

BUILDING A DIGITAL PRESERVATION COMMUNITY IN PUBLIC BROADCASTING: A CASE STUDY OF THE AMERICAN ARCHIVE OF PUBLIC BROADCASTING'S NATIONAL DIGITAL STEWARDSHIP RESIDENCIES

Andrew Weaver, Eddy Colloton, Kate McManus, Selena Chau, Lorena Ramirez-Lopez, Adam Lott and Tressa Graves, National Digital Stewardship Residents, USA

1. Introduction

The National Digital Stewardship Residency (NDSR) is a competitive program which aims to foster a growing community of professionals who are proficient in the tenets of digital preservation and possess the necessary knowledge to advance efforts for the safeguarding of our shared digital heritage. It is built around a cohort model, and to date has had contingents based in Boston, New York City, and Washington D.C. For 2016-2017, the American Archive of Public Broadcasting (AAPB) received grant funding from the Institute of Museum and Library Services (IMLS) to create an NDSR cohort of audiovisual archivists placed in geographically diverse public broadcasting stations not limited to just the Northeast region of the United States. These consist of both television and radio stations and represent a wide range of sizes as well as current capacities for digital preservation.

This range allows insight into models of best practices at different scales for different regions. Residents assess broadcast and production workflows, which are often complex and not well documented. These residents work collaboratively with staff to incorporate preservation standards that lead to increasing access to digital content, maintaining effective workflows, and reducing digital preservation threats while improving authenticity, understandability, persistence, and renderability of digital content. The work done in these residencies may offer similar local broadcast environments ideas to consider for safeguarding their history while simultaneously creating it.

Digital stewardship for audiovisual collections is knowledge that may not easily disseminate to all groups with audiovisual collections. To that end while each resident's project varies from processes and goals, the overall mission to preserve their organizations' content strongly resonates in each of their proposals.

Information on each host site's project is available at <http://ndsr.americanarchive.org>.

2. City University of New York Television, Andrew Weaver

The library and archives at City University of New York Television (CUNYTV) exist in a uniquely central position within the station. All content that is produced goes through the library, and the library itself is involved with broadcast scheduling. Archival materials are stored on LTO tapes, with a migration from LTO 5 to LTO 7 currently underway. Consequently, CUNY TV already has in place a robust digital workflow built on a system of shell scripts (or microservices) that enable the efficient creation of archival packages containing a vast spectrum of preservation metadata.¹ This makes my NDSR project somewhat unique amongst this cohort, as I am dealing more with refining relatively advanced systems as opposed to implementing more basic tenets of digital preservation.

To date, I have been focused on raising CUNYTV's compliance with the recommendations put forth in the NDSA levels of digital preservation (Phillips, et al., 2013). As such, I am working extensively with this metadata to make it more centralized and easy to use, while also assessing possible enhancements to the types of metadata gathered. When archival packages are generated by CUNY's microservices a broad spectrum of fixity and technical information is created and stored as sidecar files. There currently is a system for adding some of this information

¹ These scripts are publicly available on Github at <https://github.com/mediamicroservices/mm>.

into an existing FileMaker database, but this has been found to be prohibitively time consuming, with the reports from a single LTO tape taking several hours to upload. As a result of my preliminary NDSA levels assessment I have concluded that this data needs to be more accessible for queries as well as more explicitly integrated with metadata pertaining to PREMIS events.

In order to accomplish this I have been working on developing a MySQL database to store fixity and PREMIS metadata, as well as modifying existing microservices to intelligently communicate with this database. In the current prototype, this is accomplished in two ways. CUNY TV has several microservices that generate fixity metadata on the package level as well as on the collection level (such as for an entire LTO tape). These scripts have been modified so that after execution, the generated manifest will be cycled through in an iterative manner to report all checksums and file paths to the database. This is used in conjunction with a database table for PREMIS information that records event types and event outcomes both for fixity generation as well as for all other microservices involved in package creation. I have found the PREMIS Data Dictionary (PREMIS Editorial Committee, 2008) a useful document for providing guidance on developing a standardized vocabulary for events, as well as for what type of information to prioritize.

