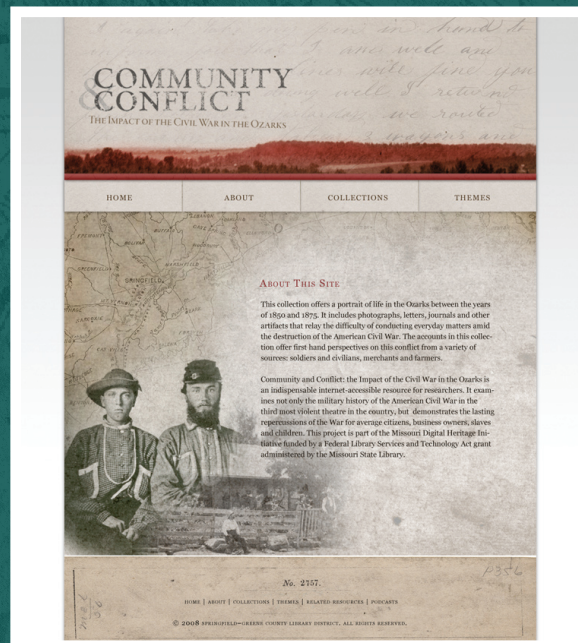


PLANNING AND DEVELOPMENT MANUAL FOR DIGITIZATION PROJECTS

Springfield-Greene County Library District

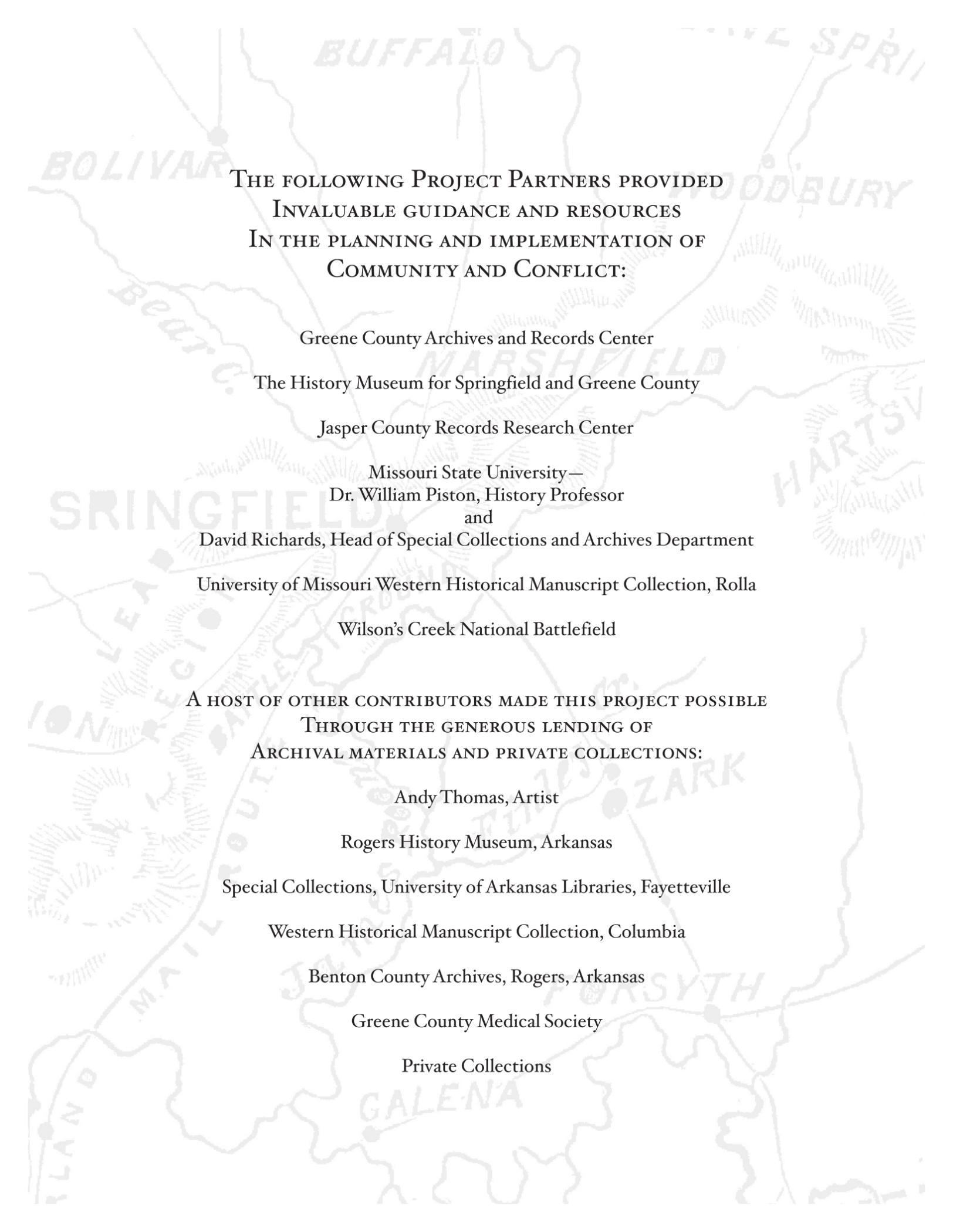
“Community and Conflict: The Impact of the Civil War in the Ozarks”



Community and Conflict: The Impact of the Civil War in the Ozarks is a Library Services and Technology Act Digital Imaging Grant funded by the Institute of Museum and Library Services and coordinated by the Missouri State Library.



Springfield-Greene County Library District • P.O. Box 760 • Springfield, MO 65801-0760 • thelibrary.org



THE FOLLOWING PROJECT PARTNERS PROVIDED
INVALUABLE GUIDANCE AND RESOURCES
IN THE PLANNING AND IMPLEMENTATION OF
COMMUNITY AND CONFLICT:

Greene County Archives and Records Center

The History Museum for Springfield and Greene County

Jasper County Records Research Center

Missouri State University—
Dr. William Piston, History Professor
and

David Richards, Head of Special Collections and Archives Department

University of Missouri Western Historical Manuscript Collection, Rolla

Wilson's Creek National Battlefield

A HOST OF OTHER CONTRIBUTORS MADE THIS PROJECT POSSIBLE
THROUGH THE GENEROUS LENDING OF
ARCHIVAL MATERIALS AND PRIVATE COLLECTIONS:

Andy Thomas, Artist

Rogers History Museum, Arkansas

Special Collections, University of Arkansas Libraries, Fayetteville

Western Historical Manuscript Collection, Columbia

Benton County Archives, Rogers, Arkansas

Greene County Medical Society

Private Collections

TABLE OF CONTENTS:

Introduction	I
Defining the Project.....	2
Cooperating Partners.....	4
Defining the Target Audience.....	7
Developing a Workflow Strategy	9
Identifying Collections and Promoting the Project	13
Creating the Metadata Scheme	16
Selecting and Evaluating Collections.....	19
Organizing and Storing Digital Material	25
Scanning Standards and Guidelines	27
Transcription Policies and Procedures	33
Conclusion	39
Appendix.....	41-59



Introduction

The Springfield-Greene County Library District staff prepared this handbook while working on *Community and Conflict: the Impact of the Civil War in the Ozarks*, a digital planning Cooperation grant funded by the Library Services and Technology Act, administered by the Missouri State Library. It is a collaborative project involving scholars, librarians, archivists and museum staff interested in expanding and enhancing the online resources available to disseminate information about the Civil War in the Ozarks. The partners offered invaluable assistance, advice and generously shared their knowledge of the Civil War, the key to an outstanding collection of materials to be digitized in the next phase.

Digital imaging provides enhanced access and is a key component of conservation and preservation. Customers and staff can use digital surrogates to access rare and fragile documents, greatly reducing the damage that naturally occurs as a result of frequent handling and lengthy exhibit cycles; however, there is a larger issue addressed by digitization. The Institute of Museums and Library Services has launched *Connecting to Collections: A Call to Action*, a national initiative to raise public awareness of the importance of caring for our treasures and to underscore the fact that these collections are essential to American history. Heritage Preservation conducted a study that found that nearly 190 million objects in U.S. collections are in immediate danger. Once these collections are lost, we cannot get them back. *Community and Conflict* brings together a collection of valuable materials to explore and interpret a catastrophic time in the history of the Ozarks.

This document is designed to incorporate the requirements of Missouri Digital Heritage and CONTENTdm with best practices in digitization. While these requirements are emphasized, the standards and procedures suggested here can be adapted to any project regardless of affiliation. Throughout the handbook, users will find examples from the *Community and Conflict* project with tips and suggestions that provide insight into the decision making process.

There are many complex steps involved in digitization which go beyond the mere act of scanning. Proper digitization also requires extensive work in collecting, indexing, and processing metadata, implementing quality control measures, creating transcriptions, providing secure methods of storage, and incorporating these resources into supported software designed for online access. These steps have been assessed, revised, tested and published by a multitude of institutions. There are many digitization handbooks online to assist any project; this guide seeks to alleviate the complexity and provide users with a path to making any project a success. The standards suggested in this handbook represent a compilation of best practices from several different guidelines and numerous sources are cited to provide more detailed information.

Defining the Project

Defining the scope and content of the digitization project is arguably one of the most significant steps in digitization. Depending on the nature of the project, this step may be simple or relatively complex. Institutions with a defined collection to digitize will have an easier time establishing the scope of the project than an institution that is pulling together multiple collections.

The key to this step is to identify the primary goals of the project.

- What will this project accomplish?
- Why will this project be important?
- What needs will it fill; for what audiences?
- What makes this project unique?
- Who would be interested in this project?
- What collections can be used in this project?
- Are other organizations working on similar projects; if so, can your organization partner with them?

The *Community and Conflict* project began as a result of the renewed interest in the Civil War due to the upcoming sesquicentennial. The Springfield-Greene County Library District formed a network of partners from previous collaborations who were interested in preserving and disseminating Civil War resources about the Ozarks. The partners consisted of museum, library, archive and university professionals who have extensive knowledge of Civil War collections and are familiar with resources available in the region. Surprisingly, limited research has been conducted on the impact of the Civil War in the Ozarks and only a few collections were available online. Project partners were assembled under this broad category; then, during the course of the planning project they worked together to refine the scope, establish themes and subjects, and identify collections that fit within those guidelines.

The scope was narrowed by determining the targeted time period, geographic area, themes and subjects. The planning project partners decided to focus on the Civil War in the Trans-Mississippi Theater, specifically the Ozarks region, which they defined as: the southwest corner of Missouri from Bates County south and Phelps County west, and two counties “deep” into southeast Kansas, northeast Oklahoma, and northwest Arkansas. The time period was established between the years 1850 and 1875. Project partners decided this time period was flexible to accommodate unique collections from the region.¹

¹ The selection committee evaluated individual collections outside the project’s time period. Based on their evaluation of the importance of the individual collection, special allowances were made to include some collections beyond the established project years. For more information about the selection and evaluation process, consult the “Selection and Evaluation of Materials” chapter.

Establishing Project Themes, Subjects and Identifying Potential Collections

Identifying the themes, subjects and potential collections for digitization is an evolutionary process. Original themes and subjects may shift or expand as the work progresses. As staff reviews collections, they may find that a single collection holds several items that fit into multiple themes. This may lead to searches for other collections with similar content, further defining the themes and subjects. Some projects may begin with a perceived need or by identifying content areas in a collection that have not been digitized, resulting in a search for additional collections to support these themes and subjects.

Partners for the *Community and Conflict* project decided they wanted to focus on more than just the military aspects of the Civil War; they wanted to explore how the War affected and influenced life in the Ozarks. They defined themes such as: urban development, politics, social consequences of the War, experiences of women, children, and minorities, and many more. Project staff consulted the *Library of Congress Subject Headings* and the *Ozarks Periodical Index* to develop a set of controlled subject headings to describe the themes and subjects, providing a standardized structure. Using the Library of Congress and the regional index together provided terms which covered the Civil War-related topics and those specific to the Ozarks region.

