National Digital Stewardship Residency | New York
Final Report
Dinah Handel

Host: CUNY Television

Harnessing Media Micro-Services for Stewardship of Digital Assets at CUNY TV

Description of the Project

The CUNY Television Library and Archives archiving and preservation initiatives were formally established in 2011, and have grown exponentially since. Realizing the need to standardize, enhance, and document these procedures, the CUNY Television Library and Archives applied and was selected to host a National Digital Stewardship Resident. Additionally, the Library and Archives were in the process of implementing a new Digital Asset Management System (DAMS), and needed to verify and migrate 1 petabyte of data from LTO 5 to LTO 7.

Thus, the NDSR project at CUNY Television had three major components:

1. **Refine, enhance, and standardize media archiving and preservation workflows.** The CUNY Television Library and Archives uses the microservices framework to process, transcode, create metadata for, deliver, and package its media assets. The microservices approach means that archiving and preservation actions are broken down into individual processes that accomplish specific actions, and these microservices are written in a scripting language called bash, hosted on the site GitHub, and installed using the homebrew package manager. For example, one micro-service script transcodes a video file input into an access copy suitable for broadcast, another micro-service script creates technical metadata for each of the video files within an archival information package using ffprobe and mediainfo. For the National Digital Stewardship Residency, I was asked to interview staff and relevant stakeholders in their daily workflows, consult with the Manager of the Library and Archives, and propose and implement changes and enhancements to the scripts based on the suggestions posed in these interviews. Additionally, I was responsible for investigating Digital Preservation standards and proposing different possible ways to implement them into the Library and Archive's workflows. Finally, I was tasked with creating documentation for the media micro-services used in the Library in Archives, as well as internal policies and procedures.
2. **Assist with the implementation of the new Digital Asset Management System.** At the start of the Residency, the Library and Archives was in the process of implementing a new Digital Asset Management System (DAMS) to house both broadcast media assets as well as a possible solution to making production material (b-roll, raw, and remote footage) searchable. I was asked to collaborate with the Manager of the Library and Archives and the Associate Archivist to assist in the implementation of the DAMS.
3. **Install and test LTO 7 hardware and software, and create workflow for the data migration from LTO 5 to LTO 7.** The CUNY Television Library and Archives utilizes LTO storage technology for the long-term storage and access of its digital media content. Until the start of the residency, the CUNY Television Library and Archives stored data on LTO 5 tape. In order to guarantee the long-term access of this content, it was necessary to verify and migrate the data to the newest generation of LTO tape. I was responsible for updating the open source software used to mount and write LTO tapes, installing and testing the new LTO 7 decks, and create and document a workflow for the migration of data from LTO 5 to LTO 7 tape.

Project Partners

The primary project partners were Dave Rice, the Manager of the CUNY Television Library and Archives, and Catriona Schlosser, the Assistant Archivist at the CUNY Television Library and Archives. Additional staff at CUNY Television that assisted with the project's goals were Annie Murphy, Amy Monte, Oksana Israilova, and Saul Spicer. The project was also completed with the assistance and consultation of Nicole Martin, Archivist at Human Rights Watch, the Association of Moving Image Archivists listserv and webinar series "Introduction to Digital Formats" and fellow NDSR NY cohort members and members of previous NDSR cohorts.

Project Execution

As outlined in the project description, my project at CUNY Television had three major components. Thus, I will present the work accomplished for each of these components.

Microservices/archiving and preservation workflows at CUNY TV

Documentation

Prior to my residency at CUNY Television, there was little documentation of internal archiving and preservation policies and practices. During my residency, I wrote documentation that outlined the functionalities of the media microservices, which is open source software used to process media. This included creating extensive directions for download and use in the readme document, as part of the media microservices repository on GitHub. Further, I collaborated with the Assistant Archivist to document various internal practices and procedures of the archive. This documentation was created as part of a private repository, and included text on the technological infrastructure within the archive, processes relating to metadata creation and scheduling of TV programs, the workflow for captioning of programs, and instructions for the digitization or reformatting of analog media. Specifically, I worked on documentation outlining the technological infrastructure of the library and archives.

