



9-19-2019

## Thinking About Digital Archives: One Tool at a Time

Jeannine Keefer  
*University of Richmond*

Greta Bahnemann

Follow this and additional works at: <https://scholarship.richmond.edu/art-faculty-publications>



Part of the [Arts and Humanities Commons](#), and the [Library and Information Science Commons](#)

---

### Recommended Citation

Bahnemann, Greta, and Jeannine Keefer. "Thinking About Digital Archives: One Tool at a Time." In #DLFteach Toolkit 1.0: Lesson Plans for Digital Library Instruction. PubPub, 2019. <http://doi.org/10.21428/65a6243c.b6e5b193>.

This Article is brought to you for free and open access by the Art and Art History at UR Scholarship Repository. It has been accepted for inclusion in Art and Art History Faculty Publications by an authorized administrator of UR Scholarship Repository. For more information, please contact [scholarshiprepository@richmond.edu](mailto:scholarshiprepository@richmond.edu).

**#DLFteach • #DLFTeach Toolkit: Lesson Plans for Digital Library Instruction**

# **Thinking About Digital Archives: One Tool at a Time**

**Greta Bahnemann, Jeannine Keefer**

**Published on:** Sep 19, 2019

**DOI:** 10.21428/65a6243c.b6e5b193

**License:** [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

## **Session Specifics**

Can be used for workshop or for introductory lessons in a semester-long class

## **Instructional Partners**

Archivists, Digital Humanities Librarians, Digital Humanities faculty, discipline-specific faculty

## **Audience**

- In the academic world: mid-track undergraduate students, senior capstone students, graduate student methods, and faculty interested in Digital Collections.
- Staff working with cultural heritage and archival collections in museums, libraries, and historical societies.

## **Curricular Context**

Archival research, Digital Humanities methods using primary source materials, and digital collection building

## **Learning Outcomes**

Participants/students will have been exposed to a variety of practices and methods applied to digital collection development and exploitation. The lessons can be used to introduce digital collection methods to a variety of audiences including: undergraduates, first year graduate students, and faculty in an academic setting, as well as staff at cultural heritage organizations, archivists, and librarians. Participants can build out a comprehensive Digital Humanities project using an archive or improve skills in any of the areas covered.

## **Preparation**

The lesson plan is conceived as a kit of resources. It includes glossaries, links to sources, and a sample archive, as well as individual lessons. The instructor should have a baseline comfort with the key concepts and best practices associated with digital collection work. In preparation for teaching these modules, they should familiarize themselves with the key concepts and the tools provided, and understand the wider application of the lessons included. Audience members require no prior knowledge or preparation, but it is assumed that audience members possess basic research skills

## **Materials**

All materials are included in the kit. The strategies can then be applied to local archives and digital collections or any other analog or digital collections. It is assumed that the instructor will have prior

knowledge of the tools and concepts. The kit is intended to be a first introduction to these tools and concepts.

## Session Outline

The kit below includes an introduction, basic definitions, project and tool examples, a sample archive, and exercises that can be performed on those materials using the tools and concepts.

## Thinking About a Digital Archive: One Tool at a Time

*A road trip is just a road trip until it's a digital project*



Created by

Greta Bahnemann and Jeannine Keefer

October 1, 2018

## Table of Contents

Introduction

- Using the Kit
- Thinking About Digital Collections: Purpose & Applications

Techniques and Tools

1. Text Analysis
2. Image Analysis
3. Online Exhibits

4. Mapping and Geospatial Projects
5. Sharing with an Aggregate
6. Creating Metadata
7. Copyright & Rights Management
8. Transcription & OCR

## A Sample Digitized Archive

- Resources 1-12: Digital photographs
- Resources 13-14: Postcards
- Resources 15: Journal
- Resource 16-17: Brochures
- Resource 18: Map
- Metadata table for sample digitized archive

## Metadata Mapping Tables

- Table 1. Field Mapping to the Dublin Core Metadata Element Set
- Table 2. Crosswalks: Dublin Core - MODS - VRA CORE 4.0

## Exercises and Activities

- Exercise #1: Text Analysis
- Exercise #2: Image Analysis
- Exercise #3: Image Analysis
- Exercise #4: Online Exhibits
- Exercise #5: Mapping and Geospatial Projects
- Exercise #6: Sharing with and Aggregate
- Exercise #7: Creating Metadata
- Exercise #8: Creating Metadata
- Exercise #9: Creating Metadata
- Exercise #10: Copyright and Rights Management
- Exercise #11: Transcription and OCR

## Introduction: Using the Kit

This resource kit is designed to help users create, explore, and leverage digital collections. It provides a myriad of ideas on pedagogical and practical approaches to the various issues associated with digital collections, including metadata creation, geospatial metadata, copyright, image and text analysis, and more.

The kit offers concrete starting points for skill building, finding vetted and reliable examples, and viewing creative applications in multiple disciplines. Each section includes: a definition, a brief overview of the resource, links to examples of the idea or concept, and provides links to vetted content which provide additional information. Note: The definition is derived from the resource’s website or from current scholarship relating to this resource. The examples show that particular component or part in context of a real-world application. The list of additional resources and tools enables users to learn more about different areas of digital collection practice.

The projects, websites, and software mentioned in the “Examples” and “Additional Resources” sections are provided by the authors to serve as illustrative examples. They are in no way an endorsement of said products, projects, and services.

## Thinking About Digital Collections: Purpose & Applications

Digital collections are created to preserve rare or fragile materials, provide access to unique or valuable materials, and to increase the exposure of all collections. But these reasons are just the tip of the iceberg in terms of understanding the usefulness of digital collections. Digital collections can also serve as the content source and provide structure for classes and teaching; exhibits and storytelling; in promotional materials for your department or organization; and serve as the foundations of digital methods of inquiry in small and large scale digital humanities projects.

## Techniques & Tools

The following resources and exercises will help users understand the concepts, considerations, issues, and problems involved in building and exploiting digital collections. They will familiarize users with ways in which archive and collection content can be interrogated, shared, and understood.

### 1. Text Analysis

Text Analysis is a broad term covering the many different processes by which language-based documents examine the various ways in which people make sense of who they are and how they fit into the larger world. Texts are parsed to extract machine-readable data and facts. This creates sets of structured data out of unstructured documents.

Text analysis is performed on text-based materials, which includes everything from books, films, television programs, magazines, advertisements, and more. Text analysis allows users to see both similarities in patterns, as well as differences in how individuals interpret the world around them and the world’s cultures and subcultures. . Text Analysis is is sometimes referred to as Text Mining. Text Analysis tools help users both visualize and interpret the resulting analysis. In addition to tools ready

for use, programming languages are often employed to perform text analysis activities including Python and R.

### **Examples:**

- [America's Public Bible](#) (Lincoln Mullen.com)

Explores trends and frequency in use of biblical quotations in newspapers mined from Chronicling America: Historic American Newspapers (LoC).

- [Early Modern Print: Text Mining Early Printed English](#) (Washington University, St. Louis)

Provides an introduction and explanation of the various tools and visualizations that are possible using the EEBO Text Creation Partnership using the XML/SGML encoded transcriptions of early printed books in Early English Books Online. Presents examples with tools and visualizations, such as the N-gram browser and corpus analysis of English print culture before 1700 in EEBO-TCP.

- [Martha Ballard's Diary](#) (Cameron Blevins)

Text mining, analysis, and topic modeling of Martha Ballard's diary (1785-1812) who documented daily life as a midwife in Maine.

- [Mining Biodiversity](#) (Biodiversity Heritage Library)

Text mining methods are being used to produce semantic metadata and index, along with visualization, crowdsourcing, and social media, to provide enhanced access to Biodiversity Heritage Library documents.

- [Mining the Dispatch](#) (Digital Scholarship Lab, University of Richmond)

The full run of the Richmond Daily Dispatch from November 1860 to April 1865 is text mined and presented with topic models based on prominent topics found within articles of this newspaper. Charts display the topic proportions by month in all articles containing that topic, transcriptions of articles can be viewed with additional topics identified

### **Additional Resources:**

- [Textual Analysis: A Beginner's Guide](#) by Alan McKee

[Chapter 1, "What is Textual Analysis"](#) guides students away from finding the "correct" interpretation of a text and explains the various nuances and meanings found in text-based materials.

