National Digital Stewardship Residency | New York
Final Report
Mary Kidd

Host: New York Public Radio

A Digital Preservation Roadmap for Public Media Institutions

Mary Kidd
NDSR Final Report
Submitted on June 13, 2016

Description of the Project

1) Overview: Provide a very brief overview describing the need, problem, or challenge addressed by your project, the audience served, your project design, your intended results, and the extent to which you achieved your intended results.

NYPR produces two main types of programming: traditional "terrestrial" content distributed over the air, and digitally distributed content such as podcasts, video and streaming media. The ways which digital files flow through the various content management and archival systems to long-term storage vary, and depend upon department, content type and/or individual staff. Current workflows must be updated to standardize the submission and ingest processes into the digital repository. Overall, The Archive is looking for solutions and strategies to enhance their role preserving and making NYPR's digital legacy accessible and discoverable.

The Archive's aim in applying to become an NDSR host was not only to help better understand and improve upon processes but to also put forth the NDSR resident's findings publically so that other media archivists could use it as a framework or guide for their own digital archiving endeavors. The resident, in attending media and technological conferences throughout the country, and by conferring with current/past NDSR residents who worked on similar projects, discovered that digital stewardship within a "live" production environment poses special and unique challenges to an archive.

The resident researched the current state of digital production at NYPR in a variety of ways:
- First, mentors provided the resident documentation detailing how The Archive migrated from a Microsoft Access database to the current PBCore-backed catalog. These documents were reviewed closely by the resident while being trained to use the web CMS, CAVAFY and DAVID.
- Interviews were scheduled with various content creators at NYPR, as well as key leaders throughout national media archives. The interviews were recorded, and in most cases hand-transcribed; the transcriptions were shared with The Archive for future reference. Here, the resident gathered information about how people used NYPR's web content management system (CMS), the PBCore-backed catalog (CAVAFY) and the production software/archival repository (DAVID).
- Data from the interviews was also input into a Visio workflow chart to show how various content types flowed from creation through various production/archival systems.
- The Archive purchased a license to TreeSize Professional, which was used to perform an MD5-based duplication scan throughout all folders in the DAVID production system. The resident also set up the program to perform a nightly scan across the DAVID folders to measure file growth over time, and export data in an Excel spreadsheet.
- The research conducted was aggregated into a final "Digital Preservation Roadmap Report" submitted to The Archive in May 2016.

One of the resident's major findings was that The Archive lacks a digital file collection policy document. The Roadmap Report, first and foremost, recommends that The Archive create a

digital collection policy document and work actively with various production departments to ensure that its tenants are reflected throughout the various systems currently in place.

In addition to the written report, the resident also prepared a presentation tailored specifically to the Broadcast Engineering Department. The intent of this talk was to discuss findings and recommendations that required the Broadcast Engineering Department's help, involvement and support. In addition, the presentation aimed to give visibility to the NDSR project's findings, and take the first step forward towards garnering buy-in for its recommendations.

2) Description of Project Partners: List the names and describe the roles and contributions of any project partners as identified in your project proposal.

The following individuals were interviewed in order for the resident to gain qualitative data regarding systems usage. Partners outside of NYPR were consulted in order to understand how other media archives set up and advocated for current digital archival workflows/systems.

Name	Role	Department/Show	Organization
Megan Ryan	Senior Producer	Brian Lehrer Show	NYPR
Vickie Allen	Director, Media Operations & Media Library	Media Operations	PBS
Marine Boudeau	Director, Product Development and Audience Experience		NYPR
Valentina Powers	Director of Digital Operations		NYPR
David Mc Lean	Video Producer	The Greene Space	NYPR
Glenn Mohre	Interactive Media		NYPR
Kasia Mychajlowycz	Producer	Freakonomics	NYPR
Arianne Wack	Multimedia Production Assistant	Radiolab	NYPR
Richard Yeh	Producer	News	NYPR
Amy Pearl	Producer	Interactive Content	NYPR
Jennifer Hsu	Web Producer	Digital Media	NYPR
Gregg Gasperino	Associate Director of Broadcast Technology	Broadcast Technology	NYPR
David Satkowski	Digital Audio Network Engineer	Broadcast Technology	NYPR
Rob Christiansen	Digital Audio Engineer	Broadcast Technology	NYPR
Andy Lanset	Director of Archives	Archives	NYPR
John Passmore	Archive Manager	Archives	NYPR
Margaret Bresnahan	Archive Manager?	Digital Archive	MPR
Richard Hake	Host	News	NYPR
Cindy Rodriguez	Reporter II	News	NYPR

