF is governed by a 19-member board of directors. Total BnF expenses for 2019 exceeded 240 million euros. (Bibliotheque Nationale de France, 2019a).

Section 6.2.3.2 Bildarchiv Foto Marburg: German Documentation Center for Art History - Photo Archive Photo Marburg

Supported by the Philipps University of Marburg and founded in 1913 by art historian Richard Hamann, the Bildarchiv Foto Marburg (Bildarchiv Foto Marburg, n.d.) is a national and international visual cultural photo research facility. The center collects, indexes, documents, and displays a collection of around 1.7 million photographs, one of the world's most extensive image archives of European art and architecture. Additionally, the center collects research into the history, practice, and communication of visual arts. An eight-member scientific advisory board supports the center.

Section 6.2.3.3 Finnish National Gallery

The Finnish National Gallery (Finnish National Gallery, n.d.), launched on January 1, 2014, is the largest art museum in Finland. It encompasses three museums, the Ateneum Art Museum, the Museum of Contemporary Art Kiasma, and the Sinebrychoff Art Museum, with combined visual art holdings of approximately 42,000 items that encompass art, artifacts, and archival materials dating from the 19th century to modern. "The Finnish National Gallery's values; Together, Transparent, Professional." (Finnish National Gallery, n.d.) It receives more than 600,000 visitors per year. Financed by the Finnish government and governed by the Ministry of Education and Culture.

Section 6.2.3.4 Gemäldegalerie, Staatliche Museen zu Berlin

The Gemäldegalerie (Gemäldegalerie, Staatliche Museen zu Berlin, 2020) possesses the world's most impressive collection of European paintings, and approximately 1.000 are on display at any given time. Located in Berlin, the museum's collection spans from the 13th to the 18th century and includes paintings by Albrecht Dürer, Peter Paul Rubens, Jan Vermeer van Delft, and many others. Unlike many museums in Europe, the Gemäldegalerie's collection was not acquired from former royal dynastic collections but the Prussian government starting in 1815. It is one of the many German state museums spread across Berlin and boasts 72 main galleries. It opened its doors to the public in 1830 and gained an international reputation primarily under the directorship of Wilhelm van Bode (1890-1929). The German federal government funds the museum.

Section 6.2.3.5 German Historical Museum

Located in Berlin's historical center, the German Historical Museum (MfDG) (German Historical Museum, n.d.) serves as the national history museum for Germany. It prides itself as "a place for lively communication and discussion of history." (About page). Founded in 1987 by the Federal Republic of Germany and the State of Berlin on the city of Berlin's 750th anniversary, it became the central history museum of Germany. After reunification in 1990, the museum included all German history. MfDG acquires, collects, and exhibits artifacts of German history and the former Royal Prussian Armory. It holds approximately 900,000 objects and makes many of those accessible online through its databases and Europeana. Partnering with the Living Museum Online (LeMO), the online database presents resources on "European history from 1815 to the present." (About page)

Section 6.2.3.6 Kunsthistorisches Museum

The Kunsthistorisches Museum (KHM) (Kunsthistorisches Museum, n.d.) is the largest and most famous fine arts museum of Vienna, Austria. The museum and six other museums are administered by the KHM Museum Association, whose mission is to store, present, and mediate "works of art and culture from the Republic of Austria." (KHM-Museumsverband, n.d.). The museum and the Natural History Museum opened around 1891 by Emperor Franz Joseph I of Austria-Hungary. The original intent of the museum building was to house and make public the Habsburgs' art collection. The museum is funded by various sources, including the Austrian taxpayers, the KHM Museum Association, entrance fees, donations, and the museum retail shop.

Section 6.2.3.7 Kupferstichkabinett, Staatliche Museen zu Berlin

The Kupferstichkabinett Museum (Kupferstichkabinett, Staatliche Museen zu Berlin, 2020) contains 660,000 prints, drawings, watercolors, pastels, and oil sketches by famous artists such as Botticelli, Rembrandt, and Vincent van Gogh. It boasts the most extensive graphic arts collection in Germany. The collection began in 1652 with the original purchase of approximately 2,500 drawings by Friedrich Wilhelm of Brandenburg for the royal court library. Though, routine collecting did not actively begin until 1831, when the museum was officially established. The collection is permanently housed at the Kulturforum and is close to the Staatliche Museen zu Berlin. The German federal government funds the museum.

Section 6.2.3.8 Library and Archives of Canada

The Library and Archives of Canada is mandated to collect, preserve, provide access, and display Canada's past and present heritage (Library and Archives of Canada, 2017). It strives to help Canadians understand who they are as Canadians and provide lasting memories of the Canadian government and its institutions. The collection, spanning a 140-year history, includes 20 million books, government and private documents, drawings, plans, maps, 30 million photographs, 90,000

films, 425,000 works of art, the largest collection of Canadian sheet music, and other archival materials. For the fiscal year 2019-2020, the Library and Archives of Canada's website received 4,322,323 visits. (Library and Archives of Canada, 2020) It is governed by the Librarian and Archivist of Canada and a network of advisory committees.

Section 6.2.3.9 The Mauritshuis

The Mauritshuis is an independent museum housed in a 17th -century monument that presents works by Dutch and Flemish artists from the 17th -century, including many works from Rembrandt, Vermeer, Steen, and Rubens. The Mauritshuis manages the Foundation Royal Picture Gallery of Mauritshuis and the Prince William V Gallery buildings and collections. The entire collection of 911 paintings can be viewed online as high-resolution digital images. In 2019 the museum received over 480,000 visitors (Mauritshuis, 2020b-f, Home, History of the Building, Collection, Organization, & Search pages). The museum is funded by several government funds, private funds, and private benefactors (Mauritshuis, 2020a). A Supervisory Board oversees the work of two directors and more than 70 employees.

Section 6.2.3.10 Musée d'Orsay

The national museum of the Musée d'Orsay (Musée d'Orsay, 2016-2020) is one of three museums (Musée d'Orsay, Musée de l'Orangerie, and Musée Hébert) operated since 2004 under the auspices of the French Ministry of Culture. The museum opened its doors in 1986 and declared its mission to showcase the "artistic creation in the western world between 1848 and 1914." (History of the collections, para. 1). Its collections derive from the Louvre, the Musée du Jeu de Paume, and the National Museum of the Modern Art. In 2019 the museum welcomed over 3 million visitors. Managed by a board of directors, this group determines the programming, exhibitions, and other cultural activities. Acquisitions in 2018 derived much of its almost 3 million euros funding from the sale of tickets (Musée d'Orsay, 2019).

Section 6.2.3.11 The National Museum of Art, Architecture and Design

The National Museum of Art, Architecture and Design in Norway was created in 2003 with the Museum of Decorative Arts and Design, the Museum of Contemporary Art, and the National Gallery of Norway (National Museum of Art, Architecture and Design, 2020, About). The museum's mission is to collect, hold, exhibit, conserve, digitize, and document its collection. Furthermore, it promotes knowledge about and increases social awareness of Norway's cultural heritage. In 2019, "the collections consisted of more than 230,000 original drawings and photographs at the incorporation into the new National Museum." (Faltin, 2019. para. 12).

Section 6.2.3.12 Nationalmuseum

Sweden's Nationalmuseum is mandated to collect, preserve, and make accessible the cultural heritage of Sweden (Nationalmuseum, n.d.). Its directive is to promote art and to create an awareness of and knowledge of art. The collections, totaling more than 16,000 works, covers a wide range of materials including sculpture, paintings, prints, and drawings spanning four centuries (1500-1900) from a variety of artists such as Renoir, Rembrandt, Degas, Swedish artists Ernst Josephson and Anders Zorn, as well as famous Dutch and French artists. Additionally, the corpus comprises approximately 530.00 works from the Middle Ages to the present. There are two other collections, the National Portrait Gallery and the Royal Castles Collections (about 5.000 masterpieces) and the Gustavsberg Collection (roughly 45.000 objects). The Swedish government annually provides the 164 million Swedish Krona (SEK) operating budget. The collection budget relies solely on private donations and interest earned from foundation endowments.

Section 6.2.3.13 Statens Museum for Kunst

The most significant art museum in Denmark, the Statens Museum for Kunst (SMK) (Statens Museum for Kunst, n.d.). It holds about 260.000 works of visual art. Initially acquired by Danish kings, the collection was given to Denmark's people in the mid-1800s when the country became a democracy. The SMK is responsible for the creation and development of the central museum district. Another obligation is the continued partnerships with national and international museums focused on the visual arts. Their current goal is to make the entire art collection available online for free use.

Section 6.2.3.14 Portal to Texas History

The University of North Texas Libraries conceptualized and created the Portal to Texas History in the early 2000s (Portal to Texas History, 2020). It gives access to Texas-related primary sources, rare and historical materials. Its statewide partner content creators, including private collectors, libraries, museums, historical centers, and archives, create engaging, fact-based, and stimulating research materials. In 2019, the Portal to Texas History included over 1.5 million digital images from some 426 partners and had received more than 77 million uses. In 2005, It launched as a resource for educators expanded its offerings in 2007 through a grant from the United States National Endowment for the Humanities (NEH) to include historical newspapers. Two collections are featured in the "File naming assessment investigation" dataset, the Austin History Center and the Dallas Municipal Archives. The Austin History Center showcases the beginnings of the city to the present. It has 1,130 materials displayed on the Portal of Texas History. Available to the public since 1985, the Dallas Municipal Archives comprises documents, manuscripts, maps, photographs that document the history of Dallas, Texas. There are 6,996 matching results on the Portal of Texas History from the Dallas Municipal Archives.

Section 6.2.3.15 Rijksmuseum

For detailed information about the Rijksmuseum Museum, see section 3.3.2.2.

Section 6.2.3.16 Slovak National Gallery

The Slovak National Gallery established in 1948 by the Slovak National Council, is a traditional fine arts museum (Slovak National Gallery, 2020). The museum prides itself on fulfilling its role as the custodian of the Slovak national cultural heritage. Over the next thirty years, it grew in its buildings, collection, and collection scope. The over 70,000 item collection of 20th -century artworks includes applied art, design, photography, stage design, and naïve art.

Section 6.2.3.17 Wellcome Collection

The Wellcome Collection located in central London since 2004, was established by the Wellcome Trust, the largest charity in the United Kingdom, upon the death of Sir Henry Wellcome in 1936 (Wellcome Collection, n.d.). Sir Wellcome acquired most of the collection between 1890 and 1936. Its mission is to create a museum that supports researchers and the general public to think creatively about the connections between science, medicine, health, art, and life. It exhibits, collects, and displays both physically and digitally its many collections. The Wellcome Collection includes a museum and a library containing over two million items related to the history and progress of medicine. In 2013 the museum expanded its physical site and saw visitors increase up to 700,000 per year. (Wellcome Collection, 2017).

Section 6.2.3.18 Library of Congress

For information on LC, see section 2.5.5.

Section 6.2.4 File naming assessment investigation image selection

The researcher hypothesized that images highlighted on social media, such as Twitter and Facebook primarily, by the owning institution or an aggregator such as Europeana would have the most views and possibility of far-reaching reuse. More specifically, the researcher prioritized selected images posted on social media platforms for at least 30 days and the number of responses such as "like" or "comments" on the social media posts. As a result of this chapter's selection criteria and novel research agenda, its digital library and image dataset are much more extensive and comprehensive than the previous chapter.

