

Preserving Digital Public Television: Is There Life After Broadcasting?

by Nan Rubin,

Project Director, *Preserving Digital Public Television*, NDIPP of the Library of Congress, USA

Television production has shifted rapidly from an analog process to one where virtually all programs are created and finalized as digital files. Such productions in public television are at great risk of being lost, because practices for long-term preservation of digital video are just now emerging, and because there is no mandate for preservation within the U.S. public broadcasting system. The National Digital Information and Infrastructure Preservation Program (NDIIPP) of the Library of Congress funded *Preserving Digital Public Television*, a partnership between WNET-TV in New York, WGBH-TV in Boston, the Public Broadcasting Service, and New York University, to build a model preservation repository for digital video files and to examine broader issues related to operating such a repository. In addition to designing the repository itself, the project became a lead advocate for adopting technical and metadata standards across the television field and is pioneering preservation-compliant digital production workflows. The project also successfully advanced the position that the U.S. public television system recognizes the necessity of preservation to keep digital productions alive. This led to national funding being allocated for the first time to create a pilot initiative with a long-term goal of properly managing the collective archival holdings of public broadcasting in the U.S.

From Analog to Digital: The Transformation of Television Production

Public Television is responsible for the production, broadcast and dissemination of programs which form the richest audiovisual source of cultural history in the United States.

(Librarian of Congress, 1997)

In the very first set of reports commissioned by the National Digital Information Infrastructure Preservation Program (NDIIPP), the Library of Congress identified the challenges of preserving digital television productions:

By nature and necessity, public broadcasting is a hodge-podge of media types and formats... In whatever manifestations these objects previously existed, they become bits and bytes before they reach the public eye. That is an enormous amount of digital information to manage over time. As we move into the increasingly complex digital world, those charged with preserving our television heritage have the opportunity to develop and establish better coordinated and standardized preservation policies and practices to ensure what television programs and related assets survive.

(Ide, MacCarn, Shepard, & Weisse, 2002)

When this was written, it was not yet evident that television broadcast operations would be altered so profoundly or so rapidly by digital production and distribution technologies. In less than a decade, broadcast television has shifted largely from a manual process based on videotape, to one that is almost entirely file-based. Virtually all programs are now shot and edited digitally, and completed programs are finalized as digital files. Television distribution and transmission have been equally transformed, as tape-based submissions to the Public Broadcasting Service (PBS) and other national program services in the U.S. are being replaced by digital file transfers, and nearly all local broadcast playback is now tapeless, with programs stored, assembled and aired as files directly from a server.

The viewer experience has shifted as well. The explosion of on-line broadcast content, coupled with a constantly changing array of viewing devices, have created a fundamentally altered video environment which requires programming to be viewable on everything from the very smallest iPod screen to giant wall-size flat panels. By summer 2009, all U.S. full-power television stations will have turned off their analog transmitters and be broadcasting exclusively on digital channels. The all-digital television chain will be complete, from program producer at the start of production to the mobile viewer at the end of the line.

What do these changes mean for television archives? In a culture that expects broadcast media to be available whenever it chooses, the notion of a video archive takes on new meaning: not as a gatekeeper to accessing older content, but rather as a guardian protecting that content and keeping it vital.

Unlike videotape, it isn't enough to close a digital file and put it on a virtual shelf. Leaving archiving to the end of the production process opens the door to a host of threats including dissociation, migration problems and obsolescence. For video in particular, acceptable practices to save and access very large files, manage ever-changing file formats, and maintain rich metadata are just now emerging. By using digital asset management systems and related tools, there is a great opportunity to create value and revenue from assets. However, it requires adopting a new approach that incorporates preservation practices into the entire digital production chain.

Bringing Digital Preservation to Public Television

The non-profit Corporation for Public Broadcasting (CPB) was created by the U.S. Congress through the Public Broadcasting Act of 1967, specifically to dispense Congressional funding to America's public broadcasters, now numbering more than 350 television stations and over 600 radio broadcasters. Annual funding to CPB provides direct grants to stations and supports system-wide needs including program funds, copyright royalty fees, and national program distribution.