- [Introduction to Text Analysis: About Text Analysis](#), Duke University Libraries

Text analysis Library Guide from Duke University including a concept overview, web scraping, and analysis methods and tools.

- [Voyant Tools](#)

Voyant Tools is a web-based digital text reading and analysis environment. It is intended to facilitate reading and interpretive practices for digital humanists and the general public. Users can type or paste in text or URLs and obtain analyzed results.

- [Mallet](#)

Java based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction and other machine learning applications to text.

### [TAPoR 3](#)

TAPoR 3 is a gateway for tools for sophisticated analysis and retrieval, along with representative texts for experimentation. The gateway has a curated list of tools for Web Scrapers, Data modeling and management, Voyant Tools, social media analysis, and more.

## 2. Image Analysis

Image Analysis is the extraction of any meaningful data from images. Images can take on a multitude of meanings and subjects. Digital images' meaning is primarily extracted by means of visual analysis conducted by humans, though there are algorithms that allow a machine to discern patterns such as faces, objects, etc. in digital images.

The human eye, however, can discern layers of meaning and intent beyond the simple content of a digital item. Mechanical image analysis continues to evolve and improve with recent improvement in pattern recognition and facial recognition, but as of this writing it remains an inexact science. While the examples below come out of the discipline of art history, the methods are the same, no matter the content of the image. Cataloguers will likely need to conduct research on the image to accurately describe it. These descriptions can be as simple as “a baby sitting on a woman’s lap” or as nuanced as “Christ sitting on the lap of the Virgin Mary holding a pomegranate and giving the blessing.” Both are correct, but one provides more context and layers of possible meaning.

### **Examples:**

- [Introduction to Art Historical Analysis](#) by Dr. Robert Glass

While this source focuses on the analysis of art, the methods employed to research and describe a work of art can be applied to any object or image. Dr. Glass also covers the object, visual (formal)

analysis, style, cultural artifacts, subject, iconography, and function.

- [\*\*Picturing America\*\*](#)

Picturing America was a project of the National Endowment for the Humanities (NEH) that brought masterpieces of American art into classrooms and libraries nationwide. The project concluded in 2009; but many of the educational materials created for the program are still available for use by students, teachers, and lifelong learners.

- [\*\*Beyond the Taj: Architectural Traditions and Landscape Experience in South Asia\*\*](#)

Beyond the Taj is a collection of visual and written materials assembled to support instruction and research on South Asian architectural expression in the joint perspectives of architecture and ethnography. The visual core of the collection consists of approximately 7000 photographs of works of architecture, pilgrimage locales and domestic life taken in India and Sri Lanka by Professors Robert D. Scotty MacDougall (1940-1987) and Bonnie G. MacDougall (1941-2017), an anthropologist and an architect.

**Additional Resources:**

- [\*\*Online Reference Sources for Cataloging Visual Materials\*\*](#)

The selected resources represent free Web sites used by Library of Congress staff to catalog material in a general documentary picture collection. Many of the sites feature visual images for easy comparison to what you are cataloging.

- [\*\*HistoryProfessor.org Image Analysis\*\*](#)

Website is maintained by Zachary M. Schrag, Professor of History at George Mason University. His image analysis advice is designed primarily for undergraduate and graduate students studying United States history.

- [\*\*SymbolDictionary.net: A Visual Dictionary\*\*](#)

Includes multiple entries on symbolism in material culture. There are pages devoted to Symbols in Christian Art, Ancient Greek and Roman Symbols, Symbols of Death, Masonic, and Norse Symbols, etc.

- [\*\*Proceedings of the National Academy of Science\*\*](#)

Article on face recognition accuracy of forensic examiners, super recognizers, and facial recognition algorithms.

### 3. Online Exhibits & Digital Storytelling

Online exhibits provide a space to display, interpret, and create a narrative for digital collections. Exhibits can serve any and/or all of the following goals: preservation, access, promotion/awareness, and pedagogy. The platform used to create an online exhibit might be as simple as a web page or as complex as a full digital arts or humanities project. Exhibit platforms range from expensive software for purchase to open source options freely available online. Keep in mind that free, open source options often require maintenance that may result in additional costs as well as demand staff time and expertise.

Online exhibits are just one form of digital storytelling. There are other ways to add context, meaning, and access points to digital content, including collection guides, finding aids, and primary source sets. These methods often result in less complex results than a narrative-based exhibit and are faster and easier to create; but they can engage users with digital content in new and innovative ways.

#### Examples

- [Digital Public Library of America Exhibitions page](#)

Stories of national significance drawn from source materials in Libraries, Archives, and Museums across the United States.

- [Digital Public Library of America Primary Source Sets](#)

Explore various topics in history, literature, and culture via sets of primary source materials developed by educators, librarians, and museum professionals complete with teaching guides for classroom use.

- [Minnesota Reflections' Primary Source Sets](#)

Modeled from the DPLA's Primary Source Sets, these resources are designed to help students develop critical thinking skills by exploring a variety of topics related to Minnesota history and culture. Using both primary and secondary sources, these sets bring together different resources in new ways to help students better understand historic events and people in their context.

- ["Making Modern America: Discovering the Great Depression and New Deal"](#)

This site was created by students at the University of Oklahoma enrolled in the Presidential Dream Course, "Making Modern America: Discovering the Great Depression and New Deal" during Fall 2015. During the course, students became immersed in history, politics, and culture of the 1930s through lectures, readings, film screenings, and field trips. Note: uses Omeka.

- ["When Nixon Met Elvis"](#)

Online exhibit at the National Archives that tells the behind-the-scenes story of how and why the meeting between Elvis Presley and President Richard Nixon in the Oval Office occurred - told through original photographs, letters, and memos.

## Additional Resources

- [Omeka](#)

A leading open source web publishing platform for digital collections. First funded by the Institute of Museum and Library Services (IMLS) from 2007-2010, Omeka provides an option for museums, libraries, and archives wishing to publish collections and narrative exhibits to the web.

- [Collective Access](#)

CollectiveAccess is software for describing all manner of things, and allows users to create catalogues that closely conform to user needs without custom programming.

- [Collection Space](#)

CollectionSpace is a free, open-source collections management application that meets the needs of museums, historical societies, biological collections, and other collections-holding organizations. CollectionSpace is designed to be configurable to each organization's needs, serving as a gateway to digital and physical assets across an institution.

- [Open Exhibits](#) <http://openexhibits.org>

Open Exhibits is an initiative that looks to transform the way in which museums and other informal learning institutions produce and share computer-based exhibits. Open Exhibits is both a collection of software and a growing community of practice.

- [Drupal](#)

Drupal is a content management software. It's used to make many websites and digital applications. Drupal is open source software. Anyone can download, use, work on, and share it with others. It is built on the principles of collaboration, globalism, and innovation. It is distributed under the terms of the GNU General Public License (GPL).

- [Knight Lab](#)

Northwestern University Knight Lab is a community of designers, developers, students, and educators working on experiments designed to push journalism into new spaces. Easy to use

storytelling tools such as Storymap, TimelineJS, SoundciteJS, and Juxtapose allow users to create narratives using digital media, text, maps, and timelines.

- [Scalar](#)

Created by the Alliance for Networking Visual Culture and housed at the University of Southern California, Scalar is an open source digital authoring and publishing platform created for long-form, born-digital scholarship online. The platform allows authors to utilize nested, recursive, and non-linear narrative formats. Authors can assemble media from a variety of sources including YouTube, Vimeo, and [scholarly] digital collections with minimal technical expertise required. Scalar features a built-in API which allows authors to share their Scalar content with other applications.

## 4. Mapping and Geospatial Projects

A number of different standards currently exist for creating and capturing geographic metadata. Adding Geographic Metadata or Geolocations to digital collections can aid in user discovery, access, and use of digital items. Adding geolocation metadata can make a digital collection ready for future geospatial based projects and activities. Mapping and Geospatial data can also be gleaned from primary texts, reports, and other scholarly resources. Some mapping and geospatial projects might serve only as research while others can be digital humanities projects.