Using specified selection criteria, the author selected 100 images from European and North American digital libraries. The selection criteria were to identify images that were popular enough to be shared over the web. To determine this, the author prioritized images that organizations highlighted on their homepages and shared with others via their social media accounts, including Europeana's Twitter and blogs. The author also selected images from notable artists, such as Monet and Van Gogh, whose artwork has been digitized by digital libraries that allow users to download and share images. Finally, the author cross-referenced several of these images with Wikipedia to prove they had been shared across the web at least once.

Section 6.2.5 File naming assessment investigation data collection

Section 6.2.5.1 Data collection - original file name

The author downloaded images from the original Digital library source and recorded the following metadata for each image: object file name provided by the digital library, title, creator, digital library, and original work creation date. The author used the original file name generated during the download and not any subsequent file name that may have been placed on it by an aggregator. The author compiled all other information from digital library-provided metadata at the time of download. Finally, the author documented these values in a spreadsheet.

Section 6.2.5.2 Data collection - user-generated file name

The author collected user-created file names in two steps. First, to identify shared and/or reused images, the author uploaded images into the Google Images search home page. Then, the author used results from the "Pages that include matching images" section to capture the user-generated file name. Second, the author documented the user-generated file name by simulating a "download" or "save as" of the image. Then, the author placed the first ten eligible file names from each image into a spreadsheet. Upon completion, the author generated 1,000 data points for analysis.

The author created four exclusion criteria that determined the eligibility for a reused images' file name in the final 1,000 data points:

1. Duplicate URLs. The author determined that any Google Images search result that included links to the same web page multiple times was redundant. Consequently, the author would include the first instance in the final results and ignore subsequent links.
2. Digital library holding the image. The author excluded any Google Images search results from the digital library from which the image resides with the understanding that this would not yield a user-generated file name.
3. Websites that did not allow for direct download. The author excluded any sites that prevented using the "download" or "save as" features, such as Google Arts and Culture sites.
4. Social media sites. The author excluded any results from social media websites since the embedded metadata discussion in the chapter five. demonstrated that these sites remove possible user-generated file names and replace them with non-descriptive file names upon upload.

Section 6.2.5.3 File naming assessment investigation data normalization

Typically, European and North American digital libraries that held images used in this study translated object metadata into English, including object title and creator. However, this translation was not always the case for user-generated file names. Users created file names in a variety of

languages. As a result, the author used Google Translate to convert file names into English equivalents when the user-generated file name was in a language other than English. This process made it possible for the author to (a) identify the frequency at which file names are changing and (b) generate the aggregated file name based on the English translation. This translation was necessary to compare search results between original and user-generated file names.

As an example, the author translated the user-generated file name

> "777px-Redningsbåden_køres_gennem_klitterne_(high_resolution).jpg"
>
> to
>
> "777px-The lifeboat is driven through the dunes_(high_resolution).jpg".

The author recorded both the untranslated and translated versions of the file name in the spreadsheet for transparency. Figure 21 displays the process map for this procedure.

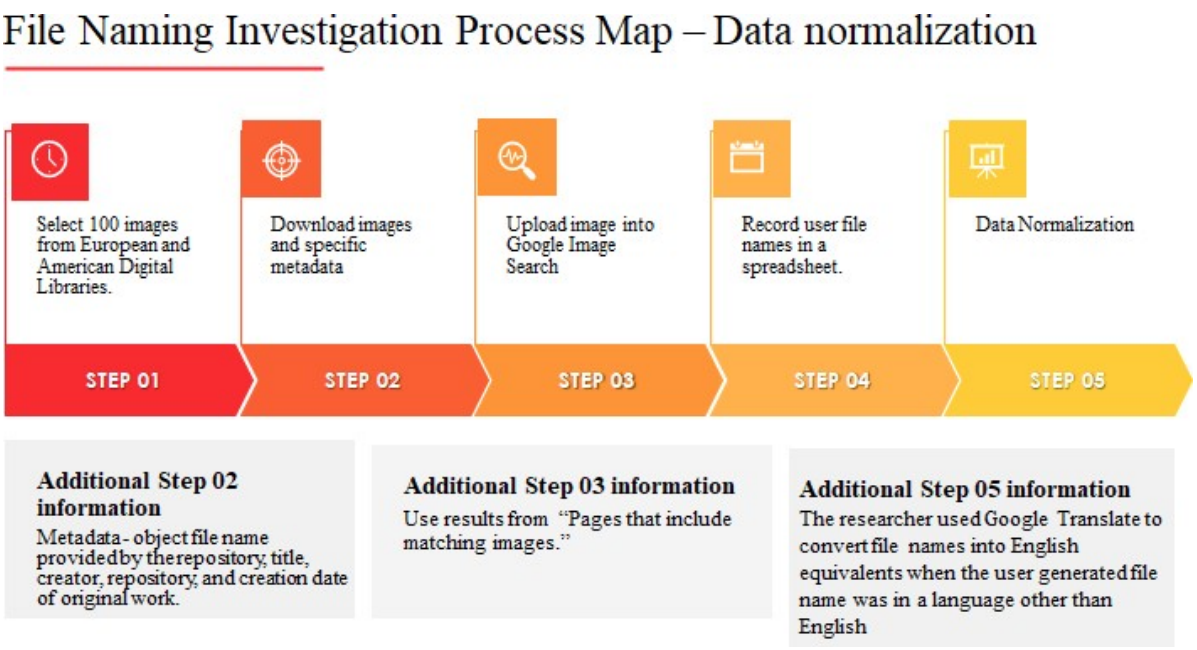

Figure 21 - File naming investigation process map - data normalization

Section 6.2.6 File naming assessment investigation data analysis

The author utilized the sum and percentage functions in a spreadsheet to analyze data compiled in the spreadsheet. These totals provided the comparison data needed to answer questions one and two.

Section 6.2.6.1 Data analysis - frequency of attributes in user-generated file names

The author compared the composition of user-generated file names against two values: the original file name and the combination of the object's title, creator, and/or date. To do this, the author counted the number of times a user retained part or all of the original file name and the number of times a user changed the file name to match some combination of the object's title, creator, and/or date. The author documented these values in a spreadsheet. Figure 22 displays the process map for this procedure.

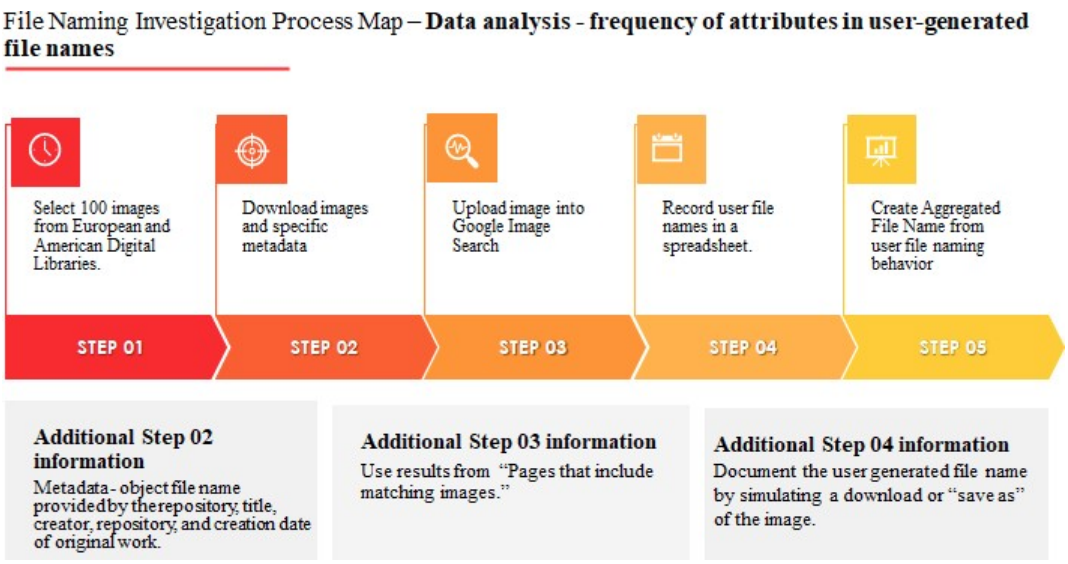

Figure 22 - File naming investigation process map - data analysis - frequency of attributes in user-generated file names

Section 5.2.6.2 Data analysis - aggregated file name and original file name search

To understand which file naming convention produced more Google Images search results, the author used the text string values for the aggregated file name and the original file names as the search query. The author documented in a spreadsheet the number of times that exact matches occurred in the top 50 search results for both the aggregated file name and original file names. This approach had the potential to produce 5,000 results for each query. The author limited the number of results to 50 because occurrences after that were less likely to match the exact image queried. To remain consistent with other exclusion criteria, the author eliminated data that Google placed into the "Sponsored Images" section of the search results because Google displays this data for monetary reasons, not necessarily because it strongly matches the search query. Figure 23 displays the process map for this procedure.

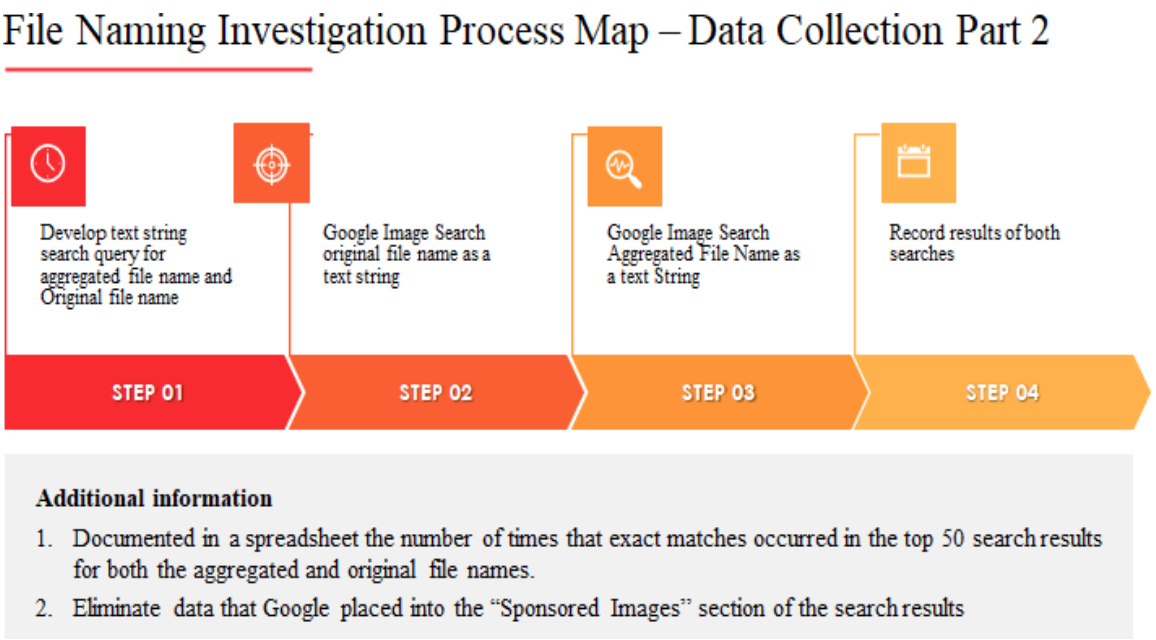

Figure 23 - File naming investigation process map - data collection part 2

Section 6.2.7 File naming assessment investigation results

Section 6.2.7.1 Frequency of attributes in user-generated file names

After collecting file name data, the author obtained 920 total research results, out of a possible 1,000, to analyze. The author found that not every image queried resulted in at least ten results. Thus, accounting for the discrepancy between the actual number of hits (920) and the total possible hits (1,000). Upon analyzing the data, the author found that

- 92% of the file names were changed by the user,
- 65% of user-generated file names included some or all of the title, creator, date,
- approximately nine percent of user-generated file names contained part or all of the original file names.

