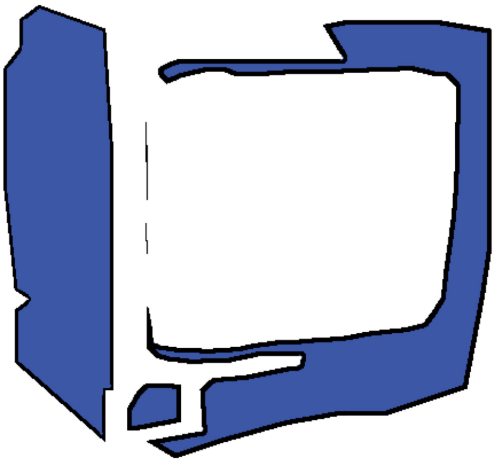

Strategies for Sustainable Preservation of Born Digital Public Television

A Report by the **Preserving Digital Public Television Project**
Part of the National Digital Information Infrastructure and Preservation Program (NDIIPP) of the Library of Congress

February 2010

Written and Edited by
Yvonne Ng (NYU), Nan Rubin (WNET), and Kara Van Malssen (NYU)
With Howard Besser (NYU), Karen Cariani (WGBH), Irene Taylor (PBS),
Winter Shanck (WNET), and Jeff Ubois
Thanks to Joe Pawletko (NYU), David Ackerman (NYU), Alicia Kubes (NYU), and B Morse (PBS)

Table of Contents

Executive Summary

Digital technology has revolutionized media creation. Because of digital technologies, media can be created, transmitted, and distributed much quicker, cheaper, and more broadly than ever before. Along with other broadcasters, public television is now producing and distributing its programs digitally. Today, program materials are increasingly less likely to be on physical tapes or reels of film, but are created, edited, and stored solely as digital files.

New tools have created the potential for immediate access to digital content, and such access has added growing value and benefits to saving the media and being able to reuse it. But long-standing practices used for conserving physical videotapes and films are no longer viable, which means preserving large digital video files brings new challenges and many questions. What does it take for this content to be cared for and preserved? What resources are necessary to ensure its longevity? And how much will it cost to maintain access to these fragile files well into the future?

Making the investment in digital preservation is critical for the survival of public television programming. Although the scope of the task might seem daunting, it is manageable with good planning. The goal of **Strategies for Sustainable Preservation of Born Digital Public Television** is to show that supporting a preservation repository is not only necessary, but can be highly beneficial to public television. Moreover, because preserved materials have growing value, the benefits of preservation will ultimately outweigh the initial and ongoing costs of establishing a preservation repository.

A preservation repository of public television programming will make it possible to:

- Encourage greatly increased demand and new uses for audiovisual content because it is made findable and available.
- Lower access costs to help meet the demand for this content.
- Support the use of content that would otherwise be lost without preservation.
- Generate new income for copyright holders and others as a result of increased uses.
- Preserve culture and heritage for future generations.

This report examines the requirements for long-term preservation of born digital video files. It analyzes the costs and potential income sources associated with maintaining a preservation repository, and argues that the operating costs can be self-sustaining. While it does not specifically focus on technologies or infrastructure scenarios that could support long-term preservation, it provides a high-level look at ALL the requirements of sustainable digital preservation of public television, well beyond simply storage.

Chapter 1 lays out the recent shift in the technical landscape of broadcast television and how the transition from tape to file-based program production has created a new set of functional problems that have to be solved for successful archiving and preservation. **Chapter 2** outlines the arguments for saving production files, identifies different stakeholders who have an interest in keeping the content, and discusses the value created through preservation. We also look at the concept of “sustainability” and explain why it involves more than mere data storage, but also requires a *managed environment* that encompasses long-term institutional resources and commitment.

In **Chapter 3**, the elements required for sustained digital preservation are examined and are presented against the criteria for creating and operating a trusted digital repository. **Chapter 4** introduces the economics of planning for digital preservation, and the implications of applying different economic models to repository costs, based on diverse income sources and stakeholder demands. Extending the economics further, **Chapter 5** presents a detailed analysis of one repository’s costs by dissecting the Preserving Digital Public Television repository design as a case study.

This report challenges the popular notion that the economics of preservation are limited to storage costs and operating income. Instead, it supports the concept of a trusted digital repository that incorporates not only the resources needed to sustain the materials being preserved, but also the systems, activities and staffing required to sustain the repository itself.

This approach to digital preservation is necessary because:

- Sustainable preservation requires not only sound technological infrastructure, but also management practices that support long-term access.

- Sustainable preservation requires ongoing, reliable, and sufficient financial support. These costs can be affordable, especially if they are seen as a long-term investment.

Likewise, managed digital preservation is feasible because:

- The largest costs associated with preservation are related to staffing and personnel, not technology or facilities.

- There are a variety of successful operating models that already exist within public broadcasting that could be used to structure repository operations and finances.

- The needs of preservation should not be seen as competing with those of program production. Digital preservation is not an optional "add-on" expense to production, but can be incorporated into production budgets as a necessary requirement for the ongoing usability of the materials.

Viewers keep reminding us that public television programming is important, and they are beginning to expect access to archival programming. But until very recently, there has been no mandate for public broadcasting to support audiovisual preservation, leaving preservation to be done locally on an ad-hoc and inconsistent basis.

In the digital era, however, such 'hit-or-miss' practices will not suffice to keep program materials viable. Instead, support for program preservation and long-term access must be a system-wide responsibility that builds on the program investments that have already been made, and is structured so that both the costs and the benefits of preservation will be shared across the public broadcasting system.

With well-supported and secure technologies, properly managed metadata, and sound preservation strategies, a preservation repository can provide a wide range of tangible and intangible benefits for public broadcasting. We must continue fostering a commitment to sustain digital preservation across the entire system, with the understanding that saving our digital assets over time serves not only the public broadcasting system, but benefits the American public as a whole.

* * * * *

Chapter 1

INTRODUCTION AND OVERVIEW: A Sustainable Approach to Preserving Digital Public Television

In less than a decade, television production, distribution and preservation have undergone a radical shift. Today, the number of programs shot, edited, and shared as digital files is increasing rapidly. 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.

What do these changes mean for television archives? Practices to conserve and protect videotape recordings 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 by the National Digital Information and Infrastructure Program of the Library of Congress (NDIIPP)[1] set out to solve some of these difficult problems by designing a model repository for public television. In the process, the project also examined other issues closely related to repository operations, such as standards for metadata and operating costs.

This report is the result of research conducted by the *Preserving Digital Public Television* Sustainability Working Group. It explores the requirements for the long-term sustainability of digital public television program files, and specifically, how these impact the creation of a system necessary to operate a reliable

1 http://digitalpreservation.gov

preservation repository.

Entering the Era of Digital Preservation for Programming

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, 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.

At its outset, this report recognizes that preserving “born-digital" materials presents significant obstacles, and that maintaining large quantities of public television content might seem daunting. While archivists working with videotape are well aware of the problems of obsolescence and decay, digital media presents an entirely new set of challenges. Where techniques such as climate-controlled storage, technological monitoring, and reformatting could safeguard videotape for many years, in a digital environment, these methods are no longer relevant.

The effective lifespan of a digital file can be much shorter than its analog counterpart, as there are far more threats to its long-term usability over time. Unlike safeguarding videotapes, the activities required to properly archive digital files cannot be left until the end of the production process. How will these media be managed over the long-term? What needs to be done, how will it be accomplished, and who will be responsible for ensuring that this part of our cultural heritage survives into the future?

Saving digital public television requires a new approach that brings preservation principles into emerging all-digital production and delivery environments, storage systems, and media asset management. 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.

The NDIIPP Project: Preserving Digital 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."[2] However, until 2008, CPB had never allocated any funds to support this charge nor implemented any program for system-wide preservation.[3] Currently, public television allocates very few resources for preservation. Only a few stations and PBS have formal archives in place, which leaves program preservation largely as an afterthought and is performed locally on an *ad hoc* basis.

In 2004, public television stations Thirteen/WNET in New York and WGBH in Boston, in partnership with the Public Broadcasting Service (PBS) and New York University, organized Preserving Digital Public Television (PDPTV)[4] as a collaboration to introduce digital preservation issues and practices to the public television system. The project is funded by the National Digital Information Infrastructure and Preservation Program (NDIIPP) of the Library of Congress.[5] It has been aimed specifically towards preserving *born-digital* program files, and has not been engaged in the digitization of analog materials. It is scheduled to be completed in 2010.

---

2 Public Broadcasting Act of 1967, US Code Title 47 (1967), S. 396

3 In 2008, CPB launched *the American Archive* initiative, which is developing a system-wide program preservation plan.

4 More information is available on the PDPTV website, http://www.thirteen.org/ptvdigitalarchive/

5 More information is available on the NDIIPP website, http://digitalpreservation.gov.

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;
- And examine issues of long-term content accessibility and methods for sustaining digital preservation of public television materials.

In keeping with these goals, along with designing a preservation repository, the PDPTV project has also analyzed the questions of what is required to sustain digital preservation. The primary methodologies included documenting the experiences of the PDPTV team in building the prototype repository; examining production and distribution workflows; and reviewing existing current research on digital sustainability and parallel efforts by our colleagues in various fields, such as higher education, library science, information technology, and broadcast television in the US and around the world.