The tools available for these projects vary greatly in terms of cost, complexity, and user friendliness. Those listed under “Additional Resources” address tools for formatting and creating the metadata to those employing that metadata to make a map. Some mapping programs focus on narrative, while others are powerful engines for processing large amounts of geospatial data into legible and useful maps.

### Examples:

- [Minnesota Digital Library](#)

MDL has applied GeoNames to all items in their public collection “Minnesota Reflections.” Each item in Minnesota Reflections has been assigned a geolocation using GeoNames and includes a GeoNames URI.

- [Middle Eastern Geodatabase for Antiquities \(MEGA\)](#)

The MEGA-Jordan project was launched as a collaboration of the Getty Conservation Institute (GCI) and the World Monuments Fund (WMF) with the DoA for the development and implementation of a geospatial information system to inventory and manage Jordan's numerous archaeological sites.

- [Global Crossroads: The World's 25 Busiest Airports](#)

View the world's 25 busiest airports using geospatial data to obtain a sense of their gargantuan scale and global significance.

- [Renewing Inequality](#) by Digital Scholarship Lab, University of Richmond

*Renewing Inequality* visualizes the displacement of hundreds of thousands of families under the urban renewal initiatives of the 1950s and 1960s.

Viewing them from above gives a sense of their gargantuan

#### **Additional Resources:**

- [Getty Thesaurus of Geographic Names \(TGN\)](#)

TGN is intended to aid cataloging, research, and discovery of art historical, archaeological, and other scholarly information. However, its unique thesaurus structure and emphasis on historical places make it useful for other disciplines in the broader Linked Open Data cloud. For GIS information, TGN may be linked to existing major, general-purpose, geographic databases.

- [GeoNames](#)

The GeoNames geographical database is available for download free of charge under a creative commons attribution license. It contains over 25 million geographical names and consists of over 11 million unique features whereof 4.8 million populated places and 13 million alternate names. All features are categorized into one of 9 feature classes and further subcategorized into one out of 645 feature codes.

- [Best Practices for Geospatial Metadata Creation for MWDL Cultural Heritage Digital Repositories](#)

In July of 2013, the Mountain West Digital Library Geospatial Discovery Task Force convened with the purpose of developing a standard format for recording geospatial metadata.

- [Knight Lab StoryMaps](#)

StoryMapJS is a free tool to help users tell stories on the web that highlight the locations of a series of events.

- [ArcGIS](#)

Mapping and analysis: location intelligence for everyone. ArcGIS Online is a mapping and analysis product. Users can use it on its own or expand their work with other ArcGIS products.

Users can share and integrate their work with other ArcGIS users. In addition to interactive maps, you can create StoryMaps to add narrative to your work.

- [QGIS](#)

QGIS is a free and open source geographic information system that allows you to create, edit, visualize, analyze, and publish geospatial information on Windows, Mac, and Linux platforms.

- [CARTO Location Intelligence Software](#)

CARTO is the platform to build powerful location intelligence apps with the best data streams available.

- [Neatline](#)

Neatline is a geo-temporal exhibit builder that allows you to create complex maps, image annotations, and narrative sequences from Omeka collections of archives and artifacts, and to connect your maps and narratives with timelines that are more-than-usually sensitive to ambiguity and nuance. You can import these documents (georeferenced historical maps, manuscripts, high-res photographs, etc.) from an existing collection, or create a new digital archive, yourself. Neatline offers the user the ability to not just place dots on maps, but to conceive and create complex meaningful visual representations of scholarship. Neatline only works with the installed version of Omeka. See [Neatline Quickstart Guide](#) for information on how to get started.

## 5. Sharing with an Aggregated Collection

The digital collection landscape is ever changing and evolving. No longer must digital collection projects stand alone. In today's world more and more institutions and organizations are choosing to participate in aggregations of collections. These aggregations can happen across departments and divisions at large institutions, across organizations to create regional or local collaborations, and can also include large-scale projects such as participating in a statewide aggregation such as the Minnesota Digital Library, the Digital Library of Georgia, the California Digital Library, and many others. Aggregated collections can also cross state, regional, or national boundaries. Examples are: the Mountain West Digital Library, the Digital Public Library of America, and Europeana.

### Examples:

- [Digital Public Library of America Professional Community](#)

Resource page for DPLA Service and Content hubs as well as prospective members.

- [Minnesota Digital Library](#)

The Minnesota Digital Library (MDL) supports discovery and education through access to unique digital collections shared by cultural heritage organizations from across the state of Minnesota. MDL works with 188 contributing partners from cultural heritage organizations across Minnesota.

- [Digital Library of Georgia](#)

The Digital Library of Georgia is a GALILEO initiative based at the University of Georgia Libraries that collaborates with Georgia's Libraries, archives, museums, and other institutions of education and culture to provide access to key information resources on Georgia history, culture, and life.

- [Mountain West Digital Library](#)

Mountain West Digital Library provides free access to over 960,000 resources from universities, colleges, public libraries, museums, historical societies, and government agencies, counties, and municipalities in Utah, Nevada, Idaho, Montana, Hawaii, and other parts of the U.S. West.

- [Europeana](#)

Europeana works with thousands of European archives, libraries, and museums to share cultural heritage for general consumption and research. Users have access to over 50 million digitised items including books, music, artworks, and more.

#### **Additional Resources:**

- [Digital Public Library of America Terms and Conditions](#)

DPLA's Terms of Service including DPLA's Privacy Policy, account creation, restrictions on use, and licensed content.

- [Digital Public Library of America Metadata Application Profile](#)

The DPLA Metadata Application Profile (MAP) is the basis for how metadata is structured and validated in DPLA, and guides how metadata is stored, serialized, and made available through our API in JSON-LD.

## 6. Creating Metadata

Metadata. No term is more tied to the creation of digital content; and no single term is more confusing, intimidating, or misunderstood. Metadata is essentially information; and refers to any kind of content: analog or digital. Metadata can describe single objects such as photographs and maps, and it can also include describe objects made up of multiple parts such as books, pamphlets, archaeological sites, and

the built environment. Other examples of objects made up of component parts include: a set of historical postcards, a set of video-taped course lectures, or an oral history made up of an audio file, transcript, and image of the person being interviewed.

Metadata can be embedded into the file OR it can be in a system that is external to the file, such as a database. It is important to realize that metadata provides users with multiple access points into a digital collection. It identifies and describes content; and it disambiguates content. Metadata also enables content to be discovered and facilitates user searching and browsing.

**Metadata can be further divided into three major categories:**

1. Descriptive Metadata
2. Structural Metadata
3. Administrative Metadata -

Administrative Metadata is further differentiated into three sub-parts:

1. Technical Metadata
2. Preservation Metadata
3. Rights Metadata

**Metadata Definitions:**

- **Descriptive Metadata** - Information that refers to the physical attributes and intellectual content of material. Descriptive metadata aids in the discovery and identification of such materials.
- **Structural Metadata** - Information about the relationship between the intellectual or physical elements of a digital object. Structural Metadata denote whether a resource is a simple/single page or compound/complex object.
- **Administrative Metadata** - Data necessary to manage and use information resources. Typically external to the “informational content” of the resource.
- **Technical Metadata** - Data captures the information about the PROCESS of digitizing an item and provides the requirements for using a digital item.
- **Preservation Metadata** - Information about an object used to protect the digital object from harm, injury, deterioration, or destruction.
- **Rights Metadata** - Information about how, when, and where an object can be used/consumed, shared, reused and repurposed and/or transformed.

Metadata serves many purposes. While it can describe, identify, and disambiguate content, it can also relate information associated with the physical ownership of the originals as well as the rights holders. Metadata can also provide valuable information about the digital production of materials, including

file size, equipment used for format conversion, and information on the master files and web access files.

Metadata can also provide more sophisticated forms of access and understanding. It can help users understand the meaning associated with the content. The meaning associated with content can be tied to specific creators, styles, and time periods.

- Is it a painting of a mother and child OR is it a painting of the Virgin Mary and the infant Christ?
- Is it a black and white image of a house OR is it a photograph of “Fallingwater,” an example of modern, organic architecture designed by Frank Lloyd Wright?
- Is it a drawing of a table OR is it a drawing of the altar at St. Stephen’s Cathedral in Vienna, Austria?

A parade is just a parade until we are given the tools to understand more. A user’s understanding of such an image changes when the user knows it is a Temperance Parade or a 4th of July Parade, or a worker’s strike parade, etc.