Table 14 expresses the number of times and percentage of times the user changed the file name upon reuse.

File Name Change Results						
Original file name given by the digital library	Number of results of Google Images search file name hits	Number of time users changed the file name to be part of or some combination of the creator, title, or date	Part or all of the original file name used in users file name	Part 1. Percentage of times that a user changed the file name to be part of or some combination of the creator, title, or date	Part 1. Percentage of time that a user kept Part or all of the original file name used in users file name	Part 1. Percentage of time that file name changed
Totals	920	600	85	65.22%	9.24%	92.00%

Table 14 - File Name Change Results

The results demonstrated that one-quarter of user-generated file names contained some other values, such as alphanumeric strings that did not include part or all of the original file name or the

aggregated file name criteria: image title, creator, and/or date. Highlighted in Table 15 is an example of one of these other values.

Original file name given by the repository	Original title	Original creator	Digital library	Creation date of original work	User file name
RP-P-2001-734	Reiger in sneeuw Heron in snow	Ohara Koson	Rijksmuseum	1920 - 1930	il_794xN.1831812295_rvyn.jpg

Table 15 - Example of other value: alphanumeric strings

Section 6.2.7.2 Aggregated file name and original file name search

Out of a possible 5,000 results for each search type, the author recorded 1,321 exact image results using the aggregated file name string and 83 using the original file name string. The author found that the aggregated file names generated 26% of matches in search results. Conversely, there was a less than two percent success rate finding the original image using the original file name. This result is expressed in table 16 below.

Aggregated vs Original File Name Results				
Original file name given by the digital library	Part 2. hits/total results for image search using aggregated file names only (no image) 50	Part 2. hits/total results for image search using original file name as text (no image) 50	Part 2. Percentage of total hits using the aggregated file names	Part 2. Percentage of total hits using the original file name
Totals	1321	83	26.42%	1.66%

Table 16 - Aggregated vs Original File Name Results

Section 6.3 Discussion

The file naming assessment investigation asked two questions:

1. What is the user-generated file naming behavior? Does it match the combination of title, creator, or date based on the embedded metadata verification study outcome in chapter five?

2. Does user-generated file naming behavior yield more extensive search results than the original file name?

To answer question one, the author discovered that the users are changing the file names when reusing images, a vast majority of the time, 92%. Additionally, the user was more likely to adopt some combination of title, creator, and/or date when generating reused image file names. While the author does not investigate the intent of this behavior and can only speculate based on existing literature (file naming to serve as a memory device), the results indicate that typical file naming behavior embraces the employment of descriptive file naming, in particular a pattern of title, creator and/or date. The author believes that further study of this combination would lead to an even better understanding of user file naming behavior.

To answer question two, the author compared search results for the aggregated file names and the original file names and found that the aggregated file name search queries are 25% more likely to produce hits of the original image than the original file name queries.

The findings demonstrate that GLAM organizations represented in the study do not follow the descriptive file naming best practices discussed in section 6.1, nor do they leverage file naming as a search engine optimization tactic as discussed above. Of the numerous institutions and 100 digital images chosen, all original file names consisted of alphanumeric text strings that serve an administrative purpose. Twenty-four out of 100 contained human-readable text. Of these, 24, twenty-one retained the digital library name, and one included a descriptor of the image's original format. Only two file names contained the descriptive "title" element.

Based on these results, the author contends that employing an "aggregated file name" text string search offers a more practical approach to tracking the reuse of images than using image files (as in the image "drop and drag" method") as the search query with Google Images. Additionally, the author contends that practitioners with knowledge of search engine indexing techniques and the file naming investigation results should be using the aggregated file name strategy when naming files. As Smith (2017) writes, "a digitised object without **descriptive, human readable (including file name)** metadata is invisible to search engines and therefore will be undiscoverable, rendering the digitisation of the item redundant" (note: emphasis added) (p. 9). Nieuwenhuysen (2018b) further emphasizes the concept of the importance of descriptive, human readable file naming,

> "In the case of a search for images, this situation is even worse, as the words in a query can by definition not directly be matched with an image, but only with the name of the file that carries the image" (p. 17).

Finally, a future study might experiment with file naming queries to determine which sequence of values would derive the most positive matches in Google Images search results. Likewise, this future study could also investigate the impact of file naming formatting, such as camel case, dashes, or underscores, on engendering the most relevant search results. Finally, a future study could test opportunities to automate the aggregated file naming search strategy process. This research would require developing programming scripts that query the metadata (title, creator, and date) of digital images within a digital library to create an aggregated file name list. This list will then be queried through a reverse image lookup service such as Google Images or Tineye Reverse Image.

Chapter 7: Research findings, limitation, and conclusions

This dissertation posed three research questions:

1. Can practitioners use Google Images and embedded metadata as practical tools to track digital object reuse?
2. Are there patterns in descriptive and administrative embedded metadata from reused digital images that practitioners can use to inform reuse discovery strategy?
3. If yes to question two, what is a strategy that practitioners can implement to discover and trace the reuse of digital objects more effectively?

Section 7.1 Research findings

To answer question one, the author performed two verification studies. After completing the CBIR/RIL (Google Images) verification study, the author determined that the Google Images search engine remains a powerful tool for tracing commercial or blog-related reuses of images. This tool is beneficial if one is trying to determine copyright infringement. However, if digital library practitioners are trying to assess how other audiences, including educators and general users, are discovering, using, and sharing digitized images, it is much less helpful. This unhelpfulness is due, in part, to the algorithm change employed by Google Images search after the Getty Institute's legal action. Given the limited results derived by the Google Images search, the author contends that it does not stand as a viable assessment tool beyond evaluating copyright compliance for cultural heritage institutions.

After completing the embedded metadata verification study, the author found that embedded metadata would not be a reliable assessment strategy. This strategy failure is because many social media platforms, including Pinterest, Facebook, and Twitter, change or overwrite embedded metadata upon upload into their website, making embedded metadata non-persistent.

To answer question two, the author performed a file naming assessment investigation that found that users change file names most of the time to correspond with some combination of the original object's title, creator, or date. As Nardi et al. (1995) note, "people attempt to name their files in a way that will help them remember the name" (para. 25).

To answer question three, the author contends that practitioners should change their current file naming practices and reuse search strategies to take advantage of the relationships between digital image reuse, descriptive human-readable file naming, and reuse discovery. Chapter six's file naming assessment investigation results challenge many of the existing digital library file naming customs, which typically prioritize administrative-based file names over human-readable file names. The investigation indicates digital image reuse discoverability increases when practitioners follow the user's file naming behavior and incorporate elements of the object's title, creator, and/or date into the file name when creating files or uploading them to blogs and websites especially Wikipedia sites. This approach aligns with Inglis and Prosser's (2014) emphasis on mimicking user behavior. They write, "as custodians of digital content, we need to think more from the researchers' point of view and less like curators: people come looking for resources in ways we often haven't even imagined" (p. 3). Beyond researchers, practitioners should start "thinking" like different user groups, including genealogists, K-12 teachers and students, and general, everyday users, when file naming.

Additionally, when trying to discover the reuse of digital images by users, practitioners should use the "aggregate file name" as a search strategy because this produces a more diverse and more significant number of search results, as discussed in chapter six.

Section 7.2 Research limitations

The verification studies focus on CBIR/RIL (Google Images) search and embedded metadata because previous literature has proven these tools to be potential reuse assessment solutions. However, as the results indicate, these tools are not ideal given the current limitations that exist. Also, the methodology focuses only on analyzing the output from Google Images search results. Other studies might benefit from alternative data collection methods, including focus groups, surveys, or data analysis derived from log files as examples. Another limitation is that the author focused exclusively on static images. The study did not consider moving images or PDF documents (including images embedded in PDF files) because the Google Images search engine does not allow these file types to be queried. Google Images only supports static, image-based file formats including jpg, .gif, .png, .bmp, .tif, or .webp. Finally, additional factors can impact the results of a Google Images search for reuse assessment purposes. According to Perrin, Yang, Barba, and Winkler (2017),

> "Collections that are complete and not updated (called static collections) will be treated differently than actively growing collections by search engines. Search engines prioritize web sites that update frequently (Google, 2010), which means growing collections will get prioritized over static collections and can receive more use as a result. Use for static digital collections tends to go down from year to year as the collection reaches its target audience. Someone using a digital collection once might never need to revisit the collection, so year to year use could drop off. This drop off can be confusing to digital collection evaluators, as it looks like the collection is failing and in need of more marketing, but, in fact, the collection has reached its intended audience and may have fulfilled the majority of its purpose." (p. 192).

The author also acknowledges that data from chapters four and five of this study are more prominently from United States GLAM institutions. Approximately two-thirds of the images are from United States sources. It is unknown and difficult to assess if this is a limitation, but the author concedes that it may have influenced some of the final results of these chapters. Any limitations from the number of US images compared to International images can be attributed to Google Images search algorithms and cultural factors. The author tried to counteract these factors by striving for equivalency. Additionally, the author notes that several commercial browsers were employed to conduct studies in chapters four and five but contend that browser usage difference would not impact how Google Images operates.

Section 7.3 Research conclusion

As this dissertation has shown, there continue to be significant limitations in the tools and methods used for assessing digital image reuse. The author's focus on file naming practices, specifically descriptive file naming and aggregated file naming search strategy, presents an alternative approach to reuse assessment data collection. However, for descriptive file naming to become part of a viable assessment, significant shifts in practitioner approaches to file naming must occur.

As more GLAM institutions launch and manage digital image collections to showcase their rare and unique materials, the need to understand a digital library's service success and relevance becomes increasingly essential. Discovering and measuring reuse is an integral part of digital collections assessment. Recognizing how the general public uses digital images through effective assessment strategies, such as changing practitioners' file naming customs and using an aggregated file naming search discovery approach, will help digital library practitioners better assess reuses "in the wild" and demonstrate their value, success, and importance.

References

References follow the *Publication Manual of the American Psychological Association, Seventh Edition (2019)* citation style sheet as provided by Purdue University's Online Writing Lab: https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/general_format.html.

Adobe. (2020). *XMP Adding Intelligence to Media*. Retrieved April 21, 2021 from https://www.adobe.com/products/xmp.html

Agosti, M., Di Nunzio, G. M., Ferro, N., & Silvello, G. (2018, January). Thirty years of digital libraries research at the University of Padua: The systems side. In *Italian Research Conference on Digital Libraries* (pp. 30-41). Springer, Cham.