As the themes and subjects began to unfold during planning meetings, partners were able to suggest collections that could support these headings. They also identified themes that lacked collections and helped direct project staff to fill these gaps. Through this process, the preliminary list was expanded and partners prioritized the most important themes and subjects, fulfilling the promise behind the name: *Community and Conflict: the Impact of the Civil War in the Ozarks*.

Organizations that recognize a need to provide digital content in a particular area, but have not yet identified specific collections or established themes should ask the following:

- What collections do customers frequently request for research?
- Which collections are most commonly accessed by staff?
- Which collections pose the most preservation concerns?
- Which collections could benefit the organization and its customers by being digitized?
- What topics would most interest customers and which collections represent those topics?

The decisions made during this step will assist the development of the project and will form the basis for future decisions.

Cooperating Partners

There are many advantages to working with multiple partners to plan and implement digital imaging projects. Building local, regional or statewide relationships produces exponential benefits as the circle of partners among community organizations grows. Partners at every level will enrich the end product by assisting to:

- Leverage resources and expertise
- Design projects to meet the needs of customer base
- Effectively reach target audiences
- Measure quality assurance and evaluate outcomes
- Promote and publicize the project
- Accomplish common missions of multiple agencies

The list of organizations to consider as partners depends on the subject areas of a particular project; however, it is important to brainstorm all potential stakeholders. While recognizing individuals or organizations in the immediate local area is important, consider experts such as professors, county archivists or museum curators throughout the region or state. After fully developing the list, contact the organizations to assess their interest in participating and, if so, to determine their role in the future project. Ask them how this project would benefit them and their institution and if they are aware of others who should be contacted.

Community and Conflict project partnerships

- Libraries
 - Wilson's Creek National Battlefield
 - Springfield-Greene County Library District
- Universities
 - Two Missouri State University faculty members
- Archives
 - Jasper and Greene County archives
 - Western Historical Manuscript Collection at Rolla
- Museums
 - History Museum of Springfield and Greene County
 - Wilson's Creek National Battlefield Museum

Organizations that are not a part of the project team can also assist by providing information and collections. Official partnership may not be formed with these associates for various reasons such as technicalities, legalities or simply distance. While these associates may not be official partners, they can provide a multitude of benefits to the project.

The *Community and Conflict* project worked with several institutions in addition to the official planning partners. These associates provided access to their collections for digitization, consultations and historical information, all of which were extremely valuable to the project's success.

- Historical Societies and Museums
 - Missouri State Archives
 - Historical Society of Missouri
 - Missouri History Museum
 - Rogers Historical Museum, Arkansas
- Civil War Roundtable of the Ozarks
- Lebanon-Laclede Genealogical Society
- National Park Sites
- Special Collections, University of Arkansas Libraries, Fayetteville
- Pea Ridge National Military Park, Arkansas
- Western Historical Manuscript Collection in Columbia
- Benton County Archives in Rogers, Arkansas
- Private Collectors

The Missouri State Library Development Office is also an important partner. The staff provided guidance and support during project development and implementation.

Establishing a smoothly functioning, effective coalition of partners takes time, and determining mutually understood roles is an important first step. Interested partners should be asked to complete partnership statements after they have agreed to participate. The Institute of Library and Museum Services offers a partnership statement to help assess the level of commitment from the partners' perspective. It includes:

- Name of the Organization
- Project Contact
 - Name
 - Telephone
 - E-mail address
- DUNS Number
- Mailing Address
- Web Address
- The mission of the organization
- The service area of the organization
- The partner's key roles and responsibilities in the project

Support from partners might include a financial contribution or in-kind donations such as: staff time, equipment, promotional materials, advertising or supplies. Experience tells us that historians and archivists are a very dedicated group. They care about the retrieval and dissemination of the information found in collections everywhere.

The *Community and Conflict* partners assisted by:

- Defining the geographic and contextual scope of the project
- Developing the themes
- Locating materials
- Designing the survey instruments
- Providing information about target audiences, promoting the project
- Evaluating the collections

The Springfield-Greene County Library District working alone would not have accomplished a collection with the in-depth and broad coverage of the history of the region, nor would they have been able to identify all of the relevant materials. The knowledge and expertise of the cooperating partners who worked on the project have been extremely valuable, and there is no doubt that *Community and Conflict* is a much richer collection as a result of their involvement.

Defining the Target Audience

Successful digitization projects are based on the needs and interests of a target audience or multiple target audiences. The questions to ask during the design phase are:

- Who among the general public is interested in this topic?
- Who would benefit from this information?
- Are there special interest groups related to this topic?
- Are there others who could be inspired to become interested in the topic?
- Is this something that high school or college students are interested in? If not, how could it be designed to appeal to this audience?
- Is this topic of interest to other groups?

Community and Conflict was developed as a collection of primary source materials about the Civil War in the Ozarks. Asking these questions revealed that there were some obvious audiences such as university scholars and high school students who would be interested in that material.

Staff of the *Community and Conflict* project identified their primary target audience as scholars, students and lifelong learners. Research on the Civil War in the Ozarks has been relatively unexplored, and little digitized material on the region is available. Knowing this, staff set out to encourage scholarship not only among professional scholars but to also inspire emerging researchers and high school students to conduct research on untouched resources. By addressing the questions above, project staff and partners identified those who would be most interested in the subject and then they brainstormed ideas for additional audiences. The Civil War has a wide following among special interest groups such as Civil War Roundtables, Sons and Daughters of Union and Confederate Veterans, metal detector groups, genealogists, local historians, and many others.

Surveying the Target Audiences

Once the target audiences have been determined it is important to gauge their level of interest and identify their specific research needs. Conducting a survey and tabulating the results will allow the project staff to decide how to best design the content, and allocate time, funding and expertise to ensure maximum audience usability.

The steps for developing a survey are:

- Identify goals of the survey
 - What does the project staff hope to learn from the survey?
- Create a list of questions to accomplish goals of the survey
 - Open-ended questions
 - Scaled Response Questions
- Assess suggested questions and select those best suited for the target audience and survey goals
 - Administer sample survey
 - Evaluate results of the sample and determine the quality of data returned
 - Make changes to the survey
- Administer survey on larger scale
- Evaluate responses and create usable data

Open-ended questions are valuable because they allow maximum input by the surveyed audience; however, some audiences find surveys annoying and only apply minimal effort in answering the questions. Scaled-response questions are designed to allow quick assessment and response time for users, and the data is easy to calculate. The *Community and Conflict* project staff created 20 different questions that would address the goals of the survey. Project partners then evaluated each of the questions and selected seven for inclusion. Project staff decided to use a combination of open-ended and scaled-response questions, allowing for ease of evaluation, a high response rate, and options for longer explanation.

Community and Conflict staff hosted the survey online through a program called *Survey Monkey*. The electronic survey was distributed through e-mail and a web page on the Library's web site. The survey was also administered using hard copies in person. Response rate for the electronic survey was lower than surveys administered in person. Project staff and partners projected that the quality of responses to online surveys would be higher than in-person surveys, because only truly interested respondents would complete the online version. They also felt that surveys administered in person to groups might be rushed because some individuals might have little interest in the topic and would be eager to complete the survey. While surveys conducted in person are entirely optional, participants may feel required to complete the survey when the administrator is in the room. In the end, staff found that there was not a noticeable difference in the quality of response between the electronic and paper survey.

The *Community and Conflict* survey indicated that 61.3% of the surveyed audience was interested in military engagements. With this knowledge, staff worked to ensure the topic of military engagements was well represented in the selected collections. The level of interest for all selected themes was identified and collections were selected according to the expressed level of interest in a particular content area. The survey also asked audiences to identify search functions and organizational structure that they preferred. This information was incorporated into the planning and development of the web page. Defining and understanding the target audience and their expectations are important steps in the development of a digitization project. A copy of the *Community and Conflict* survey can be found in the appendix (1 - A).

Developing a Workflow Strategy

A workflow strategy is a concise plan designed to outline the tasks and formulate a feasible methodology for implementing a successful project. An effective workflow strategy will prioritize goals, establish a timeline and clearly define the allocation of staff and other resources required to complete the entire project.

The process begins with the overarching goals of the project. What does the project set out to accomplish? How many pages, objects or images will the project digitize? What will be the extent of the end product; will it involve uploading the images and metadata to CONTENTdm and the development of a local web page? Or only the CONTENTdm work? Depending on the nature of the project, the overarching goal may or may not be clearly defined at the beginning. This chapter will discuss two different approaches to creating a workflow strategy. The first example will follow a project of one collection, *The John A. Mack Letters*, which had a clearly defined set of materials to digitize. The second example will explore the workflow strategy for the *Community and Conflict* project which began with an undefined amount of material to digitize.

Projects that begin with a defined set of material for digitization are more manageable. The selected collection will have an established number of documents to scan.

An organization may choose to digitize only selected portions of a collection. *Community and Conflict* project staff chose to digitize three of five diaries from the Lyman Bennett collection, because two of the diaries were not within the scope of the project.

When the collection size has been established, staff should determine the timeframe for the project. What are the time restrictions for the grant or other funders? When will the material be made available to the public?

Some workflow strategies are based on a fixed timeline with a firm deadline; others have an undefined timeline. The first approach described here is based on a time sensitive project and the second approach has an undefined deadline. Project staff should identify the project timeline and compare that to the amount of material to be digitized.

The John A. Mack collection contains 49 letters totaling 240 scans. The Library has three months to complete the digitization project. 240 scans divided by a three month time period will require 80 scans to be processed each month to meet the deadline. The digitization process includes: scanning, transcribing, collecting metadata, and hosting the content online. The processing count can be further broken down into twenty scans a week or four scans in a single business day. Four scans a day sounds reasonable, but remember there are other duties such as quality control checks and content development that must also be completed in the three month period.

When the collection size and timeframe have been established, staff can identify the tasks that need to be accomplished within that time period. List each task, place the tasks in order of accomplishment and identify when the task must be completed.

**Example:
Digitization Steps**

1. Select / Evaluate Collections
2. Scan Collection
3. Transcribe
4. Collect Metadata
5. Check Quality Control
6. Write Web Site Content
7. Host Collection Online

Each of the steps must be evaluated and ranked in an order of completion as shown at the left. After the collection is scanned staff will transcribe the documents and collect metadata, etc. When the steps have been outlined, test each step for problematic areas. A walkthrough will identify timing issues and other unforeseen pitfalls. By timing each step staff can estimate how long it will take to complete the entire project. This estimation can then be compared to the project deadline and necessary adjustments made. Once the timing and steps have been established, set weekly or monthly goals and when each step will take place to meet those goals, then assign each task to specific staff.

Multiple year projects, or those with an undefined deadline, can follow a similar process to establish the workflow. The priority for this type of project is to define structured steps for digitization, assign specific tasks to staff members and document each responsibility in the workflow strategy. The entire project team will then understand individual responsibilities, the order of the steps and how their role impacts the project as a whole. Projects with an undefined deadline may take multiple years to complete, which can be overwhelming; so, it is important to demonstrate incremental progress. Establishing a general timeline, goals and measurable benchmarks will keep staff members on task and ensure that the project is moving forward at a steady pace.

This process begins with a time assessment. Select a test collection of documents, and ask staff to record the processing time to complete selected digitization steps, such as transcribing or collecting metadata.

The *Community and Conflict* staff recorded how long it took to scan, collect metadata and transcribe an individual document. They chose documents with varying levels of condition and quality of handwriting. Two staff members measured their processing time for each document. An average was taken between the two staff members and the varying times to process the 15 documents. These calculations resulted in an average of 50 minutes to process one four-page letter. It is better to slightly overestimate the amount of time required to process the material. Underestimating may result in insufficient time and budget to complete the project.

When the average processing time has been identified, calculate the total hours of staff time dedicated to the project in a single week; then determine how many documents could be processed in that time period.

In the example at the right, project staff could complete 144 four page letters in one work week, if the staff spent all of their time processing letters. This estimation is not realistic because many other factors interfere with the work of staff throughout the course of the day; however, this calculation shows the maximum number of letters that could be completed in one week and can be used as a basis for estimating the total time required by cutting the maximum estimation in half or thirds. Some documents may take a longer or shorter time to process depending on the condition, length, and legibility of the handwriting. All of these factors will impact the actual amount of work that is completed.