This understanding of both content and context can inform how we describe the digital resource. Descriptive Metadata can provide clues as to the meaning, the context, and the relationships associated with the analog original. Without metadata, the original meaning, understanding of relationships, and intent are essentially lost.

Metadata is structured into a schema or order. A metadata schema is essentially a logical plan showing the relationships between the metadata elements. It establishes the rules for both the use and management of metadata (note: sometimes a metadata schema is referred to as a “metadata element set” or a “data structure standard”). A metadata schema might be homegrown, or it might conform to one of several standardized schemas available such as Dublin Core, MODS, and VRA Core. Some schema are purposefully broad such as Dublin Core to accommodate all types of content, while others, such as the VRA Core, have been developed with a specific content type in mind (i.e. visual content of art, architecture, urban planning, and cultural objects). These discipline-specific schemas are used to address specific elements needed by a discipline. Decisions about the structure of metadata for a digital collection should be determined by the digital collection, its content, and its users.

Creating a more detailed metadata schema allows metadata creators to create detailed, precise, and unambiguous metadata records. In common speak, creating distinct fields for creators, descriptions, dates, and subjects, etc. in your metadata schema allows metadata creators to know what specific information is populated in each field. Ideally metadata should also conform to controlled vocabularies. Using controlled vocabularies provides standardized language for names, types, formats, and subjects. This use of standardized language vs. natural language enables content to be searched,

browsed, and shared. Metadata schemas can be mapped to one another in order to share information between databases, collections, and institutions. See *Metadata Mapping Tables: Table 1 and Table 2*.

### **Relational vs. Non-Relational Cataloging**

The creators of digital collections can approach the naming, identifying, and the structuring of relationships in two different ways. The first approach is referred to as “Relational Cataloging.” This approach is predicated on the creation of separate work and image records that establish and connect relationships.

Separate records are created for different creators, place names, and repositories, and individual works of art/architecture, etc. This type of cataloging creates a structure that is built by linking these different records to each other. A painting for Vincent van Gogh will utilize a creator record for van Gogh, a work record for the individual painting *The Yellow House*, and a repository record for the museum which holds the original (in this case the Vincent van Gogh Museum in Amsterdam). This approach results in highly interconnected digital content which can easily show data patterns and relationships. Relational cataloging reuses data so that complex relationships and connections between works and multiple images of that work, as well as creators and repositories can be leveraged and displayed. This type of cataloging lends itself to works of art, architecture, urban planning, decorative arts, archaeology, etc. A schema that provides for relational cataloging is VRA Core and the companion work *Cataloging Cultural Objects*, which provides thorough explanations on how to use the schema.

The second approach to cataloging digital content is less relationship-based and is more frequently employed by the creators of digital collections of cultural heritage materials. The difference can be partially accounted for by the differences in the collection types. Cultural Heritage digital collections are comprised of a vast amount of unique and unpublished content, such as photographs, diaries, letters, and journals. This content is highly specific to times, places, and communities, with little connectedness to other things (for example: a photograph of Main Street, Richmond, Virginia has little connection to a photograph of Main Street, International Falls, Minnesota).

Finally, Dublin Core does not fully allow for or acknowledge complex and interconnected relationships in its metadata. While it does have a simple relationship field it does not express the interconnected relationships found in disciplines such as archeology and art history. An example that shows that hierarchical relationship is a pair of bronze candlesticks that are displayed on an altar. That altar is located in a chapel, which is in turn one part of a large cathedral. Each of those components should have a metadata record that relates it to the other parts and shows the hierarchy. When using Dublin Core, digital content creators must build relationships in other ways using controlled vocabularies, subject headings, etc. This web of connectedness is created via mindful and intentional metadata work.

**Examples:**

- [Digital Public Library of America Metadata Application Profile](#)

The DPLA Metadata Application Profile (MAP) is the basis for how metadata is structured and validated in DPLA, and guides how metadata is stored, serialized, and made available through our API in JSON-LD. The MAP was originally developed in 2012 and has been updated occasionally since. It is based on the Europeana Data Model (EDM), and integrates the experience and specific needs for aggregating the metadata of America's cultural heritage institutions. The current version is 4.0.

- [Library of Congress Digital Library Standards](#)

Web page with numerous links to current information on resource description formats and digital library standards.

- [Metadata Data Structure Standards, Government of Canada](#)

Metadata standards from the Canadian perspective including specific management and documentation guidelines for various collection types including: Humanities, Art, Visual Resources, Architecture, Ethnological and Archaeological Collections, and Sites and Monuments.

- [Minnesota Digital Library Metadata Guidelines](#)

The MDL Metadata Guidelines provide organizations contributing collections to Minnesota Reflections with detailed information and assistance on completing the data entry process of their projects.

- [Metropolitan Museum of Art \(Open Access Collections\)](#)

The Metropolitan Museum of Art's online collection includes over 375,000 hi-res images of public-domain works of art which can be downloaded, shared, and remixed without restriction.

- [Field Museum \(Anthropological Collections\)](#)

“Anthropology is all about what makes us human, our place in nature, our common concerns, and our differences. We explore these ideas through laboratory and collections-based research at the Museum and at field sites throughout the world. We build and maintain the Museum's world class collection, which now includes more than a million and a half objects, documenting the diversity and accomplishments of humankind. Through registration, conservation, collections management, and curation we preserve this collection and its documentation in order to connect communities, researchers, and the public to our shared global heritage.

Due to the size of our collection, please be aware that there may be errors and inconsistencies in the data presented here. We are continually updating and correcting our data. We welcome scholars, members of descendant communities, and others to contact us for confirmation or clarification on data you find here. If you would like to request a confirmation, a correction, or send an update to a record, please contact us.”

- [University of Massachusetts Amherst Libraries Metadata Guidelines](#)

These Metadata Guidelines were written to better position UMass Amherst Libraries' Digital Collections for optimal indexing and display in an aggregated environment.

#### **Additional Resources:**

- [Dublin Core Metadata Initiative](#)

The Dublin Core Metadata Initiative, or "DCMI", is an open organization supporting innovation in metadata design and best practices across the metadata ecology. DCMI's activities include work on architecture and modeling, discussions and collaborative work in DCMI Communities and DCMI Task Groups, global conferences, meetings and workshops, and educational efforts to promote widespread acceptance of metadata standards and best practices.

- [METS Metadata Encoding & Transmission Standard](#)

The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation.

- [MODS Metadata Object Description Schema](#)

The MODS schema for a bibliographic element set may be used for a variety of purposes, and particularly for library applications. The standard is maintained by the Network Development and MARC Standards Office of the Library of Congress with input from users.

- [Encoded Archival Description](#)

Encoded Archival Description is an XML standard for encoding archival finding aids. EAD is maintained by the Technical Subcommittee for Encoded Archival Standards of the Society of American Archivists, in partnership with the Library of Congress.

- [Library of Congress Authorities](#)

Use Library of Congress Authorities to browse and view authority headings for Subject, Name, Title and Name/Title combinations, and download authority records in MARC format for use in a local library system. This service is offered free of charge.

- [The Getty Research Institute/Getty Vocabularies](#)

The Getty vocabularies contain structured terminology for art, architecture, decorative arts, archival materials, visual surrogates, conservation, and bibliographic materials. Compliant with international standards, they provide authoritative information for catalogers, researchers, and data providers. The vocabularies grow through contributions. In the new linked, open environments, they provide a powerful conduit for research and discovery for digital art history.

## 7. Copyright & Rights Management

Information about the rights that may be held in and over a digital item is referred to as a copyright or rights management. In recent years, the digital collections community has moved away from collection-specific statements and towards more standardized statements, such as the work outlined at RightsStatements.org. This move towards understanding and assigning standardized rights statements aims to establish as rights as just another form of standardized metadata. Creating standardized data about the use and reuse of digital materials relates to larger issues of usability and Linked Open Data.

Copyright and rights management are governed and determined by the nation of origin and vary greatly. Copyright in the United States is determined on a case by case basis.