Although the Act authorized CPB "to establish and maintain, or contribute to, a library and archives of noncommercial educational and cultural radio and television programs and related materials," CPB never allocated any funds to support this charge. Consequently, no mandate for system-wide preservation exists at any major U.S. public television institution.

Without funding very few formal archiving activities exist. Only a small number of producers are capable of taking on the responsibility and costs of preserving their own materials, and preservation planning is largely an afterthought. The rapid changes in digital technology are rendering recording and playback systems obsolete at breakneck speeds, and while tools for managing large and complex video files are fast evolving, they are not yet perfected. This puts digitally produced programs at great risk of being lost, and potentially leaves a large gap in saving America's public television legacy.

U.S. public television stations WNET in New York and WGBH in Boston recognized this challenge. Between them, the two stations produce roughly 60% of the national prime time public television series in the U.S. including *Frontline*, *NOVA*, *American Masters* and *Great Performances*. Moreover, because WNET and WGBH each maintain its own archives, both stations had a demonstrated commitment to long-term program preservation. Both knew that solving the demands of digital preservation would be costly and that no station could do it alone – it would require a collaborative effort.

When the Library of Congress invited proposals through NDIIPP to promote digital preservation practices, WNET and WGBH formed a partnership with the Public Broadcasting Service (PBS) to develop a model repository for digital video. PBS operates the national network that distributes public television programs to more than 300 local stations across the country, and it is the principle de facto repository for national programming. The PBS warehouse holds more than 150,000 analog tapes going back more than 50 years.

WNET, WGBH and PBS are all directly engaged in producing, distributing and archiving television programs, and they shared a recognition that public television had to take steps to begin protecting its rapidly growing collection of digital assets. Because they are primarily broadcasters, however, they had little experience developing a preservation repository.

New York University provided the expertise that was lacking. The NYU Digital Library Technology Services team had extensive experience designing repository systems specifically for transferring and preserving large, video files wrapped in rich metadata. The project further benefited from a relationship with NYU's Moving Image Archiving and Preservation Masters Degree Program, whose students have produced excellent research as part of the project and whose graduates have become full-time project staff.

Project Goals: Build and Test a Model Repository

Understandably, the priorities of public broadcasting are program production and broadcast delivery, not saving program assets. Together, WNET, WGBH, PBS and NYU organized *Preserving Digital Public Television*¹ as a collaboration to introduce digital preservation issues, practices and benefits to the public television system.

The project formally began in September 2004 and will be completed in 2009 with a main focus to develop a model preservation repository for large digital video files. Because public broadcasting as a whole has little exposure to preservation issues, the project also proposed examining issues related to content selection and appraisal, studying copyright impediments, and examining relevant financial and governance models for repository operations. An underlying goal was to promote an understanding within public broadcasting that to exploit its programming well into the future, digital preservation should be a new priority worthy of investment.

To be successful, *Preserving Digital Public Television* had to demonstrate that building a repository was technically possible, and that operating a repository was functionally and economically feasible. In this context the initial set of activities were:

- Designing a test repository for born-digital public television content.
- Developing standards for metadata, file formats, wrappers and production workflow practices.
- Drafting recommendations for appraisal and selection policies for public television content.
- Examining issues of long-term content accessibility and operational sustainability.

In planning for the NDIIPP project, the public television partners understood that identifying commonly used file formats and production protocols, determining appropriate metadata requirements, and adopting technical standards would be key. The project naively assumed that both commercial television networks and large collecting institutions like the Library of Congress, which was completing its Packard Campus of the National Audio-Visual Conservation Center, were making progress on solving these same problems and public television could benefit by joining in with work already underway.

The project quickly learned, however, that this was not the case. Other video producers, including the networks and the Library itself, were in fact struggling with the same technical issues. Instead of tagging along, the project found itself in the unanticipated position of leading the effort in the television industry to create a standard for video file wrappers, and adopting one of the first sets of metadata schema appropriate for long-term video preservation. Both of these outcomes were unexpected.

As for bringing a new consciousness to the system, our project has sparked widespread support to launch local as well as national preservation activities, and we helped prompt CPB to propose creating *The American Archive*, its first funding ever to support preservation services for U.S. public radio and television programming.