Sample Calculations

2 full-time staff: 80 hrs /week
total

2 part-time staff: 40 hrs /week

Total staff time: 120 hrs/week

120 hrs = 7,200 minutes

If staff requires 50 minutes to process one four page letter, then in 7,200 minutes they can process 144 four page letters.
(7,200 / 50 = 144)

**Realistic projection:
48 letters a week**

A realistic projection of capacity is based on staff skill level, resources, the condition of the materials being processed and a number of other factors. The example above is based on a one-third projected rate, or 48 letters per week. Monthly, quarterly, and yearly timeframes are then estimated to establish benchmarks and goals. For example:

48 letters per week	48 x 52 weeks	2,496 completed in 1 year
---------------------	---------------	---------------------------

Staff can now identify 2,500 letters to be processed during the year. With the collection size and rough time period established, staff can determine other project tasks to be completed, such as writing interpretation and content, designing the web site, hosting images on the web site, checking for quality control, reports and evaluation. The tasks and the order in which they are undertaken are factored into the project timeline. Each task is then assigned to an individual or group. By approaching the workflow with a clearly defined strategy, issues involving time management, work assignments and other time sensitive challenges can be recognized and properly addressed.

The collections for the *Community and Conflict* project were scattered throughout the Ozarks region, some in public institutions and others in private collections. None of the material had been preselected for digitization. Staff conducted an extensive survey by traveling throughout Missouri, Arkansas and parts of Kansas looking for collections relevant to the project.

Due to the enormous scale of the digitization project being planned, it was necessary to determine how many scans could be identified and processed in a 12-month period, the time allowed for a Library Services and Technology Act funded project. Considering the time traveling and coordinating other aspects of the project, staff formulated a four-step processing strategy. It was decided that scanning would be completed in the field, and metadata and transcriptions would be collected at the office. Material would then be uploaded to CONTENTdm after quality control checks. This strategy was designed to eliminate unnecessary travel and reduce expenses by expediting the scanning process.

After establishing this processing strategy and determining the steps involved, staff evaluated the time required to accomplish the entire process by timing each step. By doing so, they were able to establish and formulate a timetable and workflow for completing 2,500 scans.

As the project continues to develop, staff may periodically find that changes to the workflow strategy are required to ensure the project functions smoothly. The workflow strategy and timeline that were developed are included in the appendix (Workflow 2 – A, Timeline 3 – A).

Identifying Collections and Promoting the Project

This chapter explains the process of identifying collections for community or regional digitization projects, and locating private and public collections by promoting the project to a broad audience. It further illustrates the importance of choosing planning partners who have expertise in the topic and a working knowledge of libraries, museums, archives, private individuals and other institutions that may hold potential collections of interest, and also serve as distribution points for publicity. The first agenda items for planning partners should be:

- Defining the scope of the project
- Identifying collections

There are several additional approaches to locating collections. Telephone surveys of related organizations such as public libraries, historical societies, archives and genealogical groups in the region can produce valuable leads. Even if these institutions do not hold relevant collections, staff may know where primary source materials can be found.

The Internet is another great place to begin searching. Most archives and library special collections offer descriptive finding aids, many of which are available online, that provide a wealth of information about the items in their collections. These resources can be used to develop a preliminary list of materials to consider reviewing during an onsite visit.

Promoting the Project and Identifying Private Collectors

- Create a well-designed flyer or brochure with a professional appearance and have enough copies printed for a wide distribution. The Springfield-Greene County Library District included an offer of a free digitized copy of the scans for those who contributed materials. See appendix (4 - A).
- Contact area organizations with a similar focus such as special interest groups, historical societies and genealogical organizations; then, develop a presentation about the project and request an opportunity to speak at a meeting, or ask for time to make an announcement.
- Distribute flyers and public surveys at meetings, museums, archives and libraries.
- Check calendars of related organizations for upcoming conferences or regional meetings and determine the most effective ways to establish a presence at the event; possibilities include booths, table talks or providing information registration packets.
- Write press releases for the news media and organizational publications to be distributed by the host institution and partners.

Site Visits

As collections are identified, develop a schedule for visiting sites and scanning collections. Before travelling to a site:

- Review the web sites and search indexes and finding aids
- Develop a preliminary list of collections of interest
- Personally call to explain the project to the curator/manager/director/archivist, and seek permission to scan collections
- Request an appointment for a site visit and share the list of items of interest. If this information is not available online, ask for suggestions of materials relevant to the project
- Confirm permission to view and scan collections and ask for assistance with access
- Clarify institutional scanning policies
- Ask about space and power available for scanner and laptop
- Send a follow-up letter confirming arrangements; include a list of collections and ask that the selected items be pulled in advance of the scheduled visit.

Private Collectors

It is important to be flexible when working with private collectors. Some collectors do not want to move their materials, but prefer to have collections scanned on site. In other cases, collectors may feel more comfortable meeting in a public place, such as a library. If visiting the collector to scan materials on site, describe the scanner and the space needed to work. If meeting at a public space make the necessary arrangements in advance with the appropriate staff to ensure that resources are available to complete the work at that particular site.

Legal Release Forms

Consult with legal counsel to review current policies and laws on digitizing, publishing and the fair use of digital items. Then draft a legal release form to fit the project specifications. Before an outside collection is scanned, the contributor or owner (*releaser*) providing the collection must read, agree to and sign the release form.

The release form ensures that the institution will not be held legally liable for using the master files for their own publications, research, exhibitions, presentations or other online databases. The *releaser* does not need to relinquish copyrights to the materials provided; however, they will not have control over the uses of the digital surrogates by the project institution. The release statement should clarify that fair use will be established in accordance with policies and project goals of the institution, and it also provides that the items will not be used for monetary profits. A signed release form avoids potential litigation should the *releaser* decide later to object to the use of the materials. A copy of the *Community and Conflict* release form can be found in the appendix (5 - A).

Generally, copyright laws are complex and can be confusing when determining if a work is in the public domain. The American Library Association Office for Information Technology has developed a *Copyright Slider* to answer these questions. Information is available at <http://www.wo.ala.org/districtdispatch/?p=421>.

Creating the Metadata Scheme

Metadata is defined as data that describes other data, or data that describes digital scans, and it is essential to effectively access digital collections. Sarah Shreeves, Coordinator of the Illinois Digital Environment for Access to Learning and Scholarships, states, “The potential of shareable metadata is only possible if the metadata is of good quality. Institutions must provide metadata that is understandable outside of a local institutional context.”² All collections require some form of descriptions to be used in finding aids or for cataloging purposes. There are three types of metadata: administrative, structural and descriptive.

Metadata		
Structural	Administrative	Descriptive
<ul style="list-style-type: none">• Data about arrangement of scanned images• Shows relationships between scanned images	<ul style="list-style-type: none">• Technical data<ul style="list-style-type: none">• Date scanned• Date altered• Rights• Embedded in image• Viewed in image properties	<ul style="list-style-type: none">• Describes scanned image• Catalog type entries• Keyword search terms• Created and stored in a separate file from scanned image

Most administrative data will be generated through computer software such as the scanning program or CONTENTdm. Structural metadata is managed through CONTENTdm; the technician only needs to apply a unique identifier to the descriptive metadata. Descriptive metadata is created and managed by a metadata technician, so this chapter will focus on descriptive metadata.

Descriptive Metadata

Descriptive metadata defines elements for discovery and identification of scanned images. It includes all of the terminology that enables keyword searches, and it explains the significance of the image to the user; therefore, generating high quality descriptive metadata is extremely important. Rigorous quality control is mandatory because incorrect descriptive metadata leads to irretrievable images. Metadata descriptions also provide meaning and significance to the end user. The basis of the Missouri State Library’s Missouri Digital Heritage project (MDH) is Dublin Core metadata, which consists of 15 essential elements used for resource description. The name Dublin Core derives from the scheme’s origins at a 1995 workshop in **Dublin**, Ohio, in which these **core** standards were selected to describe a wide range of resources.

² Shreeves, Sarah. *Stewardship in the Digital Age: Managing Museum and Library Collections for Preservation and Use*. Institute of Museum and Library Services. Page 15 of Web Wise 2007 Conference Proceedings. <http://www.imls.gov/pdf/WebWiseProceedings2007.pdf>

The scheme established by Dublin Core and suggested by MDH is designed to be adaptable by various projects. Each project should develop its own metadata structure based on the standards offered in the appendix (6 - A). Note that there are required elements that must be included, and there are optional fields. The metadata scheme developed by the *Community and Conflict* project can be found in the appendix (7 –A). Not all of the optional fields suggested by MDH applied to the *Community and Conflict* project. For example, the **Title. Alternative** field was not applicable to the *Community and Conflict* project, so it was not included. Staff also found that they needed a few additional fields that were not offered in the MDH scheme; such as, site accession number, volume and **Format.Original**.

When developing a metadata scheme, it is important to identify the field name, Dublin Core field relations, a detailed explanation of each field and a sample entry. The Dublin Core fields can be interpreted differently by multiple institutions; therefore, each field explanation must be detailed and specific. If not, fields might be misunderstood by different project members resulting in conflicting data entry.³

³ More information about each field can be found at: <http://dublincore.org/documents/dces/>

Metadata can be collected on multiple levels: collection, object or page level. Depending on the nature of the project, different levels of metadata will be appropriate.

Below are hypothetical examples of both collection and item level metadata. This information would be entered into the “description” field, but aspects of the description would be entered into other appropriate field such as the date.

Collection level metadata	Object level metadata
<p>Johnny Rebel Papers, 1861-1862—</p> <p>This collection consists of letters and diary kept by Johnny Rebel during his military career, February 1861 to November 1862. Johnny Rebel was originally from Rogers, Arkansas and enlisted in the 1st Arkansas unit at Fayetteville. During his service, he marched to Missouri where he was involved in the Battle of Wilson’s Creek and numerous other skirmishes. In his diary he describes the land in Missouri, military actions, camp life, and his hopes that the war will end soon so that he may return home to continue farming with his father. Also, included in this collection are five handwritten letters to Johnny Rebel from his mother, Dixie Rebel, from the family farm in Rogers, Arkansas. These letters are dated from April 1861-September 1862. In her letters, Mrs. Rebel speaks of the hardships at home, the failing crops, the sudden illness of Johnny’s father, and the troops moving through their county. Johnny Rebel was killed in a battle along the border of Missouri and Arkansas in December 1862 as his unit was marching towards Pea Ridge.</p>	<p>Rebel, Dixie, 1862-08-06. Rogers, Arkansas. Letter to Johnny Rebel, “Dear son.” CSA0186065, The Rogers Historical Society, Rogers, Arkansas.</p> <p>Letter to Johnny Rebel, from his mother, Dixie Rebel. The letter is dated 1862-08-06, and was written from the family farm in Rogers, Arkansas. Mrs. Rebel writes to her son that his father has suddenly taken ill and has been confined to bed with “paralyzing effects.” As a result of his father’s illness, there is no one to tend to the crops and fears the crops will be lost. Mrs. Rebel speaks of Union troops that have passed through the county and stopped to loot the neighboring farms. She inquires when Johnny will be returning to Arkansas, and if he might be able to get a furlough to come home to help salvage the remaining crops and to help her tend to his father.</p>

The *Community and Conflict* project choose the following approach to metadata.

- Collection level metadata will be used to provide users with an introduction to the collection, and help them understand its significance.
- Object level metadata will be collected for documents; such as, letters, court cases, depositions and other similar documents. Staff determined that the metadata collected on each page would not be distinctively different from what could be collected from the document as a whole; so, they decided that a better understanding of the letter would be achieved at the object level.
- Page level metadata will be collected for objects; such as, diaries or tax record books with separate entries on each page. This approach will provide more detail and enhance the user’s understanding of the object.

Selecting and Evaluating Collections

The goal of digitization is to provide greater access to information that is not widely available to the general public. It is important to select high caliber material for the projects because digital collections improve the research value of a database, generate interest in the collection and entice visits to the host institutions.

