PRESERVING DIGITAL PUBLIC TELEVISION
Final Report
June 2010

Prepared by Nan Rubin, Project Director

PRESERVING DIGITAL PUBLIC TELEVISION
Final Report

TABLE OF CONTENTS

APPENDICES

    A.   **Project Staff**

    B.   **Executive Summary: Repository Design**

    C.   **Key Findings: Sustainable Preservation of Digital Public Television**

    D.   **Dissecting the IP Issues of Television Production**

    E.   **Article: CURRENT Newspaper**

    F.   **Article: Library Trends Journal**

    G.   ***The American Archive***

PRESERVING DIGITAL PUBLIC TELEVISION
Final Report

Prepared by Nan Rubin

"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)

NEW PRESERVATION PRACTICES FOR TELEVISION ARCHIVES

In less than a decade, television production, distribution and preservation have undergone a radical shift. The costs of video recording and editing systems are now well within the means of most members of the public, and the ubiquity of media on the Internet, coupled with the mass deployment of hand-held devices, have transformed not only the medium of television but the entire environment for creating and watching video.

Distribution and transmission have been equally transformed, as all over-the-air television broadcasting is digital, and tape-based program distribution is rapidly being replaced by digital file transfers, and distribution and transmission have been equally transformed, as tape-based submissions to the Public Broadcasting Service (PBS) and other national program services are being replaced by digital file transfers. This has all been spurred on not only by advances in digital technology, but also by the new conditions ushered in when all on-air television broadcast signals were required to become digital in February 2009.

What do these changes mean for television archives? Practices to conserve and protect recordings on videotape are well established, and the costs for maintaining and storing physical media are easily calculated. However, in an age of digital files, the requirements for preserving television programs are far different from storing videotape. It is not enough to close a digital file and put it on a virtual shelf. For video in particular, acceptable practices to save and access very large files, manage ever-changing file formats, maintain rich and consistent metadata, and protect file view-ability into the future are just now emerging.

Preserving Digital Public Television, a project funded in 2004 by The National Digital Information and Infrastructure Program of the Library of Congress (NDIIPP), [http://digitalpreservation.gov] set out to solve some of these difficult problems by designing a model repository for public television. In the process, the project also determined standards for metadata, explored rights issues relating to video archives, analyzed operating costs, and brought a new consciousness about the importance of digital preservation to the public television system.

Public television is now confronting new challenges as it enters the digital era. The requirements for preserving born-digital programs and ancillary materials, such as source footage, un-aired segments, animations, transcripts and program run-downs in electronic formats, present much different issues than the familiar practices used to maintain content on videotape, film, and paper.

Saving digital public television requires a new framework that brings preservation principles into emerging all-digital production and delivery environments, storage systems, and media asset management applications. It also demands a shift in public television's approach to preservation, based on the characteristics and behavior of digital files. With a collective effort, the public broadcasting system has a unique opportunity to stand at the forefront of this new domain, and to take on this challenge on behalf of future generations.

<div align="center">* * *</div>

PROJECT GOALS: BRINGING DIGITAL PRESERVATION TO PUBLIC TELEVISION

In the Public Broadcasting Act of 1967, Congress authorized the Corporation for Public Broadcasting (CPB) "to establish and maintain, or contribute to, a library and archives of noncommercial educational and cultural radio and television programs and related materials." However, CPB never allocated any funds to support this charge, and no demand for system-wide preservation was implemented. Consequently, only a handful of stations and entities like PBS have formal archiving activities in place.

Without a preservation mandate, digitally produced programs in public television are at great risk of being lost. The rapid changes in digital technology are rendering recording and playback systems obsolete at breakneck speeds, at the same time adequate tools for managing large and complex video files are not yet perfected. This has left a very large gap in the preservation of America's public television legacy.

Public television stations WNET in New York and WGBH in Boston, which produce roughly 60% of the national prime time series including *Frontline* and *NOVA* at WGBH, and *American Masters* and *Great Performances* at WNET, recognized this challenge early. Because WNET and WGBH each maintain its own archives, the stations were already committed 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 take a collaborative effort.

When the Library of Congress invited proposals under NDIIPP, WNET and WGBH created a partnership with PBS to build a model preservation repository for "born-digital" public television programs. PBS operates the network that distributes public television programs to more than 300 stations, and because most national programs pass through PBS before they are aired, it is the principle *de facto* repository for these programs. (The PBS warehouse holds more than 150,000 videotapes of programs going back more than 40 years.)

These institutions understood that public television had to take steps to protect its rapidly growing collection of digital assets. As broadcasters, however, they had little experience building a preservation repository. New York University provided the expertise that was lacking. The NYU Digital Library team had extensive experience designing repository systems specifically for large digital files wrapped in metadata. The project further benefited from a relationship with NYU's Moving Image Archiving and Preservation Masters Degree Program, whose students provided excellent research and whose graduates became full-time project staff.

Together, WNET, WGBH, PBS and NYU organized *Preserving Digital Public Television* (PDPTV) [http://wwwthirteen.org/ptvdigitalarchive/] as a collaboration to introduce digital preservation issues and practices to the public television system. The project was aimed specifically towards preserving *born-digital* program files and was not engaged in digitizing any analog materials. Under the strong leadership of Ken Devine from WNET, and Dr. Howard Besser at NYU, activities formally began in September 2004 and were completed in March 2010.

The goals of the PDPTV project were to:
- Design and build a prototype preservation repository for born-digital public television content;
- Develop a set of standards for metadata, file and encoding formats, and production workflow practices;
- Recommend selection criteria for long-term retention;
- Examine issues of long-term content accessibility and methods for sustaining digital preservation of public television materials, including IP concerns.
- Introduce the importance of digital preservation to the public broadcasting community.

Understandably, the priorities of public broadcasting are program production and broadcast delivery, not saving program assets. Most program preservation is handled as an afterthought. 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. To this end, along with designing a preservation repository, the PDPTV project also analyzed many basic questions of what is required to sustain a digital preservation environment in the context of U.S. public broadcasting program production and distribution.

Project Structure and Organization

The project was coordinated by Nan Rubin at WNET/Thirteen, which was the lead institution, with roughly 20 people (3-6 from each partner institution) at any given time comprising the core staff. These included key leadership staff from each entity, with relatively low turnover of personnel. The project was also fortunate to have assistance from a number of outstanding graduate students from the Tisch School MIAPP program at NYU, who provided quality research and support services to the project. . [Complete staff list in **Appendix A**]

The work was organized by teams which included members from each organization. The teams divided the work into two primary areas, with some overlap:
  ▪ Technical Issues, Repository Design and Operations;
  ▪ Content Concerns, including Selection and IP Issues.

Discussions and tasks were conducted primarily by ongoing email exchanges and conference calls, with occasional in-person team meetings. News and informal progress reports were circulated frequently to share information on current developments, and there was significant participation of team members in public broadcasting and professional conferences, NDIIPP partner meetings and similar gatherings. The entire group was convened each year for a regular 'annual meeting' to review progress and map out the tasks and timelines for the coming year.

Throughout, the work was performed with a high degree of collaboration and collegiality. Reports and other materials were prepared with a great deal of participation from team members, and although it was not required, most decisions were made with agreement of the team members or by consensus of the group as a whole. Generally, nothing was released in final form without incorporating the comments and concerns of all those working on the report.

The primary methodologies included documenting the experiences of the PDPTV team in building the prototype repository; detailed examination of operational issues such as program production workflows and distribution requirements; and gathering the most current research or experiences on the topics under consideration. This included reviewing materials such as publications, bibliographies, and conference presentations; surveying other collections and repositories for policies and practices; and collecting information directly by organizing focus groups, special gatherings, and through conversations and interviews with select individuals. During the course of the project, there were also many site visits and participation in a wide range of events and gatherings, which also provided rich opportunities to collect information and resources.

This project was enormously successful. We produced a significant body of reports; published articles in key journals and other publication; and made popular presentations at dozens of conferences, symposia and special events in the U.S, Canada and abroad. Much to our surprise, this project emerged as a respected leader nationally and internationally in approaching technology issues relating to preserving digital video.

Most importantly, by promoting the importance of digital preservation to public broadcasting, we were instrumental in helping to create the *American Archive*, a new initiative at the Corporation for Public Broadcasting, which is its first genuine investment in long-term preservation and access of U.S. public radio and television programming.

All reports and papers produced on these and related issues were published on the project website and have been circulated and promoted broadly.

http://www.thirteen.org/ptvdigitalarchive/

We are extremely proud of the quality and scope of the activities performed with NDIIPP support. What we did accomplish was much more than we planned, and we had a much greater impact than we could have imagined. As such, ***Preserving Digital Public Television*** more than exceeded the project goals and well surpassed the expectations of the Library of Congress as an original NDIIPP project.

* * * * *

DELIVERABLES AND ACTIVITIES
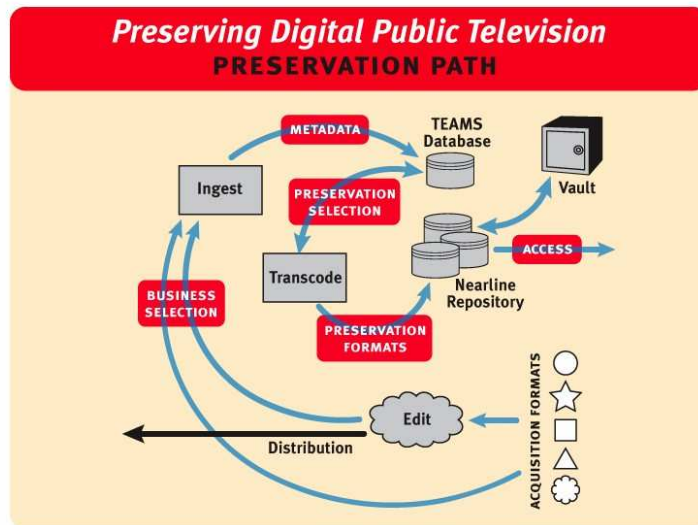
Original Concept: Preservation Repository Path, 2004

The original proposal conceived of the repository as a series of discrete technical tasks in a lab-type environment, and initial deliverables focused on:

- Appraisal and selection criteria.
- Technical standards, file formats and related concerns.
- Descriptive and technical metadata standards.
- Repository architecture and design functionality.

The approach was to identify commonly used file formats, determine appropriate metadata requirements, and examine technical standards that would be critical to repository functionality. The project naively assumed that commercial television networks and large collecting institutions such as the Library of Congress (completing its Packard Campus - National Audio-Visual Conservation Center) were already making progress solving these same problems, and that public television could simply "tag along" with work underway.

However, we quickly learned that this was not the case. Instead, we found that relatively little progress was being made in these areas, even by the major broadcast networks. This threw attention on our efforts, especially in metadata standards, which rapidly took on the reputation of being at the leading edge of video repository planning. Consequently, much of the work that the project produced was well-received and has been widely circulated, in the U.S. and abroad.

1.   Selection and Appraisal

Library of Congress Television Card Catalog

Given the understanding that digital recording technology was becoming so inexpensive that huge volumes of raw footage were being recorded, it was evident from the beginning that the question of how to select materials for preservation would become increasingly important. Some of our earliest research and publications were on this topic.

A number of reports were produced that explored this issue, based in part on work that had already been published at WGBH, under the excellent leadership of long-time WGBH Archivist Mary Ide, and her colleague Leah Weisse.

>   Mary Ide and Leah Weisse (WGBH), "Developing Preservation Appraisal Criteria for a Public Broadcasting Station." The Moving Image, Vol 3, No 1, Spring 2003.
>
>   "Recommended Appraisal Guidelines for Selecting Born-Digital Master Programs for Preservation and Deposit with the Library of Congress" By Mary Ide and Leah Weisse (WGBH Archives). May 25, 2006.

Recommendations for selection policies were also drawn from original research using focus groups that were organized by the project, and by surveying best practices and policies followed by a selection of moving image archives.

>   Focus Group Report: Producers
>
>   Focus Group Report: Scholars
>
>   Focus Group Report: History Professors

Focus Group Report: Social Studies Teachers

Focus Group Report: TV writers

Library and Archive Best Practices Survey, December 2005

Appraisal Bibliography, December 2004

We did not meet the deliverable of creating an *Advisory Committee on Selection of Content* to assist selecting materials for the repository. Initially, we thought we would be able to submit a large collection of content to the repository, and that Advisors would be helpful making the selections.