Using the OAIS Reference Model

Digital objects must undergo ongoing transformations within ever-changing technological systems to remain usable. File formats and software upgrades happen frequently, and once a digital object has been put in a repository or saved on a server, it is by no means «preserved.» Instead, digital files require duplication, auditing, migration and ongoing maintenance to remain accessible and usable. As Ken Thibodeau, Director of the Electronic Records Archives program at NARA, notes, «empirically, you cannot actually preserve an electronic record. You can only preserve the ability to reproduce the record.”²

The clear implication is that digital preservation is not merely file storage, but requires adequate infrastructure, facilities and resources to ensure that the files will be protected and remain readable over time. *Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC)*³ is the currently recognized standard in the U.S. by which to measure a repository’s capability of acting as trustworthy, sustainable, digital caretaker.⁴

Developed as a joint effort between the National Archives and Records Administration and the Research Libraries Group, TRAC provides guidelines for institutions that offer long term preservation services. The TRAC criteria follow the well-established, standardized functional requirements of a digital repository as outlined in the ISO Standard *Reference Model for an Open Archival Information System (OAIS)*.⁵

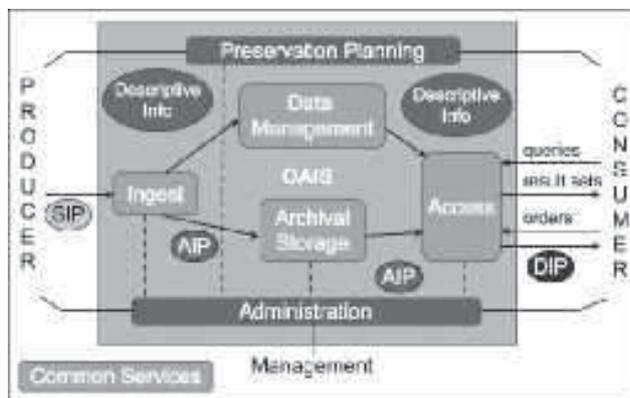


Figure 1. Diagram for a Digital Repository based on the OAIS Reference Model.

ity as outlined in the ISO Standard *Reference Model for an Open Archival Information System (OAIS)*.⁵ The OAIS Reference Model provides a generic vocabulary for discussing digital preservation repository concepts based on broadly-supported standards and operational practices.

To meet the expectation of long-term file sustainability, the NYU Digital Library Technology Services team intended from the start that the model repository would become an OAIS-compliant Trusted Digital Repository, following the TRAC criteria. Functionality was designed around DSpace, an open-source software application used at NYU to store and manage other moving image collections, and technical issues rested primarily on how best to organize test files and metadata to create Submission Information Packages (SIPs) and Archival Information Packages (AIPs) following the OAIS construct.⁶

Testing Program Files

To test ingest and retrieval for the repository, the team used a sample of 35 hours of programming, both HD and standard definition files drawn from the national series *Nature*, *Frontline* and *Religion and Ethics Newsweekly*, plus a selection from the local program *New York Voices*. The test files originated from three sources – high resolution program masters from WNET and WGBH, and lower resolution distribution versions of the same programs from PBS. This collection of programs allowed the project to test of a mix of both high and low resolution program file formats. It also required the files to be accompanied by a wide range of metadata that was collected from multiple sources.

Collecting and Managing Metadata

Because the model repository does not input any new descriptive metadata, to create a useful Archival Information Packages (AIPs), sufficient information about each program file had to be packaged and sent along with the video as part of the Submission Information Package (SIP). The program SIPs therefore had to include the requisite metadata as well as the program files themselves.

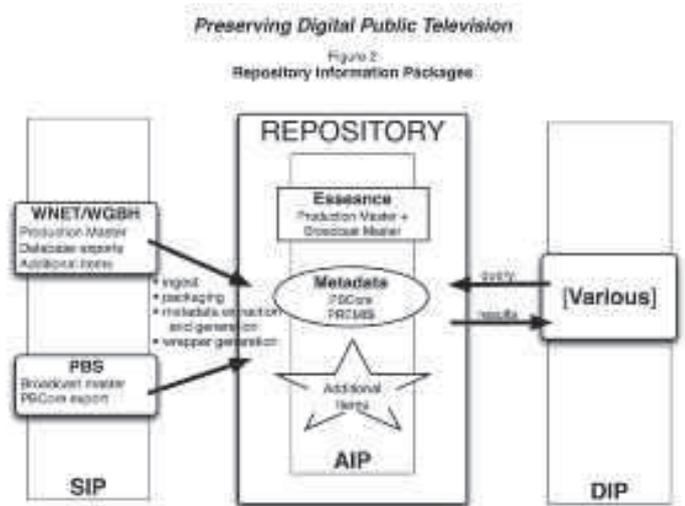


Figure 2. Repository Information Packages.