Sometimes selection is a straightforward process when dealing with one or two small collections; however, selecting materials from multiple collections of significant size can be an overwhelming process. It is easy to lose focus of the project goals and scope by becoming submerged in the amount of material from which to choose. When selecting collections, ask the following questions:

- How does this object relate to the predefined scope and themes of the project?
- What can be learned from this object?
- What can be taught to others about this object?
- Why will others want to research this object?

Addressing these questions helps preselect high quality material to be included in the digitization project.

The Northeast Document Conservation Center has established a three-step selection process for choosing materials to be included in a digitization project. This process has been adapted to fit the needs of the *Community and Conflict* project, and it is suggested that each institution follow the basic structure but alter the process as needed. The *Community and Conflict* project included the following phases:

- **Nomination**

Collection managers, curators, researchers, historians, local experts and many others identify and nominate material for inclusion; they also document the reasons for including the material in the project.

- **Evaluation**

An evaluation committee reviews the nominations and, based on the scope of the project, makes determination about including or excluding materials.

- **Prioritization**

A selection committee reviews and assesses collections proposed by the evaluation committee based upon value, risk, demand and existing coverage; a score is assigned to each collection, and each collection is ranked based on that score.

Nomination

During this phase suggestions are made by various individuals to identify collections that will be included in the digitization process. The Northeast Document Conservation Center recommends that an organization use the form shown in the appendix (8 - A).

The nominator should consider the following questions:

- How much of the collection is accurately and well-documented?
- Are there finding aids available for the collection?
- What is the condition of the collection?
- Where is the collection stored, and is it easily accessible?
- Are there restrictions on any parts of this collection?
- Does the institution retain copyright for the collection?
- How frequently is this collection requested by patrons or used by institutional staff?
- Are these materials unique to the institution, or are they located in other repositories?
- How does this collection relate to the scope and themes of the project?

All nominations are reviewed by an evaluation and selection committee before they are chosen for digitization; nomination does not guarantee inclusion.

Evaluation

It is important to evaluate each collection nominated. This evaluation process may require a committee, or individual, to physically review each nominated collection. In other cases, evaluations can be made from the information provided on the nomination form. The evaluation process includes a review of the material presented in the nomination form and answering a series of questions about the collection:

- How well does this collection relate to the scope and themes of the project?
- What is the condition of the collection?
- Are the materials in the collection oversized? If so, does the organization have the resources to scan this type of material?
- Are these materials unique to the institution, or are they located in other repositories?
- Does this collection provide new insight to the topic?
- Have these materials, or similar materials, already been scanned by other organizations?
- Are there similar collections available to researchers online?
- How can this collection contribute to the scholarly research on the topic?
- Does this collection have any restrictions such as copyright or donor restrictions? If so, these records may need to be held back until restrictions have lapsed.

- Does the collection require substantial research to be useful? If so, can the project provide the contextualization?
- Can this collection be paired with other collections to provide additional value to the overall understanding of particular project themes?
- Can this material be used to create educational programs or be helpful to other aspects of the project?
- Is the target audience interested in this collection or subjects?

Prioritization

After collections have been approved by the evaluation committee, a final assessment is made by a selection committee that will score each collection based on value, risk, demand and existing coverage.

Informational, Artifactual, Associational, and Evidential Value

Collections are rated on different types of value and each collection will be scored based upon the specific values that apply to it.

Informational Value

Informational value refers to a collection's relationship to the content scope of the digitized project.

- High value offers significant content and important details regarding people, places and events that reflect directly to the scope of project.
- Moderate value offers only minor details or content about topics and themes.
- Low value provides little or no relevant details or content to topics or themes.

Artifactual Value

Artifactual value refers to the nature of the original materials.

- High value collections are in good condition, are rare and unique, and/or can provide cultural significance. Examples are: diaries, journals, documents in rare historic formats such as platinum prints or materials in unusual genres.
- Moderate value materials that are mass produced or are widely available. Examples are: library bound books, stereographs or post cards.
- Low value materials are in poor condition, or they are copies, reproductions or duplicates of the original.

Associational Value

Associational value refers to original materials linked to an individual, place or events. Examples are: letters or documents signed by a notable individual, photographs taken by a soldier or a canon from a battle.

- High value materials include personal papers of a notable person, items discovered in an archeological excavation or other artifacts linked to a specific person, place or event.
- Moderate value materials include county, state or federal agency documents; such as, circuit clerk records, birth certificates or land deeds.
- Low value collections consist of copies, duplicates or modified reproductions.

Evidential value

Evidential value refers to the document's ability to prove an event or activity occurred, thereby providing "evidence" of such event.

- High value documents are the originals in original form.
- Moderate value documents may include legal records from county, state or federal agencies.
- Low value documents have been modified, copied or reproduced.

How to Score Value

- High value - score 6 points if 25% or more of the collection is deemed to have high value in any of the above categories.
- Moderate value - score 3 points if less than 25% of the collection has high value and meets the criteria of moderate value in any of the above categories.
- Low value - score 1 point, if it fails to meet the criteria of high-moderate values.

Risk

Risk refers to the preservation needs and concerns of the materials within a collection.

- High risk – chemically unstable, self-destructing and can be damaging to nearby materials; they can pose health hazards to staff and researchers who come into contact with them. Examples of damage to high risk material are: mold, mildew, water damage, fire damage, nitrate films and negatives, thermal copies and severe acid migration to the point of paper deterioration.
- Moderate risk – materials have received primarily mechanical or physical damage due to improper housing or handling. This also includes materials that are losing informational content naturally or gradually due to their composition; such as, compact discs or VHS cassette tapes. Examples of damage to moderate risk material are: tears, iron gall ink damage and lamination. Tightly rolled documents are considered to be at moderate risk and must undergo humidification treatments before scanning.
- Low risk – materials are located in areas with adequate conditions; they have undergone needed conservation treatments and have been encapsulated in

Mylar. They are housed in pH neutral materials and can be found in environmentally stable facilities. They are not in danger of self-destruction and they pose no threat to nearby documents; however, they may have foxing, slight fading, adhesive residue, insect damage and smell like mold or mildew.

How to Score Risk

- High risk – score 6 points if 50% or more of the material meets the high-risk criteria.
- Moderate risk – score 3 points if it contains less than 50% high-risk materials and 50% or more of the collection meets the moderate risk criteria.
- Low risk – score 1 point if 50% or more of the materials have adequate storage conditions or if all of the major conservation concerns of the collection have been treated or addressed.

Always be prepared to handle rare and fragile documents. Different organizations will have varying handling policies, but always wear cotton gloves when working with collections. Always pick up a document with two hands by the strongest section; never pick up a document by a fragile section as the section may tear when handled. If necessary, place the document on an oversized archival board and lift the board to provide ample support to the document. It may be necessary to use an overhead scanner to capture fragile books and documents. Never send an historical document through a paper feeder, because the feeder may jam causing further damage. If unsure how to scan or handle and scan fragile documents, stop and seek professional consultation before proceeding.

Demand

Demand refers to how frequently the materials within the collection are requested by researchers and how useful the material will be to researchers.

- High demand – score 6 points for materials that will be frequently requested and very useful to researchers.
- Moderate demand – score 3 points for materials that will be requested on an average basis and useful to researchers.
- Low demand – score 1 point for materials that will have low request rate and will not be very useful to researchers.

How to Score Demand

- Each collection is evaluated based on its potential demand value. Consult local historians to assist in evaluating collections..

Existing Coverage

Existing coverage refers to the amount of documentation of similar material already available to the public through venues such as publications or online databases.

- High coverage materials are available online or in another format.
- Moderate coverage indicates that limited portions of the collection are available online. It also refers to collections that have content similar to other collections available online or another format.
- Low coverage materials are not available online or in another format and there are limited collections of similar nature readily available.

How to Score Existing Coverage

- High coverage – score 1 point if more than 50% of the collection is already available.
- Moderate coverage – score 3 points if less than 25% of the collection is already available, and less than 50% of the collection is reflective of similar content already online or another format.
- Low coverage – score 6 points if less than 10% of the collection is available online or another format, and there are few sources that contain content of similar nature.

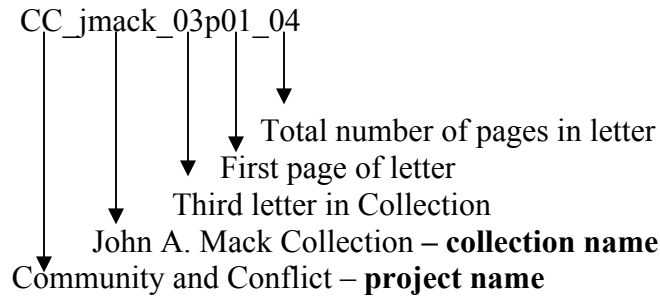
How Rank Collections

A score is assigned to each collection based on the criteria of the four evaluation categories. The scores from each of the categories are totaled, giving a numerical rating to the collection as a whole. The scanning prioritization of the collection is then determined by the highest resulting numerical score. In the case of an identical score, compare the actual preservation concerns to determine which collection receives higher priority.

The prioritization form from the Community and Conflict project can be found in the appendix (9 – A). For additional information about the Northeast Document Conservation Center's three-step selection process, consult the digitization handbook at <http://nedcc.org/oldnedccsite/digital/dighome.htm>.

Organizing and Storing Digital Material

The naming structure of the files is very important to the organization of the digital materials. In order for the material to be successfully uploaded into CONTENTdm, the metadata file name and the scanned image file name must match. The naming system can originate from multiple sources, and can be as complex or as simple as the institution would like. Below is an example of a file naming system from the *Community and Conflict* project:



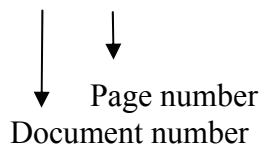
CC_jmack_03 – **object level**

CC_jmack_03p01_04 – **page level**

The file naming system for the *Community and Conflict* project was designed to provide an organizational description of the document. The file name CC_jmack_03p01_04 explains that file relates to the first page of the third letter in the John A. Mack collection in the *Community and Conflict* project, and that there are four total pages in that letter.

Initially project staff explored the use of a simple numerical system.

Example: 00001.01



This system was developed to provide the simplest means of numbering files. This numerical system only shows the document number and page number, and does not provide any additional information that would be helpful when trying to identify a specific document. A user would be forced to memorize each file number, open each file or consult a file index to identify any contextual information. It was quickly determined that this numbering system was not practical and abandoned. Staff found that too much time was spent hunting for a specific file and that they needed a numbering system that would provide a means of quickly locating files.

Metadata for MDH CONTENTdm projects can be entered directly into the acquisition station or imported through a tab delimited .txt file. For additional information about setting up a tab delimited .txt file, consult the handbooks distributed by CONTENTdm.⁴

Separate file folders are recommended for storing all of the images and the metadata files. For the *Community and Conflict* project, staff created a folder for each object. Each folder is named with the object level identification (CC_jmack_03). Within each folder is the metadata tab delimited file and another folder containing all of the object images. This allows easy access to the desired information and scans when uploading into CONTENTdm or another program.

Digital Archives

Digital collections require the same level of attention to preservation as traditional materials. “Successfully preserving digital collections requires a holistic, life-cycle management strategy that emphasizes creation, appraisal, documentation and reuse. You cannot just preserve a digital object: You must preserve the entire digital ecosystem.”⁵ The Missouri State Library serves as the “trusted digital repository,” which Priscilla Caplan described in her speech as the organization that provides reliable, long-term access to managed digital resources.

Some methods for local digital storage are:

- A single computer
- Store on or backed up to a CD or DVD
- An external hard drive
- Network Server
 - NAS: Network Attached Storage
 - SAN: Storage Area Network

Regardless of the local storage option chosen, it is important that the records are maintained and routinely backed up. It is good practice to have off-site and off-line storage for the master copies to protect the digital archives from fires or other natural disasters. Files should be backed up to a DVD or other transportable device and stored at a separate location from the digital archives. Since technology is always changing, it is difficult to know which storage devices will be compatible with future programs. As new technology develops it will be important to convert the files to the new format and software.