Scripts

Much of my project work resulted in edits to the media microservices code used to process audiovisual files. These edits consisted of changes and additions to enhance the functionality of the scripts, remove redundancies and depreciated code, and standardize syntax across all the scripts. To that end, the following changes were made to the media microservices code:

- Added new functions and formulas: get channel layout, added capability to handle sd, replaced videofilterchain with add_video_filter function, created filter_to_middle_option function, created unset_variables function, created _write_error_log function to increase error logging and checking, forcing sd to be hd option in makebroadcast, created migratefiles to better log rsync activities.
- Created a graphical user interface (GUI) for mmconfig.
- Added options to ingestfile, the primary script used to process, transcode, package, and deliver audiovisual files: added PSA option to create packages of public service announcements and deliver PSAs to omneon for broadcast and long-term storage, added preservation option which allows for creation of AIP of digitized materials and accounts for additional logs created during format migration, made the graphical user

interface the default option for users, integrated the script makeslate as an option in ingestfile, removed unnecessary submission documentation directory, added code so that script exits when person clicks “cancel” button in the GUI, added code so that script exits when trying to run in a directory that doesn’t exist.
- Standardized variables and functions across all scripts, wrote more comments into the code, updated usage section of scripts to correctly identify options, deleted depreciated functions and code.
- Metadata: wrote script to document structure of archival information package which was integrated into makemets, began draft of script to check and verify mets, created package type categories and added logging to capture log of package type.
- Wrote QA testing script, created option for interlacement testing and integrated into ingestfile.
- Modified restructureforcompliance based on external community needs.

DAM

When I arrived at CUNY Television, the library and archives was in the process of implementing an open source digital asset management system (DAM). The DAM was intended to serve two purposes: first, it would house access copies of all of the television programs broadcast on CUNY Television, and second, it would serve as a searchable and centralized repository for material used by the production department, with the intention of cutting down on storage bloat resulting from redundancy of audiovisual files. To assist with getting materials uploaded into the DAM, I wrote a script that queries the DAM and our broadcast server to determine which files have not been uploaded to the DAM, transcodes the file if it hasn’t been uploaded, and moves it to a directory that is “watched” by the DAM, which then automatically uploads the file periodically. My project was less focused on the implementation of the DAM for the production department, although I conducted early interviews with production staff to determine if implementation was possible, and participated in meetings between the library and archives, the web and new media team, and the production department.

LTO migration

The third component of my project was to research, build, and test a workflow for the migration of data from LTO 5 to LTO 7 tape. To begin, I spent time researching and testing different file transfer commands, including bbcp, rsync, and gcp, to determine which command would transfer the files most quickly, with the least number of errors. Then, I developed the workflow to move the files from LTO 5 tape to a staging hard drive, perform fixity verification on the files, write them to LTO 7 tape, and read them back to perform an additional fixity check to ensure that everything had written properly and without errors. Developing this workflow included writing two scripts: aipupgrade, which restructured older packages based on current package standards, and collectionchecksum, which aggregated the contents of individual checksum files from Archival Information Packages and created checksums for metadata files. Additionally I made modifications to the open source software used to write LTOs, ltopers, to automatically read back and create checksums for the contents of the tape immediately after writing. I also created extensive directions outlining the workflow as part of the library’s documentation, and

spent time training the associate archivist on the process. During my time as a resident, I successfully verified and migrated 30 TB of content from LTO 5 tape, upgraded and wrote 15 TB from the library's acquisitions, and wrote 5 TB of digitized materials.