Another component of my project that I have been investigating beyond standard preservation metadata is the implementation of perceptual hashing within our workflows. Perceptual hashing is similar to normal fixity metadata in that it uses algorithms to analyze content, but while fixity is seeking to identify identical data, perceptual hashing seeks to create “fingerprints” that can be used to establish connections between similar data. If this data were to be collected, it could potentially aid in the identification of derivative materials as well as help automate some elements of cataloging. Due to the wider availability of tools for creating these hashes for audio materials², so far I have been testing the efficacy of creating fingerprints from the audio component of video programs. These tests have been relatively successful at establishing links between content, and the next stages of my project will involve establishing both how to efficiently store and effectively use this information.

3. Louisiana Public Broadcasting, Eddy Colloton

The archiving initiative at Louisiana Public Broadcasting (LPB), as it exists now, began just under a decade ago, in 2009. Through the American Archive of Public Broadcasting (AAPB) project and an IMLS funded project, LPB embarked on the difficult task of preserving and providing public access to the station's thousands of video assets. The station's archive contains a wealth of culturally and politically significant content, from governor's debates to interviews with Louisiana artists, musicians and writers. Not to mention, the archive has tremendous value to the station and its producers. Archival footage is now often employed in LPB's weekly news magazine program, *Louisiana: The State We're In*, illustrating the relevance of the state's history through the lens of today's issues.

LPB works in partnership with the Louisiana State Archives to provide public access to moving images in these two institutions' collections through the Louisiana Digital Media Archive (LDMA), located at ladigitalmedia.org. The public facing website of the LDMA is built on top of a custom designed Microsoft SQL database that both institutions use to catalog and track their materials. The database is used contemporaneously to describe content that has recently been produced by LPB, as well as the analog video materials from the stations 40+ year history. LPB's analog video legacy is being migrated to digital formats in-house. As an established TV station, LPB is in the fortunate position to have access to analog video equipment and expertise in using, maintaining, and repairing this equipment. As digital video files are created, whether they be recently produced programs or documentaries from the 1980s, an access video file encoded for web streaming is automatically created and sent to the station's web server via open source scripting designed and maintained by the station's IT department. These access files are then

2 Most of my experimentation has used the Chromaprint library: <https://acoustid.org/chromaprint>

linked to their corresponding record in the database, and, copyright permitting, are published to the LDMA website.

Needless to say, developing and managing this workflow has taken significant time and effort from station employees' whose primary occupation is facilitating and creating local broadcasting. Given this situation, policies, procedures, and responsibilities have developed as a matter of course, and were not fully documented. One of the goals of LPB's AAPB National Digital Stewardship Residency (NDSR), which began in July of 2016, has been to create workflow documentation.

As the resident at LPB, I began developing this documentation by interviewing LPB staff. My goal was essentially to follow the "life-cycle" of a particular program, from creation, to broadcast, and finally to deposit in the archive. Not surprisingly, I often found that an individual involved in a particular part of the workflow would be very familiar with their own tasks and responsibilities, but have little knowledge about the other aspects of the program's journey to deposit. Untangling some of the complexities of this system was time consuming and at points confusing, but has been very valuable as we move toward the second phase of the project, which will focus on making improvements to the workflow to promote digital preservation.

Given that the workflow created for archiving LPB's audiovisual legacy has been successful in describing and providing access to both contemporary and historic material, coupled with the challenge of making large-scale shifts in institutional practices, my recommendations for improving the workflow will involve more "tweaks" than sweeping change. Much of this work will involve drafting and implementing policy documents that re-iterate many of the responsibilities that engineers, IT staff, and producers already perform, simply providing a template to perform those responsibilities in a more uniform manner. Similarly, digital preservation practices like fixity checks and metadata extraction can be added to these ongoing responsibilities, hopefully, without disruption. Clearly articulated policies can also provide a mechanism for further development of the archival workflow.

