

Accessible History: Putting a Century of *The Chronicles of Oklahoma* Online

Whitney Johnson-Freeman
Presenter

Megan Scott
Presenter

Hannah Carroll
Presenter

Abstract

Since 1921, the Oklahoma Historical Society has published *The Chronicles of Oklahoma*, a nonfiction journal focused on topics related to the state's history. In 2020, the Oklahoma Historical Society partnered with the University of North Texas (UNT) Libraries to make the extensive back catalog of quarterly-published issues of the *Chronicles* publicly available online through the Gateway to Oklahoma History (The Gateway) digital library. The scope of the project eventually expanded to include preserving individual articles in the Gateway to promote their discovery and impact for researchers. This presentation describes the project management practices, workflows, and metadata standards that were used to support the project. It also explores the challenges and highlights the successes of this multi-year preservation project.

Keywords: digitization, metadata, Oklahoma history, workflows, project management

Introduction

At the University of North Texas (UNT) Libraries, the Digital Library division supports a variety of digitization projects and collections. In the fall of 2020, we began working with the Oklahoma Historical Society (OHS) to include their regional publication, *The Chronicles of Oklahoma*, available digitally through The Gateway to Oklahoma History (The Gateway). The goal was to make this publication easier to access and increase its value to researchers. Once the majority of the issues were online, the OHS decided to expand the scope of the project to include creating records for the individual articles from each issue, so researchers could discover and cite articles more easily. While Phase One (The Issues) and Phase Two (The Articles) of the project had similar workflows, Phase One reflected a file processing workflow and Phase Two was more focused on metadata. To better discuss this project, we will first provide some context.

The UNT Libraries Digital Library division created and currently oversees a total of three different digital archives, including The Portal to Texas History, The UNT Digital Library, and The Gateway to Oklahoma History. The Gateway was created in 2012 and hosts a variety of material related to Oklahoma history.¹ It was developed with support from the OHS, and it is populated and maintained through an ongoing partnership between OHS and UNT Libraries.

The goal of the OHS “is to collect, preserve, and share the history and culture of the state of Oklahoma and its people,” and they have published *The Chronicles of Oklahoma* for over a century.² *The Chronicles* documents notable events and history of the Oklahoma area from researcher and personal perspectives. *The Chronicles* also includes book reviews, state policy news, meeting minutes, and membership information for the OHS.

Phase One Overview

The initial goal of the project was to get the 1921 to 2019 issues of *The Chronicles* online. *The Chronicles* has been a quarterly publication

since its initial release, so this phase of the project has nearly 400 total items. Most of the issues had already been digitized as a result of previous archival initiatives, so the project focused on processing these digital files and creating metadata.

Phase One Project Planning and Management

The team overseeing *The Chronicles* project was already experienced with digital archiving tools and principles. However, working with historical content and scanned content from outside sources provided some unique challenges. Initial steps included identifying resources already available and resources that would need to be created to support the project. The first was reaching out to knowledgeable individuals from other areas of the UNT Digital Library division who had experience with this type of material. This knowledge provided valuable guidance for navigating the challenging aspects of the project. We also already had access to Scan Tailor, software that processes scanned images of text to ensure the content is clear and easy to read. Finally, personnel from the UNT Libraries' Digital Projects Lab were available to assist the team with scanning missing content, including pages that had been missed, damaged, or distorted during their initial scan.

The next step was to document workflows and details of project management. Since the team was working remotely, we needed to start with a system to track the progress of the project and document missing content. We decided to use Google Sheets because it allowed the whole team to communicate directly with the OHS.

The next component was documentation, which ensured we were consistent in our workflows. This was especially important because some of these processes were new to most of the team. Our aim was to have documentation that was incredibly detailed with pictures or descriptions and notes on why a certain action was taken. We emphasized the importance of continually updating the documentation. This style of documentation was especially important for the graduate students working on the project because it helped them gain the skills and confidence to make decisions on their own.

Another resource we needed to create was what the UNT Digital Library division calls their folder workflow. Essentially this is an internal infrastructure to guide items through the processing workflow. This is something used across the Digital Libraries division, but each project has its own unique application of the folder workflow.³ The folder workflow for *The Chronicles* included quality control and a temporary archive which provided some checks and balances in the project.

