

OSSArcFlow

Digital Dossier

By Kari Smith and Joe Carrano

Massachusetts Institute of Technology Archives and Special Collections

OVERVIEW

The MIT Institute Archives and Special Collections (IASC) serves as the "memory" of MIT, collecting and preserving records that document MIT's history and the people who have been a part of that history. The collections include both published and unpublished materials in various formats. This information does not reflect digital curation practices outside of IASC.

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

- Digital Archivist: 95%
- Institute Archivist and Program Head for Digital Archives: 50%
- Processing Archivist: 5%
- Director for Digital Preservation: 100%
- Digital Operations Coordinator: 50%

Describe information technology support available for digital curation activities:

- Most of technology support is done by the digital archivist and program head.
- As of 2017, the archives is now receiving some technology support for developing new tools as well as ongoing maintenance and tweaking of core systems (1 engineer (developer)).
- Enterprise systems support is still being determined for digital curation systems.

- We host ArchivesSpace through LYRASIS and we have an SLA with Artefactual staff who provide our IT support for our locally installed Archivematica instances.

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

- We are well along with our comprehensive digital preservation storage program. Curated content is currently stored on institutionally managed network drives.
- We currently do not have an active discovery system for most archival material, but the intention is to use the ArchivesSpace public user interface.
- Some of our finding aids are available online as html/pdf's on the archives subdomain of the libraries' website. A subset of these finding aids and archival publications have catalog records searchable through BARTON, our OPAC.
- Some of our digitized archival material is discoverable through DOME, the Libraries' DSpace instance.
- Most discovery is mediated through individual reference requests via email to the archives.

How does staff share documentation about collections and activities?

- Collections documentation is mainly shared through archival control files (physical and digital) and activities and processes are kept on local shared network drive folders. Some documentation also lives in places like MIT Dropbox, Google Drive, and the MIT Confluence wiki space.

What is the size/scope of your digital collections?

- 10.1 TB currently manage in local network storage, 2 PB on offline drives across 70 collections.

DIGITAL CURATION ACTIVITIES

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

- Decide to acquire digital material
- Determine Digital Preservation Requirements
- Receive Digital Content
- Ingest/Process Digital Content
- Manage Preservation Object Storage
- Make Digital Content Available

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

- For known content we use Exactly to bag and transfer materials from donors or from in- house/vendor digitization.
- Web archiving is done with Archive-It and Webrecorder.
- Email is appraised and transferred using ePADD.
- Finding aids and accession records are in ArchivesSpace.
- BitCurator is used for unknown materials or those that require examination before accessioning. This entails additional tasks like virus checking, possible disk imaging, and identification of PII; Kryoflux if need to image floppy disks.
- During the Ingest stage we use Archivematica to identify, normalize, and package digital objects for preservation and dissemination.
- Fixity tool is used to monitor file attendance and integrity of all archival files on network drives.

Who [which roles] contributes to digital curation in your organization?

- Digital Archivist
- Institute Archivist and Program Head for Digital Archives
- Director of Digital Preservation

- Digital Content and Reformatting Team
- Scanning and reformatting QC associate
- Metadata entry assistant
- System Administrators
- Digital Library Systems Engineer

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

- “Categories” can correspond to format types, but in some cases categories correspond more easily to the way in which the material was accessioned.
- “Categories” here refers to content that shares a common workflow.
- Known born-digital content
- Unknown born-digital content
- Unstructured mix of known and unknown born-digital content
- Digitized content
- Content by category:
 - Video
 - Image–Digitized
 - Image–Born Digital
 - Web
 - Audio
 - Text
 - Geospatial
 - Database
 - Structured Database
 - Text–Digitized
 - Web with Video
 - Disk Image

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

- Workflows are fairly consistent for known born-digital content and digitized content. Unknown digitized content requires more work to analyze and identify information about the files for appraisal, processing, and preservation planning, such as running additional tools and reports.

GOALS FOR DIGITAL CURATION

- Successfully implement web-based transfer tool.
- Expand on the Content Types characteristics to include level of effort and capabilities for working with it and knowing the tools that will assist with arrangement and description in the next year or two.
- Implement the integration of ArchivesSpace and Archivematica.
- Implement appropriate and sufficient digital preservation storage and services.