As this effort looks to the future of material created and archived by LPB, I have also been spending time with the content already in the archive and available on the LDMA website. I have been working with the IT manager to extract technical metadata from all of digital videos files in the archive and import that information into the database. I'm also regularly brainstorming with the station's full-time archivist, Leslie Bourgeois, looking for ways to promote and improve the LDMA website.

While digital preservation practices and audiovisual materials often appear to be frustratingly complex, I'm comforted by the fact that much of my work as a National Digital Stewardship Resident has been rooted in traditional archival practice. I am regularly thinking about technology-agnostic concepts like accession policy, documentation of provenance, and public access. Moreover, many of these concepts are prevalent in my cohort's projects too, which has given me a sounding board and support group for working through challenges in my first position as a professional archivist. I feel that, as a cohort, we are all motivated by the commonalities in our projects and our experiences in preserving public media, and that working together in this effort has improved the project for everyone involved.

4. Minnesota Public Radio, Kate McManus

Minnesota Public Radio (MPR) began as a small college radio in Collegeville, Minnesota as a classical music station in 1967 that grew to a giant entity much respected in Minnesota. MPR pushes out an impressive amount of content through its three corner stone radio stations, and more recently through its podcast networks. While the content is managed pretty well in our internally built DAMS, Eddy, it is sometimes messy and certainly not easy to make stories public through the current archive portal.

My project has several discrete goals, with the overall aim being to clean up the internal metadata, normalize fields, and make it easier for digital content to be pushed to the web. In the next few months, I'll be exploring metadata cleanup. Producers all use fields very differently here, and because of the nature of breaking news; we can't make any fields mandatory on the production side.

I'm really lucky that I have an archivist to work with. We also have awesome inhouse IT folks who are constantly building the databases that feed the websites. Right now, we have an internal content management site, Eddy. Eddy was the third attempt at an internal database after two proprietary digital asset management systems were burned down and raided to build what we have today. Eddy is useful for in-house content management, but there's no way right now to make Eddy accessible to outside users. Right now, the workflow to getting content from our play-to-air system (Dalet) is somewhat automated, ingested in batches. The metadata that is created on the production side stays with the audio when it lands in Eddy, which is good for the archives purposes, even if we get that metadata with typos, misspellings, or even raw notations that would not be appropriate to push out to the public.

We've been using a separate archive portal that is not fed by Eddy, but we're hoping in the new year to build a new skin over Eddy that can become a new Archive portal site. Right now I'm completing the documentation of Eddy use, and in the next few weeks, I should have a robust metadata field map, linking what's currently on the Archive portal to what we hope will become the new portal.

There are two major concerns:

1. Which fields can we hide, is there a way to push quality information to the "new portal" without having to touch every record?
2. Of the stuff that's pushed to the current archive website, can we cross that back to Eddy without losing the public facing information, populate the "public" or safe fields, and push it forward?

Today the answer is "I don't know yet," but I'm really excited to explore solutions in the coming months. While my project is very different from most of the other AAPB NDSR projects, the camaraderie and the shared resources are absolutely vital to my work. I need the feedback that I get from my fellow residents, and often the questions they bring to the cohort help me to think more deeply about my own project.

5. KBOO Community Radio, Selena Chau

KBOO Community Radio in Portland, Oregon (<http://kboo.fm>) went on the air in June of 1968 as an independent, member-supported, non-commercial, and volunteer-powered radio station. Volunteers do everything--create, host, engineer, board op, edit, produce, and broadcast their own shows and run membership drives. While KBOO is unique, it faces many of the same challenges at other public broadcast archives for preserving and providing public access to the station's audiovisual assets. KBOO's time-crunched staff support the urgency of news stories that affect the community, constantly fundraise through membership drives, and are challenged to recruit regular volunteers for work central to its mission. KBOO is further challenged by not having an archivist in the organizational structure. When moving towards the goal of providing access to audio, the foundational work of digital preservation is often not taken into consideration.