⁴ The handbooks are accessible at www.contentdm.com to organizations that have a CONTENTdm license. Organizations accessing CONTENTdm through the Missouri Digital Heritage can contact the Development Office at the Missouri State Library for the account access information. If not accessing CONTENTdm through Missouri Digital Heritage, visit the website directly for more information.

⁵ Caplan, Priscilla. *Stewardship in the Digital Age: Managing Museum and Library Collections for Preservation and Use*. Institute of Museum and Library Services. Page 9 of Web Wise 2007 Conference Proceedings. <http://www.ims.gov/pdf/WebWiseProceedings2007.pdf>

Scanning Standards and Guidelines

Digital imaging is the process of converting a document into an electronic file that is an exact copy of the original and depicts every detail, from physical damage to the color of the document. This chapter explores minimum guidelines for scanning success; however, institutions are encouraged to exceed them and strive to capture the best possible images allowed by the project budget. In 2001, the Missouri Digitization Project Planning Committee established and defined key guidelines essential for digitization and provided in-depth explanations for such requirements. The following information is from the *Scanning Guidelines and Recommendations*.

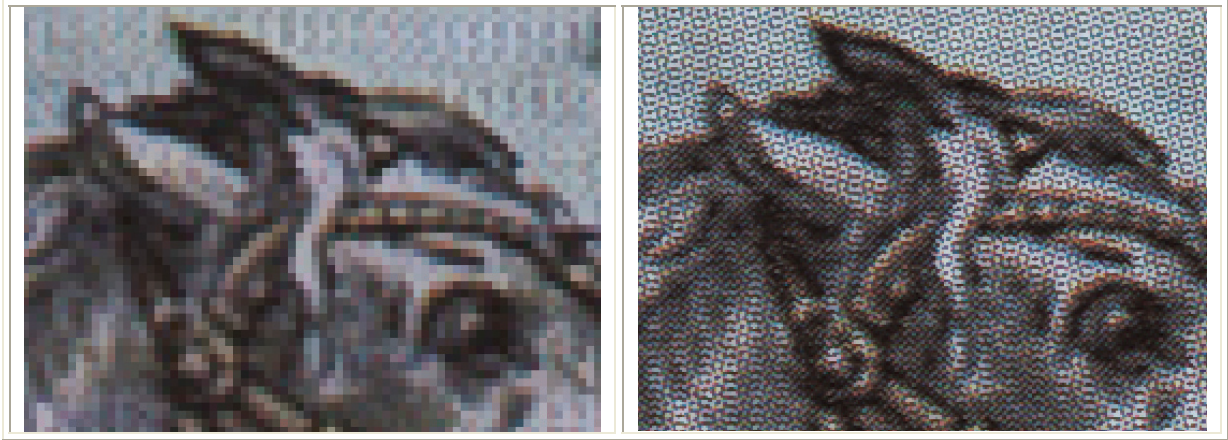
Factors Affecting Quality

The creation of high quality digital images requires the integration of several elements. The following section describes the terminology used in the process and the factors that are involved in images creation.



Each digital image consists of a series of dots or **pixels** arranged in rows and columns. Each pixel can be represented as black, white, shade of grey or color. Individually each pixel only represents a small section of the entire image, but when paired with neighboring pixels the image as a whole comes into focus. The arrangement of the pixels on the image canvas is called a **bitmap**. Each pixel is a single point on the canvas. Multiple pixels are used to represent the horse's ear above, which in reality is a small portion of the entire image. The bitmap designates where each pixel belongs to create the entire image.

Each pixel consists of bits, and the amount of bits in each pixel is measured by **bit depth**. Bit depth ranges from 1-bit per pixel to 24-bits per pixel. Higher bit depth allows for greater variations of colors. 1-bit only allows for black or white, while an 8-bit image supports 256 shades of gray or color. Finally, a 24-bit image can represent more than 16 million unique colors.



Resolution determines the clarity of an image, and resolution during scanning is measured in ***dots per inch (dpi)***. This is literally a measurement of the dots per square inch of a scanned image. Higher dpi allows for greater image quality, as more shades of color are available to represent the image. The horse on the left was scanned at 150 dpi and the horse on the right was scanned at 600 dpi. At 600 dpi, more pixels are available to show greater details in the horse's head. To see the individual pixels at 600 dpi the image must be enhanced to the point where it is difficult to distinguish what is actually being viewed.

The display quality also depends on the monitor's resolution, which is measured in ***pixels per inch (ppi)***. Each monitor is limited on the total number of pixels it can display. Higher display resolution allows for greater image presentation. Display resolution can be increased by purchasing a high quality video adapter for the computer. Images can be scanned at a high dpi, but with a low quality monitor the screen will not represent all of the pixels that were captured.

Digital Master and Derivative Images

Most digitization projects involve the creation of multiple copies of a single image. The types of images are ***master images***, ***access images*** and ***thumbnails***. Access and thumbnail images are also known as derivative images, because they are created from the master image. The master images will be an exact replica of the original document and will not be compressed; they are scanned as ***TIFF*** files.⁶ This file format allows for the greatest flexibility to merge into the next generation of technology. All of the derivative images are compressed to allow for ease of access to the image through network connections. Imaging software, such as Photoshop, can be used to enhance derivative images by editing scans of damaged documents.

⁶ TIFF – Tagged Image File Format

Creating Master Images

Create a master image of each item to be digitized. Scanning should be done at the highest practical level to achieve the best quality image possible within the budget. The objective is to have a high quality image from which to create derivative image files. The digital **master** serves as the digital source for future generations of derivative files. When changes in technology offer new opportunities or necessitate **migration**, the master will be used. Long-term preservation of digital master files requires a strategy of identification, storage and migration to new media. **Master images** should not be edited or processed for any specific output, and they should be uncompressed to ensure that no information is lost. Since the master will serve as the surrogate for the original, intensive quality control should be applied during the scanning process.

Derivative Images

The derivative files are primarily used to enhance presentation over the Internet. Derivative files are created from the master, usually involving a loss of information through compression and editing. Derivative images can be edited through image software if the original image is illegible. Uncompressed master image files will be very large. A master image will be captured at 300 dpi on average, which will produce a 24.6 MB TIFF file; 300 dpi is the minimal standard for master images. Master images captured at 500 dpi, as with the *Community and Conflict* project, result in a much larger file. Such files travel slowly through computer networks; therefore, compressed files are used for online distribution. Derivative images are normally compressed into **JPEG or GIF** files through imaging software.⁷ There are two broad types of compression, **lossy** and **lossless**.

- **Lossy** compression involves the removal of data from the image file. The data is lost and is not recoverable even if the file is reconverted to a TIFF.
- **Lossless** compression does not involve a loss of data; however, the files cannot generally be compressed as much as lossy files.

The *Community and Conflict* staff chose to create exact digital surrogates of the original document in order to accurately represent the collection. Therefore, all images, regardless of document type, were captured with 24-bit color and approximately 500 dpi. The standards shown in the next table can vary based on condition of image being scanned.

Example: Faded documents or documents in poor condition should be scanned at a higher resolution to increase zoom capabilities.

⁷ JPEG – Joint Photographic Experts Group
GIF – Graphics Interchange Format

Scanning Reference Guides

The following charts were designed to provide a quick reference guide for digitization projects. Project staff adapted these guidelines from the Missouri Digitization Project Planning Committee's *Scanning Guidelines and Recommendations*. Note that three files are created for each image: Master Image, Access Image and Thumbnail; each of these files serves a distinct role when hosting images online. CONTENTdm requires only a master image; when that image is uploaded, the access and thumbnail images are generated by the software.

Reference Guide for Master Images	<u>Text Documents</u>	<u>Photographs</u>	<u>Maps</u>
Bit – Depth	24-bit color	24-bit color	24-bit color
Capture Resolution	500 dpi	500 dpi	600 dpi
File Type	TIFF	TIFF	TIFF
Compression	None	None	None

Reference Guide for Access Images	<u>Text Documents</u>	<u>Photographs</u>	<u>Maps</u>
Bit – Depth	24-bit color	24-bit color	24-bit color
Compressed Resolution	200 dpi	200 dpi	300 dpi
File Type	JPEG	JPEG	JPEG
Compression	10:1	10:1	10:1

Reference Guide for Thumbnail Images	<u>Text Documents</u>	<u>Photographs</u>	<u>Maps</u>
Bit – Depth	24-bit color	24-bit color	24-bit color
Compressed Resolution	72 dpi	72 dpi	72 dpi
File Type	JPEG	JPEG	JPEG
Compression	Inherent to Image Software	Inherent to Image Software	Inherent to Image Software

Scanners

Proper equipment is crucial in order to meet established digitization standards and guidelines. Scanners are capable of supporting resolutions from 72 to 8800 dpi and it is important to note that resolution is established by the scanner. A ***Charged Coupled Device (CCD)*** is the hardware within the scanner that determines the highest resolution it can achieve. Low quality scanners tend to have lower quality CCDs, thus producing lower quality scans. Image software cannot increase or enhance the actual resolution of an image, but it can be used to alter the presentation of the image. Purchase high quality scanners and scan at high resolutions to produce the best images.

Scanner Suggestions for Various Material Types			
Single leaf, regular size, flat materials	Single leaf, oversized, flat materials	Bound materials	Transparent media
<ul style="list-style-type: none">• Flatbed scanner• Sheet fed scanner (if non-brittle)• Digital camera	<ul style="list-style-type: none">• Oversize flatbed scanner• Sheet fed scanner (if non-brittle)• Digital camera	<ul style="list-style-type: none">• Digital camera with book cradle• Right angle, prism, or overhead flatbed scanner	<ul style="list-style-type: none">• Slide scanner• Film scanner• Multi-format flatbed scanner• Digital camera

Quality Control

The digitization process involves multiple steps often completed by multiple staff members. Each project should develop quality control checks to ensure the final product meets the adopted standards. The National Archives and Records Administration recommends that a minimum of 10 images or 10% of each batch of digital images, whichever quality is larger, be checked for quality control standards. The following chart was developed from NARA recommendations.

Quality Control Checks for Scanning and Image Files		
Scanning Related	Image Related	File Related
<ul style="list-style-type: none"> • Correct dimensions • Correct resolution • Orientation <ul style="list-style-type: none"> o Document – landscape/horizontal o Image – horizontally or vertically flipped • Image skew • Cropping • Image completeness • Missing pages or images 	<ul style="list-style-type: none"> • Tone <ul style="list-style-type: none"> o Brightness o Contrast • Color accuracy • Saturation • Dust • Newton's rings • Missing scan lines • Loss of detail • Lack of sharpness • Flare 	<ul style="list-style-type: none"> • Files open and display • Proper format <ul style="list-style-type: none"> o TIFF, etc • Compression <ul style="list-style-type: none"> o Derivative images only • Color mode <ul style="list-style-type: none"> o RGB, grayscale, etc • Bit depth <ul style="list-style-type: none"> o 48-bits, 24-bits, etc

Additional Resources

The Colorado Digitization Program offers outstanding information about the details of scanning, types of scanning hardware, and resources for deciding which equipment to purchase:

<http://chnm.gmu.edu/digitalhistory/links/cached/chapter3/link3.45.CDPscanningguidelines.html>

The Technical Advisory Service for Images:

<http://www.tasi.ac.uk/advice/creating/scanners.html>

Scanning tips by Wayne Fulton:

www.scantips.com

Moving Theory into Practice: Digital Imaging for Libraries and Archives - Cornell University Library 2000-2003:

<http://www.library.cornell.edu/preservation/tutorial/contents.html>

Preservation of Library & Archival Materials: A Manual "Digital Technology Made Simpler" - Northeast Document Conservation Center:

<http://www.nedcc.org/resources/leaflets.list.php>

Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images – U.S. National Archives and Record Administration June 2004:

<http://www.archives.gov/preservation/technical/guidelines.pdf>

Transcription Policies and Procedures

Providing an online transcript of a handwritten original document is an invaluable research tool. The goal of transcribing is to produce a copy of the manuscript that is easily legible and accessible. The transcription process includes a complete replication of the original document in a typed format. Some organizations choose to work with .doc files which allows for various specialized formats such as ~~strike through text~~^{superscript}, subscript, and many other format types. The CONTENTdm software chosen for Missouri Digital Heritage requires that the transcription and metadata be entered in a .txt file. The encoding of the specialized script types in .doc files are not applicable to .txt files.⁸

Suggested standards for transcription in CONTENTdm are as follows:

File Type and Naming Structure

Copy every word of a document into a .doc or .txt file.⁹ When entering a transcription into CONTENTdm, the file name needs to match the file name of the scanned document.