Dan Jeselsohn	Content Distribution Administrator	Digital	NYPR
Josh Shepperd	Assistant Media Professor at Catholic University, RPTF Chair	N/A	Catholic University/RPTF
Laura Soto-Barra	Chief Librarian	Research, Archives and Data Team	NPR
Camille Salas	Artemis Product Owner	Research, Archives and Data Team	NPR
Sarah Knight	Taxonomist	Research, Archives and Data Team	NPR

Project Execution

1) List the activities completed during the project and compare them to the activities you set out to complete for the project. Whenever possible, describe activities in both quantitative and qualitative terms. If you did not complete all the activities, explain why.

In November 2015 the resident listed tasks/activities she set out to complete as part of her NDSR Interim report. The following original list is provided below, outlining tasks (in bold) and results.

Basic training in DAVID, Cavafy and the CMS.
Both Andy Lanset and John Passmore trained the resident to use these systems within the first 2 months of the residency.

Understand the basic principles of digital audio
- Completed an online Lynda learning course covering basic digital audio principles (~7 hours online course time).
- Training on physical media to digital transfers.
    - Analog to digital audio transfer training:
        - Set up a reel-to-reel tape on a Studer machine onto an uptake reel.
        - Splice, and adhere leader tape.
        - Adjust the Studer machine azimuth.
        - Set up decks to transfer audio from tape machine into ProTools.
        - Catalog and ingest digital file into DAVID.
        - Additional azimuth adjustment training was received during the ARSC Conference pre-conference workshop, where additional experience was gained on using a digital scope.
    - Removal digital media transfers:
        - CD-R:
            - Use the Ripstation machine to bulk rip over 4 dozen CD-Rs from the collection.
            - Produce and analyzing error correction reports using the Plextor drive and software for each CD-R ripped.

- Ripping Minidiscs using Sound Studio and importing into DAVID
- Use SQL to bulk rename files
- Use the command line to convert WAV metadata using an XSLT stylesheet into a PBCore compliant XML file to be ingest into CAVAFY.
  - MiniDisc
    - Basic usage of Sony MiniDisc hardware and SonicStage software.
    - Take basic course on the Python programming language.
    - Install and use ffmpeg to bulk convert and rename ATRAC data files.
- Hard drive cloning of hard drives submitted to The Archive by video producers.

Basic training in mySQL queries, the command line, and XSL.
- John Passmore trained the resident to access and use the phpMyAdmin application and perform basic SQL queries.
- Further SQL training was procured by the resident by watching and completing Lynda online course exercises.

Perform interviews of those staff the the mentor has identified as key users of DAVID, Cavafy and/or the CMS.

Twenty-four staff, including 6 individuals from organizations outside of NYPR were interviewed by the NDSR resident. The interviews were recorded and hand-transcribed by the resident and submitted to The Archive in a shared Google Drive folder for future reference. All recorded interviews, with the exception of those interviewees that wished to not be recorded, were transcribed in full by the resident.

The resident would have preferred to have performed more interviews over more internal staff, but was constrained by available time. Several interviews scheduled were cancelled and never rescheduled, due to both the interviewees time constraints, and additional scheduling conflicts posed by the NDSR's conference travel plans.

Use information gathered from interviews to build up a workflow chart that illustrates the "flow" of data, from production to storage.

The resident received access to Microsoft Visio and produced a completed workflow diagram outlining various broadcast flows. The diagram was shared with mentors at the end of the residency.

