DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2017

Collaborative Metadata Application Profile Development for DAMS Migration

Poster

Anne M. Washington
University of Houston,
USA
awashington@uh.edu

Andrew Weidner
University of Houston,
USA
ajweidner@uh.edu

Keywords: metadata; migration; application profiles; metadata schemas; digital asset management systems; CONTENTdm; Hyku

Abstract

In 2015, after an extensive review process, the University of Houston (UH) Libraries chose the open source systems Hyku (then known as the Hydra-in-a-Box project), Archivematica, and ArchivesSpace to form the Libraries' digital collections access and preservation ecosystem (Wu et al., 2016). This suite of systems, along with locally developed tools, form the Bayou City Digital Asset Management System (BCDAMS). In 2016, the BCDAMS Implementation Team began work on a multi-phase process to roll out the new systems to replace the current digital collections management system, CONTENTdm. Phase I of this process included developing fundamental models and principles as well as much of the local infrastructure and workflows (Weidner et al., 2017). Phase II of the project will involve migrating existing digital collection metadata and files to the new digital asset management system (DAMS).

This poster summarizes work done during Phase I of the project to prepare for the migration work in Phase II. This included working collaboratively to develop a Metadata Application Profile (MAP) and crosswalk for the Hyku digital collections access system, and an analysis of metadata remediation required to prepare for migration. It addresses work underway at many institutions exploring or actively migrating to a new DAMS. This poster shares the UH Libraries unique experience in preparing for the migration of UH Digital Library (UHDL) data from CONTENTdm to a new system and offers some general considerations for migrations.

To develop the MAP and crosswalk from CONTENTdm, the Descriptive Metadata Working Group (DMWG) was formed. This interdepartmental team represented the Metadata and Digitization Services Department and Special Collections. Led by the Metadata Coordinator, who is also the BCDAMS Project Manager, it included the Metadata Librarian, Metadata Unit staff members, the Coordinator of Digital Projects, the Hispanic Archivist, and the Special Collections Project Manager. It was important to have these different perspectives represented to address the metadata needs of different collections, further a shared understanding of the scope and function of the system, and gain support from stakeholders such as Special Collections curators.

Foundational migration considerations that informed the group's work were the types of content in the digital library, the purpose of the DAMS, and the technical specifications of the DAMS. As part of the DAMS evaluation process and preparation for system migration, an inventory of content types in the UHDL was completed. The Metadata Unit staff reviewed all of the digital collections individually recording the types of items it contained, e.g. single-sided images, double-sided images, documents, single-part audio, etc. In its discussions, the DMWG considered the descriptive needs of these content types, as well as content types that may be included in the digital library in the future, such as born digital content. This analysis also made clear collections' varying levels of complexity. Eight collections of low or moderate complexity were selected as test collections, used to test software and workflows.

https://doi.org/10.23106/dcmi.952137952

Early in the work of both the BCDAMS Implementation Team and the DMWG, it was important to ensure a shared understanding of the purpose and scope of each component of the digital asset management and preservation ecosystem. Hyku was defined as the access system with the primary purpose of digital object discovery and access. Maintaining archival context of digital objects is especially important to Special Collections staff and users, but extensive metadata related to that context is out of scope for the access system. Instead, ArchivesSpace was determined to be the system of record for that information and a digital object's Archival Resource Key (ARK) and ArchivesSpace identifier would maintain the connection between digital objects in the access system and the physical items managed in ArchivesSpace. Once these purposes were clear, it was easier to have conversations around metadata fields and input guidelines that were appropriate for the different systems.

The technical specifications of the access system influenced the earliest MAP decisions and the work of the DMWG. At the time the DMWG was developing the Metadata Application Profile, Hyku was in early stages of development, and the built-in metadata structure was unclear. It was known that Hyku would provide a simple way for institutions to get their content to the Digital Public Library of America (DPLA), so it was assumed there would be a way to align the Hyku metadata with the DPLA Metadata Application Profile. Because of this interoperability, as well as the appropriateness of the schema for UHDL content, the team used the DPLA MAP v4 (DPLA, 2015) as the basis for the Metadata Application Profile. Given the extensible nature of Samvera Community (formerly Hydra Community) software, the team also considered additional elements not included in the DPLA MAP as long as they were elements from a linked-data ready schema such as BIBFRAME.