In actuality, we used a very small number of programs to test the repository functions, and they were selected based solely on technical criteria and availability. Consequently, the purpose of the Advisory Committee was never triggered, and the group was not established.

In addition to Mary Ide and Leah Weisse from WGBH, Daisy Pommer and Winter Shanck from WNET, plus Bea Morse at PBS made significant contributions to this area.

2.   Inventory of At-Risk Materials

Sample of Tapes from the WNET Archive

The three participating public television organizations all manage and maintain an archive of programming and related materials. The WGBH Media Library and Archives has been in place since 1978 and holds a collection of 750,000 editorial assets. At WNET, the Archive was officially created in 1998 and currently manages its own collection of 35,000+ programs. And

although PBS produces nothing on its own, it is the *de facto* repository for national public television programs because nearly all programming distributed nationally passes through PBS. Consequently, it maintains a warehouse filled with 150,000+ videotapes going back more than 40 years.

Most of the finished programs in each of these collections are on analog videotape, encompassing aging and obsolete formats such as 2" helical scan and ¾" U-matic video cassettes, but also including more recent recordings made on digital tape formats such as D3 and Digital Betacam.

Because the basic premise of NDIIPP was to develop operational models for preserving file-based digital content considered at-risk of being lost, the funding explicitly prohibited digitizing any content and required that the materials to be preserved had to be *'born digital',* that is created originally as digital files. Despite the obvious fragility and instability of many of our analog videotapes, the Library accepted our proposition that programs recorded on a number of digital videotape formats, such as D-2 and D-3, were also highly at-risk and could be used as appropriate content to test in the repository.

To identify the extent of these digital holdings, each institution produced a comprehensive inventory of its own collections, based on its own database information and operational practices. However, with a combined set of holdings of close to 750,000 items, each managed individually on different databases, it was not practical to publish these lists or merge them into a single massive inventory. Instead, we produced a simple summary report of the findings, and refer LoC to each respective institution for details or additional information about its own specific at-risk inventory. These individual inventories are available on request. The reports were overseen by Mary Ide and Leah Weisse at WGBH; Daisy Pommer and Winter Shanck from WNET; and Bea Morse, Glenn Clatworthy and Irene Taylor at PBS.

We did use the inventory process as an opportunity to research and document the paper records, films and videotapes of the very earliest public television holdings from the pre-PBS period, which are at the Library itself. The detailed *"Report on Public Television Holdings At the Motion Picture, Broadcasting and Recorded Sound (M/B/RS) Division of The Library of Congress"* documents the history of the early public television holdings at the Library, various catalogs and lists documenting the extensive collections, and photographs of the materials in storage.

This series of reports was produced by NYU Research Assistants Tanisha Jones and Pamela Smith, and was not published, but a CD produced with all elements of this report was given to the Library.

3. Metadata and Related Topics
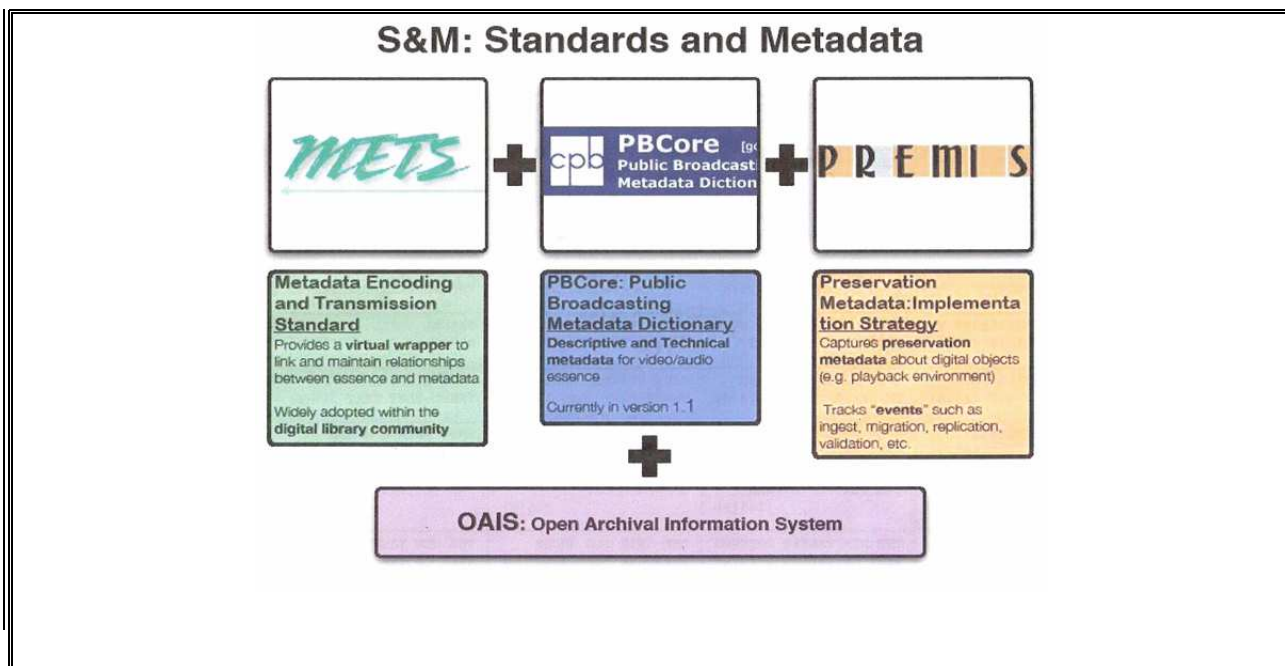
Metadata Schema Used for Preserving Program Files

One of the most important aspects of the ***Preserving Digital Public Television*** Project has been to investigate appropriate standards for descriptive, technical, rights and preservation metadata.

The importance of adequate metadata is central to long-term file access, such as the ability to search for and find content; file interoperability for successful transcoding and usability across platforms; and especially for file migration to keep file data intact as formats move forward in time. Moreover, with the transformation of program production from analog to file-based digital workflows, the sequence for capturing and creating appropriate archival metadata has shifted from a process that was added largely at the end of the production line, to workflows where the information is generated towards the front end and throughout the production process.

The project recognized this, and consequently produced detailed analyses of the (analog) production workflows used for several of the national productions used as test programs for the repository, including *Nature* and *Frontline*.   The intent was to identify what kind of metadata was created, by whom, and when during the production process.

The project also spent considerable effort analyzing and evaluating options relating to adopting metadata standards for video preservation, including a focus on PBCore, the metadata dictionary developed specifically for use by public radio and television to facilitate file interoperability.

Although PBCore was developed by public broadcasters and funded by the Corporation for Public Broadcasting, it was not fully constructed and for a period of time, CPB support was suspended. This project was especially instrumental in raising awareness across the system of its use as an appropriate metadata tool to support archiving, and because of this, CPB renewed funding for PBCore and revived the effort to further develop it as part of the *American Archive* initiative.

The schema we eventually adopted represent one of the first sets of standards designated for preserving video, and it has been widely considered by moving image archives and repositories in the U.S. and abroad. It will also be critical that the metadata standards used by public broadcasting be compatible with those adopted by LoC's Packard Campus for Audiovisual Conservation when it becomes capable of accepting program files. Kara van Malssen from NYU became our primary project metadata expert.

4.   File Format and Packages
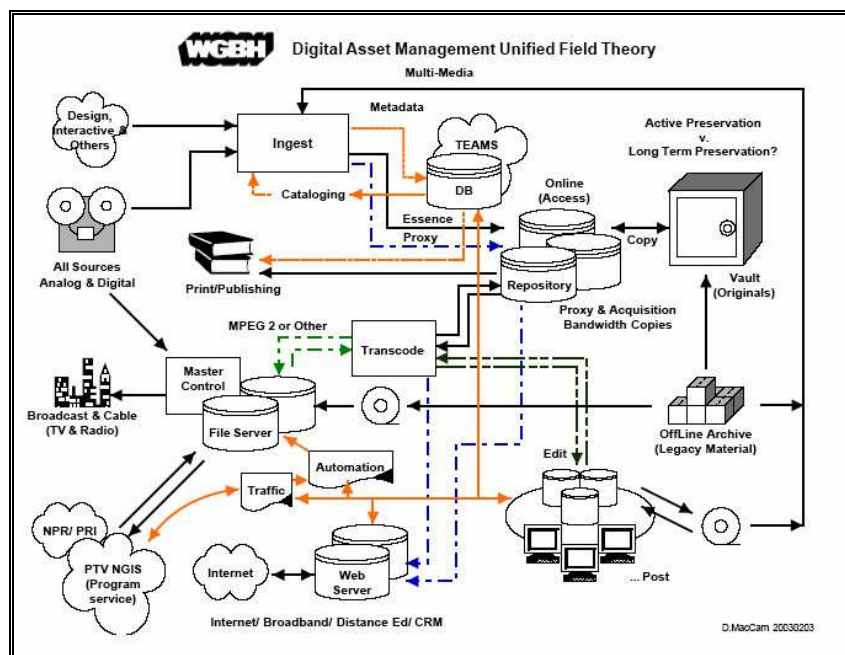
Digital Asset Unified Field Theory – D. MacCarn, WGBH 2003

With the ever-changing models of digital recording devices with their associated video file formats, plus the many different tools needed to edit them, store them and transform them into other formats, one of the key questions that the repository had to answer was how to manage the broad array of video formats and file packages. Over the period of the grant, it was difficult to keep up with all the technical changes in the market.

To advance the discussion in relation to preservation, we produced a number of reports under the direction of WGBH Chief Technologist Dave MacCarn and the excellent guidance of our NDIIPP Program Officer Carl Fleishhauer.

A related issue we also tried to address was the question of using a standard 'wrapper' as a container for the video files that needed to be preserved together. A relatively stable standardized video file wrapper can facilitate the successful long-term preservation of digital files, particularly to support future file migration and interoperability. But there was no single acceptable or accepted wrapper when the project was initiated.

To encourage a shared approach to solving the problem, the project convened a group of more than twenty distinguished technologists, digital collections designers and project managers, sponsored by Greg Lukow and the M/B/RS Division of the Library. At the "Wrapper Roundtable," the public television partners were surprised to learn that the lack of consistent video format and wrapper standards was also a major problem shared by the commercial broadcast networks.

We were not successful in designing the elusive wrapper, but our efforts did spur the efforts of the commercial sector, including vendors, to continue such development.

"Survey of Digital Formatting Practices in Public Television Program Production." Written by Dave MacCarn (Chief Technologist, WGBH), Edited by Nan Rubin. September 2007.

Notes from the Wrapper Roundtable. Prepared by Carl Fleischhauer (Library of Congress) with additions by Glenn Pearson and edited by Nan Rubin. March 9, 2006.

Factors for Evaluating Preservation File Formats for Digital Video." Prepared by Jerome McDonough (New York University) for the Preserving Digital Public Television Project, 2005.

Notes from a meeting on formats for digital video preservation related to the Preserving Digital Public Television project funded by NDIIPP. Held as an adjunct session to the PBS 2005 Technical Conference. April 17, 2005.

Carl Fleishhauer (Library of Congress), "Discussion Paper on Video File Formats and Wrappers." March 2005.

