

CWRGM PROJECT METHODOLOGY

The partnership in which CWRGM is grounded — MDAH (digitization), MDL (metadata), and USM History Program (transcription & tagging) — allows for focused and steady production represented in this workflow (below).

Digitization: The state's governors'

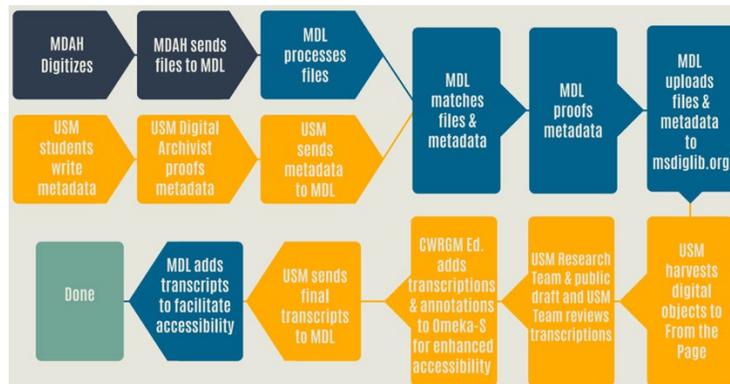
papers are already housed at

MDAH, eliminating the need to

scour regional and national

repositories for these records. As a

project partner, MDAH is entirely



responsible for the digitization process, including two-stage quality verification (review), of over 20,000 documents. MDAH transfers tiff files electronically to the CWRGM partners at MDL in regular batches of roughly 1,000 documents (usually one archival box). MDAH's digitization schedule is available in the Supplementary Materials and their digitization protocols are available [here](#). MDAH also updates or creates calendars for CWRGM collections that our team at MDL uses to draft metadata and complete first-stage verification, which involves a close review of the work and comparison against the MDAH calendar by someone other than the researcher who drafted it. One of the positions in our budget is focused on continuing a part-time qualified archival researcher at MDAH to help primarily with calendar creation. This position ensures that the project proceeds on pace despite collections that has never been calendared, such as that of Gov. John Marshall Stone — one of our largest series with an estimated 6,000 documents.

Metadata and Early Access: MDL is an online archival library based at USM that provides digital access to millions of records from repositories across the state. MDL serves

CWRGM as a digital repository for the electronic surrogates of the digitized document shared by MDAH. When MDAH sends digital access files to MDL via a secured File Transfer server, MDL assistant director Elizabeth La Beaud and MDL staff convert them to JPEG format and match, name, and folder the files with their metadata.¹ Much of La Beaud's work has been in-kind in the past, but CWRGM has started to pay her to dedicate additional, outside hours to the project because of her critical role in our productivity (see more below).

CWRGM's Visiting Digital Archivist Austin Justice oversees the creation and two-stage verification of metadata following [CWRGM metadata protocols](#) based on the Dublin Core standard (ISO 15836-1:2017).² CWRGM student researchers are trained by Justice for this work using these protocols, a customized spreadsheet template, and training videos Justice developed for the project, as well as Zoom training sessions with repeated feedback on work during early training stages. These researchers draft metadata MDAH calendars, updating item descriptions with language that is appropriate in the modern age. In weekly batches, the first reviewer, who cannot also be the person who drafted the metadata (to ensure a fresh set of eyes on the work), analyzes the completed metadata using Google-refine's text facet and Microsoft Excel's pivot tables to identify any inputs that do not conform to MDL's Metadata Best Practices and CWRGM Metadata Protocols. Any concerns are recorded in an issues spreadsheet and returned to the metadata creator for correction. Once all the records in a batch are approved, the metadata goes on to a second reviewer (different from the first and the creator) who double-checks the metadata for conformity and accuracy, again using Google-refine's text facet and Microsoft

¹ La Beaud is a permanent staff member at USM Libraries/USM.

² Justice's position was originally funded by the Watson-Brown Foundation, and is currently funded through June 2022 through CWRGM's NEH grant.

Excel's pivot tables. This reviewer also compares the reviewed metadata against the actual digitized document to increase accuracy.

CWRGM benefits from a graduate research assistant pulled from the USM History program fully dedicated to the project. This individual helps with metadata review and assists with uploading digital document files with metadata to MDL, which then allows for uploading to From the Page where all transcription and enhanced subject tagging takes place. When time allows, the GRA also assists with transcription and tagging. This, too, has proven to be a critically important position that helps us avoid log-jams in the workflow.

When digitized documents have verified metadata, MDL provides the public with the earliest possible access to these items, earlier than if the project had to wait for transcriptions and enhanced subject tagging features to be drafted and reviewed. MDL and/or the CWRGM GRA uploads new batches of documents at a minimum of every other week, while the discovery index updates automatically on a nightly basis. But MDL cannot host enrichments such as enhanced subject tagging (discussed below) that dramatically improve document discoverability. That is why CWRGM digital developer Anneliese Dehner designed an Omeka-S site for the project that launched with our first 2,000 documents in June 2021.³ When MDL has 250 more documents at its site than we do at CWRGM.org, we include a link at the Omeka-S site called "Sneak Peek" and publicize this via social media to alert users that they can view digital files, without transcriptions or subject tagging, at MDL. Usually, though, the team is able to keep up with transcriptions, and we turn off "Sneak Peek" to avoid confusing people about the best location to view files.

³ This work by Dehner in 2020-2021 was NEH funded. She proved so critical to the project for continued discoverability features for the site that we continued her funding since July 2021 through NHPRC.

Transcription and Tagging/Enhanced subject tagging features: Once MDL provides a CSV file for uploading digital document files and verified metadata, CWRGM Senior Assistant Editor Lindsey Peterson uploads these to [CWRGM's home at From the Page](#) (FtP).