Ajinkye. (2019, December 20). *Best reverse image search engines, apps and its uses* (2020). Beebom Media Private Limited. Retrieved March 16, 2020 from http://beebom.com/reverse-image-search-engines-apps-uses/

American Library Association. (2018, May 9). *The Nation's Largest Public Libraries: Top 25 public libraries by holdings.* Retrieved February 29, 2020 from http://libguides.ala.org/libraryfacts/largestlibs#s-lg-box-wrapper-16931167

American Memory. (n.d.). Library of Congress. *History and Mission.* Retrieved February 29, 2020 from http://memory.loc.gov/ammem/about/index.html

Ames, S. (2021). Transparency, Provenance and Collections as Data: the National Library of Scotland's Data Foundry. *LIBER Quarterly*, *31*(1). Accessed on April 29, 2021 from https://www.liberquarterly.eu/articles/10371/print/

Arora, J. (2018). *Digital Libraries*. INFLIBNET Centre. Retrieved February 1, 2020 from http://lisp8.epgpbooks.inflibnet.ac.in/

Baldacci, C. (2019) 'Recirculation: The Wandering of Digital Images in Post-Internet Art', in *Re-: An Errant Glossary*, ed. by Christoph F. E. Holzhey and Arnd Wedemeyer, Cultural Inquiry, 15 (Berlin: ICI Berlin, 2019), pp. 25–33 https://doi.org/10.25620/ci-15_04

Balog, A. (2011). Testing a multidimensional and hierarchical quality assessment model for digital libraries. *Studies in Informatics and Control, 20(3)*, 233-246.

Barreau, D. (1995). Context as a factor in personal information management systems. *Journal of the American Society of Information Science, 46(5)*, 327-339.

Barrett, C. (2017, September 28). *Managing privacy and access with digital forensics tools and techniques*.[Presentation]. Right to Know: Balancing Access and Privacy Symposium at

Dalhousie University. Retrieved from https://dalspace.library.dal.ca/bitstream/handle/10222/73352/Barrett_DigitalForensics_RighttoKnowDay.pdf?sequence=2&isAllowed=y

Barton, J. (2004). Measurement, management and the digital library. *Library review.*

Beskow, D. M., & Carley, K. M. (2020). Bot-Match: Social Bot Detection with Recursive Nearest Neighbors Search. *arXiv preprint arXiv:2007.07636*.

Beaudoin, J. E. (2014). A framework of image use among archaeologists, architects, art historians and artists. *Journal of Documentation, 70(1)*, 119-147.

Beaudoin, J. E. (2016). Content-based image retrieval methods and professional image users. *Journal of the Association for Information Science and Technology, 67(2)*, 350-365.

Berg, T. V. D., Bommes, E., Härdle, W. K., & Petukhina, A. (2017). Computing Machines.

Bertot, J. C. (2004, March). Assessing digital library services: Approaches, issues, and considerations. [Paper]. In I*nternational Symposium on Digital Libraries and Knowledge Communities in Networked Information Society*, DLKC'04.

Besser, H. (2004). *The Past, Present, and Future of Digital Libraries. A Companion to Digital Humanities.* (E. Editor, Ed. Schreibman, Susan, Ray Siemens, and John Unsworth). Oxford: Blackwell. Retrieved from http://www.digitalhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-5-6&toc.depth=1&toc.id=ss1-5-6&brand=default

Besser, H. (2016). 25 years of digital archives of visual materials: What we've done, what we've learned, and what challenges remain. *Visual Studies, 31(2)*, 95-108. doi:10.1080/1472586X.2016.1173888

Bibliothèque Nationale de France. (2020). *About*. Retrieved August 1, 2020 from https://gallica.bnf.fr/edit/und/a-propos.

Bibliothèque Nationale de France. (n.d.). *Gallica Partners.* Retrieved March 22, 2020 from https://gallica.bnf.fr/edit/decouvrir-nos-partenaires

Bibliothèque Nationale de France. (2020). *Mission.* Retrieved March 22, 2020 from https://www.bnf.fr/fr/les-missions-de-la-bnf

Bibliothèque Nationale de France. (2019a). *Le budget 2019 en comptabilité budgétaire.* Retrieved March 22, 2020 from https://multimedia-ext.bnf.fr/pdf/rapport_2019_bnf_chiffres_budget.pdf.

Bibliothèque Nationale de France. (2019b). *Rapport d'activité 2019.* Retrieved March 22, 2020

from https://multimedia-ext.bnf.fr/pdf/rapport_2019_bnf_chiffres_publics.pdf.

Bildarchiv Foto Marburg: German Documentation Center for Art History - Photo Archive Photo Marburg. (n.d.). *Homepage.* Retrieved August 1, 2020 from https://www.uni-marburg.de/de/fotomarburg

Boididou, C., Papadopoulos, S., Apostolidis, L., & Kompatsiaris, Y. (2017, June). Learning to detect misleading content on twitter. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval* (pp. 278-286).

Borbely, T. (2013). Gallica: Digital Library of the Bibliothèque Nationale de France. *The American Archivist.* Retrieved March 22, 2020 from https://reviews.americanarchivist.org/2016/07/02/gallica-digital-library-of-the-bibliotheque-nationale-de-france/

Borgman, C. L. (1999). What are digital libraries? Competing visions. *Information Processing & Management*, *35(3)*, 227-243. http://dx.doi.org/10.1016/S0306-4573(98)00059-4. Retrieved February 3, 2020 from https://escholarship.org/uc/item/7c55m1xf

Borgman, C. L. (2000). Digital libraries and the continuum of scholarly communication. *Journal of Documentation, 56 (4)*, pp. 412-430.

Borgman, C. L. (2003a). The invisible library: Paradox of the global information infrastructure. *Library Trends, 51 (4)*, pp. 652.

Borgman, C. L. (2003b). Designing digital libraries for usability. *Digital library use: social practice in design and evaluation*, 85-118.

Bowen, J.P., Giannini T. (2019) The Digital Future for Museums. In: Giannini T., Bowen J. (eds) Museums and Digital Culture. Springer Series on Cultural Computing. Springer, Cham. https://doi.org/10.1007/978-3-319-97457-6_28

Bragg, M., DeRidder, J., Johnston, R., Junus, R., Kyrillidou, M., Chapman, J., & Stedfeld, E. (2015). Best practices for Google Analytics in digital libraries. Retrieved August 10, 2020 from https://files.osf.io/v1/resources/ct8bs/providers/osfstorage/580e75a16c613b01e2dd3ed4?action=download&version=1&direct

British Library. (n.d.) *British Library Labs.* Retrieved June 14, 2021 from https://www.bl.uk/projects/british-library-labs?mobile=on#

Brogan, M. L. (2006). *Contexts and contributions: Building the distributed library.* Digital Library Federation. http://repository.upenn.edu/library_papers/31

Brylawski, S., Lerman, M., Pike, R., & Smith, K. (Eds.). (2015). *ARSC guide to audio preservation*. Pittsburgh: Association for Recorded Sound Collections. Retrieved

December 19, 2019 from https://www.clir.org/pubs/reports/pub164/.

Buchanan, S., & Salako, A. (2009). Evaluating the usability and usefulness of a digital library. *Library Review*. Vol. 58 No. 9, pp. 638-651. https://doi.org/10.1108/00242530910997928

Bui, Y., & Park, J. R. (2006). An Assessment of Metadata Quality: A Case Study of the National Science Digital Library Metadata Repository. *Proceedings of the Annual Conference of CAIS / Actes Du congrès Annuel De l'ACSI*.

Burns, D., Sundt, A., Pumphrey, D., & Thoms, B. (2019). "What We Talk About When We Talk About Digital Libraries": UX Approaches to Labeling Online Special Collections. *Journal of Library User Experience.* Volume 2, Issue 1. DOI: http://dx.doi.org/10.3998/weave.12535642.0002.102

Bush, V. (1945). As we may think. *The Atlantic monthly*, *176*(1), 101-108.

Bushey, J. (2015, January). Trustworthy Citizen-Generated Images and Video on Social Media Platforms. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on* IEEE.

Canadian Research Knowledge Network. (2020a). *About CRKN*. Retrieved March 22, 2020 from https://www.crkn-rcdr.ca/en/about-crkn

Canadian Research Knowledge Network. (2020b). *History of Canadiana*. Retrieved March 22, 2020 https://www.crkn-rcdr.ca/en/history-canadiana

Canadian Research Knowledge Network. (2018) *Over 60 Million Pages of Digitized Canadian Documentary Heritage Soon To Be Available At No Charge.* Retrieved March 22, 2020 from https://www.crkn-rcdr.ca/en/over-60-million-pages-digitized-canadian-documentary-heritage-soon-be-available-no-charge

Canadiana. (n.d.). *Homepage.* Retrieved March 22, 2020 from http://www.canadiana.ca/

Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., Pagano, P., Ross, S., Soergel, D., Agosti, M., Dobreva, M., Katifori, V., & Schuldt, H. (2007). *The DELOS Digital Library Reference Model. Foundations for Digital Libraries*. Retrieved from https://pureportal.strath.ac.uk/en/publications/the-delos-digital-library-reference-model-foundations-for-digital

Candela, G., Sáez, M. D., Escobar Esteban, M., & Marco-Such, M. (2020). Reusing digital collections from GLAM institutions. *Journal of Information Science*, 0165551520950246)

Cardoso, D. N. M., Muller, D. J., Alexandre, F., Neves, L. A. P., Trevisani, P. M. G., & Giraldi, G. A. (2013). Iterative technique for content-based image retrieval using multiple SVM ensembles. *J Clerk Maxwell, A Treatise on Electricity and Magnetism*, *2*, 68-73.

Carr, D. (2015, June 26). *SEO How To Name Your Image Files*. Shutter Muse. Retrieved January 5, 2019 from https://shuttermuse.com/image-seo-how-to-name-files/

Ceruzzi, P. E., & Paul, E. (2003). *A history of modern computing*. MIT press.

CESSDA Training Team (2017 - 2020). *CESSDA Data Management Expert Guide.* Bergen, Norway: CESSDA ERIC. Retrieved March 23, 2020 from https://www.cessda.eu/DMGuide

Chapman, J., DeRidder, J., Hurst, M., Kelly, E. J., Kyrillidou, M., Muglia, C., O'Gara, G., Stein, A., Thompson, S., Trent, R., Woolcott, L., & Zhang, T. (2015). "Surveying the Landscape: Use and Usability Assessment of Digital Libraries," *Digital Library Federation Assessment Interest Group, User Studies Working Group*. doi:10.17605/OSF.IO/9NBQG.

Chapman, J., DeRidder, J., & Thompson, S. (2015). Developing best practices in digital library assessment: year one update. *D-Lib Magazine. 21*(11/12) https://uh-ir.tdl.org/handle/10657/1342

Chastain, S. (2018, August 9). *What Is Metadata? The Hidden Information in Photo Files*. LifeWire. Retrieved December 16, 2018 from https://www.lifewire.com/what-is-metadata-1701735

Chung, E., & Yoon, J. (2011). Image needs in the context of image use: An exploratory study. *Journal of Information Science*, *37*(2), 163-177.

Chutel, P.M., & Sakhare, A. (2014, April). Evaluation of Compact Composite Descriptors Based Reverse Image Search. In *Communications and Signal Processing (ICCSP), 2014 International Conference* (pp. 1430–1434). IEEE.

Christensen, S. O., & Dunlop, D. (2011, January). The Case for Implementing Core Descriptive Embedded Metadata at the Smithsonian. In *Archiving Conference* (Vol. 2011, No. 1, pp. 116-120). Society for Imaging Science and Technology.

Cleveland, G. (1998, March). Digital Libraries: Definitions, Issues, and Challenges [Occasional Paper]. *International Federation of Library Associations and Institutions*. Retrieved March 20, 2020 from www.ifla.org//VI/5/op/udtop8/udt-op8.pdf

Coburn, J. (2020). Defending the digital: Awareness of digital selectivity in historical research practice. *Journal of Librarianship and Information Science*, 0961000620918647.