Bibliography: Preservation File Formats for Video and Associated Metadata, March 2005

5. Repository Design
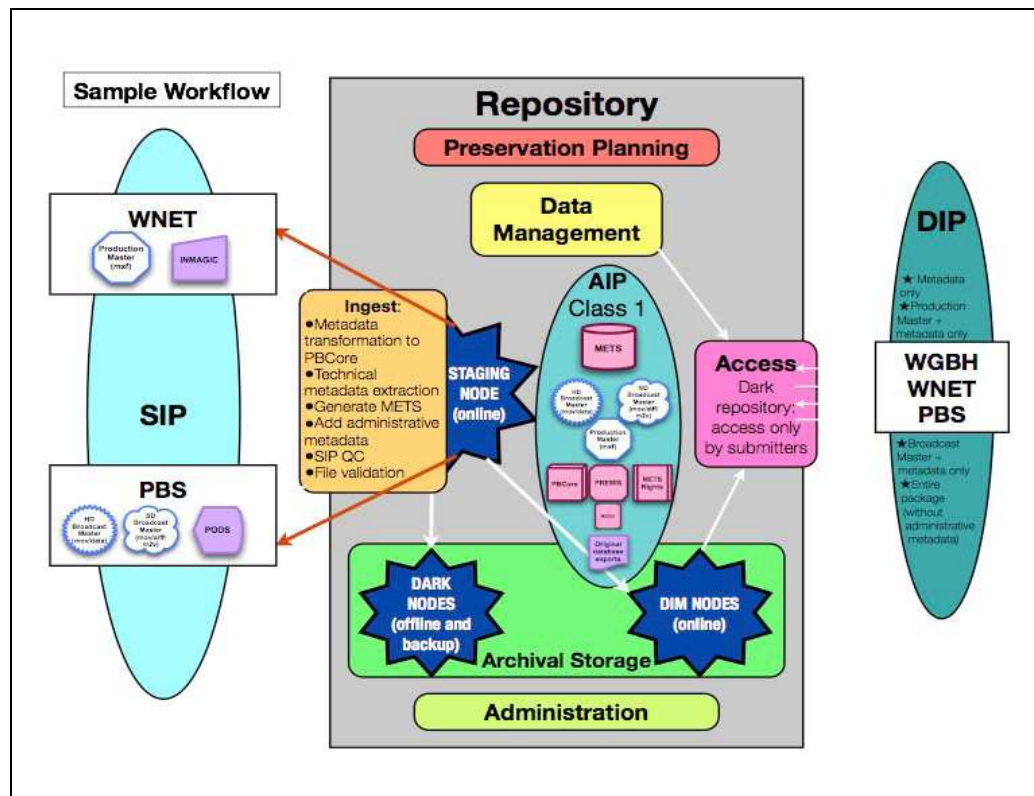
OAIS Functional Entities in the Model Repository

Creating, testing and operating an OAIS-compliant model repository was a central focus of the project, built by the Digital Library Technology Services team at New York University, who had the primary responsibility for this task.

The *Repository Design Report with Attached Metadata Plan* outlines details of repository architecture, technical functions and operational design. It describes:

- The design of the preservation environment;

- The technologies used to support preservation functions;

- The creation of Archival Information Packages for managing complex video files;

- The standards used to support aggregation of disparate sources of metadata, including METS, PBCore, and PREMIS;

- And the process used to determine needs of repository users in order to design output requirements.

This report was written by NYU Software Systems Architect Joseph Pawletko, with contributions by Nan Rubin (WNET) and Kara Van Malssen (NYU).

The Preservation Repository (PR) at NYU was designed as a "content neutral" system to preserve a range of content types from a variety of different projects. For this project, particular attention was given to the processes for creating the Submission Information Packages (SIPs), the structure of Archival Information Packages (AIPs), and determining and managing required metadata for each package including the roles of METS, PREMIS, and the emerging descriptive and technical metadata standard PBCore.

A sample of more 80 hours of programs and segment files were submitted to the repository, including high-resolution (production quality) program masters from project partners WNET and from WGBH, drawn from such national public television programs as *Nature* and *Frontline,* and lower-resolution (broadcast quality) distribution versions of the same programs supplied by PBS.

Metadata was also collected and aggregated from each partner institution for each program. This allowed the repository to collect, organize, process, and ingest a wide range of program file encoding formats, with different wrappers and metadata.

We were successful in solving many of the problems inherent in designing a digital preservation environment for broadcast files, including how to handle multiple file formats, different database exports and metadata structures.

At the same time, we learned that functions cannot be automated without uniform standards for processing metadata or file formats, and this is one of the biggest problems that public broadcasting will have to solve in order to design a successful long-term digital repository.

> *Repository Design Report with Attached Metadata Plan,* Prepared by Joseph Pawletko, Software Systems Architect, New York University
>
> http://www.thirteen.org/ptvdigitalarchive/files/2010/03/PDPTV_ReposDesign_2010-03-19.pdf

6. **Sustainability**
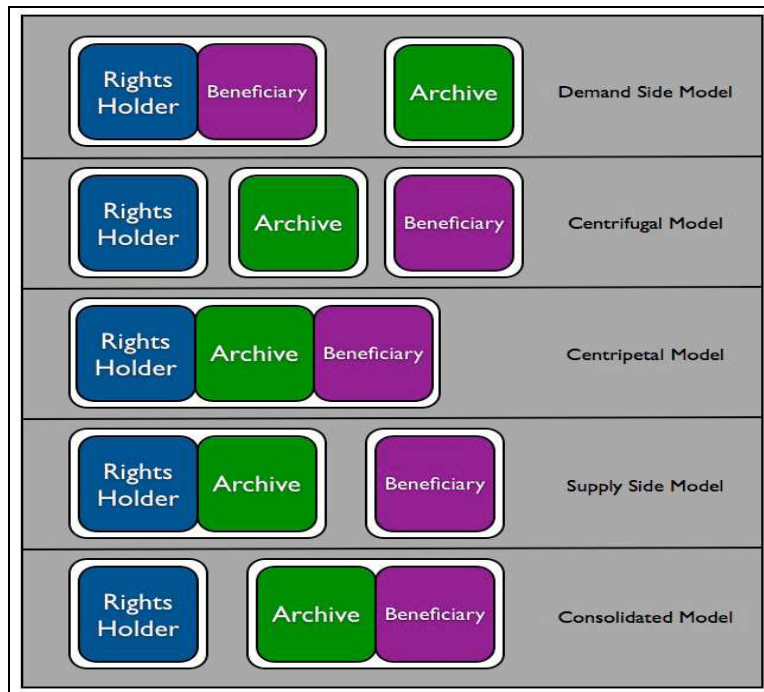
*Five Economic Scenarios Representing Different Stakeholder Interests
in Digital Public Television Preservation*

Apart from the technical architecture necessary to operate a repository, there are also a set of additional criteria that must be met in order for a repository to be viable long term. We did an extensive examination of the issues relating to long-term sustainability, in particular -

- Identifying the technological, procedural, and organizational issues that public television will face in planning a repository to safeguard its rich programming legacy;

- Discussing key aspects of economically sustainable digital preservation in a public television context;

- Offering examples of business models that can be used to support digital preservation;

- And describing and quantifying costs associated with creating the PDPTV model repository, as a study to inform planning a future facility.

Throughout the project, we based our planning on the concept that our digital content would only be sustainable if this repository is considered *trustworthy,* that is, it should meet the demands for a Trusted Repository as represented by the 'TRAC' Criteria and Checklist, developed by the National Archives and Records Administration and other library organizations.

Sustainability was also seen in the context of accepting repository practices that meet the ISO standard *Reference Model for an Open Archival Information System* (OAIS), a high-level, widely applied model that is essentially a generic blueprint for the design of a digital preservation repository applicable to any type of content.

Our resulting report offers an in-depth analysis of these issues, along with extensive discussions of financial models and stakeholder roles, each with a focus on familiar and existing models within public broadcasting that can guide the sustainability planning.

> *Strategies for Sustainable Preservation of Born Digital Public Television* was written and Edited by Yvonne Ng (NYU), Nan Rubin (WNET), and Kara Van Malssen (NYU). Assistaence was provided by With Howard Besser (NYU), Karen Cariani (WGBH), Irene Taylor (PBS), Winter Shanck (WNET), and Jeff Ubois, with additional support from Joe Pawletko (NYU), David Ackerman (NYU), Alicia Kubes (NYU), and B Morse (PBS).
>
> http://www.thirteen.org/ptvdigitalarchive/files/2009/10/PDPTV_SustainabilityStrategies.pdf

This is one of the more comprehensive reports of its kind, and one we hope will be useful to the public broadcasting system as the *American Archive* takes shape. NYU Research Assistant Yvonne Ng was responsible for most of the extensive research and writing on the report.

7.   Intellectual Property and Copyright Issues

Dr. Howard Besser at American Library Association Conference, 2009

Most television productions are a complex mix containing a broad array of elements, for example recorded and original music, still images, stock footage, animations, archival footage, and talent, and the rights for each element can be controlled by different parties. In order to make content accessible after the broadcast rights expire, all the permissions must be renewed.

The process of identifying, negotiating, and paying for new permissions from these underlying rights holders can be both labor intensive and expensive. Moreover, these difficulties are compounded when a work or element becomes "orphaned" because its copyright owner can no longer be identified or found. The great effort and cost required to obtain permissions to re-use or redistribute programs, such as to put them on-line, after initial rights expire, is a major factor that discourages the system from making an investment in long-term program preservation and an issue that public broadcasting needs to take up.

Rights issues restricting access to programs are a major obstacle to long-term preservation, and to better understand the issues, we analyzed them both in the context of restrictions on use of content, and the impact of the Digital Millennium Copyright Act on such critical preservation practices as redundancy and file duplication. Kathleen Maguire at NYU was the primary author of the report.

Intellectual Property and Copyright Issues Relating to the Preservation and Future Accessibility of Digital Public Television Programs, Written by Kathleen Maguire (NYU), edited by Nan Rubin (WNET) and Kara Van Malssen (NYU)
http://www.thirteen.org/ptvdigitalarchive/files/2009/10/IP-Report-fin.pdf

8. Corollary Content Test Activity

Snapshot of website from Nebraska ETV Network, January 2008

Websites are where the public now turns for up-to-date information, resources and background materials, and they have become required as key components for every public television production, both local and national. Public broadcast stations themselves also have active websites, and these have become major sources of original as well as additional program content.

Because these sites contain far more content and audience engagement than any single broadcast can possibly squeeze onto the air, we thought it would be important to try saving them as part of the broadcast record. In the initial proposal, we mentioned the desire to preserve non-program materials, such as Web pages associated with television programs, but the outcome was not specifically defined, and it was not certain that saving this corollary content would be possible.

However, several years into the project the Internet Archive began making their Archive-It webcrawling tool available, and we took advantage of this to capture this corollary content.

Between 2007 – 2009, we did a monthly crawl of more than 400 websites from public broadcasting organizations, stations and program projects. These files are available for viewing from our own website and have since been transferred to the Library.

http://www.thirteen.org/ptvdigitalarchive/our-work/web-crawl/

9.   Outreach to the Public Broadcasting Community

Because the public broadcasting system historically has had no overall commitment to preservation, we knew we had to build support for such values if there was to be any potential for long-term investment in a repository.

This meant introducing and promoting the concept of preservation to the system. Consequently, even though it was somewhat outside the priorities for the NDIIPP program, we drafted a number of deliverables specifically to meet this goal.

These included –
- Inviting participation from stations and other related organizations to support the mission and creation of a repository.
- Write at least one article about preservation for the public broadcasting newspaper *"Current"*.
- Make presentations at regular PTV gatherings on our progress,
- Develop a mission statement for the repository to reflect the needs and desires of the system.

Over the duration of the project, we were very successful in meeting this goal (with the exception of drafting a mission statement for the repository, a task that was transferred to the *American Archive*. [See Appendix G.] NDIIPP Team members were authors of numerous journal and periodical articles, and members became popular presenters at professional conferences and symposia, not only in the U.S. but also at international meetings focusing on technology of television and video archiving.

Presentations were made at a diverse range of gatherings, just a few examples –
- Annual PBS Technology Conference
- Iowa Digital Television Symposium
- National Library of Medicine
- Association of Moving Image Archivists
- Digital Library Federation
- Society of American Archivists
- NSF Blue Ribbon Task Force on Digital Sustainability and Access
- International Federation of Television Archives
- International Association of Sound and Audiovisual Archives
- International Society for Imaging Science and Technology
- Library of Congress National Digital Strategy Advisory Board
- Open Video Conference
- Joint Technical Symposium of the International Coordinating Council of Audiovisual Archives Associations

The clearest indication of our success in meeting this goal was the adoption of *The American Archive* as a brand new initiative of the Corporation for Public Broadcasting. The first effort ever to take on the challenge of preservation for both public radio and television, in 2007 CPB began planning for the *American Archive,* based in part on the example and lessons of the *Preserving Digital Public Television* project.

The initiative began taking shape in 2009, and as this is being written, momentum for the *American Archive* is now well underway with a target of launching its first operations in 2011.

IMPACT AND CONTRIBUTIONS

The impact of **Preserving Digital Public Television** has been far greater than simply designing the mechanics of a repository or analyzing operating issues. Its most exciting aspects were being able to help build a consciousness across public broadcasting promoting the value of preservation, and introducing some of the basic steps that program producers could take to begin preparing for preservation.   Given the absence of any system-wide preservation activities, this alone was a major accomplishment.

