

OSS ArcFlow

Digital Dossier

By Rebecca Russell

Woodson Research Center, Rice University

OVERVIEW

Woodson Research Center (WRC) is the Special Collections and University Archives for Rice University. Rice University is a private research university with an undergraduate focus located in Houston, Texas.

There are 5 professional Archivists in the department. 4 staff members work directly with digital preservation efforts, and have SAA-DAS certification. As a small department, we do not have dedicated curatorial foci, but do 'a bit of everything' including digital preservation. We typically spend about 20% of our time working on digital curation. A staff member could easily do this full-time.

Our institutional culture shapes our response to digital preservation. At Rice, it is very DIY and therefore we tend to use open-source tools and there is tolerance for staff time to learn and work with the DIY tools. Administration were supportive of staff taking time to become DAS trained, and there is enough staff experience to learn and document software steps, including command line interfaces.

Digital Scholarship Services (DSS) provide advisory services, guidance and review of the curatorial activities of the Woodson Research Center (including review of metadata, review of the ingest of collections, and monitoring collections within the preservation system.)

Fondren IT provides technical support by consulting with the WRC and DSS in acquiring hardware, software, and cloud storage space to manage digital collections. Our IT staff are for desktop support, not technical help with our digital preservation.

Digital Infrastructure

We have robust local and cloud storage environments, DSpace is our primary public discovery system. WRC uses an iterative process for workflows, using Google Docs as location for the initial centralized access/editing of our workflows and policies.

DIGITAL CURATION ACTIVITIES

The purpose of our Digital Preservation Policy is to establish a long-term digital preservation solution for our institution that will assure accessibility to special collections and unique resources.

- In accordance with our primary mission to support the institutional, research, and public service programs of the University, WRC plans to assure the long term access of our collections by continuing the digital preservation program which was developed in 2014.
- <https://digitalriceprojects.pbworks.com/w/page/44763477/Digital%20Preservation%20Support%20at%20Fondren%20Library>

Tools/Environment

- ArchivesSpace (collection management system - previously Archivists' Toolkit - full migration to AS in fall 2014)
- DSpace - to manage publicly available digital objects, Duracloud backup
- Google spreadsheet, tracking all AIPs (where they are stored, their size, their basic formats, where their hash value logs are)
- BitCurator (legacy media)

- DROID and Exiftool to gather key technical and possibly descriptive info
 - Droid reports give fixity, MIMEtype, file format name, file format version, PUIDsoftwareModified date (but not Date created or Accessed), file name and size
 - Exiftool (Command Line interface) reports give MAC times (context) - sometimes include created, includes Modified date, plus any embedded descriptive metadata
- Quickhash - to create and compare hash values over time (nearline) - investigating Fixity as a replacement for this tool
- BWF metadedit for audio files
- Handbrake for video files
- CERP (obsolete) investigating ePADD software 2018 for email preservation

Storage Environments

- DSpace (Institutional archive, publicly available digital objects) managed by 2 programmers
- DDN2 -- Nearline server storage of AIPs, local storage
- Amazon Glacier -- off-site storage of AIPs
- Offline -- not stored online (large files-such as uncompressed video, not stored in DSpace or DDN2) AIPs stored on 2 local external hard drives

Digital Curation Lifecycle (Major Phases)

- Digital objects → Ingest/Appraisal/Analysis (ArchivesSpace → Data Management (Digital object spreadsheet, ArchivesSpace - assign administrative, descriptive, technical, structural and preservation metadata) → Storage (either in public system or nearline system, external hard drive farm) → Finding Aid → Access (either in public system or nearline system)

Categories of Digital Content

- Publicly available materials
 - No copyright or privacy problems
 - Item level, generally
 - Workflow on public wiki
 - Least hands-on for our staff, least complicated, least documented
- Nearline/non-public materials
 - Has copyright or privacy concerns
 - Groups of files, or could be item level
 - Most hands-on, most complicated, most documented workflows

The category of materials determine the workflow. We use the same principles for each category:

- Rule of 3 for storage locations (3 copies, 2 formats, 1 offsite)
- Storage format must meet standards for AIP
- Creation, access, tracking - reasonable workflow for our staff and well documented
- Cost - must fit our budget of \$ and time

Recent Digital Preservation Activities at Rice

- We set DP goals annually and update our Digital Preservation Policy online

GOALS FOR DIGITAL CURATION

- Audio and video digitization, packaging and storage: KTRU (Rice's on-campus, student run radio station), Shepherd Music School digitization, oral histories
- At end of October, we hired a 6-month temporary DP staff member to help us with our legacy media processing backlog and contribute to the OSSArcFlow project.
- Darchive- Investigating a "dark" DSpace instance to manage our nearline materials.