The research contained in this paper is a preliminary step toward achieving system-wide preservation of and access to digital content. It defines and describes a sustainable approach to preserving digital public television, and suggests some ways in which this can be achieved, by:

- 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.

Sustainable Preservation of Digital Public Television: Key Findings

After the experience of researching and developing the PDPTV repository, we have concluded that preserved assets can be a great benefit to the public television system, and that each investment towards preserving programming can guarantee renewed value and long-term access to this material. Our discussion will highlight these findings:

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** and spread across the 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. Preservation-minded actions can in fact expedite the production process, while their incremental costs can be incorporated as necessary in the overall production budget.

This report will demonstrate that 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 governance models appropriate to guiding a repository exist within the system; sufficient technical resources to operate such a facility; and diverse funding options capable of sustaining its operations.

* * * * *

Chapter 2

Why Save Program Files? Making the Case for Preservation

> *"Public television has been responsible for the production, broadcast, and dissemination of some of the most important programs which in aggregate form the richest audiovisual source of cultural history in the United States."*
>
> - James Billington, Librarian of Congress. From "Television and Video Preservation 1997: A Report on the Current State of American Television and Video Preservation" Volume 1, (Washington, DC: Library of Congress, October 1997), 69.

History is built on the availability of evidence. If evidence ceases to exist, then the historical record and public consciousness are irrevocably altered. Spanish philosopher, George Santayana states in his book *The Life of Reason*, "Progress, far from consisting in change, depends on retentiveness. When change is absolute there remains no being to improve and no direction is set for possible improvement: and when experience is not retained, as among savages, infancy is perpetual. Those who cannot remember the past are condemned to repeat it."[6]

In a time of mass media consolidation and the loss of local media outlets, the "evidence" and historical record produced by decades of the diverse, noncommercial voices of public television are especially important. The historical significance of public television also is matched by its ability to enrich our lives culturally and aesthetically. In providing unique and thoughtful reflections of events and lives past and present with the highest creative and editorial integrity, public television not only documents and informs, but at the same time, inspires, impassions, and entertains in a way that is uniquely balanced, innovative, and reflective of the pluralism in our society.

---

6 George Santayana, *The Life of Reason: or the Phases of Human Progress*, vol. 1, chapter 12, p. 284 (1905). Quote available from http://www.bartleby.com/73/1292.html (Accessed on January 6, 2010).

Furthermore, the educational value of public television content cannot be disputed. Since its inception, public television has had a mandate to educate. It has fulfilled this mandate successfully through award-winning quality programming and services that meet the needs of its various audiences, from pre-school children to K-12 teachers, to the general public. Preservation has facilitated classroom access to materials that are highly valued as inspirational as well as instructional. Apart from this philosophical position, investment in the long-term stewardship of public television assets will require shared, realistic commitment by many groups. Although for many of these groups, the reasons for preservation will be obvious, for others, the reasons for such an ongoing investment must continuously be demonstrated and reinforced.

As Abby Smith, a historian and cultural analyst working with NDIIPP, points out, "Without clarifying to the public why it is important to society in general and to individuals in particular to make long-term commitments of resources to the collection and preservation of cultural content, it is unlikely to happen."[7] Besides informing the public, it is also important to articulate this value to the stakeholders who have the authority and capacity to undertake preservation. The National Science Foundation and Andrew W. Mellon Foundation-funded Blue Ribbon Task Force on Sustainable Digital Preservation and Access argues:

> Recognition of the benefits from preservation leads to an incentive to preserve, but different stakeholders may have different appraisals of value, leading to different incentives to preserve, and ultimately, different expressions of willingness to allocate resources to preservation activities. Sustainable economic models for digital preservation need to ensure that the over-arching public interest in long-term preservation is supported by the complex patterns of stakeholder relationships and incentives attached to a particular set of digital materials.[8]

Users Are Willing to Pay For Particular Qualities

Given the proliferation of digital content, the question is: what can a repository for public television offer that would be of exceptional value to users, whether they be public television producers, educators, or researchers, or members of the public? What value can the repository create over and above the value of the content itself?

In his blog, *Wired* Magazine co-founder Kevin Kelly identifies eight "generatives" or "uncopyable values" whose qualities are "better than free";[9] that is, these qualities are so worthwhile to users that they are willing to pay for them, even if the content can be had at no cost. They include the characteristics of:

- **Immediacy:** the value of being able to access content in a relatively timely way.
- **Personalization:** the value of being able to access content in a form that meets specifically defined needs.
- **Interpretation:** the value of the availability of guidance and support in using the content.
- **Authenticity:** the value of the knowledge that the content is authentic.
- **Accessibility:** the value of having the repository be responsible for the safekeeping of the content so that it is always accessible when the user wants it.
- **Embodiment:** the value of high-quality versions of the content.

---

7 Abby Smith, "Valuing Preservation," Library Trends 56, no. 1 (2007), http://muse.jhu.edu/journals/library_trends/v056/56.1smith.html (accessed January 6, 2010).

8 Blue Ribbon Task Force on Sustainable Digital Preservation and Access, “Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation, Interim Report” (n.p.: Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008), http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf (accessed January 6, 2010).21.

9 Kevin Kelly, "Better than Free," The Technium blog, posted on January 31, 2008, http://www.kk.org/thetechnium/archives/2008/01/better_than_fre.php (accessed January 6, 2010).

- **Patronage:** the value of the appreciation and loyalty that users feel toward the repository.
- **Findability:** the value of being able to find the content that the user is looking for.

To make the case for preservation, public television needs to elaborate the wide and far-reaching potential impact that the present and future accessibility of its content can have to all stakeholders. The intangible value of public television content needs to be expressed to stakeholders, and in some cases, it may even be possible to find creative new ways to convert that intangible value into monetary value.

Who are the Stakeholders?

A key challenge in discussing sustainability and future access is to rely not solely on numbers, but to offer a compelling rationale for making the investment in the first place. To begin, it is important to identify exactly who stands to benefit from a system-wide preservation initiative or repository.

In this age when it is possible to have instant online access to huge collections of digital video, the most common image of a user is an individual searching for online video content for his or her own entertainment. Consequently, when considering who the 'stakeholders' are in archival public television, the most immediate group that comes to mind is often simply 'the public.' In fact, there are a number of major stakeholder groups who would benefit from the preservation of public television, and who would have a long-term interest in increasing the value of this content beyond this conception of an undefined 'public.'

Copyright Holders

Program copyright holders have the most immediate interest in preservation. Because the copyright holder controls the intellectual property rights of the completed program, they make the ultimate determination of the disposition and long-term access to the program. A digital repository can protect this interest over time, and facilitate the controlled access that can lead to new revenues and recognition.

Program Producers and Filmmakers

Reuse of existing and previously-used materials in new productions is one of the most common needs of program producers and filmmakers, and they are a staple in public television documentaries. Outtakes from public television productions are a rich collection of primary source materials, ranging from interviews with experts, eyewitness accounts of historic events and world leaders, to b-roll footage. Proper preservation supports cost-effective production by ensuring that materials are available and accessible when needed. Program producers that are copyright holders also benefit from having their own works preserved (see above).

Program Distributors

Many program distributors also fulfill the role of program libraries, keeping distribution copies of programs long after their distribution rights have expired. Because of this, distributors often possess unique and historical copies of work, especially materials from individuals and small producers. A preservation repository can provide safe storage for program distributors, the capacity to easily retrieve materials for re-distribution, return to the producer and/or copyright holders, and other future uses.

Educators and Educational Media Producers

Increasingly, educators are using digital media in their classroom teaching and curriculum development. Educational media producers are already turning toward broadcast and other produced materials to create educational resources. Given its educational mission and quality, archived public television programming is in high demand and expected to be available for those uses.

Researchers and Journalists

Access to archives is a basic necessity for writers, journalists, and researchers working in all fields. An

archive of public television programs could be an especially rich and valuable source of primary materials documenting the social and cultural history of the United States and the world.

Students and Scholars in Higher Education

Like researchers and journalists, scholars can employ preserved primary resources from public television in their work. Students and academics in higher education are also using multi-media and historical materials for class assignments, projects and similar activities.

The 'Public'

A substantial percentage of support for public broadcasting comes from the public, both by way of direct contributions made to public broadcasting stations, and by way of tax dollars distributed by Congress, state legislations, local municipalities, and through such institutions as arts and humanities councils. The public, an important source of funding for public television, carries a long memory for favorite programs, and feels entitled to watch public television programs long after their brief broadcast life. Public television has an ethical responsibility to protect the public investment and keep it available.

Public Television

As a cultural institution charged with a public service mission, public television has an interest in ensuring the availability of its content for the public good and the historical record. As a business concern, public television also has an interest in potential cost savings and the ability to re-use content into the future.

These various stakeholders would all benefit from the preservation of the materials created by public television. On the production side, preservation must become a necessary component of an all-digital workflow. On the user side, digital preservation is a prerequisite to ensuring ongoing electronic access. Policies need to be enacted that can spread the cost burden among the stakeholders who benefit from the preservation of these materials, and at the same time, business models need to be designed such that those who bear the costs are adequately motivated to sustain the repository over the long term.