Computer History Museum. (2020a). *Timeline of Computer History, 1934*. Retrieved May 20, 2020 from https://www.computerhistory.org/timeline/1934/

Computer History Museum. (2020b). *Timeline of Computer History, 1957*. Retrieved May 20, 2020 from https://www.computerhistory.org/timeline/1957/

Computer History Museum. (2020c). *Timeline of Computer History, 1973.* Retrieved May 20, 2020 from https://www.computerhistory.org/timeline/1973/

Computer History Museum. (2020d). *Timeline of Computer History, 1992*. Retrieved May 20,

2020 from https://www.computerhistory.org/timeline/1992/

Computer History Museum. (2020e). *Timeline of Computer History, 1999*. Retrieved May 20, 2020 from https://www.computerhistory.org/timeline/1999/

Corrado, E. M., & Jaffe, R. (2017). Access's Unsung Hero: The [Impending] Rise of Embedded Metadata. *International Information & Library Review*, 49(2), 124-130.

Council of Europe. Directorate General of Administration. (2017). *Naming Files Good Practices* Recommended by DIT. https://rm.coe.int/naming-files-good-practices-recommended-by-dit/16806d3e5a

Cousins, J. (2017). Creating a Renaissance for the Library in the Digital Era. *LIBER Quarterly*, 26(4).

Crowder, J. W., Marion, J. S., & Reilly, M. (2015). File naming in digital media research: Examples from the humanities and social sciences. *Journal of Librarianship and Scholarly Communication*, 3(3), eP1260.

da Silva Torres, R., & Falcao, A. X. (2006). Content-based image retrieval: theory and applications. *RITA*, 13(2), 161-185.

Dang, Q. (2020). A Comparative Analysis of Visitor Satisfaction in the Digital Museum. *INTERIN*, *25*(1)

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)*, 40(2), 1-60.

Department of Cultural Resources. Office of Archives and History Government Records Branch. North Carolina. (2008). *Best Practices for File-naming*. Retrieved February 23, 2019 from http://digitalpreservation.ncdcr.gov/filenaming.pdf

Deutsche Digitale Bibliothek (DDB) (n.d.a). *Frequently Asked Questions* page. https://www.deutsche-digitale-bibliothek.de/content/ueber-uns/fragen-antworten#5240

Deutsche Digitale Bibliothek (DDB) (n.d.b). *Homepage*. https://www.deutsche-digitale-bibliothek.de/

Dinneen, J. D., & Julien, C. A. (2019). The ubiquitous digital file: A review of file management research. *Journal of the Association for Information Science and Technology*, 71(1), E1-E32.

Eakins, J. P., & Graham, M. E., (1999). Content-based image retrieval, a report to the jisc technology applications programme 1999. *Institute for Image Data Research*, University of Northumbria at Newcastle, [http://www. jisc. ac. uk/uploaded_documents/jtap-039. Doc].

Enzo, A. (2020) The Evolution of Computer Science Albert Enzo, Charles O. Connor and Walter

Curtis. *International Digital Organization for Scientific Research IDOSR JOURNAL OF EXPERIMENTAL SCIENCES* 5(1) 13-18, Computer Science Murdoch University, Australia.

Europeana. (n.d.,a). *About* page. Retrieved September 27, 2019 from https://www.europeana.eu/portal/en/about.html

Europeana. (n.d.). *Cultural Heritage Institutions* page. Retrieved September 27, 2019 from https://pro.europeana.eu/what-we-do/cultural-heritage-institutions

Europeana Pro. (n.d.). *German Digital Library*. Retrieved on March 21, 2020 from https://pro.europeana.eu/organisation/german-digital-library

Europeana. (n.d.,b). *Mission* page. Retrieved October 1, 2018 from https://pro.europeana.eu/our-mission.

Europeana. (2017). *Europeana Content Strategy*. Retrieved October 4, 2018 from https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Content%20Strategy.pdf

Faruque, J., Antani, S., Long, R., Kim, L., & Thoma, G. (2016, March). Image similarity ranking of focal computed tomography liver lesions using a 2AFC technique. In *Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment* (Vol. 9787, p. 97870N). International Society for Optics and Photonics.

Faltin, Annette. (2019). The National Museum of Art, Architecture and Design. In *Large Norwegian Encyclopedia*. Retrieved August 6, 2020 from https://www.snl.no/Nasjonalmuseet_for_kunst%2C_arkitektur_og_design

Fazal, N., Nguyen, K. Q., & Fränti, P. (2019). Efficiency of Web Crawling for Geotagged Image Retrieval. *Webology*, *16*(1).

Federal Agencies Digital Guidelines Initiative (FAGI). (2018). *Term: Metadata, embedded*. Retrieved from http://www.digitizationguidelines.gov/term.php?term=metadataembedded.

Feliciati, P. (2020) Improving the impact of digital cultural heritage: system, content and users' challenges. Published in : Meynard C. ; Lebarbé T. ; Costa S. (dir.) (2020), Patrimoine et Humanités numériques, Editions des archives contemporaines, doi : https://doi.org/10.17184/eac.9782813003843

Finnish National Gallery. (n.d.). *About Us* page. Retrieved May 29, 2020 from https://www.kansallisgalleria.fi/en/tietoa-kansallisgalleriasta

Fox, E. A. (Ed.) (1993). Sourcebook on Digital Libraries: Report for the National Science Foundation, TR-93-35 (439p.). Blacksburg, VA: VPI and SU Computer Science Department. Available by anonymous FTP from directorypub/DigitalLibrary on fox.cs.vt.edu or at http://fox.cs.vt.edu/DLSB.html.

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., & Sølvberg, I. (2001, September). Digital libraries: A generic classification and evaluation scheme. In *International Conference on Theory and Practice of Digital Libraries* (pp. 187-199). Springer, Berlin, Heidelberg.

Gaikar, Vishal. (2012, May 16). *6 Best Reverse image Search Engines*. Retrieved October 22 2019 from https://www.tricksmachine.com/2012/05/reverse-image-search-engines.html.

Gemäldegalerie, Staatliche Museen zu Berlin. (2020). *Home* page. Retrieved March 21,2020 from https://www.smb.museum/en/museums-institutions/gemaeldegalerie/home/

German Historical Museum. (n.d.). *About* page. Retrieved May 29, 2020 from https://www.dhm.de/en/about-us/about-us.html

Getty Images: COMPANY NEWS. 2016, Apr. 26. Getty Images to File Competition Law Complaint Against Google. Accessed on May 26, 2021 from https://perma.cc/L5MG-E33W

Giannini, T., & Bowen, J. (2019). Museums and Digital Culture: New perspectives and research.

GIF IT UP. (n.d.,a). *About* page. Retrieved April 30, 2021 from https://gifitup.net/about/.

GIF IT UP. (n.d.,b) *Winners 2020.* Retrieved April 30, 2021 from https://gifitup.net/.

Gilliland, A. J. (2016). Setting the Stage. In Baca.M (Ed.). In *Introduction to Metadata*. (Online) (3rd ed). Getty Publications, http://www.getty.edu/publications/intrometadata/setting-the-stage/.

Glowacka-Musial, M. (2020). Visualization and Digital Collections. *Library Technology Reports*, *57*(1), 5.

Green, H. E., & Lampron, P. (2017). User engagement with digital archives for research and teaching: A case study of Emblematica Online. *portal: Libraries and the Academy*, *17*(4), 759-775.)

Greenstein. D. (2002). Preface. In Covey, D. T. (Ed.) *Usage and Usability Assessment: Library Practices and Concerns*. Council on Library and Information Resources, Digital Library Federation.

Griffin, S. (1998). The NSF/DARPA/NASA Digital Libraries Initiative: A Program Manager's Perspective. *D-Lib Magazine*. July/August 1998. http://www.dlib.org/dlib/july98/07griffin.html

Górny, M., & Mazurek, J. (2012). Key Users of Polish Digital Libraries. *The Electronic Library* 30, no. 4 : 543-556.

Hamilton, E. (1992, September 1). *JPEG file interchange format*. Retrieved May 27,2020 from https://www.w3.org/Graphics/JPEG/jfif3.pdf

Hadro, J. (2015, January 21). NYPL Digital Collections Platform: An Introduction. *NYPL Labs*. Retrieved July 12, 2020 from https://www.nypl.org/blog/2015/01/21/digital-collections-platform-intro

Harran, M., Farrelly, W., & Curran, K. (2018). A method for verifying integrity & authenticating digital media. *Applied Computing and Informatics*, 14(2), 145-158.

Harris, V., & Hepburn, P., (2013). Trends in Image Use by Historians and the Implications for Librarians and Archivists. *College and Research Libraries* 74 (3):272–87.

Harvey, P. (2016). *Composite tags*. [Online article]. Retrieved August 22, 2020 from https://sno.phy.queensu.ca/~phil/exiftool/TagNames/Composite.html.

Hayes, Mark. (2018, March 30). *10 Must Know Image Optimization Tips*. Shopify Blogs. Retrieved January 5, 2019 from https://www.shopify.com/blog/7412852-10-must-know-image-optimization-tips on.

Hebron, T., & Mowry, A. (2021). Analytics for everybody: A literature review of analytics practices in multi-organization digital collections. *College & Undergraduate Libraries*, 1-10

Henderson, S. (2005). Genre, task, topic and time: facets of personal digital document management. In *Proceedings of the 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction: Making CHI Natural* (pp. 75-82). ACM.

Hicks, B. J., Dong, A., Palmer, R., & McAlpine, H. C. (2008). Organizing and managing personal electronic files: A mechanical engineer's perspective. *ACM Transactions on Information Systems* (TOIS), 26(4), 23.

Hillson, B. (2002). Gabriel: The Gateway to Europe's National Libraries. [Review of Gabriel Digital Library]. *American Library Association*. Retrieved March 8, 2020 from http://www.bowdoin.edu/~samato/IRA/reviews/issues/dec02/gabriel.html.

Howell, E. (2016, December). *Grace Hopper: 'First Lady of Software'*. Space.com. Retrieved March 13, 2020 from https://www.space.com/34885-grace-hopper-biography.html

Hu, X. (2018). Usability evaluation of E-Dunhuang cultural heritage digital library. *Data and Information Management*, 2(2), 57-69.

Inglis, K. & Prosser, D. Forward. Kay, D., & Stephens, O. (2014). *Improving discoverability of digitised collections: above-campus and national solutions: Recommendations from the Spotlight on the Digital co-design project*. Retrieved from http://repository.jisc.ac.uk/5569/1/JISC_REPORT_spotlight_on_digital_WEB.pdf

International Color Consortium. (n.d.) *FAQ, What is an ICC profile*. Retrieved May 27, 2020 from http://www.color.org/faqs.xalter#wh2.

International Press Telecommunications Council. (2021a). *IPTC Photo Metadata Standard.* Retrieved December, 14, 2020 from https://iptc.org/standards/photo-metadata/iptc-standard/.

International Press Telecommunications Council. (2021b). *IPTC. Metadata Uses Cases.* Retrieved December 14, 2018 from http://www.embeddedmetadata.org/metadata-use-cases-00.php.