Phase One Technical Workflows

The technical workflows were the most important component of Phase One. Because the digitization had been completed already, there were different file types and content organization to consider. These included portable document formats (PDFs) of born-digital issues and tag image file formats (TIFFs) of scanned individual issues and bound volumes.

The born-digital group was only a small portion of the overall issues, representing years around 2003–2019, but it was the best starting point since the team was most familiar with this type of file. The diagram in Figure 1 illustrates the workflow for these issues. In this workflow, each file goes through an initial review to identify any major problems. The next step is to create derivatives of the PDFs because the UNT Digital Library's digital archives allow users to view and interact with an item in the browser. Then the derivative files are renamed using Magick-Numbering, an internal standard utilized by the UNT Digital Library. MagickNumbering ensures the system displays the derivative files in the correct order and enables users to navigate to specific pages. After an additional round of quality control, files are run through optical character recognition (OCR) software. The UNT Digital Library's digital

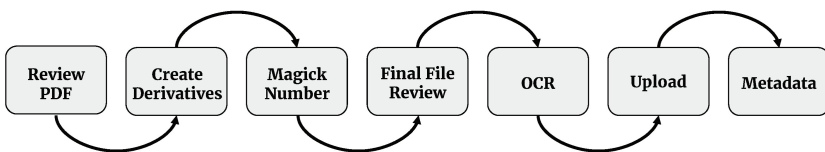


Figure 1. *Born-Digital Issues Workflow*

archives are full-text searchable, so this is a key step. Next, the whole package is uploaded to our backend archival system. Items are discoverable by authorized users but they are hidden from the public until metadata is complete, which is the final step. The metadata for the issue-level records is basic and will be explored in further detail shortly.

The bulk of the files were TIFFs of either individual issues or bound volumes. The workflow follows the same structure as the PDF workflow. However, the key differences are the inclusion of processing the files with Scan Tailor and addressing missing content. Figure 2 illustrates our workflow for the scanned content, including the additional steps for missing or flawed content.

Scan Tailor was used to clean the issues up to make them easier to read, remove marks and stamps, and create a consistent look for the content through the years. It was also important because the aging paper made some of the text difficult to read. However, our priority was to preserve the content, so images in the issues had little to no alterations done to them with Scan Tailor.

Missing or flawed content from the issues was also a significant problem. Initially, the team documented such content and requested the OHS to provide it. However, the pandemic made these requests more complicated for the OHS. Instead, we collaborated with the UNT Libraries Digital Projects Lab to scan missing content from the UNT Libraries collection of *The Chronicles*. For the issues not available in our collection, a request was sent to the OHS.

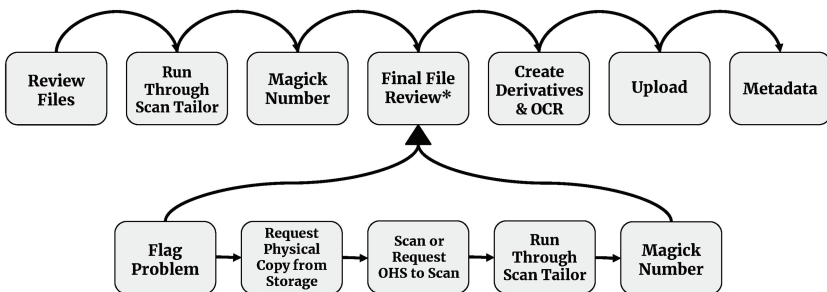


Figure 2. Scanned Issues Workflow

Phase One Metadata

The Gateway to Oklahoma History and the other digital libraries built and maintained by the UNT Digital Library division use a modified Dublin Core metadata schema. Because *The Chronicles* issues contained a variety of content, the descriptive metadata entered was simple and broad. Metadata for contributing authors, editors, and illustrators from each issue was prioritized to promote discovery of the authors' work in The Gateway. Name authority tools like the Name App, which is maintained by the UNT Digital Library division, helped streamline some of the data entry.

Phase One Discussion and Results

The team learned many valuable skills related to carrying out a long-term digital preservation project. The biggest success of Phase One was making all of the issues we had access to available online. The growing usage of the collection emphasized the value of our work to researchers. Before the start of Phase Two, the collection had reached over 3,000 visits a month.⁴ Another valuable benefit of this project was that we were able to do the bulk of it while working remotely with the help of our Digital Projects Lab. Missing content was a constant problem, and without their help, it would have taken much longer to get issues online. The documentation that we created has been invaluable not only for us during this project but has also aided in explaining our work and promoting the collection. Finally, the work we completed for Phase One made the transition to Phase Two that much easier.