Determining an appropriate set of metadata fields was a detailed and intensive task. One fundamental requirement for each program was to aggregate its content with its metadata, regardless that they originated from disparate sources, such as from the producing station (the high-resolution production master) and also from PBS (lower-resolution broadcast master.) Database exports from each source also had to be analyzed, particularly the extensive descriptive and rights metadata created by PBS for broadcast scheduling.

The initial tests of the repository revealed that, because public television program producers in the US are largely independent, each operates with different software, hardware, production procedures and recordkeeping. This resulted in production master files submitted to the repository which came in a variety of encoding and wrapper formats and which were not equally easy to manage or playback.

The project also found there was very little consistency in the way metadata for each program was recorded, as producers created certain elements of metadata, PBS generated others, and all of it was saved in different modes for different purposes (i.e. production vs. distribution). Further, metadata was not collected consistently in a centralized place, even within PBS, so collecting it to meet the needs of the repository had to be done on a time consuming, program-by-program basis.

To organize such idiosyncratic metadata, the project examined a broad range of standard metadata schema used by libraries, archives and the commercial broadcast field. Project staff paid special attention to PBCore⁷ a metadata dictionary based on Dublin Core designed specifically for public radio and television program files, and PBCore was chosen early on as the appropriate standard to capture descriptive program metadata. PBCore has been in development for several years, but remains in the early stages of system-wide adoption. Even so, the repository designed its descriptive metadata requirements around PBCore, and in the course of the project, all three of the public television partners have implemented PBCore-compliant export functionality for particular program databases. Incorporating technical metadata embedded in the video files of the AIP also proved to be a challenge. The program files were submitted to the repository in a number of wrapper formats, including MXF and QuickTime, and encoding formats, including various flavors of MPEG and DVC Pro. These required multiple tools able to extract such technical metadata as bitrate, file size, and frame size, from the file headers. Because extracting technical metadata is critical for long term digital preservation and access, this was an important operation.

Standard Metadata Schema

Transforming the submitted and extracted metadata into standard formats was a clear necessity. The solution was to encapsulate the necessary descriptive and technical metadata, plus preservation and administrative metadata, from several data dictionary schema while maintaining information unique to public television programming.

The repository developed a structure to capture all necessary metadata using PBCore for descriptive and technical metadata, PREMIS for additional technical and preservation metadata, and METSRights for rights metadata. Appropriate fields from these standards, along with virtual links to the program files themselves, are all contained within a METS (Metadata Encoding and Transmission Standard) wrapper.

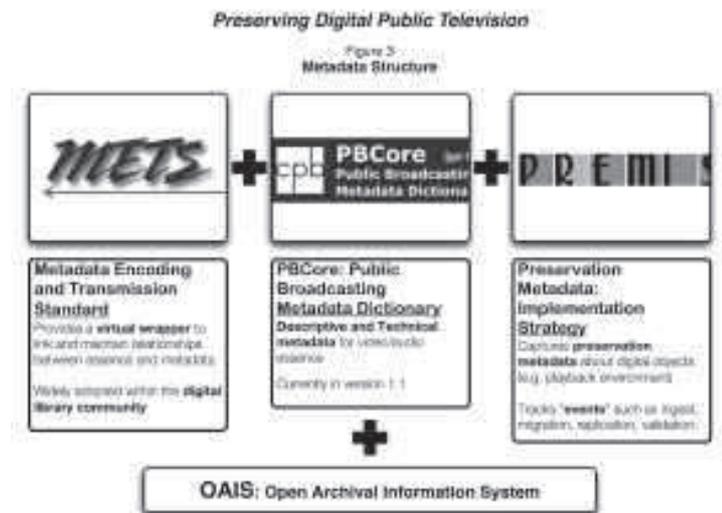


Figure 3. Metadata Schema.