Internet World Stats. (2020). *Internet Growth Statistics*. Retrieved March 18, 2020 from https://www.internetworldstats.com/emarketing.htm.

Jabeen, M., Qinjian, Y., Yihan, Z., Jabeen, M., & Imran, M., (2017). Usability Study of Digital Libraries: An Analysis of User Perception, Satisfaction, Challenges, and Opportunities at University Libraries of Nanjing, China, Library Collections, *Acquisitions, & Technical Services*, 40:1-2, 58-69

Jackie Robinson, Official Website. (2020). *Biography*. Retrieved July 17, 2020 from https://www.jackierobinson.com/biography/

Japan Search. (2021a) *About page.* Retrieved May 19, 2021 from https://jpsearch.go.jp/about.

Japan Search (2021b) *Home page.* Retrieved May 19, 2021 from https://jpsearch.go.jp/.

Jefcoate, G. (1996). "Gabriel: Gateway to Europe's National Libraries", Program: *Electronic Library and Information Systems*, Vol. 30 No. 3, pp. 229-238.

Jeng, J. (2005). Usability assessment of academic digital libraries: effectiveness, efficiency, satisfaction, and learnability. *Libri*, 55(2-3), 96-121.

Jimerson, L & Allison Leigh. (2020) Social Media in the Art History Classroom. Chapter in *The Routledge Companion to Digital Humanities and Art History.* 1st Edition.

Junior, J. R. F., Oliveira, M. C., & de Azevedo-Marques, P. M. (2017). Integrating 3D image descriptors of margin sharpness and texture on a GPU-optimized similar pulmonary nodule retrieval engine. *The Journal of Supercomputing*, 73(8), 3451-3467.

Kahn, R. E., & Cerf, V. G. (1988, March). *The Digital Library Project (Vol. 1): The world of knowbots: An open architecture for a digital library system and a plan for its development [DRAFT].* Corporation for National Research Initiatives. Retrieved October 11, 2020 from http://hdl.handle.net/4263537/2091

Kalisdha, A., & Suresh, C. (2017). Digital libraries: definitions, issues and challenges. *Science and Humanities*, 95.

Katira, C., Vora, N., Wali, K., & Medhekar, A. (2015). Advanced content based image retrieval using multiple feature extraction. *International Journal of Innovative Research in Science, Engineering and Technology*, 4(10), 9805-9812.

Kelly, E. J. (2014). Assessment of digitized library and archives materials: A literature review.

Journal of Web Librarianship, 8(4), 384-403.

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.

Kelly, E. J. (2018a). Use of Louisiana's Digital Cultural Heritage by Wikipedians. *Journal of Web Librarianship*, 12(2), 85-106.

Kelly, E. J. (2018b). Content Analysis of Google Alerts for Cultural Heritage Institutions. *Journal of Web Librarianship*, 12(1), 28-45.

Kenfield, A. S., Kelly, E., Muglia, C., O'Gara, G., Thompson, S., & Woolcott, L. (2019). Measuring reuse of institutionally-hosted Grey Literature. In *20th International Conference on Grey Literature: Research Data Fuels and Sustains Grey Literature, GL 2018 - Conference Proceedings* (pp. 83-90). (GL-Conference Series: Conference Proceedings; Vol. 2019-December). TextRelease.

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*. [Presentation]. MW2013: Museums and the Web 2013.

Kleppe, Martijn. (2015). Tracing the afterlife of iconic photographs using IPTC. *Digital Humanities*, 29.

Konkiel, S., Dalmau, M., & Scherer, D. (2015). *Altmetrics and analytics for digital special collections and institutional repositories.* Retrieved April, 15, 2019 from https://figshare.com/articles/Altmetrics_and_analytics_for_digital_special_collections_and_institutional_repositories/1392140.

Kontiza, K., Liapis, A., & Jones, C. E. (2020, September). Reliving the Experience of Visiting a Gallery: Methods for Evaluating Informal Learning in Games for Cultural Heritage. In *International Conference on the Foundations of Digital Games* (pp. 1-11).

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 and Technology* 61 (9):1734–44.

Krough, P. (2015). EXIF. Digital photography best practices and workflow. American Society of Media Photographers. Accessed May 27, 2021 from https://www.dpbestflow.org/metadata/exif

Krough, P. (2018). Metadata overview. Digital photography best practices and workflow. *American Society of Media Photographers*. Retrieved January 18, 2018 from http://www.dpbestflow.org/metadata/metadata-overview#handle.

Kunsthistorisches Museum, (n.d.). *Homepage and The Building* page. Retrieved March 21, 2020 from https://www.khm.at/en/explore/the-museum/the-building/

KHM-Museumsverband, (n.d.) *Financing notice* page. Retrieved March 1, 2021 from https://www.khm.at/entdecken/organisation/finanzierungshinweis/

Kupferstichkabinett, Staatliche Museen zu Berlin (2020). *About Us* page. Retrieved March 21, 2020 from https://www.smb.museum/en/museums-institutions/kupferstichkabinett/about-us/profile/

Lacinak, C. (2015) What to Do After Digitization. Brylawski, S., Lerman, M., Pike, R., & Smith, K. (Eds.). *ARSC Guide to Audio Preservation*. Pittsburgh: Association for Recorded Sound Collections. Retrieved December 12, 2020 from https://www.clir.org/wp-content/uploads/sites/6/pub164.pdf#page=139

Lagoze, C., Krafft, D. B., Payette, S., & Jesuroga, S. (2005). What is a digital library anymore, anyway. *D-Lib magazine*, 11(11), 1082-9873. https://www.immagic.com/eLibrary/GENERAL/DIG_LIB/D051100L.pdf

Liang, X., & Chen, Y. (2018). Libraries in internet of things (IoT) era. *Library Hi Tech*.

Lamont, M. (2014). Digital Library assessment through multiple measures. *Libraries in the Digital Age* (LIDA) Proceedings, 13.

Lesch, N. C. (2017). *A Software Framework for Image Retrieval and Visual Understanding in Dynamic and Sensor Rich Environments* (No. AFIT-ENG-MS-17-M-045). Air Force Institute of Technology Wright-Patterson AFB, OH. United States.

Library and Archives of Canada. (2017). *About Us* page. Retrieved March 21, 2020 from https://www.bac-lac.gc.ca/eng/about-us/Pages/our-mandate.aspx

Library and Archives of Canada. (2020). *Annual Report 2019-2020.* Retrieved December 12, 2020 from https://www.bac-lac.gc.ca/eng/about-us/annual-reports/annual-report-2019-2020/Pages/Annual-Report-2019%e2%80%932020.aspx

Library of Congress. (n.d.). *About the Collections*. Retrieved March 22, 2020 from https://memory.loc.gov/ammem/about/about.html

Library of Congress.(n.d.). *Digital Collections*. Retrieved October 1, 2018 from https://www.loc.gov/collections/

Library of Congress. (n.d.) Innovator in Residence Program. Retrieved June 14, 2021 from https://labs.loc.gov/about/opportunities/innovator-in-residence-program

Library of Congress. (2016, February 1). *Metadata Object Description Schema. Advantages of MODS.* Retrieved October 4, 2018 at http://www.loc.gov/standards/mods/mods-overview.html

Library of Congress. (2011). *Metadata for Digital Content (MDC). Developing institution-wide policies and standards at the Library of Congress*. Retrieved October 4, 2018 from https://www.loc.gov/standards/mdc/

Licklider, J. C. R., & Clapp, V. W. (1965). *Libraries of the Future*. Cambridge, MA: MIT Press.

Liew, C. L., Goulding, A., & Nichol, M. (2020). From shoeboxes to shared spaces: participatory cultural heritage via digital platforms. Information, Communication & Society, 1-18.

Lilly, C. (1995, December). Not just decoration: Quality graphics for the web. In *Fourth International World Wide Web Conference, Boston*.

Linder, R., Webb, A. M., & Kerne, A. (2013). Searching to Measure the Novelty of Collected Ideas. In *CHI 2013 Evaluation Methods for Creativity Support Environments Workshop* (Vol. 36).

Lloyd, Dave. (2014, April 21). *10 naming Conventions That Lead to SEO Success*. Adobe Blog. Retrieved January 5, 2019 from https://theblog.adobe.com/10-naming-conventions-lead-seo-success/ .

Mahey, M., Al-Abdulla, A., Ames, S., Bray, P., Candela, G., Chambers, S., Derven,C., Dobreva-McPherson, M., Gasser, K., Karner, S., Kokegei, K., Laursen, D., Potter, A., Straube, A., Wagner, S-C. and Wilms, L., with forewords by: Al-Emadi, T. A., Broady-Preston, J., Landry, P. and Papaioannou, G. (2019) *Open a GLAM Lab. Digital Cultural Heritage Innovation Labs*, Book Sprint, Doha, Qatar, 23-27 September, 2019.

Marchionini, G. & Plaisant, C. (1996). *User Interface for the National Digital Library Program Needs Assessment Report*. To: Library of Congress National Digital Library Program. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=DD14D93DEF5965014D5EFE1877F039EB?doi=10.1.1.20.8878&rep=rep1&type=pdf

Marchionini, G., Plaisant, C., & Komlodi, A. (1998). Interfaces and Tools for the Library of Congress National Digital Library Program. *Information Processing & Management*, 34(5), 535-555

Marchionini, G. (2000). Evaluating Digital Libraries: A Longitudinal and Multifaceted View. *Library Trends*. Vol. 49, No. 2, pp. 304333. Retrieved March 31, 2020 from IDEALS, Illinois Digital Environment for Access to Learning and Scholarship. https://www.ideals.illinois.edu/bitstream/handle/2142/8336/librarytrendsv49i2g_opt.pdf?sequence=1

Marcondes, C. H. (2020). Towards a Vocabulary to Implement Culturally Relevant Relationships Between Digital Collections in Heritage Institutions. *KO KNOWLEDGE ORGANIZATION*, *47*(2), 122-137.

Marques, O. (2016). Visual Information Retrieval: The State of the Art. *IT Professional*, 18(4), 7-9.

Matusiak, K. K. (2012). Perceptions of Usability and Usefulness of Digital Libraries. *International Journal of Humanities and Arts Computing*, 6(1-2), 133-147.

Matusiak, K. K. (2017). User Navigation in Large-Scale Distributed Digital Libraries: The Case of the Digital Public Library of America, *Journal of Web Librarianship*, 11:3-4, 157-171

Mauritshuis. (2020a). *Annual Report 2019*. Retrieved and downloaded June 1, 2020 from https://www.mauritshuis.nl/en/organisation/jaarverslag-2019/

Mauritshuis. (2020b). *Collection* page. Retrieved June 1, 2020 from https://www.mauritshuis.nl/en/explore/the-collection/

Mauritshuis. (2020c). *Home* page. Retrieved June 1, 2020 from https://www.mauritshuis.nl/en/.

Mauritshuis. (2020d). *History of the Building* page. Retrieved June 1, 2020 from https://www.mauritshuis.nl/en/discover/mauritshuis/history-of-the-building/

Mauritshuis. (2020e).*Organization* page. Retrieved June 1, 2020 from https://www.mauritshuis.nl/en/organisation/

Mauritshuis. (2020f). *Search* page. Retrieved June 1, 2020 from https://www.mauritshuis.nl/en/#

McCay-Peet, L., & Toms, E. (2009). Image use Within the Work Task Model: Images as Information and Illustration. *Journal of the American Society for Information Science and Technology*, 60(12), 2416-2429.

