

Open Source Tools for Digitization, Preservation and Access in Audiovisual Collections

Catriona Schlosser, Archivist, CUNY TV



Introduction to Uncovering CUNY TV's Audiovisual Heritage

In 2019 the Council for Library and Information Resources (CLIR) awarded CUNY Television a grant to support digitization and access to CUNY's unique collections. This project, titled "Uncovering The City University of New York's Audiovisual Heritage," digitized and made accessible audiovisual objects from a select number of CUNY campuses including Hunter College's Centro de Estudios Puertorriqueños, College of Staten Island, the School of Public and International Affairs at Baruch College, Special Collections of Bronx Community College and Medgar Evers College. These collections focus on a number of diverse topics including urban development, education, public policy and criminal justice.



Umatic tapes from the Baruch College of Public Affairs Collection



The Hall of Fame for Great Americans at Bronx Community College

In 2020, CUNY TV hired project manager Kelly Haydon to oversee the project, and she successfully implemented and adapted CUNY TV's open source workflows to process these collections. During her time here, she digitized over 1000 media objects, added descriptive and technical metadata, provided captions and made accessible 70% of these items online!



Digitization and Quality Control



The analog video objects for the Hidden Collections project were digitized using the open source software **vrecord** (found on Github at [amiopensource/vrecord](https://github.com/amiopensource/vrecord)). Vrecord gives the user more control over the digitization process because the technician can pick their video and audio input, codec, container, audio channel mapping, video bit depth and standard. The passthrough mode also provides the users with a number of different quality control views that help with setup.

The open source program, **QCTools (Quality Control Tools for Video Preservation)**, is an essential tool when reviewing digitized files. It provides the user with a visual aid created using audiovisual analytics and filtering to better understand the file. This tool helps identify head clogs, dirt and other playback errors. Along with QCTools, FFmpeg and MediaInfo also help with troubleshooting by providing the user with technical information that assists with quality control.

One major achievement from this project was the formalization of a DV workflow. This project gave the archive the opportunity to work with tools such as **DVRescue (mediarea.net/DVRescue)** to digitize DV material from the Center for Puerto Rican Studies and Baruch College collections. Our project manager Kelly Haydon experimented with and tested this software and in the process created a workflow for this format which can be adapted to other DV collections.

Tools used: vrecord, dvrescue, QCTools, FFmpeg, and MediaInfo

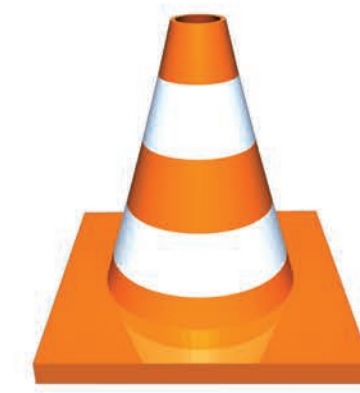
Long Term Preservation

Files created from the Uncovering CUNY's Audiovisual Heritage project were all written to LTO 7 tape using scripts available on **GitHub (amiopensource/ltopers)**. Before writing to LTO for long term preservation, each AIP was reviewed using the tools **verifypackage** and **verifytree**, which are both part of the **mediamicroservices** collection. **verifypackage** runs tests to make sure all of the files match the technical specifications they're expected to while **verifytree** validates the package's file structure and outputs any unexpected items found in the AIP in the Terminal. If the packages passed both tests, then files were written to two separate LTO tapes