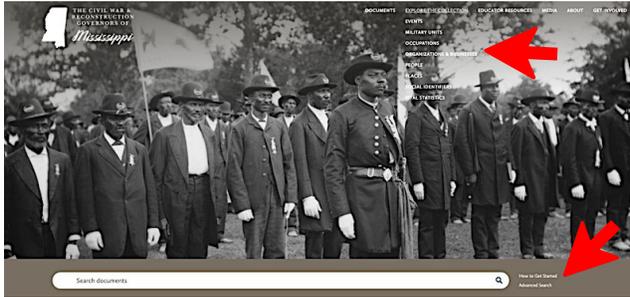
Transcriptions are drafted by volunteers and our team, and all of these go through a review process *by trained team members only* that follow CWRGM [Transcription and Tagging Protocols](#), with second-stage review conducted by a senior researcher, Peterson, and Ural. As with metadata, all levels of verification are done by different members of the research team, and all student researchers go through detailed training with Peterson that includes reviewing protocols and videos and Zoom lessons with regular feedback in their first few weeks followed by weekly meetings with Peterson and Ural throughout the semester. This training is grounded in [Documentary Editing literature](#) and in close consultation our protocols.

When CWRGM staff create or review transcriptions, they add **enhanced subject tagging features** in From the Page that significantly enhance the discoverability of our documents. These tags are tied to the following eight categories: people, places, organizations and businesses, events, occupations, military units, vital statistics, and social identifiers. Most of those categories are self-explanatory; they help users with diverse research interests understand how the collection might be of use to them and how to locate documents of interest. CWRGM director Ural created the last category, “social identifiers,” in close cooperation with Senior Editor Peterson to spotlight the experiences of individuals whose voices are often unheard in history, including the enslaved, free and freed people of color, widows, veterans (especially enlisted men), and impoverished individuals. These tags also help users think about new ways to access the collection, suggesting terms like arrest, orphans, impoverished people, and mental illness. Visitors may not think about studying criminal justice or physical disabilities at CWRGM, but

our tagging features remind users of exactly how many documents exist that address these topics and offer easy access to them. Available geocoordinates are added in the tagging process for any place, organizations and businesses, and event so these will be ready for future CWRGM mapping projects.

Selection and Arrangement: All of the 20,000 documents of the CWRGM governors' Correspondence and Papers, Military Telegrams, Letters and Petitions, and Executive Journals from 1859-1882 are included in this comprehensive digital documentary edition. In all cases, original documents are digitized to provide high-quality images for scholars and researchers. We did not know if we could include this for the tightly-bound Executive Journals, but thanks to a recent digital scanner purchase at MDAH, those, too, will feature high-resolution images that meet archival standards. The only items eliminated from this otherwise comprehensive project are duplicate documents such as copies of letters and receipts that comprise an estimated one percent of the entire collection. These duplicates are tracked internally in the calendar indexes maintained at MDAH. Because these collections are all government records that are in the public domain, they are free of copyright restrictions/required permissions, and MDAH's traditional digitization fees are part of their in-kind contributions to this grant.

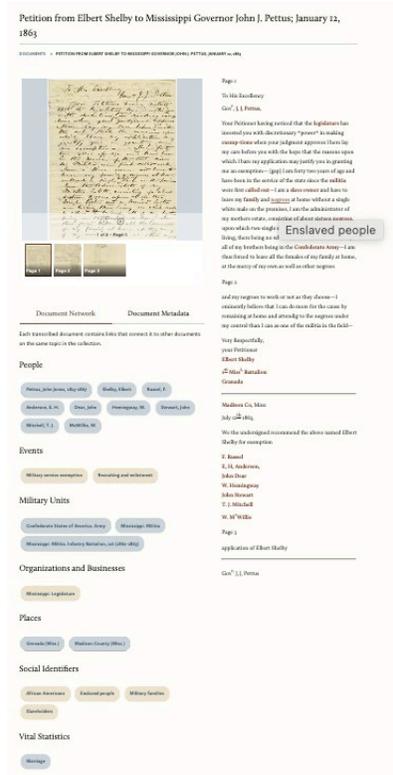
4. **Publishing Methods:** The [Omeka S home for CWRGM](http://cwrghm.org) (cwrghm.org) launched with the first 2,000 digitized documents with verified metadata, transcriptions, and enhanced subject tagging features in June 2021, and it offers users multiple approaches to accessing the collection. Digital developer Anneliese Dehner's design of the CWRGM Omeka S site (NEH funded, 2020-2021;



NHPRC funded 2020-2022) allows visitors to conduct [advanced keywords and date or date range searches](#). Users can also [browse the collection by box and folder arrangement](#), simulating a scholar's research

in the archives, including the availability of finding aids. Users can explore the eight subject categories that include occupations, organizations and businesses, vital statistics, and commonly used social identifiers, that help non-experts find items of interest (and those they may not have realized they could locate). There is also a short essay titled "[How to Get Started](#)" that is linked next to the search bar that appears on every page.

When viewing a document transcription, visitors can click on any tagged word in the transcription to connect to every other document currently at CWRGM.org that contains that term. The features also provide students with suggested word usage when encountering racial slurs in transcriptions. See, for example, where the cursor hovers over the word "negroes" and the term "Enslaved people" appears. Our suggested terminology varies based on the



historical context of the document. That term could refer to "Free people of color," "Freed peoples," or other individuals based on the date and how the term is applied by the author, and it is tagged accordingly. Keyword and most simple indexing does not provide this level of discoverability.

CWRGM's enhanced subject tagging methodology — done during the transcription process — means that users are not waiting for us to get to the time-consuming annotation phase of the process (which lies several years down the road) to benefit from these critically important discoverability features. CWRGM maximizes user engagement with a collection that continues to grow in size and in its contributions to scholars and the public.