### Examples:

- [Applying Rightsstatements.org to Digital Public Library of America](#)
- [Applying Rightsstatements.org to Europeana](#)
- The New York Public Library has applied the rightsstatements.org statements to the items in their collections in DPLA.
- [Metropolitan Museum of Art, Image and Data Resources Open Access Policy](#)

On February 7, 2017, The Metropolitan Museum of Art implemented a new policy known as Open Access, which makes images of artworks it believes to be in the public domain widely and freely available for unrestricted use, and at no cost, in accordance with the Creative Commons Zero (CC0) designation and the Terms and Conditions of this website.

### Additional Resources:

- [Code of Best Practices in Fair Use in the Visual Arts](#)

The Code of Best Practices in Fair Use was created with and for the visual arts community. Copyright protects artworks of all kinds, audiovisual materials, photographs, and texts (among other things) against unauthorized use by others, but it is subject to a number of exceptions designed to assure space for future creativity. The Code describes common situations in which there is a consensus within the visual arts community about practices to which this copyright doctrine should apply and provides a practical and reliable way of applying it.

- [RightsStatements.org](https://rightsstatements.org)

RightsStatements.org provides a set of standardized rights statements that can be used to communicate the copyright and re-use status of digital objects to the public. Our rights statements are supported by major aggregation platforms such as the Digital Public Library of America and Europeana. The rights statements have been designed with both human users and machine users (such as search engines) in mind and make use of semantic web technology.

- [Digital Image Rights Computator](https://www.digitalimagerights.com)

The Digital Image Rights Computator (DIRC) program is intended to assist the user in assessing the intellectual property status of a specific image documenting a work of art, a designed object, or a portion of the built environment. Understanding the presence or absence of rights in the various aspects of a given image will allow the user to make informed decisions regarding the intended educational uses of that image.

- [Creative Commons](https://creativecommons.org)

Creative Commons is a global nonprofit organization that enables sharing and reuse of creativity and knowledge through the provision of free legal tools. Our legal tools help those who want to encourage reuse of their works by offering them for use under generous, standardized terms; those who want to make creative uses of works; and those who want to benefit from this symbiosis. Our vision is to help others realize the full potential of the internet. CC has affiliates [all over the world](https://creativecommons.org/all-over-the-world) who help ensure our licenses work internationally and who raise awareness of our work.

## 8. Transcription & OCR

Converting documents, text, images, and sound files to digital and/or machine-readable formats is a prerequisite for many digital collection and digital humanities projects. Digitization is the process of capturing analog materials as digital images. Both transcription and OCR take digitized files one step further by creating new resources that are more accessible and understandable to a larger audience.

Transcription is the process of translating audio and/or video files into a text format. Transcription can be done manually to recover and display information that is difficult to read in an original document either due to script, printing, or scans that are difficult to read. Another option is to use Optical Character Recognition (OCR) programs to “read” digitized images of book pages, newspaper, etc. and convert them to text-based documents which can be easily and fairly accurately full-text searched. Users can then copy, edit or used this data for computational text analysis, convert to a more readily accessible format, etc. The quality of OCR can vary tremendously and is dependent on a number of factors including: resolution of the scanned document, the format and quality of the original print materials, and how well the OCR program can deal with diacritical marks. Often OCR-ed text must be fixed manually to be fully and accurately searchable. Each collection must balance imperfect OCR against fair and equal access to materials within the context of ADA compliance and how useful the text can be for those using the works for research.

**Examples:**

- [The Collegian](#)

The Collegian is the student newspaper at the University of Richmond. The archive has been digitized and OCRd, but users are invited to manually correct errors as they are searching the archive.

- [Virginia Untold: African American Narrative](#)

The names of millions of African Americans, slave and free, who lived, worked, worshiped, loved, and died in Virginia, are buried deep in the archival records and manuscript collections housed at the Library of Virginia. Untold Virginia: African American Narrative seeks to find these long silent voices. Whether contained in local court and state government records, private papers and business records, or newspapers and journals from the time, the untold narrative of a people is waiting to be discovered. Users are invited to transcribe the documents online. This is a crowdsourced transcription project and relies on the public to provide initial transcriptions.

- [Minnesota Historical Society - Transcribing, Editing, and Processing Guidelines](#)

Provides an overview for users in the standardized creation of transcripts for oral histories, as well as editing and processing guidelines.

- [Columbia University Center for Oral History Research - Oral History Transcription Style Guide](#)

The Columbia Center for Oral History Research (CCOHR) provides a range of education and research services (including the new 2018 Transcription Style Guide). CCOHR has resources to help users learn about oral history, as well as resources to help users design and construct their own projects. CCOHR

can also put users in touch with others in their city or community who might be doing similar work. CCOHR can share, upon special request, the “Telling Lives” curriculum guide developed for middle school students.

### **Additional Resources:**

- [Adobe Acrobat Pro](#)

- [ABBYY Finereader](#)

ABBYY FineReader is a robust tool for OCR. ABBYY FineReader works well with digital camera images, unusually structured text (e.g. magazine layouts, newspaper columns), offers automated workflows for conversion, and supports up to 190 different languages.

- [Tesseract](#)

Tesseract is an open source OCR engine. It can be used directly (via the command line) or with an API. Several third-party graphical user interfaces (GUI) are available for users who would like a drag-and-drop interface. Specialized packages for working with different languages and scripts, such as cuneiform and Vietnamese, are also available.

- [Google Docs](#)

Google Docs allows users to perform OCR on uploaded images and PDFs. See this blog for a walkthrough and screenshots.

- [Scripto](#)

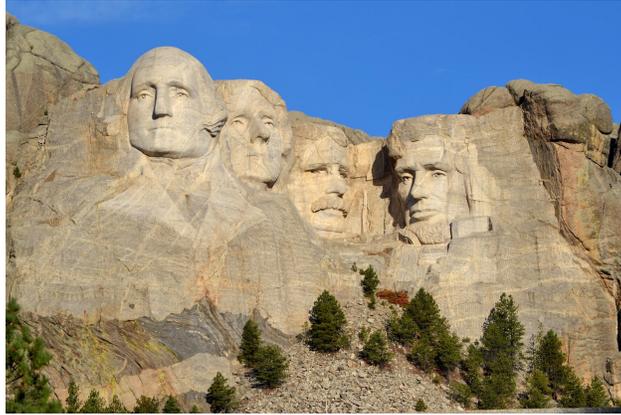
Scripto brings the power of [MediaWiki](#) to your collections. Designed to allow members of the public to transcribe a range of different kinds of files, Scripto will increase your content’s findability while building your user community through active engagement.

- [From the Page](#)

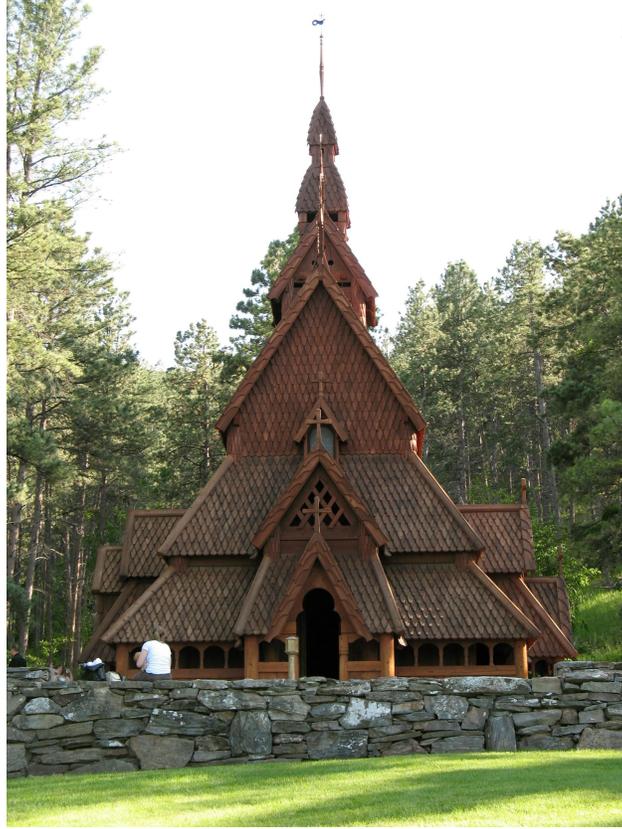
Software for transcribing documents and collaborating on transcriptions with others.