KBOO's goal is obtaining a system for making archived programs accessible to KBOO programmers, researchers, and the public. Additionally, KBOO needs help thinking about integrating the born-digital records and digitized analog items into a single search. Digital preservation frameworks, audio asset metadata management, and defining prerequisites of digitization are being scaled down to fit the staff's time, understanding, and commitment level. Working at a community radio station, I find it imperative to advocate for archives as an ongoing responsibility and incorporate values of archives and digital stewardship in the workplace.

In order to prevent issues of unknown and unorganized assets in the future, I'm encouraging change "downstream" by sharing information on filenames conventions with radio program hosts and contributing embedded metadata guidelines for audio program files for future volunteer trainings. I'm developing a modular preservation plan so that KBOO can approach a part of the plan when they know time and money is going to be available, and then execute it as determined by the overall strategy. This ensures that actions still follow the best practices defined in the digital preservation plan. Workflows and solutions underway include development of a data model that overlaps PBCore and local needs and an internal content management system that encourages group participation with the ability to export data to an integrated analog/born-digital search system.

6. Howard University Television (WHUT), Lorena Ramirez-Lopez

Howard University Television (WHUT) went on the air November 17, 1980 as WHMM (Washington Howard Multi-Media). The station's debut received considerable local and national attention as it was the first and only Public Broadcasting Station (PBS) member station licensed to and operated by a historically black university, Howard University (www.howard.edu), in Washington, DC.

The station's signal reaches over 2 million households in the greater Washington metro area. In fall 2007, WHUT also began a digital simulcast to this viewing area, making it the ninth largest media market in the United States. WHUT airs more than 3,500 hours each year of its own in-house productions of original content and reflection of the DC community as well as educational programming, national and local issues in the United States such as #BlackLivesMatter-Protect and Serve!, a multi-media campaign to raise awareness of the #BlackLivesMatter movement and highlight the importance of mutual respect and understanding among ALL individuals (WHUT, 2016).

WHUT does not have a dedicated archivist to manage their content. Instead that responsibility falls onto each staff member: editors, producers, and broadcast technicians - all are aware of the importance of backups, discoverability, and access. WHUT does have a streamlined workflow where final products and raw footage are stored and backed-up onto LTO6 tapes. We are trying to adapt the current workflow and see where more metadata (technical, descriptive, administrative, as well as archival metadata) can be inserted in order to develop a more efficient broadcast/production workflow.

The bigger task at hand is the organization of WHUT's video archive. Recall that WHUT went on air in the 1980s, and by 1990, the station was producing a live weeknight edition of its flagship series, "Evening Exchange", that has been airing for over 25 years. "Evening Exchange", "@ Howard" and "Let's Talk Education" were all series that depicted the turbulence, progressiveness, and reflection of the community during the eighties and nineties. All this history is stored away in video tapes that haven't been accessed in years. Beginning in 2011, WHUT participated in the American Archive of Public Broadcasting's project where: 5,672 assets were inventoried and uploaded to an archival management system (AMS); 1,727 additional records were delivered to WGBH; 3,945 new records were created. However, WHUT doesn't have control over the AMS, and the inventory done back in 2011 lacks the tapes' locations making it difficult to reference the title listed to where the tape is in the archive. Furthermore, while over 200 hours of footage was digitized - 150 tapes in phase one and 113 tapes in phase two - the 263 tapes do not match the number listed on the inventory spreadsheet. We are trying to cross-reference the spreadsheets while re-organizing the video archive in order to have a more centralized inventory that we can place in a database that WHUT staff will be able to use to find and share their content.

Presently, WHUT generates the same dynamic content as it did back in the eighties and nineties with current series like "The Rock Newman Show", "The Mimi Georges Show", and "ARTICO". These series continue conversations about our society and demonstrate community outreach and education with science, technology, engineering, and mathematics (STEM)

camps for preschoolers, digital media arts club (DMACS) for middle and high schoolers, and “American Graduate.” While this born-digital workflow does have everything backed-up, establishing naming conventions and organization are key steps towards making sure that episodes are findable, accessed, and archived in order to also include them in WHUT’s database.

WHUT houses and creates a rich and diverse television production of African-American experience, which we hope to preserve and access for the community and future.