Münster, S., Kamposiori, C., Friedrichs, K., & Kröber, C. (2018). Image libraries and their scholarly use in the field of art and architectural history. *International journal on digital libraries*, *19*(4), 367-383.

Murray, Mike. (2017, January 31). *SEO and Website Image Optimization 10-Point Checklist for Marketers.* Content Marketing Institute. Retrieved April 25, 2019 from https://contentmarketinginstitute.com/2017/01/seo-website-image-optimization/

Musée d'Orsay, (2016-2020). *About* page. Retrieved August 1, 2020 from https://www.musee-orsay.fr/en/info/public-administrative-establishment/in-brief.html

Musée d'Orsay, (2019). *Annual Report*. Retrieved August 1, 2020 from https://www.musee-orsay.fr/fileadmin/mediatheque/integration_MO/PDF/Rapports_activ/RAMO2018.pdf

Myers, B. A. (2020). Year III, A Brief History of Human Computer Interaction Technology.

Nardi, B., Anderson, K., & Erickson, T. (1995). Filing and Finding Computer Files. *Proceedings of the East-West HCI*, Moscow, Russia.

National Digital Library of India. (2020). *Homepage search box*. Retrieved March 28, 2020 from https://ndl.iitkgp.ac.in/.

National Digital Library of India. (2020). *Sponsor* popup page. Retrieved March 28, 2020 from https://ndl.iitkgp.ac.in/.

National Library of Australia. (2020) *About Trove*. Retrieved March 21, 2020 from https://trove.nla.gov.au/about

Nationalmuseum. Sweden.(n.d.). *About* page. Retrieved July 1, 2020 from https://www.nationalmuseum.se/en/

National Museum of Art, Architecture and Design. (2020). *About* page. Retrieved August 1, 2020 from https://www.nasjonalmuseet.no/en/about-the-national-museum/.

New York Public Library. (NYPLDC). (2020). *Digital Collections*. Retrieved October 2, 2018 https://digitalcollections.nypl.org/.

New York Public Library (NYPL). (2020). *About* page. Retrieved March 20, 2020 from https://www.nypl.org/help/about-nypl/mission.

Nieuwenhuysen, P. (2018a). Image search process in the Web using image copy. *Journal of Multimedia Processing and Technologies*, *9*(4), 124-133.

Nieuwenhuysen, P. (2018b, February). Information Discovery and Images A Case Study of Google Photos. In *2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)* (pp. 16-21). IEEE.

Oliveira, L. F. M., & Kaster, D. D. S. (2017, July). Defining Similarity Spaces for Large-Scale Image Retrieval Through Scientific Workflows. In *Proceedings of the 21st International Database Engineering & Applications Symposium* (pp. 57-65). ACM.

O'Neill, R. (2017). The Rise of the Citizen Curator: Participation as Curation on the Web (Doctoral dissertation, University of Hull).

Orlandi, S. D. (2020). Museums Web Strategy at the Covid-19 Emergency Times. *DigitCult-Scientific Journal on Digital Cultures*, *5*(1), 57-66.

Padilla, T., (2017) *On a Collections as Data Imperative.* Accessed on April 28,2021 from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwj2iNLk9aHwAhUZCs0KHVnDBR4QFjAAegQIBxAD&url=https%3A%2F%2Flabs.loc.gov%2Fstatic%2Flabs%2Fwork%2Freports%2Ftpadilla_OnaCollectionsasDataImperative_final.pdf&usg=AOvVaw0MBl2Br6Lwi9X6QDypfO9B

Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E., & Varner, S. (2019). Always Already Computational: Collections as Data.Accessed on April 28, 2021 from https://zenodo.org/record/3152935#.YIncUqFOlhF

Parker-Wood, A., Long, D. D., Miller, E., Rigaux, P., & Isaacson, A. (2014, June). A File by any Other Name: Managing File Names with Metadata. In *Proceedings of International Conference on Systems and Storage* (pp. 1-11). ACM.

Patterson, Kerry. (2016). *Can I Just Google That? Orphan Works and Image Recognition Tools.*

[Paper]. (A. Wallace & R. Deazley, Eds). Display At Your Own Risk: An experimental exhibition of digital cultural heritage, 2016.

Perrin, J. M., Yang, L., Barba, S., & Winkler, H. M. (2017). All that glitters isn't gold: The complexities of usage statistics as an assessment tool for digital libraries. *The Electronic Library* Vol. 35 No. 1, pp. 185-197. https://doi.org/10.1108/EL-09-2015-0179

Petrov, M. D. (2015). A Medicine-Oriented Image Retrieval System based on Shearlet Transform. *International Journal of Computational Engineering Research (IJCER)*, 4(4), 83-87.

Pew Research Center: Internet & Technology. (2014). *World Wide Web Timeline*. Retrieved March 13, 2020 from https://www.pewresearch.org/internet/2014/03/11/world-wide-web-timeline/

Plaisant, C., Marchionini, G., Bruns, T., Komlodi, A., & Campbell, L. (1997, March). Bringing Treasures to the Surface: Iterative Design for the Library of Congress National Digital Library Program. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 518-525).

Pomerantz, J., Choemprayong, S. and Eakin, L. (2008), *The Development and Impact of Digital Library Funding in the United States.* Nitecki, D.A. and Abels, E.G. (Ed.) Influence of Funding on Advances in Librarianship (Advances in Librarianship, Vol. 31), Emerald Group Publishing Limited, Bingley, pp. 37-92. https://doi.org/10.1016/S0065-2830(08)31002-2

Portal to Texas History. (2020). *About* page. Retrieved May 29, 2020 from https://texashistory.unt.edu/about/portal/.

Praveena, K. (2019). Transforming Libraries Analog to Digital. Conference proceedings from the *National Seminar on "Education on Digital Culture and Social Media.*" S. Saileela & S. Kalaivani (Eds.). (pp. 625-628)

Princeton University Library. (2019). *Research Data Management at Princeton*. Retrieved December 13, 2020 from https://libguides.princeton.edu/c.php?g=102546&p=930626

Punzalan, R. L., Marsh, D. E., & Cools, K. (2017). Beyond Clicks, Likes, and Downloads: Identifying Meaningful Impacts for Digitized Ethnographic Archives. *Archivaria*, 84(1), 61-102.

Reeves, T. C., Buhr, S., & Barker, L. (2005, June). Evaluating Digital Libraries. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 420-420).

Rehberger, D. & Coates, B., (2019). *File Naming in the Digital Age. Oral History in the Digital Age.* Institute of Museum and Library Services. Retrieved from April 25, 2019 from http://ohda.matrix.msu.edu/2012/08/file-naming-in-the-digital-age/.

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.

Reilly, M., & Thompson, S. (2016). Reverse Image Lookup: Assessing Digital Library Users and Reuses. *Journal of Web Librarianship*, 11(1), 56-68.

Reser, G., & Bauman, J. (2009). Embedded Metadata, part I: The Basics and a History. *Images, the Newsletter of the VRA*, 6(6). http://metadatadeluxe.pbworks.com/w/page/20792223/Basics%20and%20a%20History

Reuters. (2009) *TIMELINE: Key dates in the history of the personal computer.* Technology News Retrieved January 6, 2009 from https://www.reuters.com/article/us-laptop-sb/timeline-key-dates-in-the-history-of-the-personal-computer-idUSTRE50601V20090107

Riggs, C., Douglas, T., & Gagneja, K. (2018, October). Image Mapping Through Metadata. In *2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications (SSIC)* (pp. 1-8). IEEE.

Rijksmuseum. (n.d.). *The Rijksmuseum Open Data Policy*. Retrieved May 29, 2020 from https://www.rijksmuseum.nl/en/data/policy

Rijksmuseum, (n.d.). *Vision and Mission*. Retrieved May 29, 2020 from https://www.rijksmuseum.nl/en/organisation/vision-and-mission

Riley, J. (2017). Understanding Metadata: What is Metadata, and What is it For?. *National Information Standards Organization*, 2017. iii, 45 p. ISBN: 978-1-937522-72-8. https://groups.niso.org/apps/group_public/download.php/17443/understanding-metadata

Rogers, J. (2014). File Naming Standards for Digital Collections. *U of I SLIS Journal*. Retrieved from https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1048&context=bsides

Saleh, E. I. (2018). Image Embedded Metadata in Cultural Heritage Digital Collections on the Web. *Library Hi Tech,* Vol. 36 No. 2, pp. 339-357. https://doi.org/10.1108/LHT-03-2017-0053

Saracevic, T. (2000). Digital Library Evaluation: Toward an Evolution of Concepts. *Library Trends*, Vol. 49, Issue 2. Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/8343/librarytrendsv49i2i_opt.pdf

Saracevic, T. (2004, September). Evaluation of Digital Libraries: An Overview. In *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries*. Padua, Italy.

Schreibman, S., Siemens, R., & Unsworth, J. A. (Eds.). (2004). *Companion to Digital Humanities.* Oxford: Blackwell. http://www.digitalhumanities.org/companion/

Schwartz, C. (2000). Digital Libraries: An Overview. *The Journal of Academic Librarianship*, 26(6), 385–393.

Siefring J. (2019) Democratizing Discovery: The Impact of Digital Culture on the Research Library. In: Giannini T., Bowen J. (eds) Museums and Digital Culture. Springer Series on

Cultural Computing. Springer, Cham. https://doi.org/10.1007/978-3-319-97457-6_25

Sfakakis, M., & Kapidakis, S. (2002, September). User Behavior Tendencies on Data Collections in a Digital Library. In *International Conference on Theory and Practice of Digital Libraries* (pp. 550-559). Springer, Berlin, Heidelberg.

Shapiro, L. G., & Stockman, G. C. (2001). *Computer Vision*. Prentice Hall.

Sharma, M., & Batra, A. (2014). Analysis of Distance Measures in Content Based Image Retrieval. *Global Journal of Computer Science and Technology.* Vol. 14, Issue 2, Version 1.0. Retrieved from https://core.ac.uk/download/pdf/231149922.pdf.

Sharma, V. K., & Chauhan, S. K. (2019). Digital Library Challenges and Opportunities: An Overview. *Library Philosophy and Practice (e-journal). Retrieved August*, *28*, 2020 from https://core.ac.uk/download/pdf/286730262.pdf

Shiri, A. (2003). Digital Library Research: Current Developments and Trends. *Library Review* 52 (5): 198 – 202.

Shiri, A., Kelly, E. J., Kenfield, A. S., Masood, K., Muglia, C., Thompson, S., & Woolcott, L. (2020, November). A Faceted Conceptualization of Digital Object Reuse in Digital Repositories. In *Knowledge Organization at the Interface* (pp. 402-410). Ergon-Verlag.

Siegler, M.G., (2010, July 20). *Google Image Search: Over 10 Billion Images, 1 Billion Pageviews A Day.* TechCrunch. Retrieved October 28, 2020 from https://techcrunch.com/2010/07/20/google-image-search/

Slovak National Gallery. (2020) *History SNG* webpage. Retrieved May 29, 2020 from https://www.sng.sk/en/about-us/history.

Smith, A. (1999). Why Digitize?. *Council on Library and Information Resources*. Retrieved September 27, 2018 from https://www.clir.org/pubs/reports/pub80-smith/pub80-2/

Smith, B. (2017). *A Qualitative Analysis of the Issue of Discoverability of Irish Digital Collections; with a Specific Focus on Medieval Irish Manuscripts* [Doctoral dissertation]. Dublin Business School.