From the beginning, project partners promoted a position within public broadcasting that planning for digital preservation was no longer optional – it was a necessity. 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.

Amid such extremely fluid technology, the project emphasized the importance of "S & M" - adopting *standards* for technical operations, plus consistently collecting critical *metadata*. Because these are the very same factors necessary for successful multi-platform digital distribution, the project was able to tie digital preservation directly to effective reuse of program

content. The concept of digital preservation thus became highly relevant to stations, elevating it from a marginal concern to a major subject in the public television debate on how to make content available to reach more viewers.

As existing digital repositories mature, operating costs are being documented by such institutions as The National Science Foundation, which commissioned the *Blue Ribbon Task Force on Sustainable Digital Preservation and Access* in 2007 specifically to study cost models for large database repositories. The contribution of our work has been to focus on how to maintain very large digital video files cost-effectively, keeping in mind the particular needs and operating environments of public broadcast stations. The message has been that the costs of preservation are manageable, and instead of competing with production expenses, they should be integrated into program budgets as an essential element in the life-cycle of a program.

Lessons Learned
Over the course of the project, a number of important lessons became evident.

- Technical problems will eventually be solved and standards will be adopted when private industry agrees to collaborate. But this is a slow and bumpy process.

- With producers beginning to use all-digital production workflows, now is the moment to introduce preservation compliant metadata requirements into the process, and use it as an opportunity to improve and perfect PBCore. This should be done soon or the opportunity might be lost.

- Prompted by the preservation message, stations around the country are actively exploring partnerships with other local cultural heritage institutions to share resources for preserving their respective digital collections. This is an exciting development that should be encouraged and supported by the system.

- Despite a great deal of progress, a system-based commitment to preservation must be reinforced as an important national investment. Instead of being seen as overwhelming, costs need to be presented as feasible and manageable. The creation of the *American Archive* is a big step in meeting this goal.

- Although there are some aggressive efforts to tackle the thicket of rights issues, especially for educational use, overall public television seems unwilling to push boundaries for wider access to archival content. Much more can be done in this area, and we hope the *American Archive* will take up the need to make archival materials much more accessible.

- As the Packard Campus for Audio-Visual Conservation continues to perfect its procedures to accept, ingest, catalog and manage digital video files, we expect that public television will be a major partner in strengthening the national capacity for preserving these files, centrally and through a functional decentralized network structure.

Maintaining Momentum
Since *Preserving Digital Public Television* began, broadcasting has shed its analog systems and moved completely into a digital universe. This project has been able to impress on the public television system the message that digital preservation is not an optional "add-on" cost, but a requirement for any future use of the materials. In this, the project has been instrumental in transforming an attitude of indifference to one that acknowledges the value of properly managing our collective archival holdings.

In a further indication of support, for the very first time CPB allocated preservation funding to pilot *The American Archive. The American Archive* will develop a repository for public radio and television, and we anticipate that our work can make a significant contribution to this initiative.

Viewers keep reminding us that public television programming is precious and has made an indelible imprint. What remains is to continue building commitment across the entire system, so the critical responsibility for saving this American media legacy will be shared, sustained and nurtured over time.

We must also thank the NDIIPP program of the Office of Strategic initiatives at the Library of Congress, in particular Associate Librarian Laura Campbell, Martha Anderson, Director of Program Management, and our wonderful Program Officer Carl Fleishhauer. First, for having the vision to create the NDIIPP program; and then for showing their confidence in our proposal by giving us the first NDIIPP award made to a non-academic institution. This took a leap of faith on both our parts. We are very proud of the significant accomplishments that were made possible through this generous support, and of being a strong partner who contributed to the overall success of the entire NDIIPP venture.

On behalf of the public broadcasting system, we want to express our gratitude to the Library. We hope that through the *American Archive* and a host of other preservation activities now underway, the relationship between public broadcasting and the Library will continue to flourish and grow into the (digital!) future.

<p style="text-align:center">* * * * *</p>

APPENDIX A

PROJECT STAFF

The individuals listed below worked on the NDIIPP Project at some point during the duration of the project. Dates following their names indicate when they left the project.

WNET

Ken Devine, Co -Principle Investigator – VP Media Operations & CTO (2009)
Nan Rubin, Project Director – Special Projects, Technology Planning

Michael Boeglin – Director, IT (2007)
Jonathan Marmor – Media Asset Manager Specialist
Daisy Pommer – Archivist (2007)
Winter Shanck - Archivist

WGBH

Karen Cariani – Director, Media Archives & Library
Mary Ide – Director, Media Archives (2007)
Dave MacCarn – Chief Technologist
Leah Weisse – Archivist, Media Archives & Library

Public Broadcasting Service

Bea Morse – Director, Business & Communications, Interconnection Replacement Office

Glenn Clatworthy – Director, Program Data & Analysis
Julie Fenderson – Sr. Associate, Program Data & Analysis
Jim Kutzner – Chief Engineer
Irene Taylor – Project Archivist (2009)

New York University

Dr. Howard Besser, Co-Principle Investigator – Professor & Director,
　　　Moving Image Archive & Preservation Program
Kara van Malssen – Sr. Research Scholar

David Ackerman – Executive Director, edu Services & Digital Library Program
James Bullen – Digital Library Program Head (2009)
Brian Hoffman - Digital Library Publication and Access Manager
Alicia Kubes – MIAPP Administrative Assistant
Chih-Mei Lin – Digital Library Programmer (2005)

Jerome McDonough – Digital Library Program Head (2005)
David Millman– Digital Library Program Head
Joe Pawletko – Software Systems Architect, Digital Library Technology Services
Unni Pillau - Digital Library Programmer (2008)
Rasan Rasch – Digital Library Programmer

NYU Research Assistants
     Paula Felix-Didier
     Tanisha Jones
     Kathleen Maguire
     Yvonne Ng
     Caroline Rubens
     Pamela Smith

Project Assistance

Jeff Ubois – Archive TV

<p align="center">* * * * *</p>

REPOSITORY DESIGN

EXECUTIVE SUMMARY

This report describes the design and implementation of the prototype Digital Preservation Repository created for the Library of Congress NDIIPP project *Preserving Digital Public Television*.   The prototype Preservation Repository (PR) was developed at New York University by the Digital Library Technology Services team between 2006 and 2009. This report describes the PR design, content aggregation and processing techniques, and metadata plan.

The goals of the *Preserving Digital Public Television* (PDPTV) project were to:
- Build a model preservation repository for "born-digital" public television programs;
- Examine operating issues related to content selection, costs, and access;
- Promote system-wide support within public broadcasting for digital preservation.

The Preservation Repository (PR) at NYU was designed as a "content neutral" system to preserve
a range of content types from a variety of different projects, with "project specific" processes developed as required. This report describes both the "project independent" aspects of the PR as well as the "project specific" components developed for PDPTV.

For this project, particular attention was given to the processes for creating the Submission Information Packages (SIPs), the structure of Archival Information Packages (AIPs), and determining and managing required metadata for each package including the roles of METS, PREMIS, and the emerging descriptive and technical metadata standard PBCore.

A sample of over 80 hours of program and segment files were submitted to the repository, including high-resolution (production quality) program masters from project partners WNET and from WGBH, drawn from such national public television programs as *Nature* and *Frontline,* and lower-resolution (broadcast quality) distribution versions of the same programs supplied by PBS. Metadata was also collected and aggregated from each partner institution for each program. This allowed the repository to collect, organize, process, and ingest a wide range of program file encoding formats, with different wrappers and metadata.

Over the 5 years that this project has been in place, we have learned many lessons from building and operating the repository model. We were successful in solving many of the problems inherent in designing a digital preservation environment for broadcast files, including how to handle multiple file formats, different database exports and metadata structures. At the same time, we learned that functions cannot be automated without uniform standards for processing metadata or file formats, and this is one of the biggest problems that public broadcasting will have to solve in order to design a successful long-term digital repository.

Note that the solutions and processes described here are very specific to New York University's context. While they certainly may provide guidance to a future system-wide preservation repository, actual implementations will inevitably differ.

* * * * *

APPENDIX C

SUSTAINABLE PRESERVATION OF DIGITAL PUBLIC TELEVISION

KEY FINDINGS

Preserved assets can be a great benefit to the public television system, and each investment towards preserving programming can guarantee renewed value and long-term access to this material.

Key findings related to the value of a preservation repository:

1. Sustainable preservation requires both a sound technological infrastructure as well as creation and distribution practices that support long-term access.

2. Proper preservation, including rich metadata creation and management, will continue to add value to public television assets over time.

3. If properly managed, the benefits of preservation will ultimately outweigh the initial and ongoing costs of establishing a preservation repository. Preservation enables future access to materials, which not only serves the public television mission, but can result in cost savings for producers and increased revenue from users.

4. The costs associated with long-term preservation of digital public television content can be affordable, especially if they are seen as a very long-term investment. Moreover, if public television is interested in using its content in the future, the cost of implementing preservation practices now will be far less expensive than trying to recover lost or damaged materials later.

5. Sustainable preservation requires ongoing, reliable, and sufficient financial support.

6. There are a variety of operating models that exist within public broadcasting that can maximize efficiency and allow costs to be shared across various constituents, interest groups, and potential users of the materials.

7. The needs of preservation are not in competition with those of program production. Because preservation is *necessary* for future access to digital content, preservation-compliant production workflows must become integral to an all-digital production process. This point signals an important paradigm shift, where digital preservation is not an optional "add-on" cost to production, but a requirement for the ongoing usability of the materials.

Supporting a sustainable preservation environment for the public broadcasting system is both *feasible* and *manageable.* The system already possesses the resources and structures that can make this possible -- existing appropriate governance models; sufficient technical resources to operate such a facility; and diverse funding options capable of sustaining its operations.

* * * * *

**INTELLECTUAL PROPERTY AND COPYRIGHT ISSUES
RELATING TO THE PRESERVATION AND FUTURE ACCESSIBILITY
OF DIGITAL PUBLIC TELEVISION PROGRAMS**

DISSECTING THE IP ISSUES OF TELEVISION PRODUCTION

Public television's far-reaching influence and irrefutable cultural, educational, and entertainment value makes it a critical component of the American broadcasting landscape. The need for long-term preservation and access to public television's programming is obvious. Unfortunately, preserving and providing access to this content is complicated by a variety of technical, material and legal problems.

By analyzing the overlay of copyright permissions and legislative proscriptions that apply to using and preserving digital program content, the research contained in this paper aims to elucidate issues that may serve as impediments towards achieving system-wide preservation and access of digital content.

Our report on Intellectual Property and Copyright Issues --
- Identifies the context of rights agreements, procedures and economics by which public television programs are produced;

- Provides a detailed explanation of the many types of rights and permissions that can exist for any individual program, particularly the critically important issue of underlying rights;

- Discusses the technology sections of the copyright law and their impact on digital preservation;

- Looks at current efforts to amend the laws;

- And presents case studies of two different public television series that had to address major copyright issues before they could be reissued for re-broadcast and nonbroadcast distribution.

The report focuses on the issues associated with U.S. intellectual property and copyright law that can compromise the archiving, preservation and future access to public television content. We present this discussion to support efforts within the public broadcasting system and elsewhere to solve the complex IP issues which can keep content locked up and inaccessible, despite the expanding opportunities to reach totally new users made possible by the technology of today.

With the immediacy and ease of use of digital technologies, the opportunities for viewing video online have grown exponentially. Users now expect that both recent and older broadcast content be easy to find and viewable. However, apart from the cost of transferring

older analog video into digital files and preserving the new digital content, there are significant intellectual property and related issues that both complicate and restrain long-term access to public television content.

In its examination, PDPTV has found two major areas relating to intellectual property that have a direct impact on the ability and interest of public television to preserve digital content and make it accessible to current and future users.

- **Constraints on usage based on complex expired permissions** – Public television programming can incorporate elements from a broad range of sources, governed by a wide array of license agreements and contractual obligations. These rights permit broadcast distribution by public television stations for a limited period of time, but most of these program contracts do not include permission to make the materials available for non-commercial use after the broadcast rights expire.

  It can be very burdensome to make this content available again after the initial broadcast period has expired (such as to put them online), because renegotiating new permissions can be time-consuming and prohibitively expensive. Broadcasters may not see the point of investing in preservation if they can't reuse or provide access to the material, and may be reluctant to even continue incurring the ongoing costs of storing the material. Not having clear permission to re-use older productions is a major barrier that leaves the majority of the broadcasters' archival content largely inaccessible after broadcast.