Tools used: verifypackage, verifytree, writelto and verifylto

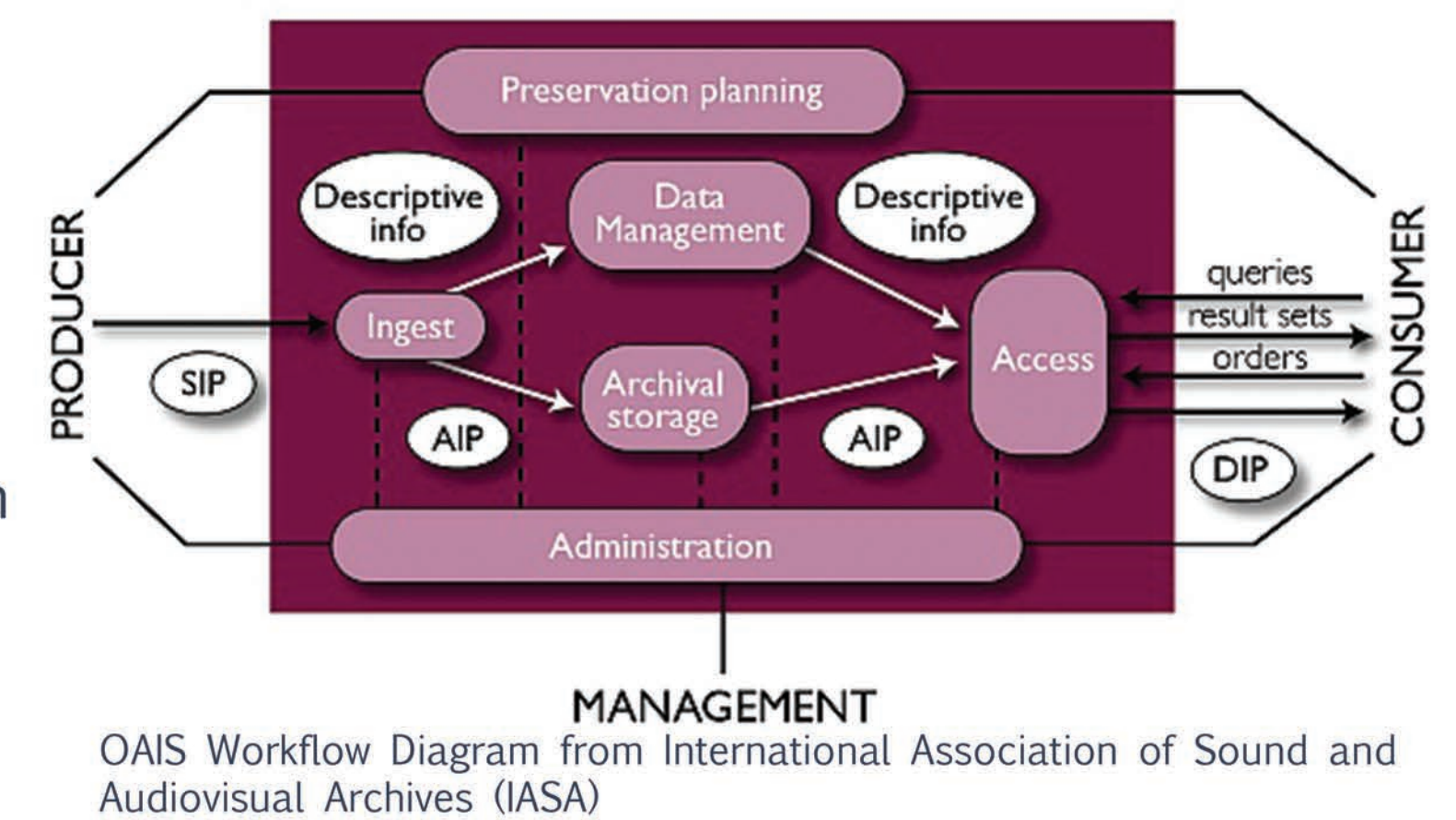
Why Use Open Source Tools?



- **Software obsolescence** is a major obstacle when managing, processing and preserving audiovisual material. Archivists are forced to find creative ways to process and preserve digital collections that don't solely rely on proprietary software
- The ability to access source code, make adjustments to it, file tickets with developers, and ensure that the tools work for an institution's needs, gives an archivist much more **control** over their collections.
- Open source programs and tools are **flexible**. Archivists can use the tools and software that work best for their institution and they can tailor it to their needs.
- Open Source software builds **community** by creating a space for people to collaborate, identify bugs, and suggest improvements. Even if you don't know how to code, groups like the **amiopensource** on Github, allows any archivist to contribute by opening up tickets, providing feedback or submitting workflows and documentation.