Export information gathered from TreeSize to analyze and identify duplication amount and behavior across the various DAVID folders, and assess file size/volume and rate of growth across folders and production units.

The resident completed a duplication MD5 checksum analysis across all folders in $media using TreeSize and Excel in order to pinpoint those folders in DAVID that have extensive duplication issues.

In addition to a duplication report, the resident also set up TreeSize to perform nightly scans over the files in the DAVID media folder. The data from these reports was aggregated into a spreadsheet which was used to analyze specific anomalies that arose charting growth of files (there were sudden peaks and drop-offs between February and April 2016). The anomalies proved useful for the Roadmap Reports overall recommendations, demonstrating the absolute need for The Archive and/or the Broadcast Engineering Department to perform file fixity and MD5 checksum scans and comparison before and after massive file migrations between folders in DAVID.

Complete a DAVID folder map
A shared Google Document prepared by the resident lists each DAVID media folder, charting purge cycles, behaviors, and size. The file was shared with The Archive at the end of the residency for their future reference.

2) Did the project need to be altered from the original project that was proposed? If so, explain why.

There were two differences between the original project proposal and the final project.
1. An additional deliverable was added at the end of the residency: prepare an hour-long presentation for Broadcast Engineers at the end of the residency to garner their support or "buy-in" of some of the recommendations put forth by the resident in the Roadmap Report. This additional task was proposed by the resident to The Archive.
2. Transcription of interviews took an extraordinary amount of the resident's time between months 1-6. Although automatic transcription services were explored, the transcriptions were generally inaccurate and cumbersome to correct.

3) Significant accomplishments in terms of project work.

● The resident connected with key leaders in national media archives work and scholarship in a series of phone and email interviews to support the project's research. This includes speaking with three archival staff at National Public Radio (NPR) Archive, an archivist at Minnesota Public Radio (MPR), a former mentor at Public Broadcasting Service (PBS) and Radio Preservation Task Force board member.
● All 20 internal interviews were hand-transcribed, and copies of these transcriptions were provided to archival staff at NYPR.
● A complete DAVID media folder list was provided to The Archive, which lists each of the 104 folders, its main function, whether or not the folder was setup to purge or move files (and if so, the rate) as well as whether or not the folder was setup to bitrate convert files.
● A Visio workflow chart that illustrates various workflows from creation (focusing on news, podcast producers, external producers, video producers and archivists) to dissemination on air or on a NYPR website.

- The resident designed and distributed a survey to all internal NYPR staff by email. With a bit of encouragement through incentive (the resident had an artist paint a realistic portrait of an archival tape as a reward to a survey taker drawn at random) 103 respondents took the survey. The survey revealed a number of perceptions that were used to justify recommendations put forth in the Roadmap Report. Namely, it revealed that staff preferred to use the web CMS to search archival content, rather than the PBCore-backed catalog CAVAFY. It also revealed confusion over how files are ingest into the archival repository in DAVID.

4) Outreach or dissemination activities (conferences, etc.)

The resident attended and participated in several national audio/visual, sound, preservation and technology conferences:
- Attended the Associated for Moving Image Archivists (AMIA) conference in Portland, OR in November, 2015.
- Attended and spoke on the Metadata Committee for the Radio Preservation Task Force (RPTF) conference in Washington, DC in February, 2016 and received a speaker's honorarium.
- Attended the code4lib conference in Philadelphia, PA in March, 2016. Attended a pre-conference workshop to build a Ruby on Rails application.
- Received a travel grant in the amount of $750 to attend the ARSC pre-conference and conference in Bloomington, IN. In addition to this, the resident was paired up with an ARSC mentor who introduced her to long-standing members.