7. Wisconsin Public Television, Adam Lott

WPT lies in the heart of the UW campus and has been broadcasting to the greater Wisconsin area since the early 1950’s. From the beginning, WPT has been dedicated to the creation of content which not only educates, but entertains the public. In 2004, Ann Wilkens was hired as the station’s first media archivist in an effort to preserve their 60+ years worth of content. Her efforts have greatly informed the ongoing digitization of legacy assets, laid the architecture for longterm storage of digital content, and produced metadata for thousands of assets. She continues to be WPT’s sole archivist to this day.

As a resident, I feel lucky to be placed in an institution with active preservation practices. Our station has the luxury of owning a full suite of VTRs for capturing our diverse collection of tapes, a fully staffed engineering department which maintains the upkeep of these tools, and a knowledgeable IT department that has developed numerous micro-services which support our preservation practices. However, despite our good fortune, the archive is still working to map itself onto a production heavy environment.

Currently, our largest bottleneck concerns the creation and management of metadata. Our assets are described across several databases and spreadsheets which must be cross-referenced throughout our workflow. Microprocesses which push our content onto LTO have become a double-edged sword, as we are given a time limit to accurately describe assets, lest we want to restore them from backup. Luckily, plans are in place to adopt the Archival Management System, a metadata management tool jointly created by AVPreserve and WGBH. The system streamlines the creation of PBCore compliant XML documents, assists in the planning of digitization projects, and provides a clean interface for inputting data. The “big crunch” has already begun as we prep our metadata for large-scale ingest.

My eventual hope is that the production and archival workflows follow the same preservation streams. All content at WPT eventually finds its way onto LTO tape, so it would be ideal to treat all content as archival content right from the get-go. We’ve had success in the past pushing archival content onto social media, and I would love to continue this practice. We are well on our way to revamping our archive, which will hopefully lead to great changes within the station at large.

8. WYSO, Tressa Graves

WYSO started as a college radio station for Antioch College in 1958. Currently, the college still holds the license to the station but WYSO has expanded to becoming the public radio voice for southwest Ohio. With a small staff like KBOO, WYSO relies heavily on its volunteers for membership drives, community events, programming, and other station related tasks. IT services are provided by the college.

The WYSO Archive was created in 2009 and consists of analog audio recordings, digital audio recordings, and other WYSO related materials. The station does not currently have an archivist but hopes to hire someone in the future. All current and past archival work has been done through volunteering, internships, and through my current residency.

The bulk of my residency responsibilities are making more WYSO audio accessible to the public. This consists of a born digital civil rights oral history project and audiotape digitized

through previous AAPB grants. WYSO has a partnership with the Greene County Public Library (the station's local county library) allowing the station to put materials online (<http://www.greenecountyroom.info/cdm/landingpage/collection/WYSOProgram>). Currently there is approximately 35 hours of local civil rights oral histories and 77 hours of digitized material that require metadata to be completed before they can go online. The oral histories also require transcripts to be created.

Currently the station does not have a digital asset management system but rather has hard drives of material duplicated and located around the station and in staffs' homes. While it would be wonderful for WYSO to upgrade to a digital asset management system so the staff could easily access previously aired broadcasts, the implementation of such software is not part of my residency goals (though suggestions of possible software is). The station first needs institution wide file naming conventions that can be used on all future created materials. Current workflows and digital preservation standards also need to be examined so they can improved. These are two important tasks that I also plan to complete during my residency at WYSO.

9. References

- Phillips, Megan, Jefferson Bailey, Andrea Goethals, and Trevor Owens. (2013). "The NDSA Levels of Digital Preservation: An Explanation and Uses." http://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf. Date Accessed: 2016-12-28.
- PREMIS Editorial Committee. (2008). "PREMIS data dictionary for preservation metadata," Library of Congress. <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>. Date Accessed: 2016-12-28.
- WHUT (2016). "#BlackLivesMatter-Protect and Serve! A WHUT Multimedia Campaign." WHUT Howard University, July 14, 2016. <http://www.whut.org/whut/blacklivesmatter/> Date Accessed: 2016-11-14.