Phase Two Overview

In the fall of 2022, the Oklahoma Historical Society funded a second phase of the project for *The Chronicles of Oklahoma*. The goal of Phase Two was to create additional separate metadata records at the

article level for the contents within each issue in *The Gateway to Oklahoma History*, to promote discovery and usability of the collection.

Phase Two Project Planning and Management

Creating article-level records necessitated splitting, moving, tracking, and describing nearly 4,000 files. To help visualize this division of content of *The Chronicles*, the contents of Volume 59, Number 2, Summer 1981 were split in the following fashion:

- “Woman with a Hatchet”: Carry Nation Comes to Oklahoma Territory
- The Right to be Served: Oklahoma City’s Lunch Counter Sit-Ins, 1958–1964
- Stand Watie and the Killing of James Foreman
- The Jim Thorpe Family: Part II
- Creating an Atmosphere of Suppression, 1914–1917
- No Time to Quibble: The Jones Conspiracy Trial of 1917
- Notes and Documents, *Chronicles of Oklahoma*, Volume 59, Number 2, Summer 1981
- For the Record, *Chronicles of Oklahoma*, Volume 59, Number 2, Summer 1981⁵

This list consists of the six unique articles that appeared in that issue, as well as the “Notes and Documents” section and “For the Record” section, which contains meeting minutes and membership information related to the OHS. Each of these items listed received its own record and accompanying metadata in *The Gateway*. Other sections that appeared over the years, including “Necrologies” or “Corrections,” are also given their own records in *The Gateway* if present. However, “Indices” and “Book Reviews” are excluded and did not receive individual records during this phase of the project.

The volume and complexity of this project emphasized the importance of documentation. The project team created three major

documents to organize the workflows of this project. This included a general master spreadsheet, a detailed metadata workflows document, and a controlled vocabulary list. The master spreadsheet fields consisted of article title, volume information, the section of *The Chronicles* that contained the article, and corresponding uniform resource identifiers (URIs) to easily access records and track progress. The detailed metadata workflows document explains the metadata for each field and how its values should be structured for the different types of article level sections. The metadata workflows document ensures that metadata for the collection is consistent, no matter who enters the information, which helped tremendously to streamline the process when graduate student assistants began contributing to the project. Finally, a controlled vocabulary spreadsheet serves as a reference tool for documenting subject metadata and ensuring consistency among subject terms that commonly appear throughout the collection. Both the metadata workflows and the controlled vocabulary list were shared with the OHS to include them in the process, solicit feedback, and ensure that the collection was aligned with the overall vision for this project.

Phase Two Metadata

The technical requirements for file processing in this phase were minimal. Most processing was completed in the first phase of the project and the files only needed to be split into their respective parts for uploading with the article-level records. Hence, creating the descriptive metadata for the article-level records became the most detailed and time-consuming task. The process of creating metadata for a project such as *The Chronicles of Oklahoma* differed greatly from the team's usual work depositing faculty research in the institutional repository and required willingness from team members to learn and adapt.

The key metadata fields for this phase were the content description and subject fields. Since this project made *The Chronicles of Oklahoma* publicly available online at the article level for the first

time, the team needed to create original subject metadata for each item. At the issue level, subjects were broad and remained static. However, at the article level of description, the subjects were more unique to the item and required more consideration. After much discussion, the team decided to represent subjects using a combination of keywords, Library of Congress Subject Headings (LCSH), and the University of North Texas Libraries Browse Subjects (UNTL-BS), a controlled vocabulary that already existed within the UNT Digital Library.

The largest consideration the team faced during this stage of the project was creating original, consistent subject metadata for challenging topics.⁶ We had to consider the most appropriate way to create accurate, respectful, and inclusive metadata that represented the original content. This was a major concern due to the large amount of content within *The Chronicles* relating to indigenous peoples and general race relations in Oklahoma throughout the state's history. The team used the UNTL-BS vocabulary in addition to keywords and LCSH for subjects related to Indigenous peoples and tribes in Oklahoma when the LCSH did not provide the same level of specificity as UNTL-BS or the most updated term preferred by the group.