- **Legal limits on making copies of programs for preservation purposes** – U.S. copyright laws forbid making unauthorized copies of any copyrighted materials, with a few very narrow exceptions. Copying by libraries and archives is one such exception, but the current law allows a maximum of only three copies to be made for preservation purposes, and only by libraries or archives that are open to outsiders. As such, it can restrict digital archives from following proper contemporary archival protocols, which require routinely creating multiple copies of files through basic operations, and relying on file redundancy to assure long-term digital survival. These restrictions have potential to impact a repository's preservation mandates, and add further complications to legal questions around providing access to archival materials.

When archival collections are made available for online access and other uses, the value of preserving the materials becomes evident. Overcoming the obstacles to providing access is central to expanding support for program preservation across public broadcasting, not only for the long-term survival of public television's collections, but also to strengthen the commitment of public television to protect its substantial body of work for future use.

* * * * *

CURRENT NEWSPAPER

Everything old can be new again

Preserving and archiving your local video legacy is the first step toward the perpetual access that people now expect

By Nan Rubin

Have you noticed that kids—and many adults, too—think every article ever written and every song ever sung is on the Internet?

It won't be long now before young people will grow up assuming that every *TV program* ever made is online, too. That's what they will expect.

It's not true, of course. But in terms of real-life impact, it might as well be. Massive numbers of Americans already look no further than their favorite search engines to find a huge selection of audio, music, articles and video.

They have no idea what is missing—what *isn't* online.

"There's an illusion being created that all the world's knowledge is on the Web," said Edward L. Ayers, a grad school dean at the University of Virginia, "but we haven't begun to glimpse what is out there in local archives and libraries . . . materials that are not digitized risk being neglected, virtually lost to the great majority of potential users."

James J. Hastings, director of access programs at the National Archives, goes even further. "If researchers conclude that the only valuable records they need are those online, they will be missing major parts of the story. And in some cases they will miss the story altogether."

If the common expectation is that everything is online now or will be soon, where is public broadcasting? Sadly, we are virtually absent.

Despite public television's mandate to "inform, inspire and educate," most of our important and memorable recorded treasures, produced at significant cost, are never seen or heard again after their brief, shining moments on the air.

Some people in our audiences don't forget, however. With great affection, our programs live for years, even decades, in the memories of viewers who wonder why they have completely disappeared.
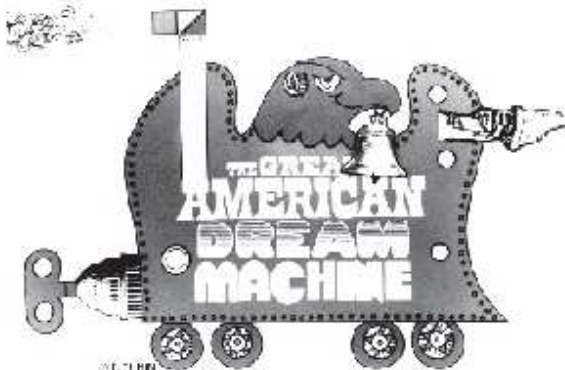
The first of TV's magazine series is stuck in the minds of viewers who remember it but stuck on the shelf, out of viewers' sight.

A legacy of "lost" programs

In the dynamic digital environment, we can live up to our mandate by making these programs viewable again. Doing so will transform the image and service of public broadcasting, especially to the millions of potential new supporters who aren't attached to public broadcasting and never saw these programs. They would be delighted to watch them for the first time.

Most of the requests to bring programs back come from an older public and their memories of television.

Eyes on the Prize, for example. Originally aired in 1987 and 1990, this inspiring history of America's civil rights movement quickly became one of the most acclaimed documentary programs ever aired on our stations.

After it had been unavailable for nearly two decades in broadcast or video, a new generation of students was demanding to see it. Only because of the sustained public outcry and with special fundraising efforts was the series re-released.

Another example: *Lathe of Heaven*. This dramatization of a well-known science-fiction story was broadcast in 1980 as a product of WNET's Television Lab. Its large following asked repeatedly to see it again. Finally, more than 15 years later, WNET cleared the

rights in 2000 and, with great fanfare, the drama was rebroadcast.

But most programs aren't so lucky. The classic 1970s series *Great American Dream Machine* is stuck in the public's mind and still stuck on the shelf. As one of the very first television magazine programs, *Dream Machine* was grandparent to every irreverent sketch comedy, music video and political satire program broadcast since. Its structure of short segments would make it ideal for online viewing.

But three dozen reels of 2-inch videotape are sitting in a warehouse, labels peeling off and magnetic coating flaking. Its satire is completely unknown to a younger audience that would likely find the series enormously entertaining and revelatory.

Ten years ago, Ron Hull, a former director of CPB's Television Program Fund, reviewed the PBS tape inventory and pulled together a list of more than 1,500 programs worthy of preserving *at that time*, such as the Nixon/Kennedy debates, Chicago Conspiracy Trial, Henry Fonda as Clarence Darrow, the *Dick Cavett Show* and *Cosmos*. Just think about the priceless content that's been added since then!

National programs aren't the only ones worth saving. Nearly every station and production unit across the system has its own video legacy to preserve and revive. A

number of organizations are in various stages of assessing and organizing preservation and access plans for their own program holdings. Among them are WTIU at Indiana University; Rocky Mountain PBS in Denver; WILL at the University of Illinois; Native American Public Telecommunications in Lincoln, Neb.; and KQED in San Francisco.

A good time to jump in

Advocates for local efforts like these are looking for assistance and ways to learn from each other. Combined with the potential of PBS's planned digital interconnection system to support shared storage, and the repository planning project Preserving Digital Public Television that I manage at WNET, resources for archiving are evolving rapidly.

The American Archive is a new initiative proposing to bring these elements together to assist public broadcasting as a whole. The archive, as proposed by the Association of Public Television Stations, would "harness the power of digital technology and telecommunications to preserve public broadcasting's audio, film and video history and make it available to the American people."

It aims to give people universal, transparent access. A viewer could simply search, point, click and view. On the other side of the interface, of course, media professionals know that creating and maintaining such a repository is not nearly that simple. Archiving digital productions is more complex than putting tapes on a shelf. There are many technical, operational and legal problems to solve. With eventual preservation in mind, producers must introduce new procedures at the very start of production.

But because stations are completing their shift to digital distribution, our procedures are already changing. With some thought, we can plan for preservation *and* access to lost programs at the same time we're going digital.

Can we afford it?

First, a word about costs.

Many people assume the biggest costs for preservation is in file storage. But this is not so, as storage capacities keep growing and costs drop. In fact, it's cataloging that is really the biggest expense—assigning the metadata to programs so we'll be able to search for and retrieve them.

There is no way around the need for metadata. Digital files are simply useless without it. Even though automated systems are being developed to do some of the work, some manual cataloging is necessary, which is costly because it's labor-intensive. We have to find more cost-effective means to

solve this problem.

Second, a word about monetizing our collections.

Don't get carried away with visions of making lots of money from your archives. Certainly a few programs will be able to earn impressive sums if we can make them available again, and some will be able to earn money online. Others can be repurposed for educational use, which also has the potential for steady sales. Realistically, though, our archival tapes are not going to generate a huge income.

Taking steps

Though our Preserving Digital Public Television project is well underway and the American Archive is being planned, stations and producers shouldn't wait for directions from national organizations.

Most of the work for access and preservation will have to done by each institution at home. Indeed, this local experience will help shape and inspire the national effort, which has always been seen as a cooperative and collaborative venture.

Here are some steps you can take now or at least prepare for. To help, our project offers resources at www.ptvdigitalarchive.org.

Assess your tape library. Stations, distributors and producers have hordes of tapes sitting on shelves and crammed in closets. It's time to do a first sort. Before you digitize anything, figure out what's worth keeping and get rid of the rest. If you are lucky, you may have an actual database of tapes, but it's more likely that there are only a list or rough notes somewhere. (No, for the first pass you won't have to watch the programs.)

We have resources that can help you decide how to evaluate what you have, so you can decide what to keep, what to dump, and what needs more thought (and maybe research). Two papers on our website—"Report and Recommendations on Archiving Television Assets at WTIU" by Lisa Carter, and "Recommended Appraisal Guidelines" by Mary Ide and Leah Weisse—might be helpful places to start.

Your first cut could leave the job much more manageable than it looks today. When WNET moved to a new building in 1999, we reduced a chaotic collection of 60,000 tapes by half and made close to $10,000 from recycling.

Index it with metadata. If a program can't be searched for, it can't be found. Though you can't paste labels on digital files, you can impose order by entering metadata—program descriptions and other information. In the future, your producers will enter this data much earlier in the production process, making it even more useful during the life of

the program material.

You'll need a database to organize your collection, but you can start with simple and off-the-shelf software such as Filemaker. To make your data accessible to other public broadcasters in the future, you can use the CPB-funded PBCore "dictionary" of metadata terms developed specifically for public broadcasting. The website www.pbcore.org is starting to collect highly useful data entry templates, and we encourage people to learn how to use it so it truly becomes the standard for public TV and radio.

After indexing, you can put the database online even if you can't make the programs themselves fully accessible there. People will still be able to find the shows.

Remaster it. Unfortunately, there's no way to keep analog tapes from deteriorating. Tapes in older formats still have to be sent out of house to be remastered. It still costs $300 to $700 an hour to have a 2-inch quad or ¾-inch tape properly cleaned and copied.

But these services will soon be creating files in the standard formats we specify, so newly remastered programs can be transferred directly into an asset management system at the same time. *Then* you will be able to watch the programs!

Meanwhile, some funders are keenly interested in projects that could liberate this content for the public, and they may bring new money to the task.

Digitize it. Stuffing video files into digital asset management systems is getting easier and easier. Soon, every station will have one DAM thing or another to help manage its program library. It won't take too much additional planning to set up an ingest station that can digitize tapes into uncompressed or lightly compressed video files for longer-term preservation, while low-resolution proxies are used as access copies.

Here, too, standards are important, and a few initiatives are underway to propose technical standards for file formats and storage "wrappers" that the system can use.

Look into who holds rights. Yes, rights-holders' agreements can be a costly obstacle to restoring some of our most valuable programs, especially performances. Don't wait for this to get neatly sorted out, because it won't be.

Where possible, we should use educational use rights and other distribution rights that we already possess. It may take some bold moves, but many of our programs, or at least segments, could probably be made available right away.

In the meantime, we should pursue a policy agenda such as the one outlined by APTS, which proposes to free up public access to public TV programs without taking away

reasonable compensation from rightsholders. For example, we need rights agreements that would permit ongoing noncommercial use of our programs in classrooms or for private viewing, regardless of delivery method. We even need explicit rights simply for preservation, protecting access by scholars, educators and others.

Show it off. When you're raising money to pay for remastering or other expenses, don't forget that one of the best ways to promote your holdings is to show them off. People love to watch old TV shows, and they will get quite excited at the prospect of helping you restore a library of gems. When you have a few programs remastered, give potential supporters a taste—show them a short clip reel.

Once the word gets out that you're bringing back old programs, folks with amazing memories will emerge, bringing a wealth of history, contacts and resources.

A time for new partnerships

It's time to get over our wasteful habit of letting our programs vanish forever. We've got decades of national and local productions sitting in storage, and the public is hungry

for them. Making programs accessible will generate great goodwill, new audiences and new funding.

The technology is here and increasingly affordable. Asset management systems for digitizing, organizing, storing and retrieving media are becoming more powerful, and storage costs are plummeting. On-demand viewing is developing rapidly. In public broadcasting, separate initiatives to develop standards for metadata, file formats and networks for non-real-time program distribution are coming together under the umbrella of the American Archive.

But even with all these pieces falling into place, the job in each community is too big, too costly and too important for stations to go it alone. We need new partnerships for preservation.

Approach local universities, libraries, archives, historical societies and museums to join forces, sharing expertise and resources. They face similar challenges in digitizing their collections, developing standards for access and preservation, designing storage facilities and figuring out how to pay for it. They, too, can't do it alone. There may be additional opportunities to team up with commercial entities that have similar needs

and interests.