Repository Services Add Significant Value

So, then, what are some of the different ways of valuing public television content? Business models in public television have traditionally judged the worth of content

Public television is already successfully adding value to its content with some of its current initiatives. WGBH, for example, has created a number of web portals for users to access online content that reflect many of Kelly's "uncopyable values." (see p.7)

Each of the WGBH portals is tailored to meet the specific needs of a targeted audience in terms of design, search functionality, and type of content available. For professional filmmakers, WGBH Stock Sales (http://www.wgbhstocksales.org/) offers high-quality, downloadable content that is searchable by keyword and visual description, payable through e-commerce. For amateur filmmakers, meanwhile, WGBH Lab (http://lab.wgbh.org/) offers lower quality but free content. K-12 students and teachers access WGBH content through another portal, Teacher's Domain (http://www.teachersdomain.org/), which offers free 3-5 minute clips specific to and searchable by teaching concepts. Compared to WGBH Stock Sales, it is designed to be easy to use, since teachers may have a lower tolerance for difficult technology than professional filmmakers.

Finally, WGBH's Open Vault (http://openvault.wgbh.org/) is designed to serve the general public and higher education. It features the ability to browse, contextualize results, and it is relatively easy to use, meeting the needs of this audience. These customized services create added value for end users, and encourage the dissemination and use of WGBH-owned content.

A much larger scale model of sustainable preservation and access through tailoring content to different user groups comes from L'Institut national de l'audiovisuel (Ina), the French National Audio-Visual Archives. Ina holds 1.5 million hours of sound and video from France's public television and radio stations dating back to the 1940s. After conducting research and surveying customers, Ina developed two online content portals. Inamédiapro, a powerful interface to all of Ina's digitized content is tailored to both the professional community as well as to INA staff needs. Rights licensing through this website has generated significant revenue for the organization, while ina.fr, the free public website offers a curated selection of lower resolution content for free. (Maron, et. al, 2009) The Ina model can certainly provide guidance to a future repository of public television content.

based largely on its immediate monetary value. Public television is now starting to recognize that archiving and preservation of its content can extend this monetary value over time and across wider markets, especially in the context of the internet with its never-ending hunger for content.

Tangible and intangible values are not always immediately evident. An excellent case in point is a video recently re-discovered in the WGBH vaults of a 1990 protest in support of a professor at Harvard Law School, featuring a young law student named Barack Obama as the principal spokesperson at the demonstration. Local footage such as this would have had some informational value when it was shot, but has only come to have great historical, and monetary value, in light of the 2008 Presidential election. This value, moreover, could only be uncovered with the ability to locate and identify the content, which is the result of good archival organization and cataloging done in the past.

Value, then, is not tied solely to content, but is also based on how content is made available to users. Part of building a sustainable repository for digital public television will be to create value through the ways in which *the repository manages the materials*. This includes the potential cost savings that comes from the ability to reuse existing footage to create new programming, the ability to reach new audiences through new digital distribution platforms, and the ability to market footage to other non-commercial and commercial entities.

Images for the Future, a joint project in the Netherlands between the Film Museum, the National Archives, the Institute for Sound and Vision and other institutions that aims to restore, conserve, and digitize the Dutch audiovisual heritage, has taken an approach to measuring monetary and non-monetary costs and benefits that we found particularly instructive. This consortium commissioned a quantitative study of the benefits of preserving audiovisual content in 2006 entitled *Outline of Benefits*.[10] The theoretical framework of the report follows the OEI (Research into Effects of Infrastructure) guidelines and weighs direct, indirect, and external costs and benefits. The study produced a list of the potential costs and benefits of preserving audiovisual content for both potential partners and the public.

While the findings are particular to the Dutch context, many of the benefits cited by the study are applicable to the preservation of public television here in the US. In no particular order, they include:

- Increased use of content as a result of better availability and findability.
- Lower costs to fulfill the current demand for audiovisual content.
- Ability to access and use content that would otherwise be lost without preservation.
- Payment to copyright holders and project partners by consumers of the content as a result of increased use.
- Increased usage encourages multimedia literacy.
- Recognition of being at forefront of digital preservation; ability to contribute to advancement of knowledge.
- Preservation of culture and heritage for future generations.
- Reinforcement of cultural-historical awareness and democracy.

A preservation repository can provide services that supply all of these kinds of values. It can ensure that users are able to discover and access authenticated content in the form that meets their needs in a timely fashion. It can do this by using well-supported and secure technologies, reflecting the needs of its designated communities, properly managing metadata, and employing effective preservation strategies.

Repository services can add value for content producers, too. For example, if producers are able to locate and access pre-existing B-roll footage easily and in a timely fashion as a result of good digital object management and a supportive technological infrastructure, they will not have to do extensive research to find what they are looking for, nor will they have to re-shoot, or purchase stock footage from

10 SEO Economic Research, “Outline of Benefits: Results of 'Indicators For Cost-Benefit Analysis Images for the Future” (Amsterdam: SEO, 2006), http://www.beeldenvoordetoekomst.nl/assets/documents/imagesforthefuture_outlineofbenefits_2006.pdf (accessed January 6, 2010).

another source. This translates into immediate savings in production budgets. A well-planned repository will also eliminate current waste in staff time and expenditures caused by missing or inaccessible data and inadequately interoperable systems.

Services that create value (or eliminate costs) clearly benefit repository users. Locating value in the repository, and not just in the content, also helps to justify the repository's continued existence. Yet it is important to continually re-evaluate the repository's services. One of the central findings in the July 2009 report, "Sustaining Digital Resources: An On-the-Ground View of Project's Today," was that digital projects remain sustainable by continuing to add value to resources as users expectations change. The authors warn that, "as new technologies develop and user expectations shift and grow, a resource risks fading slowly into irrelevance if it does not constantly grow and innovate in ways that continue to benefit its constituents."[11]

While the content itself may not change, the ways that people find and use public television material will certainly evolve over the years. Keeping up with this evolution by offering new and innovative means of access will certainly be important to sustaining public television preservation.

Defining "Sustainability"

What do we mean by sustainable digital preservation? Definitions of sustainability depend on context. The term *sustainable* as defined by the Oxford Dictionary of English, "able to be maintained at a certain rate or level," can be applied to a variety of disciplines, industries, and fields of study. In the field of digital preservation, sustainability refers to the "set of business, social, technological, and policy mechanisms that encourage the gathering of important information assets into digital preservation systems, and support the indefinite persistence of digital preservation systems, enable access to and use of the information assets into the long-term future."[12] This issue has become a genuine concern for those working in digital preservation and represents a new set of issues in the professional approach to keeping digital content alive over time.

Historically, digital preservationists have focused on the technical and practical requirements of maintaining digital information. Since the 1980s, they have investigated topics ranging from finding permanent storage carriers to developing universal computer systems. A debate raged for years over whether *migration* (a preservation strategy that involves moving digital information from one storage medium to another more current environment and/or transferring that information to a current file format) or *emulation* (the practice of recreating the environment that was originally used to render, display, or provide access to a digital work using current technologies) is the best approach.[13] Today, these two strategies are widely recognized as equally beneficial solutions, and that the type of digital content being preserved should be analyzed to determine the appropriate choice between the two.

In the past few years, the digital preservation community has expanded its understanding of the requirements for maintaining digital information beyond solely technical or procedural concerns. In addition to these issues, organizational and economic factors contribute equally to the longevity of digital content. Recent research has shifted to the larger issue of *digital sustainability*. In his article "Defining Digital Sustainability," Kevin Bradley of the National Library of Australia explains that, "Clearly it is not possible to preserve digital information without a sustainable organizational, economic, social,

---

11 Nancy L. Maron, K. Kirby Smith, and Matthew Loy. "Sustaining Digital Resources: An On-the-Ground View of Projects Today." (New York: Ithaka S+R, 2009), 11, http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability/report/SCA_Ithaka_SustainingDigitalResources_Report.pdf (accessed August 20, 2009), 11.

12 Blue Ribbon Task Force, "Sustaining the Digital Investment," 19.

13 For more information on the history of digital preservation, please see pages 151-156 in Kevin Bradley, "Defining Digital Sustainability," Library Trends 56, no. 1 (Summer 2007), http://muse.jhu.edu/journals/library_trends/v056/56.1bradley.html (accessed March 17, 2009).

structural, and technical infrastructure, nor is it sensible to preserve material without sustained value."[14]

Brian Lavoie, research scientist at OCLC Research further elaborates, "Technical issues are only one aspect of sustainable preservation activities. Ultimately, these technical processes must be coordinated with the economic process of marshaling and organizing sufficient resources to achieve preservation objectives."[15]

New Approaches to Preserving Public Television

> *"Rather than being at the end of the production chain, the archive is becoming an integral part of the production process, and is being absorbed into wider digital storage environments, include those distributed or used across organizational boundaries."*
>
> - M. Addis et al. "Sustainable Archiving and Storage Management of Audiovisual Digital Assets." SMPTE Motion Imaging Journal, November/ December 2009, p 27.