The team used a variety of tools in its work. The team communicated on Slack, a messaging application, between meetings. GitHub was used not only as a platform to publish the MAP code, but also to record discussions about MAP fields and input guidelines, as well as document the MAP and crosswalks. This content is openly available on the UH Libraries BCDAMS MAP GitHub repository: https://github.com/uhlibraries-digital/bcdams-map. The group also analyzed reports of existing UHDL metadata to inform decisions on the use of fields in the new MAP. These reports were created using Hunting, a locally developed Ruby gem used to access UHDL metadata through the CONTENTdm API (UH Libraries, 2016a). While there were established input guidelines in the UHDL metadata dictionary (UH Libraries, 2016b), the reports were useful in determining how the field was used in practice across different collections. The final deliverables for this group were a machine actionable MAP - the metadata element set and the input guidelines (UH Libraries, 2017a) - and a crosswalk from the existing MAP (UH Libraries, 2017b) to the new MAP.

After this work was completed, the Metadata Unit sought to understand the scope and scale of metadata remediation efforts required for migration. The Metadata Coordinator created a report for each digital collection that mapped the existing data into the new fields. The Metadata Unit staff then reviewed these reports noting where the field values did not meet the new MAP input guidelines. The most common issues recorded were: titles, subjects, dates, and fields capturing format and physical characteristics did not conform to new input guidelines. For example, previous input guidelines required unique titles, resulting in many titles ending in "Image 1" or "Image 2" to disambiguate otherwise identical titles. This is no longer a requirement in the new MAP as each resource is assigned a persistent ARK. Another new input guideline discourages the use of pre-coordinated subject headings, which were used in nearly every legacy collection. In Phase II, the Metadata Unit will begin upgrading the metadata by bringing existing data into alignment with the new MAP as well as doing authority control using the UH Libraries' local thesaurus application, Cedar (UH Libraries, 2016c).

Work has just begun on Phase II of the project which includes data migration to the new system. With its charge complete, the Descriptive Metadata Working Group has changed direction and expanded into the Data Migration Working Group. It now includes representation from the Digitization Unit to advise on file management requirements for the migration. Digital

https://doi.org/10.23106/dcmi.952137952

collection curators and other stakeholders will join the group as needed to advise on specific collection concerns or other areas of development. The initial goal of this group is to determine the workflow for migration starting with the test collections. Then, after migration begins in earnest, the team will work collection by collection to address its specific file management and metadata needs. There are challenges on the horizon including technical constraints and idiosyncratic digital collection data, but the foundational work of Phase I and the commitment to a collaborative approach to migration set the stage for success.

References

DPLA. (2015). Metadata Application Profile, version 4.0. Retrieved June 12, 2017 from http://dp.la/info/wp-content/uploads/2015/03/MAPv4.pdf.

Weidner, Andrew, Sean Watkins, Bethany Scott, Drew Krewer, Anne Washington, and Matthew Richardson. (2017). Outside the box: Building a digital asset management ecosystem for preservation and access. Code4Lib Journal, 36. Retrieved June 12, 2017, from http://journal.code4lib.org/articles/12342.

Wu, Annie, Santi Thompson, Rachel Vacek, Sean Watkins, and Andrew Weidner. (2016). Hitting the road towards a greater digital destination: Evaluating and testing DAMS at the University of Houston Libraries. Information Technology and Libraries, 35(2). Retrieved June 12, 2017, from https://doi.org/10.6017/ital.v35i2.9152.

UH Libraries. (2016a). hunting GitHub repository. Retrieved June 12, 2017 from https://github.com/uhlibraries-digital/hunting.

UH Libraries. (2016b). Metadata Dictionary. Retrieved June 12, 2017 from http://digital.lib.uh.edu/about/metadata.

UH Libraries. (2016c). University of Houston Libraries vocabularies. Retrieved June 12, 2017 from https://vocab.lib.uh.edu/en.html.

UH Libraries. (2017a). Bayou City DAMS metadata application profile. Retrieved June 12, 2017 from https://vocab.lib.uh.edu/bcdams-map.

UH Libraries. (2017b). UHDL crosswalk. Retrieved June 12, 2017 from https://github.com/uhlibraries-digital/bcdams-map/wiki/05-UHDL-Crosswalk.

https://doi.org/10.23106/dcmi.952137952