CUNY Television: AAPB NDSR Project Proposal

Date Submitted: December 7, 2015

Submitted by: Dave Rice, Directory of Media Library and Archives | CUNY Television

Project Title: Preservation at the Next Level: Utilizing Preservation Metadata and Fixity Features

Project Summary

Within CUNY Television, the Media Library and Archives functions as a hub to media workflows throughout the station. Media and metadata are created by production and creative departments and deposited into the Archives. The Archives subsequently processes material and delivers it to other departments tasked with distributions (podcasting, broadcast, web streaming, etc). The organization of CUNY Television media workflows amongst departments reflects the OAIS model of submission, archiving, and delivery. "Preservation at the Next Level" will build upon the prior work of the archive and NDSR NY and prioritize goals within a new project to focus on revealing connection, generation, and provenance within file-based preservation collections. The resident shall be supported in the use of open source software for the purposes of assessment and implementation of recommendations.

The resident will survey and assess the entirety of CUNY Television's digital and physical holdings to gain an impression of the state of digitization, digital preservation efforts, and possible future directions for digital stewardship and preservation for CUNY Television. The survey will be conducted through the creation of micro-services scripts that query pre-existing metadata records that exist separately across CUNY Television's Archival Information Packages. Assessment will take place in the form of visualization and analysis of metrics such as codecs in use across the collection, extent of collection digitized by

format and percent overall, progress of digitization over time, preservation actions taken upon physical and digital media, as well as other metrics to be determined by the resident.

Based upon the assessment and analysis conducted, the resident will then research, assess, and implement changes to the Media Library and Archive's workflow that support progress towards upper-levels of the NDSA Levels of Digital Preservation. The resident will conduct research as to the present state of CUNY Television's NSDA levels of Digital Preservation through interviews and assessment of current collections' metadata integrity, physical, and digital storage systems. The resident will then propose and implement suggestions to improve CUNY Television's NDSA levels of Digital Preservation by a suggested minimum of four points, with regards to Storage and Geographic Location, File Fixity and Data Integrity, Information Security, Metadata, and File Formats. Possible deliverables could include a risk assessment and prioritization of digitization efforts, a disaster preparedness policy, or a publicly accessible digital preservation policy.

Finally, this project is designed to continue the development of archival procedures of the CUNY TV Library, with a focus on the interaction of collection management strategy with preservation and the extension of our micro-service environment to facilitate new opportunities and forms of fixity, including framemd5s and perceptual hashing. CUNY Television's archive manages a complex mixture of broadcast and production materials that have been acquired over decades of production work. CUNY Television is committed to digitizing these materials; however, we realize that in order to navigate this increasingly large digital collection, we will need to find a quick and efficient way to describe and link our assets without having to catalog all the material by hand. The concept of perceptual hashing in an audiovisual preservation environment complements more traditional form of fixity. A perceptual hash may be used to identify visual similarities between two images even if the images differ because of minor visual adjustments or lossy compression artifacts. By generating perceptual hashes per frame per file at the point of acquisition or digitization, we may have the opportunity to link our preserved materials together by

visual similarities. Perceptual hashing could enable a highly automated method to relate source material to the final work or multiple copies of the same work together.

At CUNY Television, as in many public media organization, the collection is filled with redundant and related copies of material and it is occasionally difficult to identify which copy to focus preservation effort upon. Perceptual hashes can provide another mechanism to identify multiple instantiations as representing a single asset as well as provide a method to illuminate the production history and provenance of our audiovisual holdings.

The project would work with the micro-services currently in service in CUNY TV's Library. These micro-services are primarily comprised of bash scripts and defined workflows to accomplish specific tasks, such as transcoding, assessment, delivery, storage, metadata harvesting, logging, and digitization. The project anticipates coding, technical experimentation, feature development, and work on the library's databases in order to better align CUNY Television's Media Library and Archive with preservation standards, technological innovation, and digital sustainability. Because of the technical nature of the objectives both the project mentors and CUNY Television technology and engineering staff are available to work closely with the resident in support of these objectives.

Specific Project Goals & Objectives

- To increase CUNY TV's NDSA Digital Preservation Levels by a minimum of four points.
- Research and implement perceptual hashes in CUNY TV's audiovisual assets as a more quick and efficient way to link assets together for increased discoverability and fixity.
- Conduct an assessment of the present state of CUNY Television's digitized and born digital collections through the creation of micro-service scripts that query collection metadata.