CC_jmack_03p01.tif – scanned page
CC_jmack_03p01.txt – transcription/metadata file

Transcription Formatting and Stylization

Stylized formatting; such as, following the placement of heading, dates, salutations and other text will not convert into CONTENTdm. CONTENTdm transcriptions must be created in a block format without paragraph breaks. If the author continued the letter in the margins of the document or across multiple pages, transcribe the section with notation about it's orientation on the page.¹⁰

⁸ If including specialized text such as superscript or subscript in a transcription an additional image can be created for the document in CONTENTdm making it a compound object. This extra image will be a scanned print-out of the transcript in word, or a .pdf file. Both of these options will allow for specialized text to be displayed in CONTENTdm. It will also allow the user to read the entire transcription on one page. An example of this technique can be viewed at the William Holmes McGuffey letter at Miami University Libraries –Digital Collection at:
<http://doyle.lib.muohio.edu/cdm4/document.php?CISOROOT=/mcguffey&CISOPTR=844&REC=5>

⁹ Working with a .doc file will allow the use of features like spell check before saving the final transcript as a .txt file.

¹⁰ Following the original format of the manuscript can be accomplished in a word document. Consult footnote number eight for an example of how Miami University addresses this formatting issue.

Extreme Transcription Example

I would like to stay here till
spring. for I think this month will
be very cold, we don't expect to go to
Tennessee, but we may go to
in the spring, and probably stay
the same. Mary is ill. I am
well. I am writing that they will
be over by next June. I have
got another house. I would like to
be in the fall. I am well. I am
summers. I am well. I am
I am in the house to stay. we are
both well and enjoying life as
usual. The Dr. is writing last
evening. I think I am up to. all right
I have got a letter last week from
your mother. I am well. I am
to go. I want to write you this
time, and correct my mistakes.
I am in the house to stay. I am
but still will write, if it don't
interest you don't read it.

Transcription of this document can be found directly after the transcription guidelines.

Transcription Guidelines

Multi-directional Text and Text in Margins

Authors occasionally wrote text horizontally, vertically, and in the margins of a single sheet of paper. This is often found when the author had limited access to paper or had become inspired in the last page or two of a letter. In all cases, read the text in entirety and determine which entry was written first, and then transcribe the document in proper order.

Situation	Transcription Heading
Text written horizontally and vertically on one page (see the sample above)	[written horizontally on page]: transcription <u>followed by</u> [written vertically on page]: transcription
Text written in the margins of the document	[written sideways in margin]: transcription
Text written across two pages in the margin – i.e. text written across the margin of page 3 that follows the last sentence on page 4	[written across top margin, follows text on page 4]: transcription

Depict errors in original document as follows:

Spelling - Transcribe text with spelling errors as found in the original document; directly following the misspelled word add the correct spelling in [square brackets].
Note: add spelling corrections only to words that would be keyword search terms.

Capitalization – Replicate inconsistencies in capitalization as depicted in the original document.

Strikethrough & Underlined Text – Specialty text cannot be depicted in a .txt file; represent emphasis to the text by placing the word in parentheses.

Punctuation – Replicate punctuation as indicated in the original document.

Dates – Do not standardize date formats. Leave dates as depicted in the original document, with the exception of truncated dates. See example in Omitted Text section.

Mutilated or Illegible Documents

Documents that have been damaged by fire, water and other contaminants may be illegible. If the transcriber can draw inferences from the context of the sentence, the missing words are placed in square brackets. If no inferences can be drawn from the text, describe the type of damage in square brackets. Use the abbreviation MS for manuscript.

In cases where the text is uncertain, transcribe the text as closely as possible to the original. [MS illegible] can be inserted directly into the word if part of the word is readable.

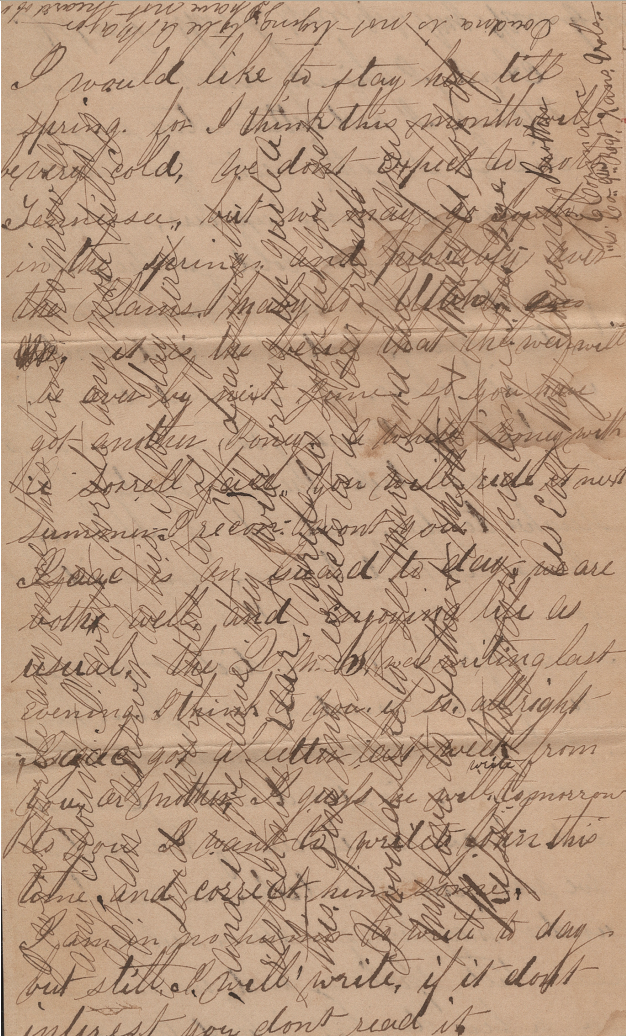
Situation	Transcription Heading
Damage to document cause by fire – part or entirety of word is illegible.	[MS burned]
Damage to document cause by water – part or entirety of word is illegible.	[MS blotted]
Text in document is completely illegible	[MS illegible]
Partial text is illegible	gr[MS illegible]ly

Omitted Text

Transcribers should refrain from entering omitted text from the document, even if the additional text would assist reading the manuscript. The goal of the transcription is to accurately portray the original text. In cases where searchable terms are omitted make sure the description metadata discusses the topic and omitted words.

Exception: Truncated Dates

Situation	Example Text	Example Transcription
Truncated Dates	Sept. 16, '61	Sept. 16, [1861]

Transcription Example	
Original Scan	Transcription ¹¹
	<p>[written horizontally on page]:</p> <p>I would like to stay here till spring. for I think this month will be very cold. we don't expect to go to Tennessee, but we may go South in the spring. and probably over the Plains. Maby to Utah as it is the belief that the war will be over by next June so you have got another Pony A white Pony with a sorrel foue [foal]. You will ride it next summer I recon won't you Issac is on Guard to day, we are both well and enjoying life as usual the [ms illegible] were writing last evening I think to you if so alright Issac got a letter last week from your mother. I guess he will write tomorrow to you. I want to write to John this time and correct him some, I am in no humor to write to day but still I will write if it don't interest you don't read it.</p> <p>[written vertically on page]:</p> <p>I won't write any more this time. No news left any account I won't write any more till I get an answer to this, if we stay here send your letters to Ft. Leavenworth. and if we leave they will bring up quite a respectable rear. Write as soon as you get this. give any respects to all our friends I would like to see Mary and all of you My love to Father & Mother & Mary A lot of Kisses. No more this time. Bye Bye.</p> <p>E. Corman 1st CO 9th Regt. Kans Vols.</p> <p>[written horizontally in the top margin]: Doudna is not trying to be a Major & I have not heard of it.</p>

¹¹ Paragraph breaks were added to this transcription for display purposes only.

Quality Control

Checking transcriptions for errors can be difficult as they often reflect errors in the original document. Each transcriber should be responsible for his/her work and double check it for accuracy; however, transcriptions should be checked by a second reviewer who may notice mistakes overlooked during the first review.

Additional Information

This transcription policy combines best practices and tested procedures. Some of these guidelines were adapted from other organizations to fit the requirements of CONTENTdm. For additional transcription policies review the following procedures:

Villanova University

<http://www24.homepage.villanova.edu/teri.incrovato/Important%20docs./Standards%20for%20Transcription.doc>

Canadian Mysteries

http://canadianmysteries.org/staff/pdf/Transcription_Conventions.pdf

Minnesota Historical Society

http://www.mnhs.org/about/departments/processing/transcribing_manuscripts.pdf

Conclusion

The final digitized product, when done well, should appear effortless. Like so many endeavors, however, professional quality digitization is far from easy and it involves a multitude of complex steps. The outcomes justify the resources expended because:

- The online collection attracts users and expands the audience for the host institutions
- Materials that were not readily available to the general public are accessible online
- Geographic barriers are eliminated
- Customers can easily and effectively find information
- The content supports educational programs and expand understanding
- Fragile items are protected from frequent handling

This manual represents the lessons learned in a single planning project. The intent in writing it was to not only fulfill the goals of the LSTA Cooperation Grant, but to provide others insight into the process and the results. Materials chosen for *Community and Conflict: the Impact of the Civil War in the Ozarks* represent a critical period in the history of the region, and make a significant contribution to the understanding of the complex issues surrounding the lives and communities of the time.

Digitization is important work. The Institute of Museum and Library Services (IMLS) has sounded an alarm through a national initiative to raise public awareness of the importance of preservation. A study by Heritage Preservation, revealed that the nation's collections of objects, documents, digital material and living collections are imperiled and in need of swift protective action. These collections are essential to American culture and history, and it is the duty of all library, museum and archive staff to preserve them for generations to come. Access the Heritage Health Index Report on the State of America's Collections at: <http://www.imls.gov/collections/about/hhi.htm>.

A very comprehensive collection of information covering all aspects of digital projects is presented by IMLS. General information, assessment tools, best practices and standards, care of digital materials, collaboration, funding and training is available at: http://www.imls.gov/collections/resources/care_dig.htm.

Another resource, *L.A. Art Online*, describes valuable lessons learned from the Getty Electronic Cataloguing Initiative: http://www.getty.edu/grants/pdfs/LA_Art_Online_Report.pdf.

APPENDIX



Target Audience Survey



COMMUNITY AND CONFLICT:

THE IMPACT OF THE CIVIL WAR IN THE OZARKS

In cooperation with other agencies, the Springfield-Greene County Library District is planning a digital preservation project focused on the Civil War in the Ozarks. Our goal is to create a database of digitized documents and make it accessible to the general public. By completing the survey below, you can help us determine the themes, content and search strategies that will make our database relevant and user-friendly.

1. When conducting research on historical events, or specifically the Civil War, what subjects are of most interest to you? Check all that apply.