This progression in thinking about sustainability stems from the recognition that there is a fundamental difference between preserving analog and digital materials. In the analog world, well-stored objects can survive undisturbed for years or even decades. In the digital world, once you turn on the lights, they cannot be turned off and then turned back on again years later. Preserving digital information requires constant, ongoing management by skilled professionals. "In this regard," Lavoie recognizes, "preservation in the twenty-first century will represent a significant departure from traditional practice."[16]

In major contrast to the best practices for conserving analog videotapes, preserving digital files demands an entirely different approach which in some ways, is just the opposite:

- **Digital files cannot remain on a 'shelf' to be dealt with at a later date when time or resources allow.** As Brian Lavoie and Lorcan Dempsey from OCLC Research indicate, "digital materials generally do not afford the luxury of procrastination. The fragility of digital storage media, combined with a high degree of technology dependence, considerably shortens the 'grace period' during which preservation decisions can be deferred. Issues of long-term persistence can arise as soon as the time digital materials are created."[17]
- **Digital preservation practices need to begin at the point of creation and continue throughout the production and distribution cycle.** Choices made in production about what file formats to use, which files to keep, and what metadata to create and capture will ultimately have an enormous impact on the longevity of digital objects.
- **Digital preservation must be seen as an ongoing process.** Once a digital object has been put on a server, it is by no means "preserved." After they are "saved," digital files remain perpetually at risk. Software upgrades and encoding format changes happen frequently, and

14 Bradley, "Defining Digital Sustainability," 157.

15 Brian F. Lavoie, "Of Mice and Memory: Economically Sustainable Preservation for the Twenty-first Century," in Access in the Future Tense (Washington, DC: Council on Library and Information Resources, April 2004), http://www.clir.org/pubs/reports/pub126/lavoie.html (accessed on January 6, 2010).

16 Lavoie, "Of Mice and Memory".

17 Brian Lavoie and Lorcan Dempsey, "Thirteen Ways of Looking at... Digital Preservation," D-Lib Magazine 10, no. 7/8 (2004), http://www.dlib.org/dlib/july04/lavoie/07lavoie.html (accessed January 6, 2010).

files require duplication, auditing, ongoing migration and maintenance according to recognized standards and best practices. A 2003 report sponsored by the National Science Foundation and the Library of Congress asserts, "Archival collections do not just happen when someone clicks on the 'save' icon."[18], otherwise they may become corrupt or obsolete and unusable. To remain usable and accessible, they have to be able to operate within ever-changing systems.

- **Doing nothing now may result in devastating loss down the road**. When a film previously thought to be lost is discovered in an attic or deep in the vaults of an archive, the world rejoices at the enrichment that the moving image artifact can once more bring to our lives. Old video footage of a leader in his university days helps form the public's collective knowledge, and brings new revenue to the copyright holder. These sorts of discoveries will become increasingly rare in the digital era. The cost of re-engineering obsolete digital video files will become unaffordable, and lack of identifying metadata means existing footage that suddenly has become valuable will remain unknown. If an appropriate preservation path is not chosen soon, much of the incredible footage acquired by public broadcasters will be doomed to loss in the digital black hole.

The digital files produced by public television are very large, complex, and come in multiple formats. They have associated metadata that also must be maintained. If born-digital public television programs and ancillary production materials are to survive and remain usable in the long-run, they require sufficient and sustained organizational, technical, social, and economic frameworks. This calls for digital content to be *managed throughout the entire life cycle* of the file, which must be done properly from the very beginning.

Many of the practices and procedures involved in digital video preservation are closely aligned with the IT concepts of **Information Lifecycle Management (ILM)** and **Enterprise Content Management (ECM)**, which have roots in business Records Management.

ILM is defined by the Storage Networking Industry Association's (SNIA) Data Management Forum as follows:

Information Lifecycle Management is comprised of the policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost effective IT infrastructure from the time information is conceived through its final disposition. Information is aligned with business requirements through management policies and service levels associated with applications, metadata, and data.

ECM is defined by the Association for Information and Image Management (AIIM) as:

Enterprise Content Management (ECM) is the strategies, methods and tools used to capture, manage, store, preserve, and deliver content and documents related to organizational processes. ECM tools and strategies allow the management of an organization's unstructured information, wherever that information exists.

As this is being written, there is a lively discussion underway among a variety of academic, scientific and technology working groups and organizations examining the economic implications of digital sustainability that had previously not received much attention[19] For example, in late 2007, the National Science Foundation and the Andrew W. Mellon Foundation, in partnership with the Library of Congress, the Joint Information Systems Committee (JISC) of the United Kingdom, the Council on Library and Information Resources (CLIR), and the National Archives and Records Administration (NARA), funded a Blue Ribbon Task Force on Sustainable Digital Preservation and Access, which was given a two-year

18 "It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation, Final Report, Workshop on Research Challenges in Digital Archiving and Long-Term Preservation, April 12-13, 2002" (sponsored by the National Science Foundation, Digital Government Program and Digital Libraries Program, Directorate for Computing and Information Sciences and Engineering, and the Library of Congress, National Digital Information Infrastructure and Preservation Program, 2003), 13. http://www.digitalpreservation.gov/library/resources/pubs/docs/about_time2003.pdf (accessed January 6, 2010).

19 For a comprehensive list, see the bibliography compiled by the Blue Ribbon Task Force on Sustainable Digital Preservation, available at http://brtf.sdsc.edu/bibliography.html (accessed March 17, 2009).

charge to look at economic and organizational issues for sustainable preservation. The BRTF recommendations will be directed towards individuals and institutions that are responsible for the long-term stewardship of large digital collections in such fields as library sciences, computing, economics, and academia.[20]

This report, aimed at public broadcasters and others managing digital media, has drawn on key publications and research by these other groups to frame our own discussion of digital sustainability in our own context.

Audience & Scope

This report is aimed at personnel within the public broadcasting system who want a better general understanding of the requirements of digital preservation. With the exception of the case study in Chapter 5, it does not discuss specific technologies or implementations. The envisioned reader is a public broadcasting station manager who wants to learn how to better ensure the survival of the station's digital assets.

Focus on Television

This report addresses the requirements for sustainable preservation of and access to public television material. The focus on television, rather than on both television and radio, is a reflection of Preserving Digital Public Television project's limit in scope to this medium, however, many of the strategies expressed here can apply equally to audio, image, text, and other multimedia content. At the same time, because video represents the largest volume of content and the least standardized in terms of production, distribution, and archival formats, television material remains most at risk.

Focus on File-Based Content

Over the years, public broadcasters have created and collected an incredible volume of material recorded on video and audiotape. Many of these media have been obsolete for years, and a sizeable percentage of them are deteriorating due to years of storage in less than optimal conditions. Because of this, we recognize that the digitization of legacy content is a high priority for the public broadcasting system. However, this urgent problem is beyond the scope of Preserving Digital Public Television, and therefore will not be covered in this report. We see the problem of preserving file-based content equally, if not more in need of attention. As more and more television programs originate as digital files, combined with newer High Definition (HD) formats and legacy content that is digitized over time, there will be a growing need for preservation environments specifically designed to manage and sustain these large, complex collections of files and associated metadata.

Focus on Preservation vs. Access

The goal of preservation is to maintain long-term access to content, despite changes in technology and user expectations. Preservation is a set of ongoing activities that enable access to content over time, despite changes in technology and user expectations. The requirements of preservation, however, are separate and distinct from the policies, procedures, and technologies needed to provide access to digital materials. Today, there are endless ways to provide access to content, and countless needs of the various user groups. Access strategies and platforms will continuously evolve over time, but in order to facilitate these changes, good initial preservation practices will result in confidence that the content will remain available, regardless of the unknown technical future. Planning appropriate preservation practices will also provide savings in time, money, and effort down the road when users require content in currently unimaginable ways.

It is exactly for these reasons that public broadcasting needs a sustainable preservation repository. The intensive, short-term, project-based approach typical of both traditional preservation and public television initiatives cannot provide that underlying infrastructure. As Jonas Palm, Director of the

20 Blue Ribbon Task Force, "Sustaining the Digital Investment."

Department of Preservation at the Riksarkivet/National Archives of Sweden demonstrates, “Without such long-term planning, digitization projects can come to behave like black holes in the sky.” [21] The same could easily be said for born-digital public broadcasting.

* * * * *

21 Jonas Palm, “The Digital Black Hole.” (n.p: Training for Audiovisual Preservation in Europe, 2008), 1. http://www.tape-online.net/docs/Palm_Black_Hole.pdf (accessed January 6, 2010)

Chapter 3

Requirements for Sustainable Preservation

The goal of digital preservation is to *maintain information*, regardless of the devices used to create the content or the file formats that the information was contained in when it was created. Preserving only the raw bits is necessary, but not sufficient. Beyond preserving a bitstream, there must also be means of finding, retrieving, and rendering the digital bits back to viewable moving images. The challenge is that digital objects and the environments they are accessed through are built on technology that is constantly evolving. In this context, sustainability requires consideration of two major elements:

- **Technical and Organizational Requirements:** the technical and procedural aspects of repository operation and management, including the organizational infrastructure that governs a repository.

- **Economic Considerations:** the potential costs and benefits of preserving public television content, plus economic and business models that can be used to sustain digital preservation activities.