- Visualize collection metadata to better understand the current state of CUNY TV's digital assets for the purpose of guiding a digital preservation policy.
- Create documentation that integrates the resident's assessment and analysis into the future work flow of CUNY TV.

Project Timeframe & Deliverables

Timeframe

- Month 1
  - An introduction to the mission, work, and collections of CUNY Television.
  - An introduction to the CUNY Television library, its staff, services, and responsibilities.
  - Introduction to relevant departments including the New Media, Broadcast, Programming/Scheduling, and IT.
  - Shadow the archivist and broadcast librarian and train on content acquisition, registration, and dissemination.
  - Interview select staff about library operations including the Control Room Supervisor, Broadcast Librarian, Captioning Manager, and Archival Technicians.
  - Under the guidance of the broadcast librarian and control room supervisor, review operational procedures between the library and broadcast groups, file delivery, transcoding, quality control, and record-keeping.
  - With the project mentors, get an overview of the library's micro-services and datasets, along with their integration and use.
  - Initial testing and training on perceptual hashing and integration planning within existing fixity workflows.
  - Review existing preservation metadata from acquisition, digitization, and processing workflows.

- Month 2-3

- Interview representative members of New Media, Broadcast, Programming/Scheduling, and IT departments regarding accessibility, library services, and information and media transactions with the library,
- Assess CUNY Television's archival practice according to the categories outlined in the [NDSA Levels of Preservation](#).
- Review current micro-service scripts and their recent development. Review project outcome from the CUNY Television's NY NDSR project.
- Use a GitHub issue tracker to detail bugs or enhancement requests pertaining to the micro-service collection.
- Prototype a micro-service to perform perceptual hashing for new acquisitions.
- Collaborate with archive and IT staff to plan development of a micro-service to conduct assessments across the collection based on existing AIP preservation metadata.
- Propose extensions or refinements in the creation of preservation metadata.
- Propose which categories of the [NDSA Levels of Preservation](#) should be addressed within the scope of the AAPB NDSR project according to perceived impact and need. The proposal should also establish particular goals as to what could reasonably be achieved by the end of the NDSR project with the collaboration with CUNY Television's archive and IT staff.

- **Month 4-5**
  - Extend implementation of perceptual hashing retroactively to existing digital preservation collections.
  - Prototype reports or inter-asset record linking based on shared perceptual hash data.
  - Determine and enact selective steps towards increasing NDSA Preservation Levels according to proposals set in months 2-3.
  - Address relevant feature requests and bug fixes for micro-services.
  - Provide an interim progress report on the progress of the residency.

- ○ Determine what preservation metadata should be copied from the Archival Information Package into a database for summarization and preservation reporting.
- ○ Extend preservation metadata procedures to communicate with the archive's database to store select preservation data (such as effort has already been enacted with LTO storage and fixity micro-services and these database reporting features could be extended to other micro-services).

- ● Month 6-7
  - ○ Design an interface to report on and summarize preservation data as a snapshot or represented over time.
  - ○ Work on goals and deliverables to improve CUNY Television's NDSA Preservation Level assessment as established in months 2-3.
  - ○ Address feature requests and bug fixes for micro-services.
  - ○ Update AIP compliance micro-service accordingly to any extensions of preservation metadata that affect the Archival Information Package.
  - ○ Demonstration of perceptual hash workflow and reporting, gathering feedback, assessing the opportunity of perceptual hashing upon digitization and preservation workflows.

- ● Month 8-10
  - ○ Ongoing and final work on goals and deliverables to improve CUNY Television's NDSA Preservation Level assessment as established in months 2-3.
  - ○ Draft final progress report culminating the discoveries and progress of the residency and outlining the remaining recommendations CUNY TV after the residency has been completed.
  - ○ Document results of new preservation reporting functions and tests and provide recommendations for scaling and extending the use of preservation metadata

- Document the results of perceptual hashing and provide recommendations for further use or consideration of the process.
- Provide final self-assessment of the archive to the NDSA Preservation Levels.
- Coordinate and draft public documentation on project results.

Deliverables

- Initial informal assessment of current archival workflow according to the NDSA Preservation Levels.
- Under collaboration with CUNY TV archivists, draft a micro-service to gather and report on perceptual hashes from media.
- Proposal for the prioritization of work and goals for increasing the overall score of an NDSA self-assessment.
- Deliverables related to project components to increase the overall NDSA Preservation Level self-assessment score are to be determined via collaboration between the Resident and CUNY Television archivists.
- Updates of micro-services to incorporate reporting functions of preservation metadata and perceptual hashing, such as pushing such data to our database.
- Design and implementation of reporting features based upon gathering preservation metadata through archival workflow.
- Design and implementation of reporting features based upon gathering perceptual hashes throughout archival workflow.
- Final formal assessment of current archival workflow according to the NDSA Preservation Levels.