The objective: a network of trusted digital repositories that can rescue our programs and keep them safe. At the same time, these archives can restore public access to these programs, introducing them to new generations and bringing a whole new set of viewers to public television. ∎

Nan Rubin manages special projects in technology planning at WNET in New York. She is project director of Preserving Digital Public Television, developed in partnership with Boston's WGBH, PBS and New York University, a leader in designing digital libraries, to design a model preservation repository.

The three-year project is funded by the Library of Congress through its National Digital Information Infrastructure Preservation Program (NDIIPP.) Rubin thanks B Morse and Irene Taylor of PBS, Kara van Malssen of NYU and consultant Jeff Ubois for contributions to this article.

For resources and information about the project, go to www.ptvdigitalarchive.org.

LIBRARY TRENDS JOURNAL, WINTER 2009

"Preserving Digital Public Television: Not Just an Archive, But a New Attitude to Preserve Public Broadcasting"
NAN RUBIN

"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)

From Analog to Digital: The Transformation of Television Production

In the 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 early in their development:

> By nature and necessity, public broadcasting is a hodgepodge 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 and production operations would be altered so profoundly or so rapidly by digital production and distribution technologies. In less than a decade, analog television has become totally obsolete by the availability of video recording and editing systems at prices within the means of most members of the public, the wholesale requirement of digital-only transmission, the ubiquity and immediacy of media on the Internet (in the form of podcasts, vodcasts, YouTube) and the mass deployment of hand-held devices to view and listen to audio and video content.

In a relatively short period, television production and distribution has shifted from a linear, sequential, analog process based on physical media, to one that is almost entirely digital, which means easily programmable, random (non-linear) access to content. Virtually all programs are now shot and edited in digital forms, and completed programs are finalized as digital files.

Distribution and transmission have been equally transformed. The Public Broadcasting Service (PBS) is replacing tape-based submissions for national program distribution with an operational system that will transfer digital files instead, using broadband networks able to handle large video files cost effectively and with integrity. Likewise, nearly all local broadcast playback is now done in a tapeless environment, in which programs are stored as files then assembled and aired directly from a server.

The viewer environment has shifted as well. *The Deficit Reduction Act of 2005* (Public Law 109-171, 109th Congress) requires all U.S. full-power television stations to turn off their analog transmitters by February 17, 2009 and begin broadcasting exclusively on digital channels. This means all analog over-the-air television signals will end, and Americans who want to watch television will need a digital

television receiver of one kind or another.   The all-digital television chain will be complete, from program producer at the front to the viewer at home at the end of the line.

Not only have production and distribution methods been radically altered, the marketplace itself has changed. Statistics showing that a third of all Internet users ages 18 to 29 watch or download a video online every day (Madden, 2007) challenge television's traditional concept of the audience. This has forced public broadcasting to examine nearly all its existing operating models, from over-the-air "appointment" program schedules and geographic market separation, to its traditional mix of support from government funding, corporate underwriting, private foundations and viewer contributions.

As viewing has shifted away from television and onto the Internet, there has been particular scrutiny of how to expand the reach of public television content beyond the broadcast schedule. This has meant not merely putting finished programs on-line, but challenging stations and producers to use the Internet to highlight more video, share other program-related materials, and invite user-generated content.

 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 barrier to accessing older content, but rather as a guardian protecting that content and keeping it vital.

Digitally produced programs in public television are at great risk of being lost, however, and not solely because of the rapid changes in technology that are rendering digital video recording and playback systems obsolete. The relatively short duration of program broadcast rights (often no more than 5 years) further complicate preservation, as rights associated with uses *after* broadcast are often absent. Combined with the great array of original source materials used in many programs and the complex rights status that may be attached to each show, the public television system faces a potentially tremendous financial burden to make programs viewable again.

Established practices that have served to archive and protect analog television programs on videotape cannot be used for the long-term preservation of digitally-produced broadcast programs. Exponentially more elements are created in myriad formats throughout the lifecycle of a program, and, critical metadata are created in databases, spreadsheets, and on paper as well as many different types of electronic records. Leaving archiving to the end of the lifecycle opens the door to a host of threats including dissociation, interoperability and migration problems, and obsolescence, not to mention managing a volume of unused materials that far exceeds that of the analog world. A new approach that incorporates preservation practices into the entire digital production chain needs to be created.

Preserving Digital Public Television: A Significant Collaboration

As part of the Public Broadcasting Act of 1967, Congress authorized the Corporation for Public Broadcasting (CPB) "to establish and maintain, or contribute to, a library and archives of noncommercial educational and cultural radio and television programs and related materials." However, CPB has never chosen to make that investment, and up to now, CPB funds have never been allocated for archival support.

Early agreements, such as those with the Library of Congress and the National Archives and Records Administration, have provided some preservation security for completed national programs aired through Public Broadcasting Service (PBS) and National Public Radio (NPR), but access is extremely limited. Moreover, local programs, which more closely reflect our daily lives, have not been included in these collections.

No mandate for system-wide preservation exists at any major public television institution, and there has been a notable lack of funds to invest in or allocate for preservation. Consequently, no single entity has

the resources or expertise to manage this task by itself. Only a small number of stations and producers have been able to take on the responsibility and costs of preserving their own materials, so that within public broadcasting, very few formal archiving activities are in place. This leaves planning for digital preservation as an afterthought in the lifecycle of public television programs.

As one of the major producers of national programs in public television, Thirteen/WNET, New York's public television station, recognized this challenge and in partnership with WGBH in Boston, was awarded an NDIIPP grant from the Library of Congress in 2004 to design a long-term preservation repository for "born-digital" public television programs.

These two television stations produce roughly 60% of the national prime time series that appear on public television, including signature public affairs and science programs such as *Frontline* and *NOVA,* which originate at WGBH, and leading cultural and historical materials, such as *American Masters* and *Great Performances*, which are produced at Thirteen/WNET. As 'content creators,' the stations control their own production units and thus have direct access to high-resolution 'master' files of completed national programs.

In addition, WGBH was already recognized as a leader in identifying issues related to digital video preservation, and it brought a demonstrated history of developing and promoting digital asset management systems within public television.

WNET and WGBH maintain the only two station-based archives in the public television system, so each was also able to provide the project with professional staff experienced in preservation practices related to both digital and analog video materials.

The third partner was the Public Broadcasting Service (PBS), which operates the national network that distributes public television programs to more than 300 stations for local broadcast. Because the majority of national programs pass through PBS before being aired locally, it is the principle *de facto* repository for this material, and the PBS warehouse holds more than 150,000 analog tapes of programs going back more than 40 years.

These three institutions are directly engaged in both managing production workflows and holding the primary collections of public television programs seen by national audiences, and their interest in participating in NDIIPP was based on a shared recognition that public television had to begin taking steps to protect its rapidly growing collection of digital assets. Because they are primarily broadcasters, however, these institutions have few resources to operate digital libraries or develop preservation repositories.

New York University provided the expertise that was lacking in these areas. The NYU Digital Library team has 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

The intent of the project was to develop a small model repository that could be scaled up and leveraged for use by the larger public television system. Because public broadcasting as a whole has little exposure to preservation issues, it was also thought that the project could offer valuable resources to the system by drafting guidelines for content selection and appraisal, studying copyright impediments, and examining relevant financial and governance models. An unstated though larger and more important 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.

In this context, *Preserving Digital Public Television* was designed as a series of discrete tasks to be tested in a lab-type environment. The initial set of activities specified:

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

Because the repository concept focused on capturing metadata during the production process, one early task was to examine production workflows. This would identify the points where key metadata was created at various stages, and see if it was carried through the program lifecycle and ultimately used for preservation. Further activities centered around researching copyright issues involved in digital public television preservation, and capturing websites that are an extension of the program content offered over the air.

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 have to be tested and collectively agreed upon. The project naively assumed that both commercial television networks and large collecting institutions like the Library of Congress, which was completing its new Packard Campus of the National Audio-Visual Conservation Center, were making progress on solving these same problems so that public television could benefit by contributing to work already underway.

The project quickly learned, however, that this was not the case. Other video producers including the networks, as well as the Library itself, were in fact struggling with the same technical issues, but no organization was making enough progress that public television could just follow along. Instead, 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.

As for the intent to bring a new consciousness to the system, our project has had a broad impact among public broadcasters by sparking widespread support for launching local as well as national preservation efforts. The culmination of these activities is leading to the creation of The American Archive, a new national entity which has been proposed to provide preservation services for public radio and television program producers for the very first time.

Project Activities

Designing the Test Repository

The expertise of the NYU Digital Library team in building other archives with similar content was leveraged to design the repository architecture for this project. To test ingest and retrieval for the repository, WNET and WGBH selected a sample of individual high-resolution production master files from national program series that included hour-long episodes of *Nature* and *Frontline,* and episodes of *Religion & Ethics Newsweekly,* a half-hour each. Several episodes of the local half-hour WNET program *New York Voices* were also selected as a sample, representing non-national programming that did not go through PBS.

Before sending a program out to stations for broadcast, PBS processes the file by adding elements such as underwriting announcements and compresses it for easier distribution by satellite or other means. To complement the production masters of the programs provided by WNET and WGBH, PBS provided the repository with the low-resolution distribution versions of the same programs (except *New York Voices.*)

Figure 1
Repository Information Packages – Source Files
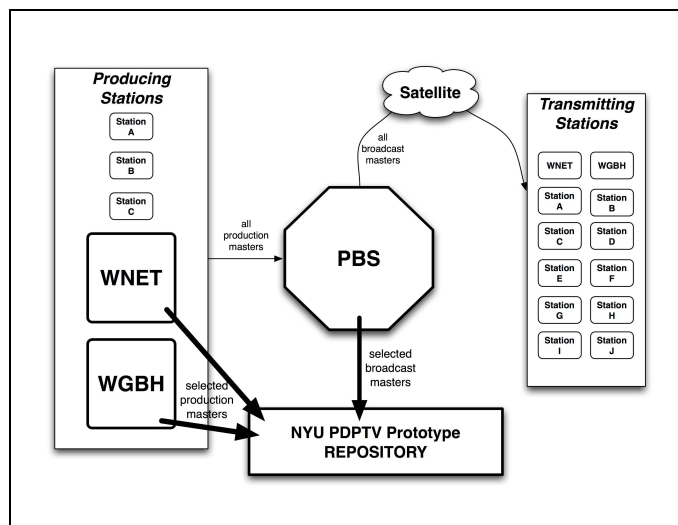
This grouping allowed the project to test of a mix of both high and low resolution program file formats created with varied workflows, accompanied by a wide range of metadata that was available but had to be collected separately from multiple sources. Altogether, more than 35 hours of programming was ingested into the repository which included multiple versions of single programs stored in different formats.

The initial tests of the repository revealed that, despite the uniform requirements from PBS for submitting completed programs for national broadcast, the finished materials coming directly from the production units had to be ingested in several different flavors of formats and wrappers (MacCarn, 2007). These were not all equally easy to extract technical, metadata from or to read.

The project also found there was very little consistency in the way metadata for each program was recorded, as production units created certain elements of metadata, PBS generated others, all of which was created in different places for different purposes (i.e. production vs. distribution.) This 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 program-by-program basis.

Determining Metadata Needs

Determining an appropriate set of metadata fields was a detailed and intensive task. Based on the assumption that the Preserving Digital Public Television repository should be OAIS-compliant, the project examined a broad range of standard metadata schema used by libraries and archives, as well as those emerging in the commercial television world. Project staff also reviewed PBCore (http://www.pbcore.org/), a metadata dictionary based on Dublin Core designed specifically for public radio and television program files.

One of the fundamental requirements of the repository was to aggregate content and metadata for a single program that came from disparate sources, both from the producing station (the high-resolution

production master and database exports) and from PBS (lower-resolution broadcast master and more database exports).

Because the model repository does not input any new metadata, in order 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 appropriate metadata as well as the program files themselves.

Figure 2
Repository Information Packages
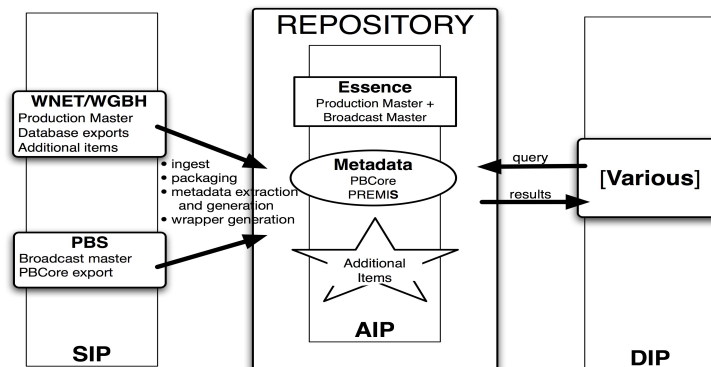
To determine the metadata components required for the AIP, database exports from both program creators and PBS had to be analyzed, particularly the extensive descriptive and rights metadata created by PBS for broadcast scheduling.