Analysis and Evaluation

1) Project Results: Analyze your entire project. Compare the actual accomplishments of the project with the intended results, outcomes, and impacts you identified in your proposal. Use quantitative data as well as qualitative examples, highlights from your evaluation, and compelling anecdotes. Identify and document significant project achievements and their value. Describe any significant unanticipated costs or circumstances that created obstacles to project success, and summarize lessons learned during the course of the project. For this report, outcomes are changes in individuals' knowledge, skills, attitudes, behaviors, or other conditions related to the purpose of the project. Impact is a large-scale and/or long-term result that affects one or more institutions, communities, or fields. Provide images, screenshots, maps, floor plans, charts, tables, publications, or datasets as appropriate to illustrate your project's achievements during the reporting period.

No advocacy without a policy
The interview data gathered from the transcribed interviews conducted by the resident suggest that most staff have little understanding how or why The Archive manages born-digital content. This has, for example, caused management in some departments to assign some staff archival tasks and responsibilities such as organizing raw audio files into folders on their desktop computers instead of consulting and enlisting the expertise of The Archive. This suggests that unique archival content may be resting outside the purview of The Archive due to the siloing of certain digital archiving tasks.

The interview data also suggests that most staff lack basic knowledge about digital archiving-specific issues, which will inevitably cause discoverability issues down the line. For example, since there are no archival submission standards in place, and no automatic task to check against any standards, a producer may submit a file to the DAVID archival repository without any descriptive metadata, such as a record title. This poses an enormous risk to records becoming lost/unfindable due to decontextualization. Additional outreach efforts will be required to educate internal staff about their role in supporting long-term preservation.

To address these issues head-on, the resident put forth a recommendation in the Roadmap Report that The Archive draft a born-digital preservation and collection policy made readily accessible to all internal staff, such as through an internal Wiki. The policy should detail the following:
- Their overall goal and mission of The Archive towards born-digital collections.
- What digital files/formats The Archive will accept into the archival repository, and which ones it will not accept.
- Minimal metadata standards for files/records submitted into the archival repository.
- Which departments or staff are crucial to upholding their the policies put forth by this document.

The Archive can use their new policy as an occasion to inspire, convince and mobilize organizational support to improve system functionality that supports the preservation of born-digital content. The creation of a new policy can also serve as an occasion to rebrand The Archive as both NYPR's physical and digital archive to the greater organization. For example, once the policy is complete, The Archive can host brown bag sessions outlining what they are putting forth, or convince Human Resources that mandatory digital archival training be necessary for all new hires.

The inspiration for this recommendation comes from the *From Theory to Action: "Good Enough" Digital Preservation Solutions for Under-Resourced Cultural Heritage Institutions* White Paper written in 2014 by the Digital POWRR Team for Institute of Museum and Library Services. This paper puts forth practical solutions that under-resourced cultural heritage institutions can put forth to implement digital archiving solutions, stressing that advocacy is "integral to digital preservation success".

Discoverability and Access
An interesting finding gathered throughout the survey distributed by the resident of over 103 respondents on systems usage qualify the idea that internal staff either require further training on archival tools available to them, or new tools entirely. For example, only 5% of respondents indicated they use CAVAFY to search for archival content. CAVAFY, a PBCore-backed web-based catalog containing metadata describing both The Archive's physical and born-digital holdings, is perhaps the most comprehensive catalog of all The Archive's holdings.

