

OSS ArcFlow

Digital Dossier

By Shaun Trujillo

Mount Holyoke College

OVERVIEW

Can you indicate a percentage of time that each library, archives and curatorial staff member spends working on digital curation?

- Head of Archives and Special Collections 1-5%
- Special Collections Archivist ~5-10%
- Metadata Librarian >40%
- Digital Projects Lead >20%
- Digital Library Applications Manager >40%
- Associate Director of Discovery and Access 1-5%

Describe information technology support available for digital curation activities.

MHC's Library, Information, & Technology Services (LITS) is a merged organization (Library + IT), which has proven to be a very productive environment for quickly prototyping and testing of technology services. We have direct access to a systems and networking associate, and can use roughly 5 hours of their time weekly. The output from our digitization program has grown at a steady rate through the years and IT support has consistently met storage and backup needs.

Characterize your organization's digital infrastructure. Do you have one discovery system or several?

Several - Islandora, the OPAC (Aleph, Ebsco Discovery), 5 college consortial finding aid site

How does staff share documentation about collections and activities?
Electronic finding aids, Islandora digital repository, library website

What is the size/scope of your digital collections?

Roughly 20,000 locally digitized objects. ~500 outsourced digitized audio/video assets. ~200 electronic records accessions. ~10 media accessions

DIGITAL CURATION ACTIVITIES

Define digital curation at your organization. What are the major phases of your digital curation lifecycle?

Primarily creation/digitization/capture, accession (for born digital records and media), description, storage, derivative creation (automated), publication

Provide an overview of the tools/environments used throughout your digital curation workflow.

- ArchivesSpace (Archives' collections management system - previously Archivists' Toolkit - full migration to AS in Winter 2016-17)
- ResourceSpace (Digital asset management system - home of master files from digitization workflows)
- Islandora (Digital collections frontend - previously CONTENTdm)
- Omeka (Archives led digital exhibitions and crowdsource transcription site)
- DSpace (Institutional archive for student and faculty generated data, as well as college wide programs and initiatives)
- Archive-It (Web archiving service for snapshotting the college websites. Strategic capture of sites based on Archives' collecting policy)
- Network Attached Storage (Acting repository of record for most of the original work produced by the Library and Archives. Currently houses ~20TB of data)

Who [which roles] contributes to digital curation in your organization?
Head of Archives and Special Collections, Special Collections Archivist, Metadata Librarian, Digital Projects Lead, Digital Library Applications Manager, Associate Director of Discovery and Access

What do you consider to be your categories of digital content?

ON-DEMAND DIGITIZATION:

- This is the most fully realized workflow for the creation, publishing, and longterm storage of digital objects. Workflow is centered around the ResourceSpace system's request and delivery features. Most of the products of this workflow are described and managed in ArchivesSpace.

PROGRAMMATIC DIGITIZATION:

- Large-scale scanning/digitization of archival collections.
- Challenges around level of description - Best practices of aggregate description vs. item level description. Input/output of description for ArchivesSpace objects has been a pain point, but we've gleaned some ideas from Duke's python scripts on Github.

BORN DIGITAL E-RECORDS:

- Current workflow was born from a 2009 National Historical Publications and Records Commission (NHPRC) grant and relies on a locally developed Digital Records Transfer system (DRXFER) and the Data Accessioner to produce SIPs and basic DIPs.

BORN DIGITAL A/V EVENT RECORDINGS:

- A steady growth of new A/V content each semester provides storage and throughput challenges, as well as questions regarding access and permissions.

BORN DIGITAL STORAGE MEDIA ACCESSIONS:

- Rare donations of storage media from staff and community. These assets provide similar issues as our born digital A/V materials. We rely on FTK Imager to produce E01 and raw disk images and Bitcurator for file system manifests and reporting.

EMAIL ACCESSIONING:

- We collect the MBox files of outgoing officers. We're interested in incorporating the epadd tool for help in processing email archives.

How variable are your workflows across different categories of digital content?

The workflows across and even within categories are fairly variable and in some instances ad hoc. Process can change depending on the context of the project. Overall, it is fair to say that each of the categories listed above has an independent/unique workflow. The two categories that are closest in terms of workflow are 'on-demand digitization' and 'programmatic digitization'.

GOALS FOR DIGITAL CURATION

- Providing better access points for digital content while maintaining archival context, i.e. better linkages between finding aids and collection guides and digital library frontend systems.
- Streamline A/V, media, and other born digital accessioning. Automation, automation, automation! Low cost, low staff processing of digital materials.
- Standardization and clarification of digital curation practice, e.g. standard and consistent definition of an AIP for long-term retention, formalize handoffs and syncing of data across systems, providing audit trails and improving availability in collections system of record.