This chapter outlines the requirements for a stable operating and administrative framework. The chapter that follows addresses the economic requirements necessary to maintain such an operation.

Technical and Organizational Requirements

Because public broadcasting is only beginning to explore the best ways to sustain its digital content, now is the appropriate time to explore reliable long-term solutions. A sustainable approach to preservation must take into account the following requirements:

1. **Bit Preservation:** Sustaining the 0s and 1s, or ensuring that the video, audio, and ancillary files remain intact over time, with no loss or corruption of bits. This practice is known in the wider digital preservation community as *bit preservation*. It requires a secure and stable technical infrastructure.

2. **Accessibility of Content:** In order for the effort of preserving bits to truly be valuable, the content must be *findable* and *usable*. Thus, mechanisms to access the content must also be sustained, which can ensure that video and audio streams can be found, retrieved, played back, and delivered to the appropriate user communities. To maintain these necessary functions over time, in the face of technology changes, a reliable system architecture, as well as policies on standards and metadata are required. The practice of preserving the content and its accessibility is called *digital object management*.

3. **Organizational Infrastructure:** The process of keeping the bits and access functions intact must be managed by a responsible entity which must persist over time. If it is to effectively carry out its mandate of keeping the content alive and accessible, this includes having a succession plan in place should the entity no longer be able to care for the content. All digital preservation activity thus relies on the organizational stability of the repository itself -- without it, the content will not survive.

4. **Reliable Funding.** Consistently attending to these requirements, of course, requires ongoing funding. Kevin Bradley states, for the digital repository, the term sustainability "is used to mean building an economically viable infrastructure, both social and technical, for maintaining valuable data without significant loss or degradation. This includes the whole socio-technical composition of the repository, the short- and long-term value of the material, the costs of undertaking an action, and the recognition that technologies do not sustain digital objects: institutions do, using the available technology."[22]

These four requirements are all highly interdependent issues which will each be examined in turn.

The Concept of a 'Trustworthy Repository'

Throughout this report, we argue that sustainable preservation of digital public television content will ideally be achieved by a reliable and responsible entity, or network of entities, which we refer to as a *preservation repository*. Moreover, successful preservation of digital public broadcast content will only be sustainable if this repository is considered *trustworthy,* in that it can ensure the long-term accessibility of the digital objects it is storing. To be trustworthy, a repository must:

- Have a secure technological infrastructure to protect the content;
- Follow good management procedures that incorporate standards and policies;
- Have sound organization and governance.

A sustainable preservation repository will have the means, policies and organizational infrastructure in place to guarantee long-term bit preservation and accessibility of the content in its care over time. The repository should also be able to demonstrate its trustworthiness in these areas to its user communities. For example, depositors will want to know that they can reliably submit their content and that it will persist viably into the future, and end-users will want to know that they can find and retrieve content that is authentic and valid.

The 'TRAC' Criteria and Checklist for a Trustworthy Repository

Various criteria for measuring a repository's capability to act as a trustworthy, sustainable, digital caretaker have been developed over the past few years. In the US, *Trustworthy Repositories Audit and*

22 Bradley, "Defining Digital Sustainability," 157.

Certification: Criteria and Checklist (TRAC)[23] is the most recent result of a joint effort between the National Archives and Records Administration (NARA) and the Research Libraries Group (RLG) with contributions from the Center for Research Libraries (CRL) and other organizations and individuals involved in digital libraries and digital curation worldwide. In Europe, the German Network of Expertise in Long-Term Storage of Digital Repositories (nestor)[24] cooperative has outlined similar criteria. Additionally, the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) toolkit, jointly developed by the Digital Curation Centre (DCC)[25] and DigitalPreservationEurope (DPE)[26] is being used by repositories around the world to gauge their own trustworthiness. A working group is currently moving these criteria toward standardization with the International Standards Organization (ISO).[27]

To be considered "trustworthy," each of these various sets of criteria follow parallel principles that provide guidelines for institutions that want to offer long term preservation services, and that are dedicated enough to have their preservation and business practices be certified by an agreed upon standard set of criteria, which, at a minimum, address all four areas identified above.

As a specific example, the introduction to TRAC states:

> In determining trustworthiness, one must look at the entire system in which the digital information is managed, including the organization running the repository: its governance; organizational structure and staffing; policies and procedures; financial fitness and sustainability; the contracts, licenses, and liabilities under which it must operate; and trusted inheritors of data, as applicable. Additionally, the digital object management practices, technological infrastructure, and data security in place must be reasonable and adequate to fulfill the mission and commitments of the repository.[28]

All digital repositories, including any that might provide stewardship for public broadcasting material, can utilize the checklist provided by TRAC as an evaluative tool. Applying these criteria to the establishment of a preservation system for public television will help guarantee that these efforts will be efficient, effective, and sustainable. Thus, the discussion that follows draws heavily from the requirements prescribed by the TRAC repository certification criteria.

Using the OAIS Model to Organize Repository Functions

The TRAC criteria were developed out of a need for repositories to demonstrate that their practices meet the responsibilities outlined in the ISO standard *Reference Model for an Open Archival Information System* (OAIS), a high-level, widely applied model that defines the functions, processes, and procedures

---

23 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories Audit and Certification: Criteria and Checklist” (Chicago and Dublin, OH: Center for Research Libraries/OCLC, 2007) http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (accessed on January 6, 2010).

24 http://www.langzeitarchivierung.de/eng/index.htm

25 http://www.dcc.ac.uk/

26 http://www.digitalpreservationeurope.eu/

27 The Digital Repository Audit and Certification Wiki can be found at http://wiki.digitalrepositoryauditandcertification.org/bin/view

28 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories,” 3.

needed to preserve and provide access to digital information.[29] It is essentially a generic blueprint for the design of a digital preservation repository, and its definitions are broadly applicable to any type of content. The model defines a set of relationships among 'functional entities' assigned to perform the various roles required by the repository [Figure 1].

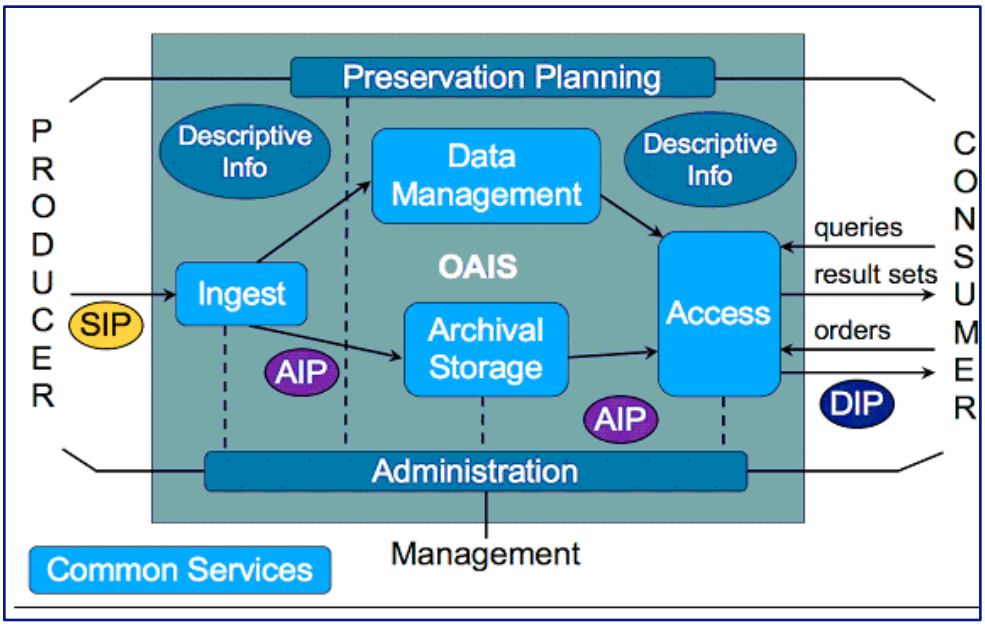

Figure 1. The OAIS Functional Entities.

We have used the following OAIS definition for our digital repository: "An archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community."[30] Furthermore, we will keep in line with the OAIS definition of long-term, being "A period of time long enough for there to be a concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future."[31]

After successfully building an OAIS-compliant prototype repository at NYU, the PDPTV team feels confident that this standard, along with the TRAC criteria, provides a valuable model for the technical and procedural aspects of sustainable digital preservation.

Applying the TRAC Operating Guidelines

Guided by the TRAC criteria, we will take a closer look at the requirements of digital preservation as they might be applied to a public broadcasting repository. In the process of examining these conditions, we will not only look at the functions of a sustainable preservation repository as defined by OAIS and the various certification criteria, but also related factors such as workflows and cost-effective preservation practices throughout the life cycle of the digital content.

These guidelines provide a meaningful way for institutions and organizations to identify the requirements

30 Ibid., 1-11.

31 Ibid.

of a sustainable preservation repository, and we feel strongly that they offer excellent, tested models for establishing reliable preservation services that can be applied to the public broadcasting context. By applying these standards as guidelines to the development of future repositories or networks of repositories, public television will be able to follow a framework that is widely accepted across sectors and has many available resources.