Early on, the project partners chose PBCore to capture 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, which has encouraged other users to design and implement PBCore-compliant exports from their various databases. As a result of this effort, the most important source of metadata for national programming, PBS's Program Offer Data Service (PODS), can now be exported directly into PBCore, making information easy to ingest, package, and disseminate.

Incorporating rich technical metadata from the video files also proved to be a challenge. Because the program files were submitted to the repository in diverse formats (including a diversity of wrappers and encoding formats such as MXF, Quicktime, and various flavors of MPEG and DVC Pro), multiple tools were required to play the videos and to extract technical metadata from the file headers such as bit-rate, filesize, and frame size.

Transforming the submitted metadata into a standard format was a clear necessity. The solution was to create a schema that encapsulated the necessary descriptive and technical metadata from a variety of different data dictionary standards, while maintaining information unique to public television programming.

To accomplish this, the repository has developed a structure to capture all necessary metadata using not only PBCore, but also PREMIS and METSRights. Appropriate fields from these standards, along with virtual links to the program files themselves, are all contained within a METS wrapper.

Figure 3
Metadata Schema
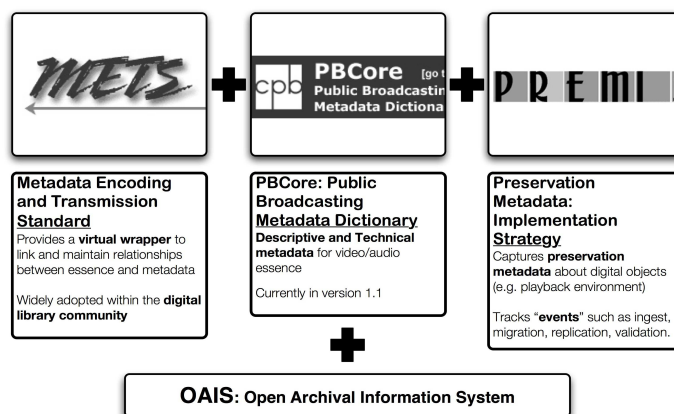
The problems that were encountered by testing these various file formats, combined with the time-consuming efforts needed to collect metadata, demonstrated the high priority for setting uniform standards as a requirement for the success of any future repository operation. Without a set of accepted standards to apply to the creation of uniformed metadata, the program file formats and descriptive metadata, automating the functions for extracting and managing the metadata of large collections will simply not be feasible.

Video File Formats and Wrappers

The creation of a relatively stable standardized video file wrapper is needed to facilitate the successful exchange and long-term preservation of digital files, particularly to support future file migration. A number of such so-called video wrapper 'standards' exist, but despite vendor claims, the files do not all actually interoperate with different equipment.

This problem was identified early as one that the project wanted to address, because there was no single acceptable wrapper when the project was initiated. To encourage a shared approach to solving the problem, the project convened a group of more than twenty distinguished technologists, digital collections designers and project managers. At the "Wrapper Roundtable," the public television partners were surprised to learn that the lack of consistent video format and wrapper standards was also a major problem shared by the commercial broadcast networks. (Fleischhauer, 2006)

To be successful, any initiative to create standards for broadcast files 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 adopt and support the desired standards.

This led members of the NDIIPP team to a series of exploratory meetings with such diverse groups as Turner Broadcasting, the Discovery Channel, and the Department of Defense, and eventually resulted in the Advanced Media Workflow Association (formerly the Advanced Authoring Format Association which represents system vendors) launching the AS-03 wrapper project specifically to ensure "…use of MXF as the file wrapper to facilitate a vendor-neutral format to support interoperability." (http://www.aafassociation.org/html/projects.html)

Analyzing Program Production Workflows

With the understanding that preservation and other valuable metadata must be captured early in the process and carried through the entire program lifecycle, another aspect of the project was to examine production workflows and identify the points where key metadata are created so that preservation practices could be integrated throughout.

The different workflows associated with the test programs were analyzed to identify how this might be done, and although they could not be altered, the workflows in place at PBS for accepting and preparing program files for distribution were also examined.

Production workflows are complex and varied, depending on the type of program and its style. For example, *Religion & Ethics Newsweekly* is a current affairs magazine comprised of several different segments packaged together to create one broadcast episode each week. It both produces is own segments and acquires materials from other sources, and like many news programs, it must be edited and assembled quickly, with programs often being completed barely by deadline. By contrast, *Nature* programs are often shot and produced by just one entity, with a deadline of a year or longer.

Each production unit has its own practices and keeps records in its own way, and to get an idea of how many different program production units exist, just look at your weekly public television schedule. There are as many units as programs. The volumes of information accumulated during a production are a mix of paper and electronic records based on legacy workflows learned or inherited by the producer and then added to by PBS, with no uniform databases and no standardized recordkeeping shared between or across any productions.

There is no 'one size fits all' solution for incorporating preservation practices into such diverse workflows, but we initially thought that a suitable approach might be adapted to test on one or two programs. This idea proved both elusive and intractable.

Apart from the inefficiencies of production workflows that largely replicate analog procedures in a digital environment, it was simply not possible to propose to any production units of our selected test programs that they alter their workflows to accommodate the needs of long-term preservation. By and large, individuals involved along the production path are not inclined to see the value of adding additional work to their already demanding tasks, especially if they are working against a deadline.

More importantly, the tools necessary to make these changes were not in place. Without a mechanism that makes it easy to ingest, store, browse and view digital video, creators do not have enough incentive to change their practices. But currently, neither WNET nor WGBH has an appropriate Digital Asset Management system (DAM) to manage the digital production workflow. Until such a system for storing program files, content elements and metadata during the production process is implemented, workflows cannot be significantly altered.

Content Selection and Appraisal

On a practical level, examining issues related to selecting public broadcasting content and drafting recommendations proved to be one of the easier aspects of the project to complete. Through a series of focus groups, teachers, educators, journalists, documentarians, academics, and others who use television in their work provided their input.

'The project partners also conducted a literature search, and collected operating policies from a number of film and video archives. This research resulted in the publication of *Recommended Appraisal Guidelines for Selecting Born-Digital Master Programs for Preservation and Deposit with the Library of Congress* (Ide & Weisse, 2006). This report outlines existing and emerging practices in moving image appraisal that are appropriate for use with national public television programming.

The project team determined that local stations wanted additional guidance in appraising their own local program productions. *Appraisal and Selection Guidelines for Public Television for Public Television Volume II: Criteria and Recommendations for Local Stations, Producers and National Productions* (Preserving Digital Public Television, 2008) focuses on specific criteria to appraise not solely national materials, but also elements and program versions of local productions.

Knowing that the Library of Congress already had a vested interest in collecting the 'best copy' of programs distributed nationally by PBS "in complete and unedited broadcast format…" (PBS-Library of Congress Donor Agreement, 1993) has also shaped the selection recommendations, so that recent programs like *The War* and *African American Live* will be able to join the ranks of long-running series like *NOVA*, *Live from Lincoln Center* and *P.O.V.*, along with *An American Family, Julia Child, The Adams Chronicles, Great American Dream Machine* and other older classics as important documentation of the U.S. experience.

Managing Program Rights

Along with issues of storage and technical architecture, the most common challenge in preserving digital video concerns intellectual property and related rights. Program productions are a complex amalgamation of original moving images, stills, unique art, scripted narrative, dialog, recorded sound, music, performance, acquired footage, and many other elements. Each part can potentially carry a long list of underlying rights -- ownership and creator rights, restrictions and authorships, union contracts, distribution agreements and other specific conditions for use. Typically, a producer obtains the right to use a clip of historic footage or music in a television production for a period of only 5 to 10 years. In addition, many public television funding models include co-ownership or shared copyright of programs to relieve the financial burdens of production. One or more entities might therefore own the copyright on a single finished program, while many others hold the mass of underlying rights.

Distribution rights allow public television programs to be broadcast a maximum number of times on non-commercial television during a finite window, for example, 6 showings over the air in 5 years. Generally, rights for a certain level of home video sales and educational use in the K-12 classroom are also part of the distribution rights package.

When these rights expire, the system no longer maintains much interest in the content unless there is an exceptional demand from viewers. Then the rights to broadcast the program again have to be "re-upped" or renewed for air by getting permission from the copyright holder as well as from all the underlying rights holders.

Reuse of the program through any new distribution method such as via the Internet, or for an additional length of time, is often discouraged because attaining the rights to do so can be very difficult. Identifying and locating all the underlying rights holders can entail a great deal of research, and paying for permissions to clear the program for reuse could easily be quite substantial.

For many productions, particularly those that use archival or stock photos or clips, music, etc., this can be very costly both in terms of research to find all the rights holders, and any direct costs that may be required. For example, the important civil rights documentary series "Eyes on the Prize" could not be broadcast or otherwise distributed for a decade because the producers could not afford the cost of renewing all the underlying rights (Dean, 2005; Bernard, 2005).

Specific authorization to preserve national public television programs is largely absent, and seeking permission to put programs into a digital repository for preservation and access is too emergent an idea to have much precedent. At the same time, prompted by public expectations that television programs should be readily available online, many rights holders are trying to create new restrictions to maintain control of their creations out of fear of losing the opportunity to exploit their own materials.

Under these conditions, the model repository would have to stay largely 'dark' unless it has explicit permission from the copyright holder(s) to allow programs to be used for anything other than archival research.

The Preserving Digital Public Television project team decided that a discussion paper relating to access and rights is critical to inform plans for a functional repository in the near future. The paper will include recommendations for drafting new model language for contracts and distribution agreements to authorize appropriate perpetual rights for long-term preservation and non-commercial access to the content.

Planning for Sustainability

A major functional issue for any digital repository is how it will be economically sustained, what expenses will be required for staffing, administration, services and maintenance along with technical facilities, storage, interconnection, migration and the like. As existing digital repositories mature, the wide ranges of these costs are beginning to be documented and analyzed (see Additional Resources). The National Science Foundation, in partnership with the Library of Congress and other large research institutions, is also studying cost models and issues of sustainability.[1]

Because the argument for program preservation is not universally supported in public broadcasting, a key challenge in discussing sustainability is to rely not solely on numbers, but to offer a compelling rationale for making this investment in the first place.

Given the relatively few sources of income traditionally available to public broadcasting, the economics of setting up and operating a preservation repository might seem overwhelming and unaffordable. Instead, the public television community must be reassured by seeing it as both feasible and viable. The case for sustaining a repository must be made without putting it in competition with the immediate demands of station operations and program production, and without being perceived as overly intimidating.

Our contribution to this effort has been to focus on solving the particular problems of maintaining very large digital video files, and to keep the projected scale manageable by describing a service that would initially provide only basic functions but system-wide benefits. The project is also closely monitoring the new research on this topic being published with growing frequency, and will incorporate the most relevant findings in our own projections.

Preserving Websites Because Programs Aren't Enough

Websites are where the public turns for up-to-date information, resources and background materials, and they have become required components for every public television production, both local and national. Websites provide expanded information about any given program, and increasingly they reflect large amounts of original video and other cultural expression. Public broadcast stations themselves have active websites, which provide media and other content that ties them directly to their communities.

Increasingly, websites contain far more video than can be contained in any single broadcast, and they are now considered an integral aspect of the broadcast record. As such, preserving them had to be considered along with saving the programs themselves. With the Internet Archive, at the end of 2007 the project began saving a large collection of websites related to the public television system, including local PBS stations across the country and individual program productions and series. Close to 400 web addresses are being captured on a monthly basis. (http://ptvdigitalarchive.org/our-work/web-crawl/)

Public television's web presence will be transformed during 2008 as the system adopts new web tools that will greatly increase the amount of video offered online. In response, many stations and producers are redesigning and reorganizing their public web sites. Our webcrawling activities will be documenting the evolution of these websites and the impact the changes are having on improving the interaction of

public television with its viewers. As such, they will provide an important additional window into the larger social landscape reflected in the content distributed over-the-air.