Resources Required for Project

CUNY Television will provide the resident with a Mac computer and workspace based in the library, which provides an open office working environment amongst the project mentors and other key staff. The resident will also have access to a workstation in CUNY TV's data center and control room. We will ensure that the resident has sufficient hardware

access for the assignment, including access to relevant networks, systems, and databases as well as decks for digital media formats such as LTO, XDCam, DV, and various digital recording formats. The resident will be supplied with installations of required software such as Final Cut Pro, FileMaker, FFmpeg, MediaInfo, and installation of CUNY TV's micro-service collection with all included dependencies. The resident will also be supplied with a station passkey and business cards.

The project mentors as well as technology staff will also actively support, train, and make resources available to support the resident's work within areas of micro-service design, audiovisual assessment, transcoding, and OAIS implementations.

Project Context

The library at CUNY Television has long served as a hub of media workflows. Newly produced or acquired content is submitted to the library for registration and cataloging and the library subsequently ensures that the content is accessible for reuse, broadcast, online access, or other uses. Formerly the library required these media transactions in videotape formats, but in 2011 the library began transitions that enabled and encouraged the submission and dissemination of digital file-based media. Also in 2011, CUNY Television expanded the staffing and services of the library to include a preservation program, research services, and new focuses on information management, media accessibility, and preservation standards.

In order to accommodate a large and steady rate of data and processing from both acquisitions and digitization efforts, the library designed an OAIS-inspired packaging standard for content management and adopted LTFS formatted LTO tapes as a digital storage medium. Whereas the library formerly managed circulation of videotapes, the library developed systems to evaluate incoming media and to transcode it to the requirements of particular uses. These operations along with other tasks (including quality control, delivery, and reporting) are performed through the application of micro-services or sets of micro-services on archival packages. Locally these micro-services exist as

registered bash scripts that coordinate open source tools such as FFmpeg, MediaInfo, ExifTool, and md5deep.

As part of a NDSR NY project running in 2015-2016, current resident Dinah Handel has conducted discovery research in order to assess the current implementation and challenges of CUNY Television's archival workflows. This research led to a series of updates and expansion of our media micro-services, the implementation of standards, documentation, and workflow refinements that are currently underway. This effort has streamlined archive and access workflows and brought an open source approach to broadcast television. This work is particularly important since it affects a living archive that strives to reduce processing time in order to quickly and accurately convert submission media into distribution media while also focusing on archival standards. More information about the project and the current progress can be seen in Dinah's NDSR blog posts at http://ndsr.nycdigital.org/author/dhandel/.

Post-Residency Project Integration Plan

As with the ongoing NY NDSR project, the resident shall work on a project that is highly integrated into the current day to day archival workflow. The work of the resident will be integrated into the live systems as they are tested and achieved. Further, the work implemented by the resident will serve as a catalyst for CUNY Television's continued growth and progress in the areas of digital preservation and stewardship.

Efforts to formalize and develop the CUNY Television archive started about four years ago with the archive growing quickly to accommodate the flow of media throughout the organization. At this later stage in development, CUNY Television is well prepared to review the status of the archive workflows in accordance with established standards such as the NDSA Preservation Levels and the mandates and goals as detailed in the OAIS Model. The role of this project supports CUNY Television's archival commitments in an active and growing archive.

Required Knowledge and Skills for Residents

Applications for the residency should have a graduate degree or equivalent in Library and Information Science, Moving Image Archiving and Preservation, or Archival Studies. Additionally the resident should have the following knowledge or qualities:

- Interest in digital archiving
- Ability to work independently or collaboratively
- Detail oriented and highly organized
- Project management skills
- Familiarity with PBCore
- Familiarity with FFmpeg
- Willingness to learn emerging digital technologies
- Familiarity with the Open Archive Information System (OAIS) framework

Preferred Knowledge or Experience of Resident

- Experience working with born-digital materials
- Knowledge of audiovisual file formats
- Experience working with bash scripts and XML
- Experience in the PREMIS metadata standard or handling process metadata
- Knowledge of FileMaker or MySQL