Criteria 1: Bit Preservation

Marilyn Deegan and Simon Tanner, authors of *Digital Preservation*, point out, "data are very bad at self-preservation. Active measures must be taken at birth to ensure that they survive longer into the future: this is known as the 'retention intention' and is essential to data preservation."[32]

The most fundamental digital preservation activity, regardless of content type, is saving the bits (the 1s and 0s) that make up the digital object. That object might be text in the form of a PDF document, or a motion picture composed of hundreds of thousands of images that represent each film frame. For public television, digital objects are usually a combination of video and audio streams that make up a digital program file, but there may also be also ancillary text and image files. Unfortunately, the bits that make up these digital objects are highly susceptible to a number of threats which can quickly render them lost forever.

Bit preservation is achieved through a combination of technologies and procedures, including *data security* and *backup*. A preservation repository will also be expected to offer services such as *content migration*, once formats are no longer supported by current software and hardware, and it should have the ability to periodically move content to new storage media, a practice known as *refreshing*. There are a variety of reasons why storage media must be refreshed, including obsolescence and mechanical failure, but also, as the European project PrestoSpace found,

> Whilst replacement of format within the broadcast world typically occurs every ten years or so, in the IT field media is replaced in cycles of five years or even less. Even without a change in the format of the component, the increases in capacity can make a replacement of components all but unavoidable within a couple of years. All forms of storage, magnetic, optical and solid state, are subject to constant revolutionary improvements in performance, and manufacturers simply cannot afford to support a range of media extending very far into the past.[33]

The requirements of the technical infrastructure to support bit preservation are not dissimilar from the best practices defined by ISO/IEC 27002 (formerly ISO 17799) *Information Technology Security Techniques – Code of Practice for Information Security Management*, which includes backup and data protection, disaster preparedness, security, automated file validation and quality assurance. Many enterprises involved in businesses other than long-term preservation may also find these requirements relevant to their business needs. These requirements are not specific to, but are critical to all digital preservation.

TRAC identifies three technical areas that a preservation repository must demonstrate competence in to be certified in data management and security. These closely follow the core principles of information security – confidentiality, integrity, and availability – known as the "CIA triad."[34]

---

32 Marilyn Deegan and Simon Tanner, "Key Issues in Digital Preservation," in *Digital Preservation*, ed. Marilyn Deegan and Simon Tanner (London: Facet, 2006), 15.

33 PrestoSpace "Tutorial: Media Migration and Obsolescence in Mass Storage: Obsolescence Issues in IT Storage Media." PrestoSpace: Digital Preservation of Audio/Visual Material. http://digitalpreservation.ssl.co.uk/storage/T2/T2-1-2.html (accessed January 6. 2010)

34 For more information, see, for example http://en.wikipedia.org/wiki/CIA_triad#Key_concepts

Technical System Infrastructure

Clearly, preservation involves more than just putting files on a server, or relying on the typical storage practices of a broadcast "archive"[35], which serves the purpose of getting programs to air, but is not designed for long-term storage. A sustainable preservation repository must maintain multiple copies of media and manage these at different locations. TRAC appropriately also requires repositories to have mechanisms for detecting and reporting bit loss. Processes defined for media and/or hardware migration, testing system changes, and reaction to software upgrades are also expected. .

Unfortunately, even the most sophisticated storage systems have a rate of failure that is unacceptable for those who are responsible for making sure that *no* bits are lost. David S. H. Rosenthal, co-founder of the distributed file replication tool LOCKSS ("Lots of Copies Keeps Stuff Safe"), observes that to compensate for these technological shortcomings, practical digital preservation systems must: "Maintain more than one copy by *replicating* their data on multiple, ideally different, storage systems." and "Audit or (scrub) the replicas to detect damage, and repair it by overwriting the known-bad copy with data from another."[36] Protection against corruption is typically ensured by creating at least three copies sorted in three geographically separate locations (ideally on different storage media. Matthew Addis, et al, recommend that, "Integrity of content is checked both periodically, e.g. every 6 months, and also when content is accessed or ingested into the system. If corruption is detected then repair takes place by replicating a known good copy from another location."[37]

Appropriate Technologies

Bit preservation also involves using technologies that are appropriate to the needs of the users who will ultimately access the bits in the form of a digital object. TRAC outlines this requirement to ensure that the repository has the software and hardware necessary to perform the services it has agreed to provide to its users. Since a public television repository would be dealing with very large audiovisual files, it must have the bandwidth to deliver content within the time period that users require. For instance, if the repository is providing content to news editors who will need high-resolution files within minutes, the repository must be able to provide that service.

Similarly, if a repository has agreed to provide users with files in formats and resolutions different than what was originally submitted, it would need the appropriate hardware and software to make derivative copies. The repository must remain abreast of the constant changes in technology and user needs, and have mechanisms in place to monitor and evaluate when hardware and software updates are needed.

Security

Disaster recovery planning, and protecting against security breeches and other threats are part of the repository's technical role. These are real threats to the content's long-term survival, beyond digital deterioration such as bit loss. Malicious attacks, from both internal and external agents, and major disasters such as earthquakes, floods, or fire can compromise an entire collection. Even simple human error poses a threat. While these events are unlikely on a day-to-day planning basis, they become more likely over a prolonged period. These disasters could have catastrophic effects both in terms of recovery costs and loss.

Security and disaster preparedness are thus additional elements essential to sustaining the bits in the preservation process. In TRAC, security refers to "IT systems, such as servers, firewalls, or routers. Fire

---

35 The concept of an archive has been appropriated by broadcasters in recent years, and has given the term a different meaning than we use throughout this paper.

36 David S. H. Rosenthal, "Bit Preservation: A Solved Problem?" (paper presented at iPRES2008, London, UK, September 2008): 5, http://www.bl.uk/ipres2008/presentations_day2/43_Rosenthal.pdf (accessed on January 6, 2010)

37 M. Addis, R. Beales, R. Lowe, L. Middleton, C. Norlung, Z. Zlatev, "Sustainable Archiving and Storage Management of Audiovisual Digital Assets." *SMPTE Motion Imaging Journal* (November/December 2009), p.29.

protection and flood detection systems are also significant, as are systems that involve actions by people."[38]

Criteria 2: Maintaining Accessibility of Content

> *"The cost of not responding to the avalanche of information can add up, yet not be immediately visible to CEOs and CFOs...We estimate that an organization employing 1,000 knowledge workers loses $5.7 million annually in just time wasted having to reformat information as they move among applications. Not finding information costs the same organization an additional $5.3 million a year."*
>
> - John Gantz, et al. *The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010*. (Framingham, MA: IDC, 2007): http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf

The second major aspect of sustainable preservation, maintaining accessibility of content, requires practices and procedures that are akin to what TRAC calls *digital object management*. This is a set of activities related to a repository's acquisition, management, and dissemination of digital objects. To ensure that it has the means to do this, a digital repository will need to collect adequate files in appropriate formats, sufficient metadata about those files, and organize them in a way that it can successfully manage and preserve them in the face of evolving technology. In the case of preserving digital public television, these are the activities that support identification, interpretation, playback, and migration of video and audio files, now and many years in the future.

The long-term accessibility of moving image content is highly dependent on good lifecycle management. This means that policies should be applied to the decisions made at the point that digital information is created; as it is being disseminated; and over the long-term. Choices of file formats, codecs and resolution, metadata creation and collection, versions, and file retention during the production and distribution stages of a digital public television program's life will have an important effect on the preservation of that program down the road. These factors will greatly impact how easily, efficiently, and reliably content can be found, retrieved, migrated, etc. These life cycle choices on long-term file preservation also have significant economic implications on repository operations and activities.

Much of the discussion that follows was informed by the experience of building the prototype repository at New York University for the preservation of born-digital public television. Working with Thirteen/WNET and WGBH, NYU analyzed production workflows, selected and collected sample programs and metadata, negotiated service agreements, and made plans to have the capacity to manage public television assets for the long-term.

Selecting and Acquiring Content

One of the Mandatory Responsibilities prescribed by OAIS is that a repository must, "negotiate for and accept appropriate information from Information Producers."[39] This requirement respects that relationships between producers and repositories vary greatly in terms of who initiates transactions and who determines what content and forms will be submitted. In a public television repository, the creators or producing entities of public television content will likely initiate transactions and be responsible for selecting the content they would like to deposit for retention. The depositor's task of selecting content

38 RLG-National Archives and Records Administration Digital Repository Certification Task Force, "Trustworthy Repositories," 43.

39 "Reference Model for an Open Archival Information System," 3-1.

to be submitted is an important part of this process,[40] and can have an enormous impact on the amount of work that must be done on the repository's end, on the overall costs involved in acquisition, and will affect what content users will ultimately be able to access.

The digital television production workflow, from pre-production to transmission, generates an enormous volume of broadcast and non-broadcast material of substantial size, especially when compared to an analog environment. In its 2008 white paper titled *The Diverse and Exploding Digital Universe*, the International Data Corporation (IDC) estimated that the broadcast, media, and entertainment industries "already generate, manage, or otherwise oversee 50% of the digital universe," which, in 2007, constituted approximately 2.25×10^{21} bits (281 exabytes or 281 billion gigabytes).[41] It would not be cost-effective or economically sustainable to acquire and preserve all of this content. In fact, on a worldwide scale, it would be technically impossible to store all of the digital content that we now create. On a smaller scale, the same holds for preserving digital public television -- we cannot save every single file!