Analysis and Evaluation

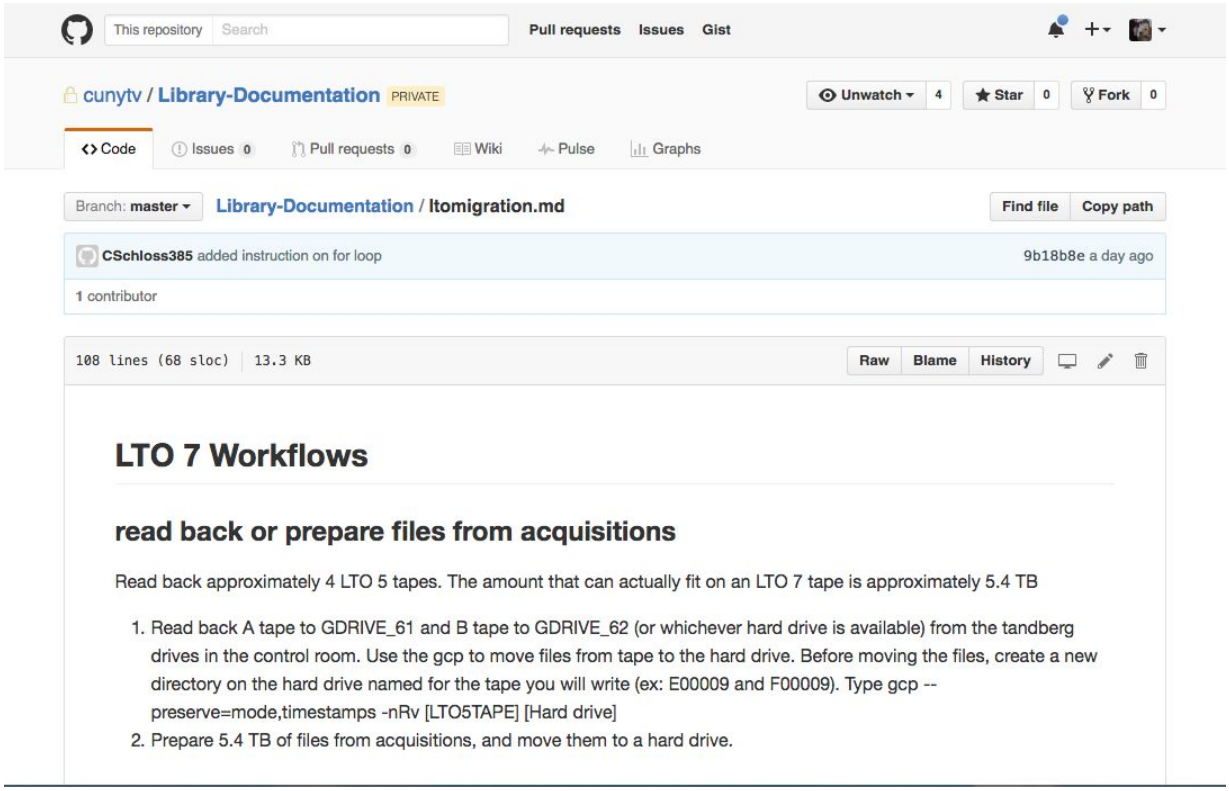
The project was successful in completing the goals stated in the project proposal. I was able to propose and implement changes to the archiving and preservation workflow through the enhancements of media microservice scripts, assist in the implementation of the Digital Asset Management System, and create and document a preliminary workflow for the data migration. Unlike other NDSR projects in which the resident prepares a report with suggestions, I was able to have my suggestions approved by my mentors, and begin implementing them immediately. However, there are some aspects of my project that will have to continue beyond the end of my residency, primarily the data migration. In order to ensure that this part of the project is continued, I made sure to spend the final weeks of my project going over the workflow with the Assistant Archivist and mentor, so that they can continue the data migration.

Deliverables

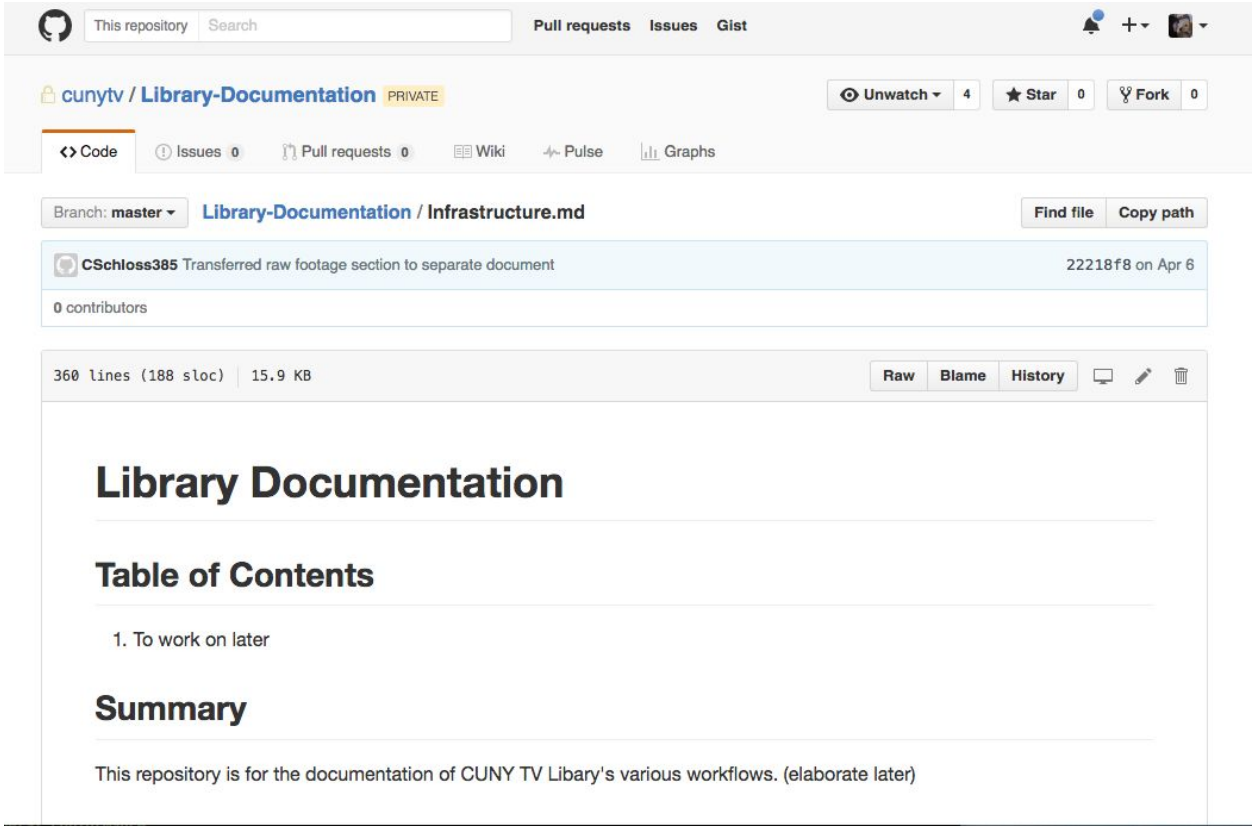
Most of my project contributions can be found via my GitHub commit history: https://github.com/dinahhandel