- | | | |
|---|---|---|
| <input type="checkbox"/> Battles/military | <input type="checkbox"/> Families/home life | <input type="checkbox"/> Letters/diaries |
| <input type="checkbox"/> Business/economics | <input type="checkbox"/> Agriculture | <input type="checkbox"/> Technology/industry advancements |
| <input type="checkbox"/> Criminal acts | <input type="checkbox"/> Social life/entertainment | <input type="checkbox"/> Politics |
| <input type="checkbox"/> Ethnic groups | <input type="checkbox"/> Gender issues (women's suffrage, etc.) | |
- Other (please list): _____

2. Rank the usefulness of each resource material below:

- | Very useful | Useful | Neutral | Not useful | Do not use | |
|--|--------------------------|--------------------------|--------------------------|--------------------------|--|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Online databases (EBSCO, J Store) |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Internet search engines (Yahoo, Google) |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Books, magazines, encyclopedias, textbooks |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Photographs |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Letters |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Maps |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Newspapers |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Personal accounts, diaries, journals |
| <input type="checkbox"/> List others _____ | | | | | |

3. What types of research material do you gather when first starting to conduct historical research?
- _____
- _____

4. If you were looking at a historical document online, which format would you prefer to see? Check one.

- ☐ Scan of the original ☐ Transcription ☐ Scan and transcription

5. List some websites that you have found most helpful for historical research:
- _____
- _____

6. Rank the usefulness of the features you prefer to use on websites:

- | Very useful | Useful | Neutral | Not useful | Do not use | |
|--|--------------------------|--------------------------|--------------------------|--------------------------|----------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Keyword searching |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Drop-down menus |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Geographical links |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Chronological organization |
| <input type="checkbox"/> Other features: _____ | | | | | |

7. Check your occupation:

- | | | |
|---|--|---|
| <input type="checkbox"/> Teacher | <input type="checkbox"/> College faculty | <input type="checkbox"/> Graduate student |
| <input type="checkbox"/> High school student | <input type="checkbox"/> Undergraduate student | |
| <input type="checkbox"/> Other. Describe: _____ | | |

Use the reverse side of this page for comments.



Springfield-Greene County Library District
P.O. Box 760 • Springfield, MO 65801-0760 • thelibrary.org

Community and Conflict Work Flow Strategy

In order to meet the digitization goals for phase two, 2,500 pages of documents will be processed within ten months. To complete this project on schedule, it is essential that staff complete a monthly four step strategy to process approximately 250 pages per month, or at a daily rate of 12 pages, in addition to completing all other tasks. This ten month plan allows for two additional months to address any unforeseen technical issues, equipment malfunctions, or other unavoidable complications that may arise. If the initial 2,500 pages are completed ahead of schedule, additional collections will be included, up to 3,000 total pages.

Project staff meetings will be regularly scheduled during the grant period to consistently evaluate the workflow processes, communicate any changes that are needed, and to assure compliance with the timeline that was established in the Digital Imaging LSTA Application.

The collection processing strategy includes the following steps:

1. Scan collections
2. Collect metadata
3. Transcribe text
4. Upload all materials into CONTENTdm

During the evaluation process, collections were organized according to location. The collections that are the furthest distance from Springfield will be digitalized first. The purpose of this prioritization is to avoid any potential delays in collection processing which may occur as a result of winter storms or treacherous road conditions.

July 1, 2008 – ongoing

Brian Grubbs and Teresa Hernandez will conduct site visits as prioritized by distance. During the site visits, selected collections will be scanned in their entirety. Two scanners will be used to expedite the scanning process and reduce excessive and unnecessary travel. All scans taken will then be stored to an external hard drive and routinely backed up to prevent any potential loss due to equipment malfunctions. As scans are captured, staff will ensure all images meet quality control standards. Ideally, the scanning process will be completed the first week of each month.

As scans of individual collections are captured, staff will complete steps 2, 3 and 4 of the strategy above, thus enabling all of the collections to be processed systematically. The project associates will gather metadata and transcribe the text; when not in the field making site visits, Grubbs and Hernandez will also participate in the metadata gathering and transcribing process. All metadata will be entered into the metadata worksheet that is generated through CONTENTdm. From these worksheets, Grubbs and the CONTENTdm technician will then import the data directly into CONTENTdm. Grubbs will conduct a

final review of all metadata and transcriptions for the purpose of quality control and to ensure that each collection meets project standards.

September, 2008

Project staff will launch a demonstration web site to test structure and functionality.

November, 2008 – May, 2009

The Springfield-Greene County Library Web site Design team (Nancy Eike, Murphy Moore, Sarah Marshall, and Danny Dye) will work with project staff to develop the architecture, function and layout components for the web site, and they will provide the expertise to address any potential server or software needs. Project staff will test format features and functions to ensure user friendly web access. Grubbs and Hernandez will work alongside graphic artist, Danny Dye, to create the concept art and design for the web site.

July, 2008 – January, 2009

The selection committee will work with Grubbs and Hernandez to develop and evaluate content for the web site. Community Relations staff will assist final editing of web site content prior to publication.

Ongoing Activities

In addition to the tasks and activities included in the collection processing strategy, Grubbs and Hernandez will also continue to perform a variety of other duties necessary for the success and future of the project. They will coordinate meetings with the planning partners and make promotional presentations about the *Community and Conflict* project. Due to the abundance of collections and materials available throughout the region, project staff will continue to identify key collections pertinent to the scope of this project for future inclusion. Staff will also continue to work with local Civil War historians, Dr. William Piston, John Rutherford, and Michael Price (Rutherford and Price are Library associates who work in the Local History Department at the Library Center) to evaluate new collections. Based on the size and scope of collections identified during the planning and development phase and the overwhelming response of the surveyed audience, staff believes this project will continue to attract more contributions and many suggestions for additional materials to be included. Multi-year funding may be required to fully develop the collection and achieve the project goals.

Project staff will strive to participate in any new training and development workshops relating to digitalization in an effort to keep up with the latest technology and advancements.

Tasks:	Assigned to:	Resources:	Timeline:											
			Jul-08	Aug-08	Sep-08	Oct-08	Nov-08	Dec-08	Jan-09	Feb-09	Mar-09	Apr-09	May-09	Jun-09
Finalize project procedures and workflow	Grubbs	Staff time, hardware, software	X											
	Grubbs, Hernandez & Project Partners	Transportation, scanners, computers, travel & staff time	X	X	X	X	X		X	X	X	X		
Site visits and scanning	Cameron, Glass, Grubbs & Hernandez	Staff time	X	X	X	X	X	X	X	X	X	X	X	
Metadata gathering & transcriptions	Dye, Grubbs, Hernandez & Computer Services						X	X	X	X	X	X	X	
Website development	Grubbs & Hernandez	Staff time	X	X	X	X	X	X	X	X				
Content development	Selection Committee	Staff time	X	X	X	X	X	X	X	X				
Evaluate content	Computer Services Staff	Staff time			X									
Demonstration web site & troubleshooting	Justice, Grubbs	Staff time			X	X	X	X		X	X	X	X	
Material uploaded to CONTENTdm	Grubbs, Hernandez & Project Partners	Transportation, food & staff time	X	X	X	X	X		X	X	X	X		
Identify and prioritize additional collections for inclusion in possible second phase	Grubbs & Hernandez	Staff time & Transportation												
Promotional public visits and presentations (as requested)	Partner Committee	Staff time	X		X									
Planning partner meetings	All Project Staff	Staff time												
Wrap up/conclusion														X

Project Flyer



IF YOU HAVE original artifacts from 1850-1875 to share, or if you would like additional information about Community and Conflict: The Impact of the Civil War in the Ozarks, contact:

Brian Grubbs at
briangr@thelibrary.org
or call 417-616-0544



Community and Conflict: The Impact of the Civil War in the Ozarks is a Library Services and Technology Act Digital Imaging Grant funded by the Institute of Museum and Library Services and coordinated by the Missouri State Library.

The following Project Partners provided invaluable guidance and resources in the planning and implementation of **Community and Conflict:**

Greene County Archives and Records Center
 History Museum for Springfield-Greene County
 Jasper County Records Center
 Missouri State University –
 Dr. William Piston, History Professor,
 and David Richards, Head of Special
 Collections and Archives Department

Western Historical
 Manuscript Collection, Rolla
 Wilson's Creek National Battlefield

A host of other contributors made this project possible through the generous lending of archival materials and private collections:

Benton County Archives, Rogers, Arkansas
 Bushwhacker Museum
 Greene County Medical Society
 Mine Creek Battlefield, Kansas SHS
 Rogers Historical Museum
 Shiloh Museum of Ozark History
 Special Collections, University of Arkansas
 Libraries, Fayetteville
 Territorial Capitol Museum, LeCompton Kansas

Andy Thomas, Artist
 Western Historical Manuscript Collection,
 Columbia
 Wide Awake Films
 Private Collectors

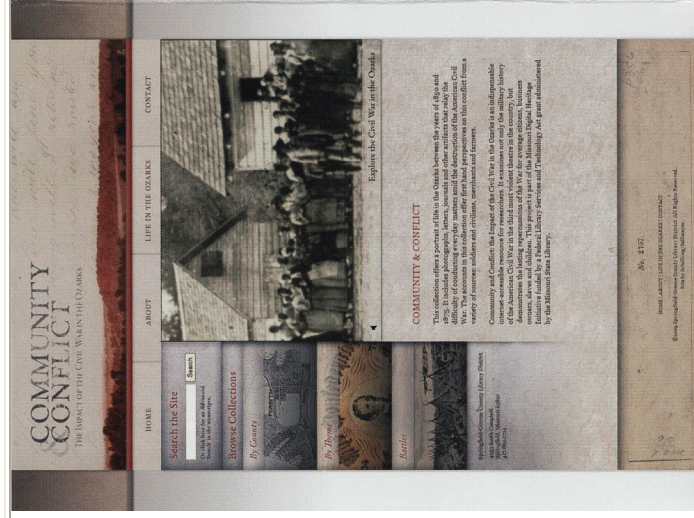

Springfield-Greene County Library District
 P.O. Box 760 • Springfield, MO 65801-0760
thelibrary.org

COMMUNITY & CONFLICT

THE IMPACT OF THE CIVIL WAR IN THE OZARKS

ozarkscivilwar.org



Research Features

- High-caliber scans of research documents
- Transcriptions with side-by-side comparisons to the scanned image
- Keyword search
- Geographic and thematic searches
- Historic interpretation
- Links to other American Civil War research sites
- Bibliographic citations

Discover Community and Conflict: The Impact of the Civil War in the Ozarks, a digitized collection of pre-Civil War, Civil War and Reconstruction Era artifacts. Learn about the effects of the War Between the States on the lives of the people living, working and fighting in one of the most contested regions of the country.

Explore life along the Border States through letters, maps, diaries, court and government documents, photographs and more, and discover the trials and tribulations of everyday Americans struggling to survive in the rugged Ozarks country as war erupted around them.

Investigate the virtually untouched resources of the Trans-Mississippi Theater from the comfort of your home, school or office computer.

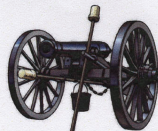
Collections gathered from institutions and individuals throughout the region document life in the Ozarks during the tempestuous period of 1850-1875 through the exploration of urban development, slavery, agriculture, military life, women and children, civilian refugees and many other issues that represent a crucial period in the history and culture of the region.

Find collections like these and others online at ozarkscivilwar.org:

The Mack Family Letters (1850-1869): These letters provide valuable insight into the lives of Union sympathizers in the Ozarks. In many ways, their experiences are typical of Southern Unionists, although their home in Missouri placed them squarely in a border region. The collection contains correspondence between the family members serving in the army and while the soldier's responses are missing, the existing letters provide researchers with a unique perspective on the civilian experience in southwest Missouri.

State of Missouri vs. Drew, Slave (1847): Court documents from a case filed in Greene County, Missouri, for which Drew is accused of murdering Lige, another slave, by giving him poisoned whiskey. This case involves the depositions of multiple witnesses, mainly other slaves, who had knowledge of a fight or disagreement between Lige and Drew. These depositions afford researchers rare accounts of the lives of slaves in the Ozarks.

Project Release Form



COMMUNITY AND CONFLICT: THE IMPACT OF THE CIVIL WAR IN THE OZARKS

RELEASE FORM

I understand that I am participating in the Community and Conflict digitization project. I understand that the purpose of the project is to document objects, photographs and other records related to the Civil War in the Ozarks (collectively "Materials"). The Materials will be used to illustrate the significance and impact the Civil War had on the Ozarks, and may be used for scholarly and educational purposes. I understand that the Springfield Greene County Library District will retain a digital copy of my Materials presented and that Materials may be used in publications, research, exhibitions and presentations and will be located on a database accessible on the Web and successor technologies, without any compensation to me. I further agree that the Materials may be used for promoting the Springfield Greene County Library District and its activities in any medium.