Impact of Non-Broadcast Digital Distribution

In designing the repository, the project began confronting important dynamics with impact well beyond the initial scope of work. One of the most significant dynamics has been the transformation of television programming distribution by digital technology, in tandem with viewers' near-universal access to the Internet. These shifts have resulted in rapid migration away from over-the-air "appointment viewing," to on-demand viewing of programs saved on DVDs, downloaded or streamed online.

There has also been an explosion of demand for older broadcast content to be findable online, and an expanding library of such material is becoming steadily available.

When coupled with a constantly changing array of viewing devices, these phenomena have created a fundamentally altered video environment which requires programming to be available as digital files viewable on the very smallest iPod screen, to wall-size flat panels.

Such factors have forced public television to face the question of how to release both current and older program materials into the ocean of online video offerings. Producers began to understand that to be successful amid such extremely fluid technology, they had to come to terms with the need for adopting standards for file formats and file storage; saving appropriate metadata; resolving certain rights issues; and acquiring effective digital content management tools.

Stations began to look for best practices and cost-effective solutions to solve these on-line distribution problems. They found that, although our NDIIPP project had approached these questions specifically through the lens of long-term preservation, most of the problems relating to creating a stable preservation environment overlapped significantly with their own efforts to get their video online.

Preserving Digital Public Television thus became highly relevant to television stations and producers across the system, and in less than two years, the whole concept of digital preservation moved from being a marginal concept directly into being a key factor in the core public television debate on how best to make content available to reach more viewers.

Leading Within the Public Broadcasting System

The steady rise of non-broadcast digital distribution led to a genuine interest in the NDIIPP project. This was not only because of the expertise the project was developing in the technology realm, but also because the project brought with it a wide array of useful resources and contacts from far outside the public broadcasting sphere. Also, through the project, public television was becoming a valuable preservation partner to such institutions as the Library of Congress, the National Science Foundation, the National Archives and Records Administration, and the Academy of Motion Picture Arts and Sciences. The project was helping position the system to participate in significant new funding opportunities.

From the beginning, the project intended to build support by engaging the public broadcasting community. In the early stages, this meant primarily going to public television's annual technical gathering and meeting with station engineers and technologists informally to solicit input. As the system shifted more firmly into an all-digital environment, we became more active promoting the value of digital preservation and organizing broader station participation in industry gatherings.

By 2007, public broadcasters began seeking us out to present at conferences and symposia, and to provide guidance on how to approach preservation planning. When the article "Everything Old Can Be

New Again" (Rubin, 2007) was published, we became the most visible effort promoting preservation, introducing stations and producers to acceptable preservation practices and advancing system-wide collaborations to support PBCore and other technical and metadata standards.

For example, because of the project's high profile, WHYY Radio in Philadelphia asked the project for advice on preparing its library of 25 years of audiotapes of the daily program *Fresh Air with Terry Gross*, for digitizing and cataloging. The station was especially interested in adopting appropriate public broadcasting standards such as PBCore, and did not want to proceed without aligning its efforts with other similar projects in public radio and television.

Major project efforts have also encouraged local broadcasters to create preservation partnerships with cultural heritage organizations in their own regions that share needs for creating and managing digital collections.   Inspired by this message, the Assistant General Manager for Content at NET (Nebraska's state-wide distance learning and public broadcasting network) organized his own regional meeting, bringing together more than 30 participants from NET's Broadcast and Web services, the University Library School, the Nebraska Regional Humanities Center, the Nebraska State Historical Society, and the Historical State Archivist, to explore interest in setting up some form of digital preservation collaboration.

The American Archive: A New National Initiative

These examples illustrate the success of Preserving Digital Public Television in inserting the framework for preserving digital video into the immediate environment of stations operations and program production. But the most dramatic response has come from the Corporation for Public Broadcasting (CPB) itself, which in 2007 initiated a new project to create what is being called "The American Archive." The concept of the American Archive, as outlined by Congressman Ed Markey, is "to harness the power of digital technology . . . to preserve public broadcasting's audio, film, and video history, and to make it available to the American people." (Speech to the Association of Public Television Stations, February 2007).

For the first time, CPB is taking steps to invest funding in preservation activities. The American Archive is being planned as a totally new entity, independent of existing public broadcasting institutions like National Public Radio or PBS, in part as a response to growing momentum within public broadcasting from projects like ours.

It will take several years to shape the American Archive, outline how it will operate, design a governing structure, and build sustainable funding streams. Nonetheless, public television stations across the country have endorsed the project and are now organizing their own resources to contribute to the collective effort.

Maintaining the Momentum

Since this project began, public broadcasting has shed its analog systems and moved completely into a digital universe. Along with the emergence of cost-effective digital technologies for content management and storage, the changes have involved more than swapping tape machines for computer servers, but also adopting totally new approaches to editing, sharing and maintaining program content.

By continuing to have high visibility, Preserving Digital Public Television has been able to impress on the public television system the message that digital preservation practices cannot be an afterthought, but are important components to be integrated into the entire program production lifecycle. The project clearly identified the most basic problems that had to be solved, with success in meeting some better than others, and tied them directly into the broader operating concerns that the system is facing.

On a more basic level, the project successfully challenged public television to discuss the need for preservation in order to keep digital productions alive at all. This had the impact of transforming an attitude of indifference across the system to one that recognized the importance of properly managing our collective digital archival holdings.

In this, Preserving Digital Public Television has played a major role preparing the groundwork for long-term collaborations that have the potential for building widespread support for digital preservation activities. But we cannot go much further without broader system involvement.

The public broadcasting community is ready to express its desire to keep our television content vibrant and useful. The technical conditions necessary to operate preservation repository will soon be solved, and the behaviors needed to add preservation-relevant metadata into program production and distribution workflows will eventually be adopted. What remains is that public broadcasting organizations continue building commitment, so that once preservation projects like ours are underway, they will be sustained and nurtured over time, not solely to serve the system, but because they benefit the American public as a whole.

* * * * *

APPENDIX G

THE AMERICAN ARCHIVE

APTS Helps Create 'American Archive'
Association urges Congressional funding to save PBS treasures

by Mark R. Smith

WASHINGTON

If you want to know why the establishment of an "American Archive" is essential to preserving the history of PBS, check out the promotional DVD that was recently created by the Association of Public Television Stations.

Those old '60s-era clips of CBS-TV news legend Edward R. Murrow welcoming WNDT to the airwaves, an early clip of Fred Rogers with X the Owl and the Beatles performing "All You Need Is Love" during the world's first global telecast quickly make the point.

The problem is that videotape that was used to acquire such treasures dries out and the oxide particles that store the information can literally drop off. So the degrading footage of a speech by President John F. Kennedy that was also included in the DVD underscores the fact that irreplaceable content could be lost forever—unless action is taken.

Today, the APTS is leading the charge for an "American Archive," a proposed digital library of PBS content. With the help of organizations like Thirteen/WNET and WGBH (which have always accounted for approximately 2/3 of the programming on PBS), the association is urging Congress to help fund the project that would determine the correct technical parameters to create a digital archive to preserve old footage, as well as new digital content that was created for the Internet and other avenues.

TALE OF THE TAPE

"The amount of content in the PBS universe is vast," said APTS President and CEO John Lawson, "and we don't have a handle on how much is out there yet."

The stations "still throw out videotape," Lawson said, noting one station in the South that allegedly tossed 40 years' worth of videotape into a landfill. Others haven't indiscriminately discarded their videotape, but haven't stored it properly, either.

"Some small stations store video in the hallway. So even if it has been kept, you don't know how well it's been kept," he said, noting storage in a fireproof vault protected by Haylon gas as one appropriate avenue.

But questions about preservation are just the first issue. The next concerns the digitizing and "metataging" of content to make it searchable and retrievable.

"We're still figuring out how to do that," Lawson said, noting that APTS is also involved in a pilot project with the Library of Congress (LOC), Thirteen/WNET and WGBH called "Preserving Digital Public Television" (PDPT) to explore various techniques for preserving video.

Lawson pointed out that since 2001, Congress has been providing funding for the digital conversion of the physical infrastructure of PBS stations by 2009. He thinks that infusion could provide the right lead-in to establish the American Archive.

"We're nearing the end of the conversion and asking Congress to allow us to repurpose more funding to invest in virtual infrastructure," he said. For fiscal 2008, APTS is requesting $40 million that will be geared toward finishing that part of the job.

"But, over time, the larger percentage of those funds could be invested in the American Archive," Lawson said. "We see this as an opportunity to start partnerships with universities, museums and libraries. There could be public-private partnerships as well."

The PDPT (which has been underway since September 2004) is being funded by the LOC's National Digital Information Infrastructure and Preservation Program (NDIIPP) grant. The project encompasses input from divisions of New York University, PBS, the LOC, Thirteen/WNET and WGBH, as well as eight other partners (*www.ptvdigitalarchive.org/partners*).

GETTING TECHNICAL

Dave MacCarn, chief technologist at WGBH in Boston, has been involved in preservation issues for 15 years. He said "getting the relevant players in the room" is a major undertaking in itself.
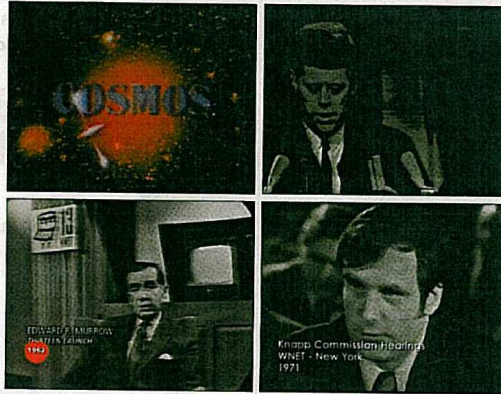
Much of the historic programming held by public broadcasters is in danger of being lost unless the material is properly digitized, catalogued and stored for future access.

"What we need to figure out," MacCarn said, "is how to package digital information, aside from getting audio and video on tape," noting that Avid's open media framework (OMF) was instrumental in the early part of the process, as it evolved into the advanced authoring format (AAF).

"Today, we have a SMPTE-approved standard called the MXF (or Material Exchange Format, a data structure intended to allow for interoperability between broadcast and postproduction platforms) that was built after working with the AAF framework," he said.

APTS

But there are different issues to deal with archiving in MXF, he said, like encoding.

"For example, if content comes in as Windows Media and I retrieve it years later, how would the decoder know what Windows Media was?" MacCarn asked.

Another concerns the fact that MXF is based on a coding format called "Key Length Value" (KLV). The key is like a keyword, the length concerns the data, and the value is the actual data, MacCarn said. "That question concerns who holds the key.

"If someone else eventually wants to access the file, they will have to know how to open the file, but with what keys? There needs to be an international registry to refer to if the information is not in your own library," he said. "That question is being researched by the Global Digital Format Registry at the Harvard University Library now."

Other issues concern how to store files and what a trusted repository for digital files would be. "And how long will it last?" MacCarn said. "A regular computer hard drive is a mechanical as well as an electrical device, so taking that route would include issues over the long term."

MAKING PROGRESS

While questions about MXF exist, PBS is due to move to a file-based distribution system rather than a real-time system via satellite and the network has developed what it calls MXF-Application Specific PBS (or ASPBS), according to Ken Devine, vice president and CTO with Thirteen/WNET.

"PBS is pretty far along in getting that adopted, so that format will work with the servers of all major manufacturers," Devine said. "So we started with that and will stay within that container with the audio and video content that will now include preservation data."

That will happen as soon as any new television program is produced. The preservation data will encompass the business, legal and technical information that will be required in the future.

> "The amount of content in the PBS universe is vast, and we don't have a handle on how much is out there yet."
> —John Lawson, APTS

"We're figuring out how to do that now and make it automatic," Devine said. "Getting humans to handle that part of the workflow is the real problem because they are not used to handling the details," adding that it is not being done elsewhere in the broadcast industry due to a lack of financial return.

But he sounds confident that, despite the hurdles, the PDPT project will evolve into the PBS archive, expand the project to other stations and producers, and eventually make it accessible to the press and the public.

"The initial access that we envision, considering copyright issues, will be through the LOC. The Congressional Reading Room and Presidential Libraries may be included at some point."

For now, Devine said, they want to focus on preserving multimedia and video content that only appears on the Web, then all of the content that has ever aired on PBS.

Then the decades of content will be saved and accessible to the people who can benefit from it most.

"We are creating a loop," Devine said, "that has never been there before." ∎