Count of Internal Staff - Tools and Resources

Total
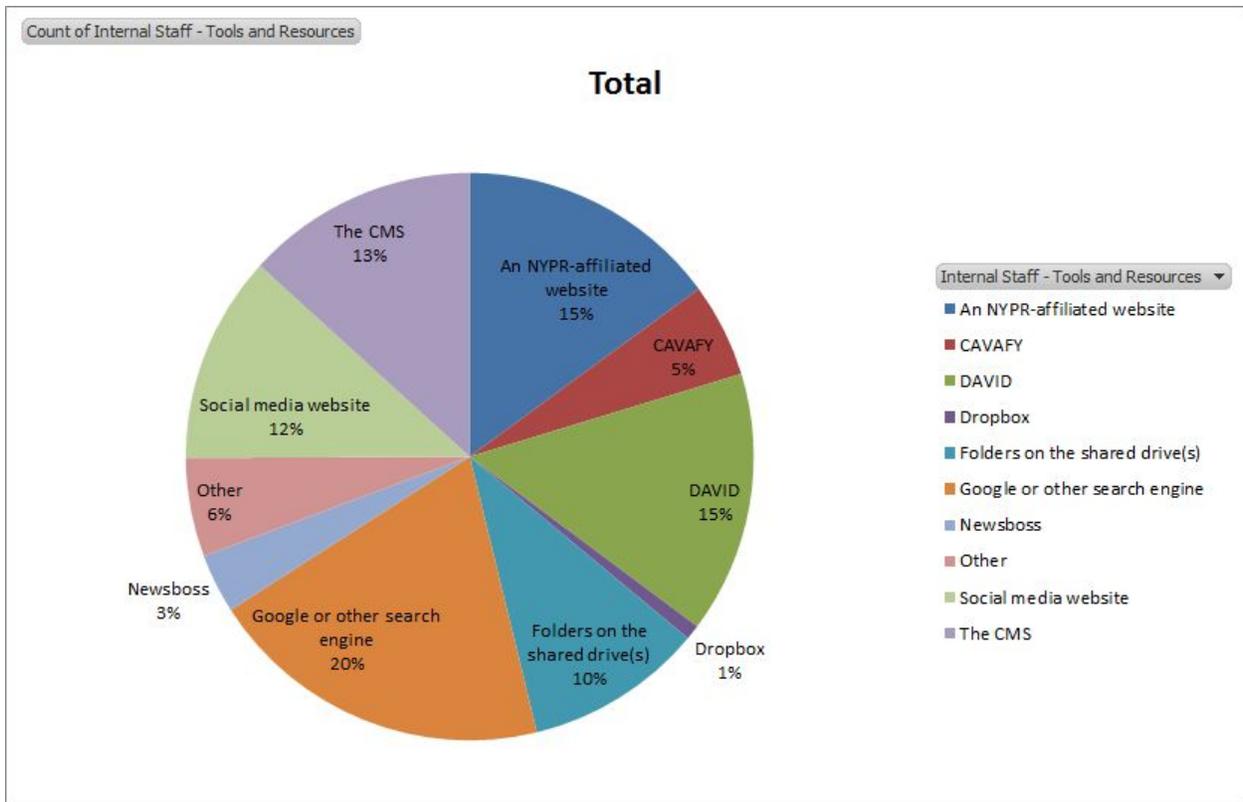
The majority of users preferred to use general search engines (20%), social media websites (12%) and NYPR-affiliated websites (15%) to perform their searches, even though these tools contain only a fraction of NYPR's legacy content. These results also makes a compelling argument that the implementation of a new digital asset management management tool is necessary to increase usage of archival tools, and allow internal staff to delve deeper into The Archive's holdings.

Test driving The Roadmap
Near the end of the resident's tenure, members of the Broadcast Engineering Department were invited a final presentation prepared by the resident. The ultimate goal of the final presentation was to share some of the Roadmap Report's findings and recommendations, focusing on those recommendations that would require Broadcast Engineering's support to establish and sustain over the long-term. After the presentation, a forum was opened to gather reactions, input and critiques from Broadcast Engineers.

One of the resident's recommendations was to alter DAVID so that airchecks are auto-cataloged and moved into the archival repository, instead of relying on scripts and manual triage of the day's finished broadcast audio files. NYPR creates airchecks by making 1:1 copies of broadcast-quality audio files which are copied into a folder in DAVID. The implementation of an aircheck to repository system would require the Broadcast Engineering Department's support, since automation would need to be built out in DAVID to support this new process.

The aircheck recommendation was received positively by Engineers, who not only agreed that such a system could be put into place, but also put forth additional suggestions for its successful

implementation after the presentation concluded. For example, at the moment, most airchecks are copied into DAVID into hour-long file blocks with little/no descriptive metadata describing how those blocks are broken down by show, host or start/end times. Broadcast Engineers confirmed that there are ways that DAVID could be programmed to break down each aircheck down to individual segments, and that each segment could be auto-cataloged with descriptive metadata to The Archive's specifications. The implementation of this solution could have a significant impact on reducing the time archivists spend vetting items into their repository. In addition to this benefit, auto-cataloging will negate the need for content creators to have to input descriptive metadata manually.