## A Sample Digitized Archive

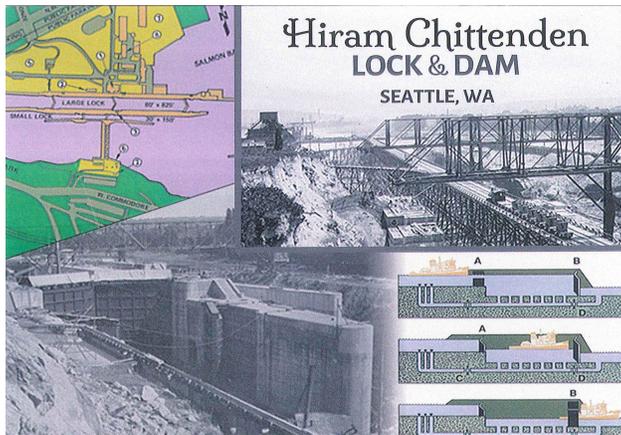












**Hiram Chittenden Lock & Dam**  
Seattle, WA

Printed in the  
**USA**



The Hiram M. Chittenden Locks, opened in 1917 and often nicknamed the Ballard Locks, provides a link for boats between the salt water of Puget Sound and the fresh water of the Ship Canal, which connects eastward to Lake Union and Lake Washington.

*Greta,  
At the end of a long  
road trip, watching ships  
embark on a different kind  
of adventure. How all is  
well in the Twin Cities  
I will see you soon  
Jeanine*

GRETA B.  
2929 UNIVERSITY AVE SE  
MINNEAPOLIS, MN  
55414



THIS AREA FOR OFFICIAL POSTAL USE ONLY

© IMPACT PHOTOGRAPHICS

78132-46



**S-M** SEATTLE ART MUSEUM

Wendy,  
What a great exhibit -  
in particular given the  
route of the road trip  
across the U.S. along  
Rte 90 - And look at  
the hair! Padmini Amidala  
has writing on them!  
Gee

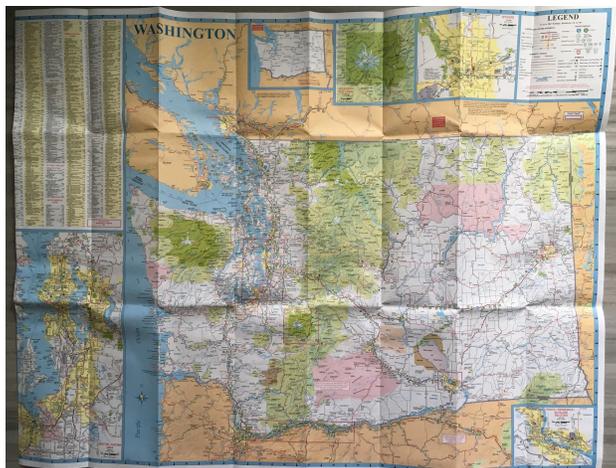
**Exposure**  
alignod

APFEX STAMP HERE

WENDY RED STAR  
200 N. Boulevard  
RICHMOND, VA  
23220

© 2002 by Seattle Art Museum. All rights reserved. Printed in the USA. For more information, please contact Seattle Art Museum, 1400 Western Avenue, Seattle, WA 98104. Tel: 206.462.3000. Fax: 206.462.3001. Email: info@artmuseum.org. Website: www.artmuseum.org.

U.S. 10-2



**Resource #1**

**Title:** “The Great Corn Palace!”

**Creator:** Image by sunnyamyk

**Description:** Exterior view of the 2013 Corn Palace in Mitchell, SD.

**Type:** Still Image

**Format:** Color photographs

**Date Created:** 2013-08-05

**Software/Hardware:** Apple iPhone 5

**Rights:** CC BY-NC-SA 2.0

**Link:** <https://flic.kr/p/gphkvg>

**GeoNames URI:** <http://sws.geonames.org/5227038/>

**Subject Headings:**

Corn Palace (Mitchell, S.D.);

Crop art;

Concert halls;

Moorish Revival;

Onion domes;

Minarets



**Resource #2**

Pink Elephant “Pinkie”, DeForest, WI

Image by: Eric Wilcox, CC BY-NC 2.0

November 17, 2009

Agfa CLACK

Image title: “My Pink Elephant”

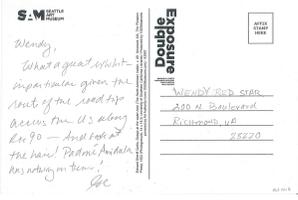
<https://flic.kr/p/7gyraG>



<p><b>Resource #3</b></p> <p>Witches Gulch, Wisconsin Dells, WI</p> <p>Image by: jmehre, CC BY-NC-SA 2.0</p> <p>May 23, 2008</p> <p>Image title: “Wisconsin Dells: Witches Falls, immediately followed by Witches Bathtub in Witches Gulch”</p> <p><a href="https://flic.kr/p/4QhE7t">https://flic.kr/p/4QhE7t</a></p>	
<p><b>Resource #4</b></p> <p>Mount Rushmore, Mount Rushmore National Park, Keystone, SD</p> <p>Image by: Diana Morales, CC BY-NC 2.0</p> <p>October 21, 2015</p> <p>Nikon D3100</p> <p>Image title: “Mount Rushmore: Mount Rushmore National Park”</p> <p><a href="https://flic.kr/p/ATqFne">https://flic.kr/p/ATqFne</a></p> <p>Color photographs</p>	
<p><b>Resource #5</b></p> <p>Jolly Green Giant, Blue Earth, MN</p> <p>Image by: Mykl Roventine, CC BY-NC-SA 2.0</p> <p>August 5, 2006</p> <p>Sony DSC-W30</p> <p>Image title: “World’s Largest Jolly Green Giant: 55-feet tall - ho ho ho! A creation of F.A.S.T Corp, Blue Earth, Minnesota”</p> <p><a href="https://flic.kr/p/KAej">https://flic.kr/p/KAej</a></p>	

<p><b>Resource #6</b></p> <p>1880 Cowboy Town, Buffalo Ridge, SD</p> <p>Image by: Nick Sherman, CC BY-NC-SA 2.0</p> <p>August 30, 2014</p> <p>Apple iPhone 5s</p> <p>Image title: "OLD WeST SALOON: 1880 Cowboy Town; Buffalo Ridge, South Dakota"</p> <p><a href="https://flic.kr/p/p2p2FL">https://flic.kr/p/p2p2FL</a></p>	
<p><b>Resource #7</b></p> <p>F.A.S.T. Corporation graveyard, Sparta, WI</p> <p>Image by: sporst, CC BY 2.0</p> <p>September 4, 2017</p> <p>Google Pixel</p> <p>Image title: "Fast Corporation graveyard of fiberglass molds, Sparta, WI"</p> <p><a href="https://flic.kr/p/XUQ7io">https://flic.kr/p/XUQ7io</a></p>	
<p><b>Resource #8</b></p> <p>Chapel in the Hills, Rapid City, SD</p> <p>Image by: Jake DeGroot, CC BY-SA 3.0</p> <p>July 5, 2008</p> <p>Image title: "The front of the Chapel in The Hills in Rapid City, SD"</p> <p><a href="https://commons.wikimedia.org/wiki/File:ChapelInTheHills3.JPG">https://commons.wikimedia.org/wiki/File:ChapelInTheHills3.JPG</a></p>	

<p><b>Resource #9</b></p> <p>Garnet Ghost Town, Garnet, MT</p> <p>Image by Kyle Freeman, CC BY-NC-SA 2.0</p> <p>September 7, 2013</p> <p>Image title: “Garnet Ghost Town - Garnet, MT: Abandoned mining town in the mountains”</p> <p><a href="https://flic.kr/p/fPbCNb">https://flic.kr/p/fPbCNb</a></p>	
<p><b>Resource #10</b></p> <p>Snoqualmie Falls, Snoqualmie, WA</p> <p>Image by: Jeannine Keefer, CC BY-NC-SA 2.0</p> <p>June 29, 2014</p> <p>Canon PowerShot S90</p> <p>Image title: “Snoqualmie Falls”</p> <p><a href="https://flic.kr/p/o7DoAG">https://flic.kr/p/o7DoAG</a></p>	
<p><b>Resource #11</b></p> <p>The Troll (Fremont Troll), Seattle, WA by Steve Badanes, Will Martin, Donna Walter, and Ross Whitehead</p> <p>Image by: Jeannine Keefer, CC BY-NC-SA 2.0</p> <p>June 14, 2018</p> <p>iPhone SE</p>	
<p><b>Resource #12</b></p> <p>Ballard Locks, Seattle, WA</p> <p>Image by: Jeannine Keefer, CC BY-NC-SA 2.0</p> <p>June 14, 2018</p> <p>iPhone SE</p>	