The rest of my contributions are housed on a private institutional GitHub account, and are as follows:

LTO workflow documentation



Infrastructure documentation



For my enrichment activity, I organized a day-trip to New Jersey, that included a trip to the RECAP facilities, the Center for Digital Humanities at Princeton University, and the Seeley Mudd Manuscript Library to visit with Jarrett Drake, the Digital Archivist at Princeton University.

During my residency, I published the following blog posts, which may be found here:
“Produceroni” in collaboration with Mary Kidd, http://ndsr.nycdigital.org/produceroni/
“LTO(woe)s” http://ndsr.nycdigital.org/ltowoes/
“Open Source Tools for Working with Audiovisual Files”
http://ndsr.nycdigital.org/open-source-tools-for-working-with-audio-visual-files/
“Cohort Field Trip to Princeton, NJ” http://ndsr.nycdigital.org/cohort-field-trip-to-princeton-nj/
“Check your AIP before you wreck your AIP”
http://ndsr.nycdigital.org/check-your-aip-before-you-wreck-your-aip/
“Media Microservices and Archival Workflows at CUNY Television”
http://ndsr.nycdigital.org/media-miscro-services-and-archival-workflows-at-cuny-television/
“Diving In Headfirst” http://ndsr.nycdigital.org/diving-in-head-first/

Professional Development
Over the course of the residency I acquired a significant amount of knowledge about the audiovisual archiving and preservation field, and shared that knowledge in a public forum

whenever possible. I completed the following public presentations and professional development activities over the course of my residency:

- Poster presentation at AMIA: In November 2015, I traveled to Portland, Oregon to present a poster at the Association of Moving Image Archivists annual conference. In addition to presenting a poster about my project, I was able to meet numerous individuals in the field and make connections with other moving image archivists from across the country.
- Speaking at METROcon: In January 2016, I presented on a panel with the rest of the NDSR NY cohort about the project thus far and some major takeaways from my work at CUNY Television.
- Talk at Code4lib: In March 2016, I presented at the Code4Lib conference with colleague Ashley Blewer, a developer at the New York Public Library. We presented on open source technology and how libraries and archives would benefit from open-sourcing their workflows and implementing a microservices approach.
- Talking at NDSR NY symposium: In April 2016, I spoke at the NDSR cohort's "Let's Get Digital" symposium, on the topic of audiovisual materials and open source technology for working with audio visual materials. I also assisted with the organization of this event.
- Blog posts/public discussions: Throughout my residency, I published numerous blog posts that sparked discussion in the comments and on Twitter.
- Over the course of my residency, I became increasingly involved in AMIA through stream participation (Digital Preservation), stream curation (Education), and task force co-leadership.
- Poster presentation at SAA: In August 2016, I will travel to Atlanta, Georgia, and present a poster with Mary Kidd and Morgan McKeehan on the utility of learning to code in archival settings.
- Paper presentation at IASA: In September 2016, I will travel to Washington D.C. to present a paper at the annual conference of the International Association for Sound and Audiovisual Archives, on LTO technology and data migration.