The problems encountered by testing these various file formats, combined with the time-consuming efforts needed to collect metadata, demonstrated the necessity of setting uniform standards across the public television system for the success of any future repository operation. Otherwise, automating the functions of managing the metadata of large collections would not be feasible, and managing the video files themselves would be overwhelming.

Developing a File-based Video Production Workflow

With the understanding that preservation and other valuable metadata must be captured early in the process, the project examined program production workflows to identify points where key metadata are created. However, this turned out to be problematic because when the project began, production workflows were largely manual and the technology and digital asset management tools necessary to manage file-based video workflows were not in place.

That changed in the October 2008, when WNET launched WORLDFOCUS, a daily half-hour international news program that is pioneering a digital production workflow end-to-end. With cost-effective digital recording, editing and file transfers finally available, WORLDFOCUS provided the opportunity to test the preservation concepts of an actual file-based workflow for the first time.

The program is produced with video files coming from P2 field cameras, studio recordings and other sources, processed

through a central ingest facility at WNET. Most are encoded as DVCPHD files (also known as DV100,) which require minimal transcoding to normalize file formats. Instead, the DV100 files remain intact but are re-wrapped from de-multiplexed MXF into a single Quicktime file. Related PBCore metadata records are created at the same time.

The DV100 files are then transferred in-house for editing (on Final Cut Pro) and when the production is finished, they are sent to the broadcast servers for play-out. For transfer to the repository, the PBCore metadata record is added to the finalized DV100 program file to create the repository SIP. The digital workflow of WORLDFOCUS has allowed us to submit genuine born-digital source footage, as well as completed program episodes, to the repository for the first time.⁸

Search for a Standard Video File Wrapper

The use of a standardized video file wrapper is considered as a requirement for successfully exchanging digital files between entities, particularly to support future file migration. A number of so-called video wrapper “standards” exist, but despite vendor claims, the files do not all actually interoperate with many equipment configurations used by public broadcasters. Consequently, the search for a functional wrapper has been persistent.

Any initiative to create technical standards for public television must dovetail with the needs of the commercial broadcasting industry, because public television on its own does not carry enough economic clout to influence hardware vendors. To examine this issue, the project convened a “Wrapper Roundtable” of technologists, digital archivists and industry leaders in hopes that the networks had made progress and would share their findings. However, the group was surprised to learn that the lack of consistent wrapper standards was also a major problem for the commercial networks.

The wrapper standard remains to be solved, but due in part to the collaborative atmosphere established by the “Wrapper Roundtable,” the Advanced Media Workflow Association (which represents vendors) has launched the AS-03 wrapper project specifically to develop an MXF standard that vendors will support.⁹

Access and Sustainability

In addition to testing the repository model, *Preserving Digital Public Television* committed to evaluate the impact of rights restrictions on digital preservation and future access, and to examine issues relating to the long term sustainability of a repository operation. Both subjects are complex, and the project has produced two comprehensive reports that provide detailed analysis of each issue in relation to public television in the U.S.

Analyzing the Impact of Rights Restrictions

Television programs are comprised of a wide array of elements governed by rights restrictions and other encumbrances. Typically, rights to use this material for broadcast are granted for a finite period, and when the broadcast rights expire, the system no longer has an interest in the program. Specific authorization to preserve public television programs or make them accessible at the end of the broadcast window is largely absent, and renewing rights agreements to permit new uses (like on-line viewing) can be tremendously costly.

This is a problem that exists across television archives and collections. *Intellectual Property and Copyright Issues Relating to the Preservation and Future Accessibility of Digital Public Television Programs* outlines the problem in-depth and presents a number of case studies to illustrate the issues.¹⁰ Not having clear permission to reuse older programs is a primary factor that discourages public television from making an investment in long-term program preservation. Until rights agreements are improved, archival content will remain largely inaccessible.

