



NDSR Project: A Digital Preservation Roadmap for Public Media Institutions

---

*Goal Summary*

New York Public Radio (NYPR) seeks an NDSR resident to aid in the creation of a robust digital preservation roadmap for its digital assets. This project will include a detailed investigation of the current landscape of the organization's digital collections and the formulation of recommendations for long-term, institution-wide digital curation policies. Our goal is to lay out a trajectory leading to a seamless and integrated approach that will leverage in-house resources, but also think of new and creative ways to capture the digital history and legacy of NYPR. As one of the nation's largest public media companies, our hope too is that the resident will produce an adaptable framework that will assist other public media companies in the creation of more comprehensive digital preservation strategies.

*Specific Objectives*

The resident would be expected to deliver a series of reports that would outline the following.

1. An assessment of the current state of digital production at NYPR (file formats, amount of materials being produced, where the materials are located, who is producing digital material). During the course of this assessment, the resident will not be expected to include an inventory of producer's personal desktop hard drives, or digital files located on non-networked storage sources. Rather, this assessment will be based on an examination of our existing institutional on-line and near-line digital storage solutions, which are managed by both the Engineering and Archive Departments.
2. A description of the various kinds of structural, descriptive, and administrative metadata being generated within the stations various departments. This portion of the project is meant to help the archives pinpoint what data is being generated by the station's various software production systems, but also in what format the information is encapsulated (XML, text, RSS, .xls). From this initial stage of discovery, the resident can make recommendations to the archives as to what metadata is important for future metadata mappings or formulate recommendations for a new metadata schema for the archives.
3. An outline of NYPR's current digital storage and production strategies. This

outline could take the form of a workflow chart or diagram, with accompanying explanation, that tracks the lifecycle of a digital object, from its creation on a networked SAN, to its storage on a station NAS, to its eventual life as a streaming access copy. It is this portion of the project where the resident will have the opportunity to contextualize digital object life within the rubric of an OAIS model and to reflect on the potential future uses and public access to the digital collections.

4. This last report will be based on a series of objectives completed in the prior three goals. Depending on the outcome of the previous investigations, the last report will include a series of recommendations that will help the archives bolster its digital stewardship strategies. The recommendations could take various forms. For example, the resident might decide that only small tweaks to the systems are needed, and that the archives can find simpler or better ways to leverage existing workflows to ensure better long term preservation. Or the resident could recommend that producers need to change the way they capture, name, and store digital information. The resident would not necessarily need to investigate new potential digital assessment systems. Although in this rapidly changing digital environment, NYPR remains open to suggestions as current systems and best practices are evolving. Depending on the skill and experience of the resident, this kind of research may be out of the scope of the residency. However, if the resident believes that entirely new systems are required at NYPR to ensure long term preservation, they can certainly point NYPR in that direction and suggest areas of future discussion.

*Timeframe &  
Deliverables*

1-2 Months  
Audio Training and Workflow Orientation

Because of the complexity of systems and workflows involved with media creation at NYPR, the resident will spend the first two months familiarizing themselves with the born-digital workflows and digitization best practices of the station's archives department. This training will include learning the analog-to-digital workflow for a variety of audio formats in the collection, including ¼" open reel tape, lacquer transcription discs, and cassette. The resident will also be trained in proper transfer of digital assets from their original carriers into appropriate file-based preservation masters. This process will include working with minidisc, DAT, CD-DA, F-1 Beta, and VHS DBX-700. The resident would not be required to process a certain amount of materials during this training. Rather, the archives feels that learning the fundamentals of audio preservation will help the resident create more informed decisions during the next stages of the residency.

3-6 Months  
Collection Assessment

The goal of next three months of the residency will be to gain a detailed understanding of the digital assets the station creates. This will be accomplished by creating a collection assessment document, which will include the kinds of digital assets producers are creating (file formats, codecs, software tools) and the total volume in terms of file size.

This assessment will require interviewing stakeholders from across departments, including IT, Engineering, Digital Operations, News, and the various show production departments within NYPR. The mentor will help facilitate these interviews and make the appropriate introductions. There is a shared interest amongst archives and engineering to codify and streamline many of NYPR's current digital preservation strategies, and these departments would welcome any fact-finding endeavors by the resident. Additionally, NYPR has recently hired a systems administrator, dedicated solely to maintaining stable and reliable audiovisual streaming capabilities. The administrator has expressed interest in working with the archives to formulate a better more efficient preservation workflow, possibly leveraging some of the solutions in NYPR's new audio tools server, which he manages.