Files that are identified to have long-term value are the ones that will be deposited in a repository. This throws the selection process onto both the content producer and the repository, who will each have to devise appropriate policies for selecting materials to be submitted to a repository that the repository will accept. This would include identifying types of files to be retained for short, medium, and long term.

The repository must also answer such questions as: How long should files be retained? Will all files be preserved in perpetuity, or will some classes of files be retained for a limited time frame? These are important questions, since, as Kevin Bradley points out, "Planned retention of digital materials for the appropriate period is part of a sustainable approach."[42] These decisions will have a significant impact on long-term management requirements and costs.

Determining The Best 'Master' Preservation File

In the public television context, one of the key operating decisions between depositors and a repository will be the format and resolution requirements of the video files to be preserved. Creators will want to submit the highest quality program files possible to a repository, which then become the preservation master files. Master files will be used to make derivative copies for all other uses, such as access copies for production, research, and public use.

The format, encoding, and resolution of the preservation master files can have a significant impact on preservation and access needs in the future. Consequently, they should be of high enough quality that they can suit the needs of each user group, now and in the future.[43]

---

40 The PDPTV project has examined the issue of selection for public television stations in detail in two reports, Recommended Appraisal Guidelines for Selecting Born-Digital Master Programs for Preservation and Deposit with the Library of Congress and a forthcoming report on selection criteria for locally produced programming. In the PDPTV model, public television depositors would then jointly determine the forms of acceptable content, such as file formats and metadata standards, with the repository.

41 John F. Gantz et al., *The Diverse and Exploding Universe: An Updated Forecast of Worldwide Information Growth Through 2011*, (Framingham, MA: IDC, 2008): 4, http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf (accessed January 6, 2010)

42 Bradley, "Defining Digital Sustainability," 159.

43 The planners of a future repository should have discussions with different user groups to determine the level of quality required. Early on in the PDPTV project, partners held a series of focus groups with potential repository customers to get their input on the types of content that should be collected, means of access, and quality of images and metadata they would need. Some of these groups had surprising requirements. The group of historians, for instance, spoke of the requirement to have high enough quality footage that would allow them to see details, such as clothing in crowd scenes, so that they could, "identify individual's political, religious and other affiliations."

The process of creating a preservation master begins when a decision is made as to what the format and resolution a digital video file will be when created, followed by what resolutions are needed when it is used for different purposes. For example, video files can be very low resolution (mesaured in terms of data rate) for online streaming (350 kbps), medium resolution for broadcast (8 mbps), and high resolution for production (50-100 mbps+). These options are constantly changing, as various manufacturers improve their recording equipment, editing systems, distribution protocols, and viewing devices.

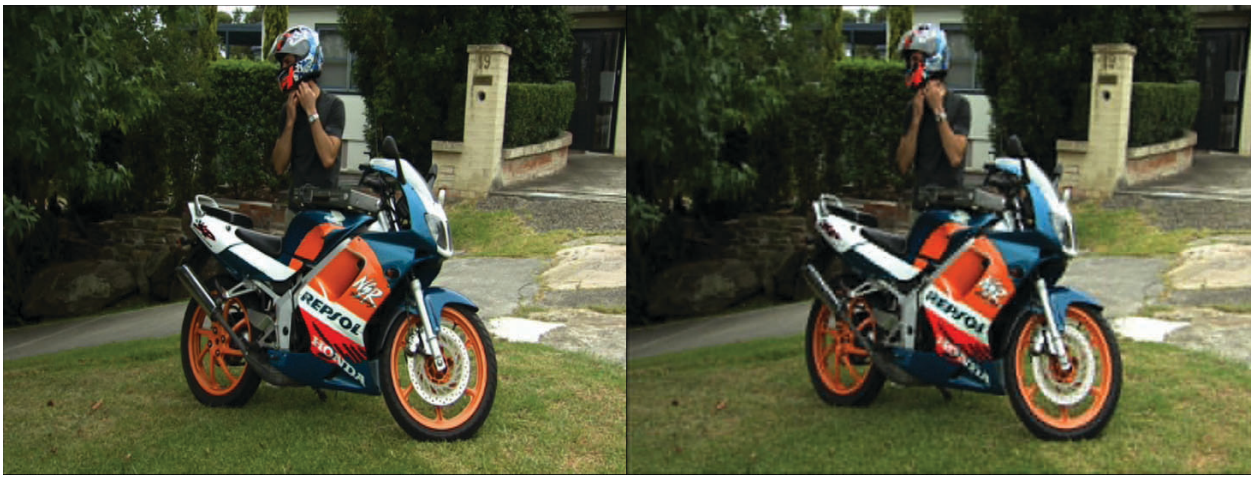

Figure 2. Same image encoded at a lower data rate (right). Notice the significant loss of image clarity. Credit: Mick Newnham (National Film and Sound Archive of Australia). Used with permission.

In order for a repository to preserve and provide sustained access to digital video, preservation master files have to be migrated to formats supported by current software and hardware, once the initial format becomes obsolete. However, when video files are converted to new file formats, there is a risk to image and sound quality. When a file is transcoded (converted from one digital encoding format to another) components of the image may be discarded due to compression algorithms acting on the image. After just a few generations, enough data chunks can be thrown away that the image becomes significantly altered or compromised.

Variety in formats, codecs, and resolution becomes an issue when files are selected for preservation, because it affects the range of technologies, skill sets, and funds needed to preserve and make these files accessible over time. Undoubtedly, a public television repository would support many file and encoding formats, but it must determine its own policies to specify realistic technical requirements, expectations for preservation, and limitations on file submissions that are appropriate to sustain within is own context.

Centrality of Metadata

> *"Searching for meaning in the content of unstructured data like images, video, clips, documents, and the numbers and characters in databases is the rocket science of the digital universe."*
>
> *- John Gantz, et al. The Diverse and Exploding Universe: An Updated Forecast of Worldwide Information Growth Through 2011. Framingham, MA: IDC, 2008. http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf*

Metadata is undoubtedly one of the most important factors in sustaining long-term access; some would

argue the most important factor. Metadata is the information about files that makes them identifiable, interpretable and retrievable as digital objects. The amount and quality of the metadata associated with digital objects will determine how well a repository can provide users with access to that content. Without metadata, preserved bits would appear only as un-interpretable strings of 1's and 0's. Metadata for public television content includes descriptive information like the title(s) of a program, credits, descriptions, and subject and genre keywords. It also refers to technical information about the file itself, such as its identifier or file name, file size, file format, codec, data rate, date of creation, aspect ratio, etc. Metadata related to usage, rights and preservation administration are also critical management tools. Comprehensive metadata makes powerful digital tools work wonders. A lack of metadata can render a multi-million dollar asset management system virtually useless.

The first TRAC condition in this area requires that the "Repository articulates minimum metadata requirements to enable the designated community(ies) to discover and identify material of interest."[44] It also states that a repository is responsible for capturing and/or creating the minimum required metadata and associating that metadata with the archived object (AIP).[45] To fulfill this important requirement, the repository, broadcasters, and users will each require an acceptable minimum set of descriptive metadata that meets each party's needs. This will likely turn out to be a significant amount of information. These stakeholders will also need to decide how that information can be supplied, and in what form.

Creating and supplying metadata can be a headache for producers, whose primary concern is to get content on the air, not fill out templates or databases. Yet in order for a piece of content to be truly searchable through keywords, descriptions, titles, and identifiers, someone must attach those terms to the content. Leaving it to a repository to generate essential metadata is risking unnecessary errors and incomplete records, not to mention added costs.

Much of the descriptive (title, creator, keyword, etc) information will need to be provided to the repository by the submitters, as they know their content the best. At the same time, a repository will need to receive metadata in a standard form, or transform it, in order to manage the data and make it searchable. Non-standardized metadata requires significant time and effort to normalize into forms appropriate for a preservation repository. This labor-intensive process adds to the overall cost of preservation, and can hinder future access.

Throughout the public broadcasting system, existing metadata comes in as many forms and flavors as do file formats. Information about legacy tapes is often found only in spreadsheets, or even old log books. Databases currently in use have an enormous variety of data models, field names and input methods, and often do not support export functionality.

Metadata collected for the PDPTV repository that originated from the two partner stations and PBS, for example, resulted in four completely different sets of information. Even when metadata were submitted by a producing station and the distributor (PBS) about the same program, the structure and value of the data was often wildly different. The NYU repository spent a considerable amount of time normalizing this data to a standard schema that was based on the Public Broadcasting Metadata Dictionary (PBCore), along with METS and PREMIS, two widely implemented digital metadata standards maintained by the Library of Congress.[46]

More and more tools and methods for improving metadata creation are emerging. Standards like PBCore can facilitate metadata creation across the public broadcasting system by providing common data definitions, controlled vocabularies, and schema. Automatic tools can be used to extract technical

---

44 RLG-National Archives and Records Administration Digital Repository Certification Task Force, "Trustworthy Repositories," 36.