Smith, K. R., Saunders, S., & Kejser, U. B. (2014, June). Making the Case for Embedded Metadata in Digital Images. In Archiving Conference (Vol. 2014, No. 1, pp. 52-57). *Society for Imaging Science and Technology*.

Somerville, M.M., & Brar, N. (2009). A User-centered and Evidence-based Approach for Digital Library Projects. *The Electronic Library*. 27, 409-425.

Soules, C. A., & Ganger, G. R. (2005, October). Connections: Using Context to Enhance File Search. In *ACM SIGOPS Operating Systems Review* (Vol. 39, No. 5, pp. 119-132). ACM.

Stafford-Fraser, Q. (1995). *The Trojan Room Coffee Pot*. Retrieved March 15, 2020 from

https://www.cl.cam.ac.uk/coffee/qsf/coffee.html.

Statens Museum for Kunst. (n.d.). *About*. Retrieved May 29, 2020 from https://www.smk.dk/en/section/about-smk/

Steinerova, J. (2007). Relevance Assessment for Digital Libraries. *Mousaion*, 25(2), 37-57.

Strzelichowska, A. (2020). *Why should you open up your digital collections for reuse - explained in GIFs.* Europeana Pro. Retrieved April 30, 2021 from https://pro.europeana.eu/post/why-should-you-open-up-your-digital-collections-for-reuse-explained-in-gifs

Surowiecki, J. (2004). *The Wisdom of Crowds : why the Many are Smarter than the few and how Collective Wisdom Shapes Business, Economies, Societies, and Nations.* 1st ed. New York: Doubleday.

Tsakonas, G., & Papatheodorou, C. (2006). Analysing and Evaluating Usefulness and Usability in Electronic Information Services. *Journal of Information Science*, 32(5), 400-419.

Terras, M., (2015). Opening Access to Collections: the Making and Using of Open Digitised Cultural Content. *Online Information Review*, Vol. 39 No. 5, pp. 733-752.

Terras, M. & Kirton, I., (2013, April 17-20). Where Do Images of Art Go Once They Go Online? A Reverse Image Lookup Study to Assess the Dissemination of Digitized Cultural Heritage. [Paper] *Museums and the Web Conference*. Portland, Oregon. http://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/

Terras, M., Coleman, S., Drost, S., Elsden, C., Helgason, I., Lechelt, S., Osborne, N., Panneels, I., Pegado, B., Schafer, B., Smyth, M., Thornton, P., & Speed, C. (2021). The value of mass-digitised cultural heritage content in creative contexts. *Big Data & Society*, *8*(1), DOI: 10.1177/20539517211006165

The University of Edinburgh. (2018, May 16). *Short but Meaningful Names*. Retrieved April 24, 2019 from https://www.ed.ac.uk/records-management/guidance/records/practical-guidance/naming-conventions/meaningful-names

The University of Edinburgh. (n.d.). *Organising Data. Records Management Section*. Retrieved from https://mantra.edina.ac.uk/organisingdata/

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 and Technology*, 68(9), 2264-2266.

Thompson, S. & Reilly, M. (2018). "Embedded Metadata Patterns Across Web Sharing Environments." *International Journal of Digital Curation*. Vol. 13, No. 1 : 223-234. https://doi.org/10.2218/ijdc.v13i1.607

Thompson, S., & Reilly, M. (2019). " Everyone's a Curator": Identifying the Everyday Curator. *International Journal of the Image*, *10*(2).

Thompson, S., Woolcott, L., Muglia, C., O'Gara, G., Kenfield, A. S., & Kelly, E. J. (2019). Assessing Transformation: Findings from the Measuring Reuse Project. Retrieved April 29, 2021 from https://osf.io/8yuxh/download.

Thong, J. Y., Hong, W., & Tam, K. Y., (2002). Understanding User Acceptance of Digital Libraries: What are the Roles of Interface Characteristics, Organizational Context, and Individual Differences?. *International Journal of Human-Computer Studies*, 57(3), 215-242.

Toevs, B. (2015, November). Processing of Metadata on Multimedia Using ExifTool: A Programming Approach in Python. [Paper]. In *Global Online Conference on Information and Computer Technology (GOCICT)* (pp. 26-30). IEEE.

Trivedi, M. (2010). Digital libraries: functionality, usability, and accessibility. Library Philosophy and Practice, 381, 2010.

Toms, E. G., Dufour, C., & Hesemeier, S. (2004, June). Measuring the User's Experience with Digital Libraries. In Proceedings of the *2004 Joint ACM/IEEE Conference on Digital Libraries*, 2004. (pp. 51-52). IEEE.

Ullrich, S., & Geis, K. (2021). Between the Extraordinary and the Everyday. *Material Image*, 117.) vol.7 iss.7

United States Code of Federal Regulations. (2019). *Title 21 – Food and Drugs. Quality System Regulations*. Sec. 820.3 Definitions. Retrieved on November 18, 2020 from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=820.3

United States National Archives and Records Administration (NARA). (2004). *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files– Raster Images*. Retrieved November 24, 2020 from https://sustainableheritagenetwork.org/system/files/atoms/file/NARA_Guidelines.pdf

Valeonti, F., Terras, M., & Hudson-Smith, A. (2019). How open is OpenGLAM? Identifying barriers to commercial and non-commercial reuse of digitised art images. *Journal of Documentation*.)

Van House, N., Butler, M., & Schiff, L. (1996a). Needs Assessment and Evaluation of a Digital Environmental Library: the Berkeley Experience. In *DL96: The First ACM International Conference on Digital Libraries.* School of Information Management and Systems, University of California, Berkeley, Bethesda, MD

Van House, N. A., Butler, M. H., Ogle, V., & Schiff, L. (1996b). User-centered Iterative Design for Digital Libraries. *D-lib Magazine*, 2(2). Retrieved March 5, 2019 from http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/february96/02vanhouse.html

Wagner, K. (2020). Random display and recommendations–exploring the web platform of the artist Ivar Arosenius and other digital collections of art. *Museum and Society*, *18*(2), 243-257.

Wan, G. G., & Liu, Z., (2008). Content-based Information Retrieval and Digital Libraries. *Information Technology and Libraries* 27 (1):41–47.

Waters, D.J. (1998). What are Digital Libraries? *CLIR Issues,* (4). Retrieved March 20, 2020 from https://www.clir.org/1998/07/clir-issues-number-4/

Waysdorf, A. S. (2021). Remix in the age of ubiquitous remix. *Convergence*, 1354856521994454.

Wellcome Collection. (n.d.) *About Wellcome Collection*. Retrieved May 29, 2020 from https://wellcomecollection.org/pages/Wuw2MSIAACtd3Stq

Wellcome Collection. (2017, October) Wellcome Collection: Who we are and what we do. *Tenth Anniversary Report.*

Wittmann, R., Neatrour, A., Cummings, R., & Myntti, J. (2019). From digital library to open datasets: Embracing a "collections as data" framework. *Information Technology and Libraries (Online), 38*(4), 49-61. Retrieved May 5, 2021 from https://search.proquest.com/scholarly-journals/digital-library-open-datasets-embracing/docview/2336304747/se-2?accountid=8361

World Digital Library. (n.d.,a). *About* page. Retrieved March 20, 2020 from https://www.wdl.org/en/about/

World Digital Library. (n.d.,b). *Filenaming Guidelines*. Retrieved September 13, 2020 from https://project.wdl.org/standards/filenaming.html

Xie, I., Babu, R., Joo, S., & Fuller, P. (2015). Using digital libraries non-visually: understanding the help-seeking situations of blind users. *Information Research: An International Electronic Journal*, 20(2), n2. Retrieved from https://files.eric.ed.gov/fulltext/EJ1065689.pdf .

Xie, I., Joo, S., & Matusiak, K. K. (2021). Digital library evaluation measures in academic settings: Perspectives from scholars and practitioners. *Journal of Librarianship and Information Science*, *53*(1), 130-152.)

York University. (n.d.). *Tip Sheet 6 - Naming Conventions for Electronic Files and Folders*. Retrieved April 24, 2019 from http://ipo.info.yorku.ca/tool-and-tips/tip-sheet-6-naming-conventions-for-electronic-files-and-folders/

Zimmermann, K. (2017). *History of Computers: A Brief Timeline*. Live Science. Retrieved March 6, 2020 from https://www.livescience.com/20718-computer-history.html.

Appendices

Appendix #1 – Institute of Museum and Library Services(IMLS) - Top 25 public libraries by holdings

Top 25 public libraries by holdings

Data from Institute of Museum and Library Services, Public Libraries Survey (PLS) Fiscal Year: 2016

For **Library Collection**, the PLS Fiscal Year 2016 Data Element Definitions state (slightly edited): This section of the survey collects data on selected types of materials. Under this category, report only items the library has acquired as part of the collection, whether purchased, leased, licensed, or donated as gifts.

There are also specific definitions for Print Materials, Electronic Books (E-Books), Audio - physical units, and Video - physical units.

- PLS 2016 Data Element Definitions

Rank	Library Name	Print Materials	Electronic Books	Audio Materials	Video Materials	Total Collection
1	New York Public Library, NY	21,935,221	1,788,493	826,360	721,149	25,271,223
2	Public Library of Cincinnati And Hamilton County, OH	4,936,677	5,889,477	362,864	532,412	11,721,430
3	Boston Public Library, MA	7,997,591	95,927	35,220	68,272	8,197,010
4	Los Angeles Public Library, CA	5,922,133	171,493	251,557	390,378	6,735,561
5	Chicago Public Library, IL	5,451,914	46,195	263,711	187,431	5,949,251
6	County of Los Angeles Public Library, CA	4,734,248	40,403	301,380	703,812	5,779,843
7	Queens Borough Public Library, NY	4,972,997	85,829	120,658	491,079	5,670,563
8	San Diego Public Library, CA	4,715,696	96,457	197,095	263,531	5,272,779
9	Dallas Public Library, TX	4,563,199	68,076	167,061	357,311	5,155,647
10	Hennepin County Library, MN	4,366,003	253,743	167,667	107,899	4,895,312
11	Dayton Metro Library, OH	871,785	3,670,432	80,361	97,096	4,719,674

Appendix #2 - CBIR/RIL verification study data

Complete data set is available in electronic format upon request. File name: CH4-Google-verification-study.pdf . Upon request.

Appendix #3 - Embedded metadata verification study data

Complete data set is available in electronic format upon request. File name: CH5-embedded-metadata-study-data.pdf . Upon request.

Appendix #4 - File naming assessment investigation original image information data.

Complete data set is available in electronic format upon request. File name: CH6-01-Original-Image-Info.pdf . Upon request.

Appendix #5 - File naming assessment investigation Google Images search information data

Complete data set is available in electronic format upon request. File name: CH6-02-Google-ImageSearch.pdf · Upon request

Appendix #6 - File naming assessment investigation File name change results data

Complete data set is available in electronic format upon request. File name: CH6-03-File-Name-change_Results.pdf . Upon request.

Appendix #7 - File naming assessment investigation aggregated results data

Complete data set is available in electronic format upon request. File name: CH6-04-Aggregated-results.pdf· Upon request

Appendix #8 – File naming assessment investigation raw data

Complete raw data set is available in pdf format upon request. File name: ch6-file-name-investigation-study-raw-data.pdf . Upon request.