Ingest

CUNY TV uses a set of scripts called **mediamicroservices** to process its audiovisual collections. All of these scripts can be found on **GitHub (mediamicroservices/mm)**. The archive follows the Open Archival Information

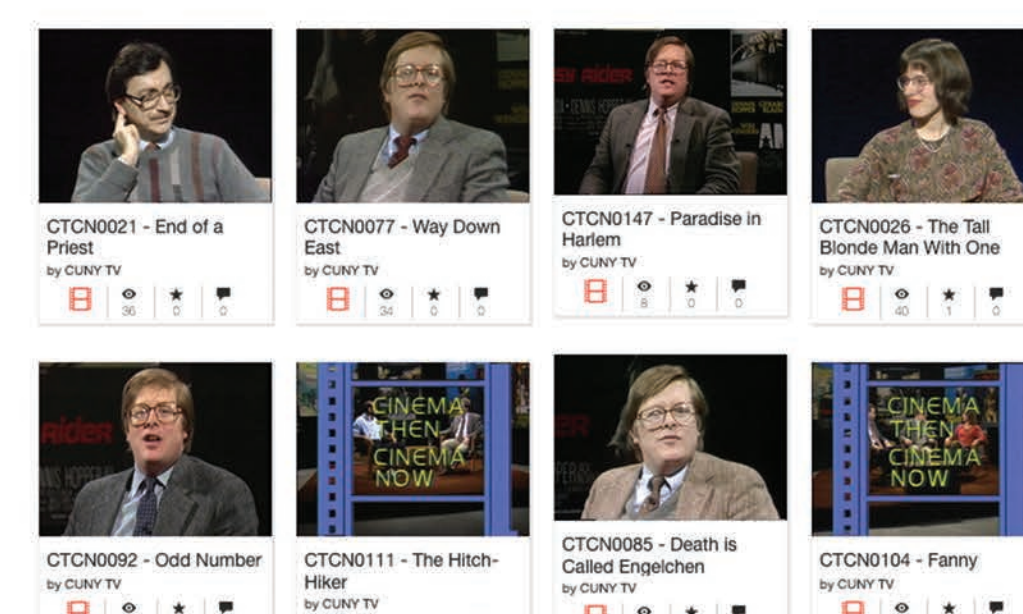


System (OAIS model) for its digital preservation workflow and runs a script called **ingestfile** to create Archival Information Packages that contain a preservation master, access files, technical metadata and checksums.

For the CLIR project, we used a preservation mode with **ingestfile** to create our derivatives. This mode allows a user to create chapter markers within a preservation file using a script called **makemkvchapters**. When working with a Matroska file, this script adds these chapters and then creates the derivatives using these chapters.

Tools used: makemkvchapters and ingestfile

Accessibility



CUNY TV's Internet Archive Page

Objects digitized for the Hidden Collections project were made accessible on CUNY TV's website (tv.cuny.edu) and the Internet Archive (archive.org/details/cunytv). Items uploaded to the internet archive relied on **archive.org** command line tools that can easily upload and add metadata to an item. CUNY TV uses a script that works with these command line tools called **uploadia** to upload the file and then to pull **pbcore** metadata from the archive's databases using scripts called **fmpbcore** and **pbprotractor**.

Captions for these files were created with a mix of auto captions pulled from Youtube using **youtube-dl** along with captions that have been human edited.

Along with uploads and captions, Kelly also worked with collections managers to give them the file types that they needed for preservation and access within their own collections by providing them with hard drives.

Tools used: youtube-dl, uploadia, ia, fmpbcore and pbprotractor.