

Metadata as Content: Navigating the Intersection of Repositories, Documentation, and Legacy Futures

Presentation

Erik Radio
University of Arizona, USA
radio@arizona.edu

Keywords: metadata; repositories; documentation

Abstract

Documentary Relations of the Southwest (DRSW) is a dataset of bibliographic metadata derived from over 1500 reels of microfilmed documents that trace the history of the southwest from the 16th century until Mexico's independence in 1821. Originally made available to scholars through a now defunct proprietary repository, DRSW's future is currently being assessed in the context of other repository solutions. While migrating content is a familiar scenario, this migration highlights key challenges in navigating the intersection of legacy design and possible futures for metadata curation and repository selection. This presentation deals with challenges revolving around three paradigms: metadata as content, system documentation generation, and metadata futures for indexing and integration.

In the repository, contextual metadata is commonly considered distinct from the content it describes. DRSW is an uncommon case, as none of the documents have been digitized; the metadata is the content. This presents unique issues since the original metadata creation for DRSW was not created under the guidance of a metadata professional and contains errors (e.g. typos, term inconsistency). As a result structure for measuring semantic loss in metadata was devised as way of preventing similar scenarios in the future, and will be discussed in this presentation.

The second paradigm revolves around the generation of documentation by systems. The selection of a system has significant effects on how metadata is processed, edited, and exported. However, while metadata can travel between systems, rarely does documentation travel along with it. In addition to contextual documentation, it appears to be increasingly critical that there be system generated metadata. While some may object this exists in change logs and similar tracking files, this are often not in an easily generatable way for digital collection managers. Further, they do not commonly include decisions as to why certain changes were made, or similar such decisions, which are critical for understanding the provenance of metadata. As a tool for mitigating semantic loss, further examination into system effects on metadata, or 'Processual Documentation', and what such a mechanism would entail will be discussed, particularly as it impacts DRSW.

The final paradigm to be explored regards what this particular scenario suggests for the future of metadata migrations and its use in repositories. As mentioned DRSW was created using a local schema. Because of the lack of documentation, local schemas such as this are not a sustainable option for migrations as multiple standards in a repository can lead to indexing troubles as well as possibly being confusing for users unless there is significant work internally towards ontology alignment. It was initially suggested that migrating the metadata to Metadata Object Description Scheme (MODS) would allow for the dataset to be more extensible into the future, but curators decided that it must adhere to the original, local standard. This provokes serious questions as to the ability of metadata to be integrated into a linked data environment. If it is critical for metadata to be preserved in its nascent form, then there must be further capacity for metadata to grown synoptically. Whether this is something that can be afforded by repository software (i.e. multiple views of an object), or through linked data projects is a topic that remains to be further discussed. This presentation will outline the roadmap for DRSW as it fits into this trajectory as a way of

facilitating a discussion following the presentation on how similar collections are being approached.