<p><b>Resource #13 a</b></p> <p><b>Title:</b> Hiram Chittenden “Ballard Locks” postcard (obverse), Seattle, Washington</p> <p><b>Creator:</b> Impact Photo Graphics</p> <p><b>Description:</b> Multi-view postcard of the Hiram Chittenden Lock and Dam. The Hiram M. Chittenden Locks opened in 1917 and are nicknamed the Ballard Locks. The Lock and Dam links Puget Sound with the freshwater Ship Canal, which connects to Lake Union and Lake Washington. The multi-view postcard includes images of: a small map, dam construction, and a cross-section diagram demonstrating how locks work.</p> <p><b>Type:</b> Still Image</p> <p><b>Format:</b> Postcards</p> <p><b>Materials:</b> paper</p> <p><b>Dimensions:</b> 10.5 x 15.7 cm</p> <p><b>Subject Headings:</b></p>	
<p><b>Resource #13b</b></p> <p>Ballard Locks Postcard: Reverse</p> <p>Materials: paper</p> <p>Measurements: 10.5 x 15.7 cm</p>	
<p><b>Resource #14a</b></p> <p>Edward Curtis Postcard: Obverse</p> <p>Materials: coated paper</p> <p>Measurements: 10.2 x 15.2 cm</p>	
<p><b>Resource #14b</b></p> <p>Edward Curtis Postcard: Reverse</p> <p>Materials: coated paper</p> <p>Measurements: 10.2 x 15.2 cm</p>	

<p><b>Resource #15a</b></p> <p>Road Trip Journal</p> <p>Creator: Jeannine Keefer</p> <p>Format: PDF</p> <p>Description: PDF of original handwritten journal</p>	 <p>The images show three pages of a handwritten journal. The first image shows two pages with text starting with 'East Sea Journal' and 'Day 1'. The second image shows two pages with text starting with 'Linnerton we had to find' and 'Day 2'. The third image shows a single page with text starting with 'Day 3'.</p>
<p><b>Resource #15b</b></p> <p>Road Trip Journal</p> <p>Creator: Jeannine Keefer</p> <p>Format: Transcript</p> <p>Description: Transcript of handwritten journal</p>	<p>Road Trip Journal</p> <p>Day 1</p> <p>◆?!Today we set off on our adventuer. We picked up the car and headed west out of Chicago along Route 90. Our destination is Seattle. I think we will take some time so we can see the weird and wonderful things along the way...</p> <p>Eleven hours later and we managed to make it to Mitchell, SD - home of the Corn Palace! We can't wait to see it in the daylight. Along the way we made sure to see an number of fiberglass roadside wonders.</p> <p>Day 2</p> <p>The Corn Palace - not just any corn palace - "The World's ONLY Corn Palace". This Moorish Revival beauty actually functions! While its decoration is made fresh each year, the structure sers as a venue for concerts, sports events, exhibits, and other community happenings. After</p>

visiting the Governor Palace we did some antiquing and had a meal at Fanny Harris's Eating Establishment. Tomorrow we head to Mount Rushmore with a stop at the Cowboy Town.

Day 3-4

Mount Rushmore. I keep expecting to see characters from *North by Northwest*. We must push on though (will update everyone w/ the pics we have taken!). Our next stop is Rapid City to find a hotel and then see the Capel in the Hills tomorrow.

Day 5

Long day on the road as we ran across Montana from Rapid City to Missoula with a stop at the Garnet Ghost Twn. If we had come in April rather than June, we could have rented a cabin there! Oh well, a quick tour will have to suffice this time.

Day 6

Missoula to Snoqualmie, WA. Mountains, Mountains, Mountains. We are stopping in Snoqualmie for a few days before we hit Seattle. Aside from, or rather in addition to, the *Twin Peaks* locations (the falls and Twede's Cafe in North Bend for cherry pie), we plan to take advantage of the spa to revive after the long drive.

Day 8

Seattle! Our itinerary: Fremont Troll, Ballard Locks, Seattle Art Museum... everything we can see before we need to return home.

**Resource #16**

Mount Rushmore brochure (1965)

[Mount Rushmore Brochure](#)

**Resource #17**

Ginkgo Petrified Forest State Park

[Ginkgo Petrified Forest State Park brochure](#)

<p><b>Resource #18</b></p> <p>Map of Interstate 90 in Washington State.</p> <p><b>Title:</b> Washington Oregon</p> <p><b>Creator:</b> AAA</p> <p><b>Date:</b> 2/17-5/18</p>	
---	--

## Mapping to the Dublin Core Metadata Element Set

<p><b>Table 1. Field Mapping to the Dublin Core Metadata Element Set</b></p> <p>The purpose of this table is to show that field names are not always the same as the field mapping name; and that not every field maps to an established element set. It also shows the distinction between Descriptive, Administrative, and Technical Metadata.</p>	
<p><b>Digital Collection</b></p> <p><b>With User-Friendly Field Names</b></p>	<p><b>Mapping to Dublin Core Metadata Element Set, Version 1.1</b></p>
<p><b>Descriptive Metadata</b></p>	
<p>Title</p>	<p>Title</p>
<p>Creator</p>	<p>Creator</p>
<p>Contributor</p>	<p>Contributor</p>
<p>Description</p>	<p>Description</p>
<p>Date Created</p>	<p>Date-Created</p>
<p>Publishing Agency</p>	<p>Source</p>
<p>Dimensions</p>	<p>Format - Extent</p>

Type	Type
Physical Format	Format - Medium
Library of Congress Subject Headings	Subject
Keywords	Subject
City or Township	Coverage - Spatial
County	Coverage - Spatial
State or Province	Coverage - Spatial
Country	Coverage - Spatial
Geographic Feature	Coverage - Spatial
GeoNames URI	Coverage - Spatial
Language	Language
<b>Administrative Metadata</b>	
Collection Name	Relations - Is Part Of
Contributing Institution	Publisher
Contact Information	None
Rights Management	Rights
Local Identifier	Identifier
MDL Identifier	Identifier
Project Affiliation	None
Fiscal Sponsor	None
<b>Technical Metadata</b>	
Date Digital	Date - Available

Item Digital Format	Identifier
Master File Format	None
Master File Size	None
Master File Bit Depth	None
Master File Resolution	None
Master File Compression	None
Master File Width	None
Master File Height	None
Master File Hardware	None
Master File Software	None
Master File System	None
Master File Checksum	None
Transcript	None

## Metadata Crosswalks

<p><b>Table 2. Crosswalks</b></p> <p><b>Dublin Core - MODS - VRA CORE 4.0</b></p> <p>Metadata Crosswalks translate element values from one schema to another. Crosswalks facilitate interoperability between different metadata schemas; and they serve as the core basis for metadata harvesting, as well as exchanging records between different projects, different collections, and different institutions. Table 2.0 shows the field crosswalk/mapping between Dublin Core (DC), Metadata Object Description Schema (MODS), and the VRA Core 4.0.</p>		
DC	MODS	VRA CORE 4.0
Title	<titleInfo><title>	Title Type
Creator	<name><namePart>	Agent Attribution Culture Dates earliestDate latestDate Name Role = Creator

Contributor	<name><namePart>	<p>Agent</p> <ul style="list-style-type: none"> <li>Attribution</li> <li>Culture</li> <li>Dates</li> <li>earliestDate</li> <li>latestDate</li> <li>Name</li> <li>Role = Contributor or Commissioner</li> </ul>
Subject	<p>&lt;subject&gt;</p> <p>&lt;topic&gt;</p> <p>&lt;name&gt;</p> <p>&lt;occupation&gt;</p>	<p>Subject</p> <ul style="list-style-type: none"> <li>Term</li> <li>Notes</li> </ul>
Description	<p>&lt;abstract&gt;</p> <p>&lt;note&gt;</p> <p>&lt;tableOfContents&gt;</p>	<p>Description</p> <p>Inscription</p>
Publisher	<p>&lt;originInfo&gt;</p> <p>&lt;Publisher&gt;</p>	<p>Agent</p> <ul style="list-style-type: none"> <li>Attribution</li> <li>Culture</li> <li>Dates</li> <li>earliestDate</li> <li>latestDate</li> <li>Name</li> <li>Role = Publisher</li> </ul>