There have been attempts in the past to create collection assessments for the station, and this part of the residency will also include a review of these past assessments in order to see what information needs to be changed, added to, or updated. Some of the ideas in past assessments have taken root in various departments. For example, there has been some progress in preserving born-digital video, but because of the rapid changes in technology many of the reports are now out of date: software systems have changed and encoding standards have shifted. NYPR has also greatly expanded its Digital Operations Department to include programmers, systems administrators, digital producers, and engineers – all of whom have had an effect on content creation and production workflows.

At the end of month 6, the resident should have completed a comprehensive report that details the state of digital preservation at NYPR.

#### 7-9 Months Recommendations and Final Reporting

In the last two months, the resident will have the opportunity to make recommendations based on their previous work. This time will be spent identifying the things the archives are doing right, but also identifying the key roadblocks and challenges the archives faces to digital preservation. A final report will be due at the end of this term, and it should include recommendations for the implementation of new digital preservation workflows. These recommendations could take a number of different forms. They could be used to evaluate available open-source or commercial digital repository solutions, but the time could also be spent formulating new ways the archives could use its current digital infrastructure.

#### *Resources Required*

Two mentors (John Passmore, Andy Lanset) and one resident. The archives will provide a dedicate workstation and the resident will have access to stakeholders throughout the institution including members of the Engineering, IT, Digital Operations, and Production departments.

The mentors will schedule weekly meetings with the resident and facilitate the interviews with stake holders in other departments. The mentors anticipate near constant contact throughout the residency, but certainly daily communication during the first 3 months of training and orientation. The archives workspace is a large communal audiovisual laboratory, so the residency will be in the center of the archives activity during the course of the project. John Passmore will be able to train the resident in basic programming skills, how to use the command line, and how to run simple queries in our various databases.

#### *Context*

The NYPR archives was founded in 2000 by Andy Lanset, but the broadcast radio stations that comprise NYPR have a history that reaches as far back as the earliest days of radio. For example, WNYC, NYPR's largest and most popular station, first broadcast on July 8th, 1924. Throughout its more than 90-year history, WNYC, New York Public Radio, has faithfully observed and documented the political, historical, scientific, and cultural events—both large and small—that have shaped New York City and the entire nation. Outside of the federal government, the WNYC Collection is the largest non-commercial collection of archival audio recordings and ephemera from an individual radio broadcaster.

The NYPR archives has spent the last 15 years repatriating, digitizing, and preserving parts of our historically older collections, but at the same time, we realize that our digital collections deserve equal care and consideration.

The archives current system of archiving digital objects is reliant on the use of disparate and largely unconnected hardware and software solutions within the station: DAVID Systems is used as the digital asset management system for the archives preservation audio files, both born-digital and digital derivatives of analog objects. The descriptive metadata for the objects stored in DAVID are captured in a separate open source PBCore-backed content management system. Various other archival digital objects (video, pdfs, photos, scanned objects) are cataloged and kept track of in a variety of ways, but primarily with Microsoft Access databases and Excel spreadsheets. Because of this workflow, the archives see a need to develop a more integrated solution – one that specifically meets the needs of a growing digital repository.

The NYPR Archives also realizes that shoehorning archival workflows into what are essentially radio production-oriented systems, is becoming inefficient. For example, DAVID Systems strips archives- embedded metadata from audio files upon upload. It also renames uploaded archival audio files to adhere to its own content model structure. In the case of digital video, DAVID supports limit file formats and codecs.

File storage is becoming an issue for the NYPR archives as well. The growing number of archives' digital preservation files stored on local servers and managed through the DAVID interface by the archives is becoming a prohibitive expense, and has the potential to slow down the system as a whole, affecting the needs of producers and reporters, and other employees at the station.

Having a NDSR resident that can dive deeply into NYPR's current digital preservation issues and make sound recommendations for a sustainable digital preservation strategy would be invaluable.

*Required Knowledge and Skills for Resident*

- MILS or degree in equivalent archives and preservation program.
- Ability to run complex MYSQL queries
- Reading and understanding XML
- Using XSLT to transform XML files
- BWF MetaEdit command line utility to run and validate checksums and embed metadata
- Using the MediaInfo command line utility to extract technical metadata
- Basic familiarity with operating in the Windows command line

*Preferred Knowledge or Experience*

- Advanced Microsoft Excel (Visual Basic for Applications and Open Database Connectivity)
- Scripting language like Ruby or Python
- Basic understanding of digital audio workflows and play back systems
- ProTools or Adobe Audition
- Understanding of current metadata standards to describe audiovisual materials (PBCore , AES)
- Knowledge or experience with open source digital asset management tools such as Collective Access and Resource Space.
- Familiarity and understanding of ISO 14721:2003 (OAIS)