I hereby grant to the Springfield Greene County Library District the right to use the Materials as stated above. By giving my permission, I understand that I do not give up any copyright that I may hold.

I also grant to the Springfield Greene County Library my absolute and irrevocable consent for any Materials provided by me to be used, published and copied by the Springfield Greene County Library District in any medium.

I release the Springfield Greene County Library District from any and all claims and demands arising out of or in connection with the copying, documenting or use of such Materials, including, but not limited to, any claims for damage, destruction, or loss of the Materials, and any claims for copyright infringement, defamation, invasion of privacy or right of publicity.

ACCEPTED AND AGREED

Signature _____ Date _____

Printed Name _____

Address _____

City _____ State _____ Zip Code _____

Telephone () - _____



Springfield-Greene County Library District
P.O. Box 760 • Springfield, MO 65801-0760 • thelibrary.org

Missouri Digital Heritage		Updated Metadata Scheme & Quick Reference Guide	
4-Mar-08		to Cataloging in Missouri Digital Heritage	
Haiying Qian			
<u>Required Elements</u> Item Level Records	Field Name	Explanation	Controlled Vocabulary or Format Required
Title	Title	A name given to the resource, usually by the resource creator or publisher, or in absence thereof, by the digitizing/cataloging agency. If no suitable title is available for the resource, the person creating the metadata record may create a title. If the item is in HTML, view the source document and title comes from meta title as viewed in source codes, also at top of screen in Netscape and Internet Explorer browsers.	Skip beginning articles like “a”, “an”, “the”, etc. Example: Title 18 th century costume resources online
	Title. Alternative	Variants of title. If the item is in html, alternative title comes from the web page itself; subsequent titles, if necessary, are expansion of portions of previous titles.	Spell out acronym, Romanized numbers and so on Example: Title. Alternative Eighteenth century costume resources online
Creator		An entity primarily responsible for the intellectual content of the resource. For example, the author of written documents; artists, photographers, the collector of natural specimens or artifacts, illustrators of visual resources, or performers or composers for sound and other audiovisual resources.	Surname, First name. Enter corporate names in full, direct form. In the case of a hierarchy, list the parts from the largest to smallest, separated by periods. Use of standardized forms of names and authority files by checking LC authority files. Use “Unknown” if creator is unknown.
Subject	Subject.LCSH	Library of Congress Subject Headings	LCSH
	Subject. Local	Keywords, place and personal names that users might search for	
Description		An account of the content of the resource.	A brief summary of the content.
Date	Date. Original	Creation date of the original resource	YYYY-MM-DD
	Date. Digital	Creation date of the digital surrogate.	YYYY-MM-DD
Format		Refers only to digital resources.	Internet Media Types (www.iana.org/assignments/media-types/) Example: “Image / Tiff” or “Text/ JP2”.
Identifier		A character string or record number that clearly and uniquely identifies a digital object or resource. The Identifier element ensures that individual digital objects can be accessed, managed, stored, recalled, and used reliably.	Most commonly, use the local digital file name. You may also use a formal identification system such as the URL, URI, or DOI. Example: Peabody_Coal#3.tif

Relation	A field describes a resource that is part of larger resource. This includes records describing individual pieces (e.g. a single photograph) and aggregations of pieces (e.g. correspondence in someone's papers) that are part of a larger collection. Generally the name of the collection is entered. A newspaper's preceding and succeeding titles can also be included in the relation field.	Input the ISSN, ISBN, other international standard numbers, and local naming conventions that describe the original in the Source field, rather than in the Relation.
Source	When applicable, use the Source element to cite any other resource from which the digital resource was derived , either in whole or in part. Some digital resources are "born digital" and derive from no pre-existing resource; in these cases, the Source element is not used.	Whenever possible, include a unique standard identifier such as an ISSN, LC call number, Dewey call number, or NTIS report number. If no standard identifier exists, use a local call number, control number, accession number, or barcode. Identify the institution associated with such locally derived numbers
Rights	Rights information often encompasses intellectual property rights (IPR), copyright, and various property rights. If the Rights Management element is absent, no assumptions can be made about the status of these and other rights with respect to the resource	Rights management or usage statement, a URL that links to a rights management statement. A rights management statement may contain information concerning accessibility, copyright holder, restrictions, securing permissions for use of text or images, etc. (examples of rights statements are shown below)
Copy Request	Information concerning reproduction of images, documents and other digital resources.	The phone number, email, and mailing address of the department or staff position that will be most helpful in getting the end-user a copy of the digital resources.
Contributing Institution	The name of institution who provided the born digital resources or the original resource which digital resources were derived	
Optional Fields: Item Level Records		
Publisher	The name of the entity that made the digital resource available in its present form , such as a corporate body, publishing house, museum, historical society or university.	

Contributor	Contributor	A person or corporate body responsible for making secondary but important contributions to the content of the resource, who/which is not already included in a Creator element field.	Example: Author of a book in creator field; illustrator of the book in contributor field
Type	Type	To describe the physical or digital manifestation of the resource, use the Format element. Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the DCMI Type Vocabulary—see 3 rd worksheet in this Excel file.	Use the DCMI Type Vocabulary (http://dublincore.org/documents/dcmi-type-vocabulary/). Exp: “image,” “text,” “collection,” “moving image,” etc.
Language	Language	A language of the intellectual content of the resource.	Use full name of language, repeat for multiple languages
Coverage	Coverage	Describes the spatial area or time frame , which intellectual content of the resource encompasses.	
Notes:			
* Alternate Subject Headings		Other thesauri may be used, such as Medical Subject Headings (MeSh), Art & Architecture Thesaurus (AAT), Getty Thesaurus of Geographic Names (GEGN). In general, any other field names or qualifiers may be used in addition to the required fields. Non-required fields should be discussed with the Missouri Digital Heritage Metadata Cataloger prior to the cataloging process to ensure optimum mapping and indexing.	Use of thesauri other than LCSH should be discussed with the database coordinator prior to cataloging in order to ensure proper mapping and indexing of subject terms.
* Books, Monographs and other Full-Text Materials.		When possible, crosswalk subject headings from MARC records if such cataloging already exists for the item. Full-text indexing of such materials should not preclude use of subject headings to improve search and retrieval.	
Collection Level Fields:		Title, Creator, Subject, Description, Identifier, Relation:	

Community and Conflict Metadata Scheme

Field Name	Dublin Core Map	Hidden (Y/N)	Searchable (Y/N)	Definition	Controlled Vocabulary / Format Requirement
Element Identifier	Identifier (M)	Y	N	A character string or record number that clearly and uniquely identifies a digital object or resource. The identifier element ensures that individual digital objects can be accessed, managed, stored, recalled, and used reliably. The element identifier is assigned by the Library staff.	CC_[Collection Name]_[Element Number] Example: CC_mosesb_03 Page Example: CC_mosesb_03p02
Title	Title (M)	N	Y	A name given to the resource, usually by the resource creator. If no title is available for the resource, the person creating the metadata record may create a title.	Skip beginning articles like “a,” an,” the,” etc. Example: John A. Mack letter to Moreau Mack
Creator	Creator (M)	N	Y	An entity primarily responsible for the intellectual content of the resource. Name of the author of the document, photographer, etc.	Surname, First name Use “Unknown” if creator is unknown.

Field Name	Dublin Core Map	Hidden (Y/N)	Searchable (Y/N)	Definition	Controlled Vocabulary / Format Requirement
Source	Source (O)	N	Y	When applicable, use the Source field to cite the resource from which the digital resource was derived , either in whole or in part. Follow bibliographic citation standards or museum practices for the physical description of the object.	Whenever possible, write the full bibliographic citation. Consult: Mills, Elizabeth Shown, <i>Evidence Explained: Citing History Sources from Artifacts to Cyberspace</i> . Baltimore: Genealogical Publishing Co, 2007. If no standard identifier exists, use a local call number or accession number. Identify the institution associated with such locally derived numbers.
Description	Description (M)	N	Y	An account of the content of the resource. Textual explanation of the content of the resource. Description also includes an explanation of the physical condition of the original document. List all conservation concerns.	A brief summary of the content. Follow guidelines and terms listed in the conservation handbook, for entries about the element's condition.
Theme	None (L)	N	Y	Themes chosen as parameters for inclusion in the Community and Conflict project. Library of Congress headings.	Follow Theme headings listed in the workbook.

Field Name	Dublin Core Map	Hidden (Y/N)	Searchable (Y/N)	Definition	Controlled Vocabulary / Format Requirement
Subject	Subject (M)	N	Y	Keywords, places and names that a user might search.	Follow Subject headings listed in workbook.
Relation	Relation (M)	N	Y	Used to identify the collection title that the original resource is associated with. Relation does not reference resources of similar themes or subject headings.	Generally the collection title is entered. Example: Moses Bradford Collection
Site Accession Number	None (L)	N	N	A character string or record number that clearly and uniquely identifies the original resource , and is assigned by the contributing institution.	Entries will vary by institution. Example: 1998.003.05
Contributing Institution	Contributing Institution (O)	N	Y	The name of the institution that provided the original resource which the digital surrogates were derived.	Example: Wilson's Creek National Battlefield Use "Privately Held" if individual wishes to remain anonymous.
Copy Request	None (L)	N	N	Information concerning reproduction of images, documents and other digital resources.	List phone number, email, mailing address of contributing institution along with accession number of requested material.

Field Name	Dublin Core Map	Hidden (Y/N)	Searchable (Y/N)	Definition	Controlled Vocabulary / Format Requirement
Rights	Rights (O)	N	N	Rights information associated with the resource (copyright and property rights). Generally a URL link to a rights management or usage statement for each contributing institution.	“The text and images contained in this collection are intended for research and educational use only. Duplication of any of these images for commercial use without express written consent is expressly prohibited.”
Coverage	Coverage (O)	Y	Y	Describes the spatial area or time frame , which the intellectual content of the resource encompasses. Use the Coverage field when the subject heading is too general or when the location is not identified in the body of the text.	Example: Missouri – Greene County – Springfield
Volume	None (L)	N	N	Listing of element identifiers for each page of an individual document.	Example: CC_mberry_03p01 CC_mberry_03p02 CC_mberry_03p03
Date.Original	Date.Original (M)	N	Y	Creation date of the original resource.	YYYY-MM-DD Use “n.d.” if date is unknown; or “ca.” for approximation date.
Date.Digital	Date.Digital (M)	N	N	Creation date of the digital surrogate.	YYYY-MM-DD

Field Name	Dublin Core Map	Hidden (Y/N)	Searchable (Y/N)	Definition	Controlled Vocabulary / Format Requirement
Format.Original	None (L)	N	Y	Description of original resource.	Example: Manuscript, newspaper, letter, etc.
Format.Digital	Format (M)	Y	N	Refers only to digital resource.	Example: tiff

Digital Project Nomination Form

1. Materials Being Nominated for Digitization (Please indicate collection number, series, number, box number, folder number, item control number or equivalent and the creator; caption of the item or a bibliographic citation to the fullest extent possible.)

2. Reason for Nomination (Describe why the materials are important, who might want to use them in a digital form, and what usages are likely if they are digitized.)

3. Potential Assistance Sources (Please indicate if you have any special knowledge or skills that might be shared with the X repository during the selection process. For example, can you provide caption information, historical background, or are you aware of potential funding sources or digital projects that are covering similar materials to those you are nominating?)

4. Restrictions (Indicate if you are aware of any reason why the specified materials should not be digitized, such as legal, ethical, or cultural sensitivities. Please be as specific as possible citing a source, such as a law or culture group and a contact name if necessary.)

5. Your Name:

6. Your Address:

7. Tel: _____

Fax : _____

8. E-Mail: _____

Note: The Selection Committee will make all final decisions on what will or will not be included in the digital project. If you have any special information you would like to share with the committee, please write it below.

Community and Conflict Collection Prioritization Form

Collection:		
Evaluation	Score	Notes
Value (overall)		
Informational		
Artifactual		
Associational		
Risk		
Demand		
Existing Coverage		
Rating		
Priority		