Although the final presentation was not written in directly into the project's deliverables, it had a profound effect of advocating for The Archives' concerns and needs and enlisting initial support of their efforts. It is the hope of the resident that The Archive may continue these sorts of advocacy efforts to foster dialogue between itself and other departments in order to garner institutional support for its digital stewardship efforts.

A copy of my presentation (along with notes) presented to Engineers can be found here (links out to a shared Google Document).

https://docs.google.com/presentation/d/1g_-8qvvat0fcjJ5YW3OKALZQ1mf7NyRWn5EFpOzGL1o/edit?usp=sharing

Lessons Learned: Transcriptions
The resident manually transcribed recorded interviews. This proved to be extraordinarily time-consuming. However, the importance of complete transcriptions proved extremely useful for narrating complex systems in the final Roadmap Report. Future hosts who plan to factor in staff interviews as an integral project task should consider whether or not transcriptions will be useful for the host after the project, and if they are deemed to be so, additional time should be factored into the project proposal timeline. In addition to this, the host should consider purchasing a license to auto-transcription software or application, as well as secure high-quality voice recording tools prior to the start of the residency.

2) What's next? Describe any plans to continue work in this area and/or to sustain the benefit of this project beyond this grant's end date.

The resident interviewed RPTF Board Member and conference planner Josh Shepperd, who later invited the resident to serve in an administrative role as a liaison between the RPTF and Association for Recorded Sound Collections (ARSC) group for their metadata aggregation project at the conclusion of the residency. The resident hopes to continue her work in radio and media preservation advocacy and efforts going forward.

3) Products: Attach one copy of any product that resulted from grant-funded activities, including final evaluation reports and instruments; research findings, publications, or manuscripts; learning resources; and other deliverables. For web-based material, provide URLs and sample screenshots.

Attached are the following documents:

- The Digital Preservation Roadmap Report submitted to the NYPR Archive in May 2016 is attached.
- The Visio flowchart in PDF format.
- Google survey results of internal staff systems usage.
- The final presentation to Broadcast Engineers prepared and presented by the resident.

Professional Development

1) Provide an overview of what you accomplished during the 20% of time devoted to professional development. What were your goals? What significant accomplishments did you achieve? What difficulties did you encounter?

The resident hoped to enhance her skills handling optical media; namely, CD Rewritables and MiniDiscs, which comprise a significant portion of The Archive's collection. The resident hoped to experiment with these objects since and discuss findings during scheduled blog posts and the {Let's Get Digital} NDSR Sympsium in order to disseminate her research to the greater preservation and archives community.

Significant accomplishments include:
- Experience using a batch CD-RW data extractor machine called a "Ripstation";
- Performing extraction error reports using a Plextor CD-RW drive and software and understanding various error reports;
- Using SQL commands to rename batches of files;
- Using XSLT to convert WAV metadata into a PBCore XML file that could be ingest into CAVAFY;
- Installing and using ffmpeg, git, python and libraries;
- Became familiar with the composition of various optical media formats and understanding the mechanics of writing/playback;
- Composing command line commands to combine/rename files;
- Using ffmpeg to batch convert audio files;
- Became familiar with the history and development of MiniDisc technology and software;
- Became aware of critical issues surrounding proprietary software and digital preservation.

2) List activities and brief descriptions of what you completed that fulfilled the following 20% time requirements:

Individual: On your own, identify, organize, and deliver/present/complete two or more activities; at least one of these activities should be organized for your Host; additional activities could be organized for other Hosts, Residents, and/or external participants.