Assessing Operating Costs

As existing digital repositories mature, operating conditions and issues are being documented by such institutions as The National Science Foundation, which commissioned the *Blue Ribbon Task Force on Sustainable Digital Preservation and Access* jointly with a number of other organizations in 2007 specifically to study economic models for maintaining large and diverse database repositories.¹¹

Instead of being seen as unaffordable, the costs to preserve digital public television had to be presented as feasible and manageable. Our report *Operating and Sustaining a Trusted Digital Repository for Public Television* discusses potential benefits of digital preservation and breaks down the technical and organizational expenses necessary to maintain a reasonable scale of repository operations.¹² It also outlines ways that public television might be able support a preservation program, based on existing income sources and operating models already familiar to the public broadcasting system. The intent is to demonstrate that with shared resources and commitment, a repository can be sustained.

Promoting System-wide Support

Since *Preserving Digital Public Television* began, broadcasting has shed its analog systems and moved completely into a digital universe. Because the project has been successful tying the concept of digital preservation to effective reuse of program content across viewing platforms, preservation has become highly relevant to stations — not an optional “add-on” cost, but a requirement for any future use of the materials.

Thus the project has been instrumental in transforming an attitude of indifference to one that acknowledges the value of properly managing our collective archival holdings. As evidence, CPB allocated funding for the very first time to pilot The American Archive, a national preservation effort for U.S. public radio and television.

Even so, despite the support from CPB, digital preservation is still not accepted system-wide as an important national investment. Viewers keep reminding us that public television programming is precious and has made an indelible imprint. What remains is to continue building commitment and enthusiasm across the entire system. The critical responsibility for saving the legacy of American public television must be shared, sustained and nurtured over time so our programs will indeed, have a life after broadcast.

Special thanks to the Preserving Digital Public Television NDIIIP Team for their assistance in preparing this article, especially Dr. Howard Besser, Professor and Director Moving Image Archive & Preservation Program, NYU; Kara van Malssen, Senior Research Scholar, NYU; Yvonne Ng, Junior Research Scholar, NYU; Karen Cariani, Director, WGBH Media Library and Archives; Bea Morse, Director, Business and Communications, PBS Interconnection Replacement Office; and Jeff Ubois, Archival TV.

Contact:

Nan Rubin
Thirteen/WNET-TV
450 W. 33rd St
New York, NY 10001
212-560-2925
RubinN@thirteen.org
www.Ptvdigitalarchive.org

Biographical Information

Nan Rubin has more than thirty years experience managing technology and facilities projects in public broadcasting in the United States. She has worked at public television station Thirteen/WNET-TV in New York City on such major efforts as restoring the station's broadcast signals after the destruction of the World Trade Center, and she planned and coordinated the creation of Thirteen's Tape Archive to preserve both analog and digital program assets. Since 2003, Ms. Rubin has been Project Director of Preserving Digital Public Television, funded by the Library of Congress's National Digital Information and Infrastructure Preservation Program (NDIIPP), where she oversees a team of 20 based at WNET, WGBH -TV in Boston, the Public Broadcasting Service, and New York University, who are building a model preservation repository for born-digital public television productions.

Footnotes:

1. <http://www.ptvdigitalarchive.org>
2. Ken Thibodeau, "Preserving Records, Saving Appearances: Perspectives on Digital Preservation", *Preserving authentic electronic records: preliminary research findings*; Proceedings from an International Symposium, February 17, 2001, University of British Columbia, edited by Luigi Sarno, Vancouver, Canada, August 2001.
3. Accessed 23 March 2009 at <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>.
4. There are other publications and projects that outline criteria for repository trustworthiness, such as the German *Network of Expertise in Long-Term Storage of Digital Resources* (nestor) project and the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit.
5. (ISO 14721:2003). Open archival information system: Reference model; Accessed 23 March 2009 at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683.
6. See report *Repository Design for Preserving Digital Public Television* at <http://www.ptvdigitalarchive.org>.
7. <http://www.pbcore.org/>
8. See *Report on the Concepts and Implementation of Preservation-Compliant Digital Production Workflow* at <http://www.ptvdigitalarchive.org>.
9. <http://www.aafassociation.org/html/projects.html>
10. See *Intellectual Property and Copyright Issues Relating to the Preservation and Future Accessibility of Digital Public Television Programs* at <http://www.ptvdigitalarchive.org>.
11. <http://brtf.sdsc.edu/>
12. See *Operating and Sustaining a Trusted Digital Repository for Public Television* at <http://www.ptvdigitalarchive.org>.