Date	<p>&lt;originInfo&gt;&lt;dateIssued&gt;</p> <p>&lt;originInfo&gt;&lt;dateCreated&gt;</p> <p>&lt;originInfo&gt;&lt;dateCaptured&gt;</p> <p>&lt;originInfo&gt;&lt;dateOther&gt;</p>	<p>Date</p> <ul style="list-style-type: none"> <li>earliestDate</li> <li>latestDate</li> <li>Type</li> </ul>
Type	<typeOfResource>	Worktype
Format	<p>&lt;physicalDescription&gt;</p> <p>&lt;internetMediaType&gt;</p> <p>&lt;extent&gt;</p> <p>&lt;form&gt;</p>	<p>Material</p> <p>Measurements</p> <p>Technique</p>
Identifier	<p>&lt;identifier&gt;</p> <p>&lt;location&gt;</p> <p>&lt;url&gt;</p>	<p>Location</p> <ul style="list-style-type: none"> <li>refid</li> </ul> <p>Textref</p>
Language	<language>	--
Relation	<relatedItem>	Relation
Coverage.Spatial	<p>&lt;geographic&gt;</p> <p>&lt;hierarchicalGeographic&gt;</p> <p>&lt;cartographics&gt;</p>	<p>Location</p> <ul style="list-style-type: none"> <li>Name</li> <li>Refid</li> <li>Type</li> </ul>
Rights	<accessCondition>	<p>Rights</p> <ul style="list-style-type: none"> <li>rightsHolder</li> <li>text</li> </ul>
--	--	culturalContext
Coverage.Temporal	--	StylePeriod

--	--	StateEdition Description Name
Source	--	Source Name refID
--	--	Inscription Author Position Text

## Exercises and Activities:

### Text Analysis

**Exercise #1:** Using transcriptions of the textual materials in the Road Trip archive or copying the text from the brochures, perform a simple text analysis to find patterns in language and to create machine readable data based on the written texts. Use any two of the tools listed in the “Additional Resources” section of “Text Analysis” to analyze the text. Compare results between platforms. What are the different ways the resulting data can be displayed? Add another text for comparison, for instance compare the journal to the post cards.

### Image Analysis

**Exercise #2:** Image Analysis and Research activity using Resources #5 and #7. Read these images for larger cultural meanings. Are these identifiable characters? With what brands or companies are these images associated? What do company mascots tell us about the company and its brand?

**Exercise #3:** Look at the Edward Curtis, #14A. Do research to better understand the complexity and controversy surrounding Curtis and his representations of Native American culture. Write up a response to this image using both image analysis and historical research. Discuss how this information would inform the metadata you create for the image in the digital collection.

## Online Exhibits

**Exercise #4:** Create an Exhibit using the images in the resources, including the map you created. Write a story or narrative using the Digitized Resources. If your story needs more information in one area, go to DPLA and find images to add to your exhibit (be sure to credit correctly, also look at rights management and be sure that the resources you selected are free to be shared and reused). Note that the online exhibit might make use of any number of online platforms for presentation, but having no platform in place does not preclude one from creating the exhibit via a storyboard including images and text that will be used to assemble the exhibit.

### Consider the following factors:

- Metadata
- Rights and Re-use
- Attribution page
- Body of text that provides context and understanding (i.e. your exhibit script) as well as image captions
- Teaching guide with discussion questions and classroom activities

## Mapping and Geospatial Projects

**Exercise #5:** Using Digitized Resource #15a and #15b, create a map of the road trip using KnightLab StoryMap or another mapping tool. Consider how a geo-temporal narrative can provide not only an interpretation of the archive, but also highlight the collection and add context. Feel free to use the other images from the collection as illustrations of items mentioned in the diary. Also feel free to use content from DPLA and the internet, but be sure to note rights and fair use parameters.

## Sharing with an Aggregate

**Exercise #6:** Create a set of criteria and requirements for submitting the digital content in this kit and its associated metadata into an aggregate at the state/regional level (such as Minnesota Digital Library or Digital Library of Georgia, etc.) and/or at the national level (such as Digital Public Library of America).

### Consider the following factors:

- Rights and Re-Use
- Metadata

## Creating Metadata

**Exercise #7:** Look at the Metadata fields in Digitized Resource #1. Use the same format and apply additional metadata fields to another resource in the collection.

**Exercise #8:** Ask users to refer to the section on Relational Cataloging and to catalog Digitized Resource #1 in both of the ways outlined: relational vs. non-relational cataloging. Assume first that the work is the Corn Palace and not the photograph of the Corn Palace. The first option will necessitate the creation of a work record for the Corn Palace, as well as a creator record, and an image record. The second approach will require the creation of a single record which should incorporate creator and work information into a single record. Next assume the photograph as the work with the Corn Palace as its subject.

Discuss the advantages and disadvantages of using the work/image model vs. the single record model. Describe the differences in metadata creation in terms of the level of granularity of data, the ability to reuse data, and the clarity of the user experience.

**Exercise #9:** Using both the Dublin Core and Generic Field schemas on the worksheets below, create metadata for Digitized Resource #1. Compare the specificity of the fields and the data created for each field. Discuss the value of using more fields which capture more granular information versus using fewer fields which capture less granular information. Discuss the display and functional implications of having multiple fields map to the same field name.

<b>Talking Points:</b>	
advantages/disadvantages	
Concept of repeating field mapping	
Broad generic schema vs. more granular fields	
<b>Digitized Resource #1</b>	<b>DC Field Name</b>
	dc: title
	dc: Creator
	dc: Contributor
	dc: Description

	dc: Subject
	dc: Coverage - Spatial

<b>Digitized Resource #1</b>	<b>Example of a Digital Collection With User-Friendly Field Names</b>
	Title
	Creator
	Contributor
	Description
	Library of Congress Subject Headings
	Keywords
	City or Township
	County
	State or Province
	Country

## Copyright & Rights Management

**Exercise #10** : A user wants to use all of the images in their publication “The History of the Old West in Modern Photographs.” Can they? Are there any items in the collection that warrant further investigation into rights ownership? How would you go about determining any restrictions? Next write up a rights statement for the collection using RightsStatements.org.

## Transcription & OCR

**Exercise #11** : Consulting the original document PDF, clean up the transcript of the journal and make a

more readable copy. Remove diacritical marks, incorrect punctuation, and correct the spelling mistakes. Determine a policy for “correcting” misspellings from the original text. This policy should account for work parameters and include a discussion of an acceptable margin of error.

## **Assessment**

There are essentially no right or wrong answers to the kit’s exercises. Each student should make informed decisions regarding the exercises based on their knowledge and experience, as well as their capacity to adopt the tools and resources presented. Exercise outcomes will vary depending on the type of student and the collection’s and/or project’s needs and purposes.

In the real world, different collections deal with the same issues differently depending on a number of factors including the size and scope of the collection, the collection’s intended audience, and the purpose of the digital project. Given the variables that exist in digital collection creation, collection developers will find a myriad of solutions to the problems all digital collections face.

## **Reflection**

Instructors should lead the participants in discussion of the outcomes of their lesson work. In particular, discussions should focus on how the tools and concepts might be applied elsewhere.

## **Contact Information**

### **Greta Bahnemann**

Metadata Librarian, Minnesota Digital Library  
Minitex, University of Minnesota  
Wilson Library, Room 60  
309 19th Ave. South  
Minneapolis, MN 55455  
telephone: 612.625.6497  
e-mail: bahne002@umn.edu

### **Dr. Jeannine Keefer**

Visual Resources Librarian  
Boatwright Memorial Library  
261 Richmond Way  
University of Richmond, VA 23173  
telephone: 804.289.8275  
e-mail: jkeef@richmond.edu