The resident's 20% time was spent experimenting with MiniDisc hardware and software. Namely, the resident explored an alternative method to using Sony's proprietary software, SonicStage, to transfer ATRAC data from a disc in a MiniDisc player to the computer over a USB connection. The resident researched and discovered netmdpython, an open-source solution developed to provide a workaround to the audio/data firewall built into MiniDisc players, encoding and software. Using netmdpython required installing ffmpeg and the lib-usb library onto The Archive's Macintosh laptop. The resulting experiments proved successful: ATRAC

data was downloaded from The Archive's MiniDiscs onto the hard drive in *.aea format: ffmpeg commands were utilized to further convert the *.aea files into *.wav files. The overall experiment allowed the resident a way to gain experience using the command line, ffmpeg and github as well as gain a deeper knowledge to the physical composition of optical media.

Group: With one or more other Residents and/or one or more other Hosts, identify, organize, and deliver/present/complete at least one activity.

The resident, along with the other four residents, helped plan the {Let's Get Digital} Symposium. The symposium was co-sponsored by the Archivist's Roundtable (A.R.T.). The resident completed the following symposium-related tasks/responsibilities:
- The resident was the main point person between NDSR and Rachel Harrison of A.R.T.'s Education Committee, and Julie May of the Brooklyn Historical Society. Met in-person with both parties throughout the time leading up to the symposium.
- Secured a donation by A.R.T. to host an additional post-symposium cocktail hour and provide alcoholic beverages free of charge.
- Maintained the NDSR information page on the NDSR Wordpress website, and updated it as needed.
- Assisted with decision-making around catering and professional sound/presentation equipment.

External: Identify, organize, and deliver/present/complete at least one activity that involves an external audience that is relevant to your project and coordinated with your Host.

The resident prepared a presentation and question/answer session for the {Let's Get Digital} Symposium that detailed her MiniDisc and netmdpython experiments completed throughout the residency. The resulting slides and other experiment's artifacts were made publically available through the NDSR github.

Professional: Participate in at least one local or regional professional activity, not including ones that are organized by program staff for all Residents to attend.

Accompanied by her mentor, the resident participated in New York Public Library's Community Oral History test run of their newly-developed Transcription Editor. Here, the resident sat with a developer who observed her using the tool to listen to and edit a time-coded transcript. The resident provided verbal feedback regarding ease of use, and suggested improvements.

3) List the conferences, workshops, memberships etc. that you selected to spend your Professional Development funds on. Describe the impact your activities had on your NDSR project, career, skillset, etc.

Memberships:
- SAA Professional Membership
- AMIA Professional Membership
- ARSC PRofessional Membership

Conferences:

- Digital POWRR Workshop in Albany, NY, 2015
  - Impact: Attended an all-day workshop discussing the NDSA Levels of Preservation document, the 3-3-3 approach to advocating for digital preservation at an institution, and received hands-on training and experience using Data Accessioner.
- Association for Moving Imaging Archiving (AMIA) Conference in Portland, OR, 2015
  - Impact: Introduced the resident to presentations addressing current media archiving practices as well as to a well-formed community addressing production-based archival processes and workflows.
- Radio Preservation Task Force (RPTF) Conference in Washington, DC, 2016
  - Impact: Introduced to a community of scholars, academics and radio broadcasters/enthusiasts to issues specific to radio and podcast preservation. The resident, along with her mentor, were selected to speak and participate in the Metadata and Digital Archiving Committee. The resident will continue in a volunteer-based administrative position serving the RPTF board on an metadata aggregation project.
- Code4lib Conference (Philadelphia, PA), 2016
  - Impact: The resident learned about how cultural heritage institutions throughout the country are using scripting and programming languages to promote greater accessibility and discoverability of digital content. A highlight includes a pre-conference workshop learning to set up and build a Ruby on Rails application.
- Association for Recorded Sound Collections (ARSC) Conference in Bloomington, IN, 2016
  - Impact: Received a travel stipend to attend the pre-conference and conference. During the pre-conference, the resident learned hands-on skills in cable soldering, Student machine azimuth adjustments and scope readings, and audio/visual transfer of a video tape. The resident's conference mentor introduced her to long-standing ARSC members. Additional presentations attended allowed the resident to learn about grassroots audio archiving and digital audio preservation initiatives taking place across the country and internationally.