We learned and adopted a few guiding principles throughout the metadata creation process:

- Include both broad and specific subjects when possible.
- Use a broader term and do not be afraid to ask for help in situations that might be more ambiguous or difficult to navigate.
- Be mindful of bias from the author and from yourself when creating descriptive metadata.
- Be open to having difficult conversations.
- Similarly, be willing to receive feedback and make corrections.
- Finally, this is not a solitary process. Collaboration and the conversations within the department and with partners at the OHS were extremely valuable for considering different perspectives when creating metadata for a historical publication that covers many different and sometimes sensitive topics over almost a century of Oklahoma history.

Phase Two Discussion and Conclusion

Building off the success of Phase One, we carried the lessons learned forward into the second phase of the project, starting with documentation. The documentation created for Phase Two such as the metadata workflow document and the controlled vocabulary list played an essential role in project management. It required the team to critically assess the metadata creation process and why the metadata was created a certain way. It also provided opportunities to discuss the work of digital projects and metadata issues with colleagues and present this experience to others. Finally, the team has been encouraged by positive user feedback we have received through emails and comments from researchers showing gratitude for making *The Chronicles* available online. Since the beginning of Phase Two, usage of the collection has increased significantly. With the continued efforts to provide more article level records, the collection saw a new peak of around 5,900 uses during March of 2023 which is almost double the usage from Phase One.⁷ Furthermore, the second phase is still ongoing. As of July 2023, there are currently 1,739 items from *The Chronicles of Oklahoma* available at the article level on the Gateway to Oklahoma History.⁸ The goal is to have content from all the issues from 1921 to the present also available at the article level, which is seeing steady progress and will likely be completed by the fall of 2023.

Contributor Notes

Whitney Johnson-Freeman is the Repository Librarian, University of North Texas, Denton, Texas.

Megan Scott is the Assistant Librarian for Digital Curation at Texas Tech University, Lubbock, Texas.

Hannah Carroll is the Graduate Student Assistant in Digital Curation, University of North Texas, Denton, Texas.

Notes

- 1 "History," The Gateway to Oklahoma History, accessed May 15, 2023, <https://gateway.okhistory.org/about/gateway/history/>.
- 2 "About the Oklahoma Historical Society," Oklahoma Historical Society, accessed May 15, 2023, <https://www.okhistory.org/about/index>.
- 3 Whitney R. Johnson-Freeman, Mark E. Phillips, and Kristy K. Phillips, "Managing an Institutional Repository Workflow with GitLab and a Folder-based Deposit System," *Code4Lib Journal*, no. 50 (February 2021), <https://journal.code4lib.org/articles/15650>.
- 4 "Statistics: The Chronicles of Oklahoma," The Gateway to Oklahoma History, accessed May 15, 2023, <https://gateway.okhistory.org/explore/collections/CRNOK/stats/>.
- 5 Oklahoma Historical Society, *The Chronicles of Oklahoma* 59, no.2 (Summer 1981), accessed May 15, 2023, <https://gateway.okhistory.org/ark:/67531/metadc1752019/>; <https://gateway.okhistory.org/ark:/67531/metadc2031396/>; <https://gateway.okhistory.org/ark:/67531/metadc2031397/>; <https://gateway.okhistory.org/ark:/67531/metadc2031398/>; <https://gateway.okhistory.org/ark:/67531/metadc2031399/>; <https://gateway.okhistory.org/ark:/67531/metadc2031400/>; <https://gateway.okhistory.org/ark:/67531/metadc2031401/>; <https://gateway.okhistory.org/ark:/67531/metadc2031403/>; <https://gateway.okhistory.org/ark:/67531/metadc2031402/>.
- 6 Hannah Carroll and Whitney Johnson-Freeman, "Creating Subject Metadata for The Chronicles of Oklahoma in The Gateway to Oklahoma History," University of North Texas Digital Library, accessed July 31, 2023, <https://digital.library.unt.edu/ark:/67531/metadc2114242/>.
- 7 "Statistics: The Chronicles of Oklahoma," The Gateway to Oklahoma History, accessed July 31, 2023, <https://gateway.okhistory.org/explore/collections/CRNOK/stats/>.
- 8 *Ibid.*