Applicant Information

Title: Using Metadata as Data for Reparative Archival Description: A Case Study
Name: Elizabeth James
MU ID: 901905714
Email: jamese@marshall.edu
College: Libraries and Online Learning, Archives and Special Collections

Project Description

Archives and special collections typically house historical materials that document places, people, and events from varying time periods. Unique letters, photographs, and unpublished manuscripts are a few of the many formats of which archives and special collections materials are composed. Conventions for applying descriptive terms to these materials, also known as metadata, so that they are available online, accessible, and searchable for use by community and academic researchers alike often rely on national standards and vocabularies created by professional organizations and government archives. However, these vocabularies are often created at the national level and can be slow to respond to changing terminology used to describe individuals and subjects relevant to the experience of marginalized community members. In Appalachia, this issue impacts regional and local archives especially, as they are more likely to document the histories of economically disadvantaged individuals, people of color, or ethnic groups. As a result, archives may be forced into using harmful or offensive terminology that their users must use to search their collection. To provide one example, the Library of Congress classification has variably used words such as “Negroes” and “Afro-Americans” to describe materials that depicted African American or Black individuals.

Archives of all sizes are currently working to mitigate the legacy of harm caused by outdated descriptive terms and metadata. However, the process of describing these materials often requires a significant amount of manual labor that has yet to be automated or systematically examined at scale beyond a few more technically intensive projects. Updating description without helpful automated tools would be nearly impossible, though many archivists are making headway using key search terms and targeted replacements. I seek to conduct a literature review of current best practices for reparative archival description, analyze the metadata that describes more than 11,000 items and collections at Marshall University Archives and Special Collections, and create a model that outlines how smaller institutions can use metadata as data to rectify these descriptive wrongs despite consistent staffing and resource allocation issues at the institutional level. Conducting reparative description will assist in democratizing the usage of the archive to be more inclusive of students and community members of all levels and areas of expertise and will make usage of archival materials less alienating to individuals from diverse backgrounds. The products created to synthesize this work will be an article and webpage sharing the results and data visualizations to make the research more accessible to a broader cultural heritage audience.

Methodology

This project will require a standard literature review of reparative archival description practices as well as a case study using the 11,000 items and collections

described in Marshall University's Archive and Special Collections. Research materials for the literature review can be acquired through Marshall databases and interlibrary library loan, and free or open source software such as Voyant Tools, Python data analysis libraries, and Tableau will be used to conduct the analysis and create data visualizations. Metadata will be downloaded from the archive's digital collections database in a standard CSV format ready for computational use.

The literature review will be used to identify problematic areas for description within the dataset and ways in which they might be remedied according to best practices. Analysis of the 11,000 descriptive metadata records will include word frequency and topic modeling approaches. Using this data, I will the create network graphs and other data visualizations to demonstrate frequency and topical spread of descriptive information and interrelations between concepts. These visualizations will be created through a combination of Voyant Tools, Tableau, and the pyvis Python library. Then, changes to descriptive metadata according to best practices will be applied to the metadata in the archive's content management system using batch import features and the metadata will be reassessed according to the metrics identified during the analysis. Once data visualizations characterizing the data have been created and data remediation has been completed, the larger process will be synthesized into a model and traditional academic article. The larger takeaways from the work will then be placed on a page of my personal website for dissemination to the larger cultural heritage community, much of which exists outside of academia.

A rough timeline is as follows:

Time Period	**Activities**
May 2021	Conduct literature review to identify descriptive problem areas and best practice solutions.
June 2021	Analyze existing Marshall University metadata to determine what descriptive issues exist in the data, model the problematic language present, and determine whether there are additional descriptive problems in the data that is not present in scholarship. Apply best practices solutions to issues. Create visualizations of data as found and data after editing.
July 2021	Begin writing article and creating webpage for more public dissemination and reuse of the case study.
August 2021	Complete article and begin submission process. Finish creating data visualizations and webpage.
August-September 2021	Write report summarizing activities conducted and products created.

Outcomes

The anticipated output of this project is an article and a webpage providing access to the original data, analysis, resulting visualizations, and model to make the research more accessible to a broader cultural heritage audience. The final article will be submitted to either the Journal of Contemporary Archival Studies or the American Archivist.

Budget

The primary cost for this project is the time necessary to conduct the research, analysis, and writing. All materials outlined in the methodology section are either free, available through Marshall, or open source.

Using Metadata as Data for Reparative Archival Description: A Case Study

What is reparative archive description?

- Distinction between iterative description and reparative description
- Is this baseline archival responsibility? Why a project approach vs. a service approach, and how can we sustain a service approach (maybe CLIR hidden collections digitization and impact of precarious labor?)
    - Scale for institution size (tiny archive w/ only a new digitization program focusing on periodic materials w/ limited metadata)

Quantifying/Qualifying metadata quality:

- View from user's perspective or the perspective of the quality of metadata?
- What's the point/goal of metadata
    - How do we measure if that goal has been met
        - Specifically, how do we measure that goal has been met in terms of equity, diversity, and inclusion issues?

What makes metadata "good"?

- Has search terms used by individuals

What makes metadata "bad"?

- Missing information (obvs)
- Harmful description
- Description that isn't accessible or meets user needs

How can we measure "good" or "bad" metadata?

- Adherence to controlled vocabularies?
- Length of description/title?

Text analysis for subjective probs

- Determine problematic words/words that need nuance
- Determine what words bracket problematic terms beyond a find/replace approach
-