45 Ibid.

46 Early on in the project PDPTV team wrote a report on the metadata standards and their use for the prototype repository, which can be accessed at http://www.thirteen.org/ptvdigitalarchive/files/2009/10/metadata-guidelines.pdf

metadata from the files and populate databases and catalog records. Audio indexing and object recognition technologies are becoming more sophisticated. Web-based cataloging tools can make it easier for any number of users to add metadata to records, from anywhere in the world. Some of these tools are even allowing users to tag time-coded sections of clips with keywords, rather than assigning terms to an entire length of video.

These standards and tools make data exchange more efficient and help to reduce the risk of errors. PBCore will hopefully gain traction throughout the public broadcasting system as its implementation at stations and distributors increases. Standardization of metadata across the public broadcasting system would free up the drudgery currently required when exchanging media and their associated metadata. It would facilitate mass submission into a repository, and would enable the repository to spend less time on metadata quality control and transformation, and more time developing tools to better preserve and provide access to the content. Standardizing metadata also serves the repository, as it allows files to be shared, exchanged, migrated or transferred to other entities. In short, as the PrestoSpace project reminds us, "Get the metadata right and the digital archive will blossom."[47]

Policies & Agreements

A public television preservation repository will only be able to perform its services by having clear policy agreements with content producers and users. As TRAC states, "Acquisition involves a crucial interaction between the repository and the depositor."[48] As content submitters, broadcasters must communicate what type of files and metadata they can provide to a preservation repository. At the same time, a preservation repository must inform submitters what types of files and metadata it will be able to preserve. Additionally, as content users, broadcasters as well as researchers, teachers, scholars, and the general public will need to communicate how they would like to find and retrieve content.

To guarantee long-term access, a repository needs a set of general operating policies that address a range of technical issues, political concerns and administrative requirements. Such policies will affect the repository's budget, business plan, technology investments, and staff skill level needs. Just a few examples of these concerns:

- **Submission**: Who can submit files? What files and data will the repository require? By what mechanism will content be submitted by the submitting entity?

- **Updates**: Should new versions of a program replace older ones stored in the repository? Should all versions of a program be maintained in perpetuity? Who has permission to take files out? By what mechanism will rights data be updated when they change?

- **Dissemination**: What user groups will be permitted to access content? By what means? Will there be cases when users will retrieve metadata only, or will they always want video and/or audio as well? Who will have permission to access different classes of content?

Submission Information Packages

The repository and the Information Producers will have to determine what content, including digital objects and accompanying metadata, will be accepted by the repository, and in what form . This bundle of information is known in OAIS terms as a *Submission Information Package (SIP*). The make-up of the SIP can also play a determining role in the form the files will be distributed, or the Dissemination Information Package (DIP), so the repository must also have some idea of the needs of end users when determining SIP requirements. The configuration of a DIP (or DIPs) is dictated, in part, by who will or can

---

47 PrestoSpace, "Metadata and Cataloging." PrestoSpace: Digital Preservation of Audio/Visual Material. http://digitalpreservation.ssl.co.uk/metadata/ (accessed on January 6, 2010)

48 RLG-National Archives and Records Administration Digital Repository Certification Task Force, "Trustworthy Repositories," 21.

have access to the files. Consequently, the criteria and policies regarding these various operations are important to be specified in advance.

Policies governing the prototype repository for the PDPTV project were informed by input of all project partners, as well as guidance from focus groups and other potential users. These discussions with station partners and PBS resulted in a concrete set of requirements for the repository to model its services on, to develop SIP criteria, and articulate dissemination guidelines.

Archival Information Packages

Once content is submitted to the repository, it needs to be packaged in an archival form to protect its integrity and authenticity, so that it may be accurately retrieved and disseminated at a later date. This is known as an *Archival Information Package (AIP.)* The AIP may be identical to how the content was submitted, but it will most likely be somewhat different.

The repository will have to verify that the correct number and types of files have been submitted, create checksums that help verify the integrity of the bits, and generate structural information that associates the different pieces of the AIP to one another. It may transform some of the submitted files or metadata, so that they are easier to manage.

TRAC provides a number of requirements that relate to the creation of AIPs from SIPs. The details are not important for this discussion, except to emphasize that the more standardized the files and metadata are that are submitted to the repository, the easier it will be to manage and preserve those files. The more complex the number and type of submitted files, metadata, and relationships, the more difficult and costly it will be to preserve them.

This area also includes a requirement that the repository have “a documented process for testing the understandability of the information content and bringing the information content up to the agreed level of understandability."[49] This implies that the repository must make the content available to users through tools that are standard to that community. A concern here might be, if files are submitted in formats that can only be read using complex and expensive hardware, the repository would be responsible for either making that hardware available, or transcoding the files to make them accessible to a wider user community. Another issue would be if sufficient metadata are not submitted, the repository will have to spend time creating it in order to provide its agreed upon service of making content searchable and identifiable.

The more the process is facilitated by the content producers, the less costly preservation services will be. As is discussed later in the case study section of this report, the highest cost for the PDPTV repository was staffing. In a future repository, this will likely also be the case, however, a more standardized, streamlined workflow, without a great number of abnormalities or need for analysis will reduce the number of staff needed, and therefore reduce the overall cost of preservation.

Preservation Planning

The core function of a preservation repository is of course, preservation. Yet, as TRAC states, "It is not enough to simply preserve information. A repository must do so in accordance with predefined, documented, preservation policies and procedures, and it must have identified mechanisms to update those policies and procedures in a response to changing technologies. Without such documentation, a repository cannot pass an audit even if its work is otherwise exemplary."[50] These policies and procedures must address the auditing of bits, refreshing of storage media, and the obsolescence of file formats, software and/or hardware. The documented preservation practice should also include monitoring the technological landscape for changes and potential file format obsolescence. For digital

---

49 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories,” 29.

50 Ibid., 31

video, especially broadcast materials, this will require close attention, as the technology *du jour* is always a moving target.

Contracts and service agreements between the repository and submitters and users will detail and document the responsibilities and obligations that the repository will undertake, as well as the requirements and costs that the submitters and users will meet. These will also help document the procedures that the repository has in place and its future plans.

Criteria 3: Organizational Infrastructure

As we have outlined, there are a number of technological and procedural functions of a repository that are necessary to ensure that content will remain intact and accessible. However, in addition to these, the organizational framework of a repository must also be stable enough to guarantee that the technologies and processes will persist over the long-term.

TRAC stresses that the organizational infrastructure of a repository is equally as critical as technological infrastructure and digital object management. TRAC measures five organizational attributes as indicators of a digital repository's "comprehensive planning, readiness, ability to address its responsibilities, and trustworthiness."[51]

- Governance and organizational viability;
- Structure and staffing;
- Procedural accountability and policy framework;
- Financial sustainability, and contracts;
- Licenses and liabilities.

Overall, the most significant organizational issues for public television will be staffing, funding and resources, and legal rights. Other relevant issues will pertain to mission, governance, policy and procedures, and the capacity for the system to sustain local and/or regional repository services, vs. building a single national entity.

Organizational Viability

In terms of governance and organizational viability, TRAC states that a repository should explicitly demonstrate its "commitment to the long-term retention of, management of, and access to digital information" through a mission statement.[52] It should also have an "appropriate formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate, or the governing or funding institution substantially changes scope."[53] These points are indeed fundamental to the sustainability of the repository's contents. Invaluable historical public broadcasting materials could potentially be lost forever if the caretaker of those materials suddenly goes out of business, and does not have any mechanism in place to transfer them to another responsible entity.[54]

The relationship between the repository and its governing or funding institution is only referred to in passing in the TRAC document, but is nonetheless important. If public television plans to assume

---

51 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories,” 9.

52 Ibid., 10

53 Ibid.

54 For an example of such a situation, see the case of Digital Railroad, a company that offered "archival" services to photographers, which suddenly went out of business in 2008. Its customers were given 24 hours notice to retrieve their work and find another place to store it (National Press Photographers Association, http://www.nppa.org/news_and_events/news/2008/10/digitalrailroad.html [accessed on January 6, 2010]).

responsibility for sustaining one or more repositories in the future, the correspondence between the repository's mission and the public television mission will likely need to be articulated. That is, public television will need to demonstrate how preservation activities serve its overall educational and informational aims to justify supporting a repository.

The access component of the repository's mission would be the key element to building this case. Access, as Ray Edmondson, Deputy Chair of the UNESCO Memory of the World program, states, is "the raison d'etre of archiving...its visible evidence [and] often its political justification."[55] Future planners of a public television repository should emphasize this connection; preservation enables ongoing re-use and availability of content. It is therefore a prerequisite to fulfilling the public broadcasting mission to "facilitate the development of, and ensure universal access to non-commercial, high-quality programming and telecommunications services."[56]

As the case of the previously unfulfilled 1967 Congressional charge to CPB to establish and maintain a library and archives for public broadcasting demonstrates, however, a stated mission or mandate does not necessarily lead to sustained resources and support. Significant changes in the funding, mission, collecting scope, or staffing of the organization that runs the repository can put data at risk of irreparable loss. As TRAC requires, any repository for public television content should identify trusted inheritors or have a plan in place to return content to depositors should the need arise.

Governance and Decision-Making

The governance of a preservation repository, whether it is maintained within the public broadcasting system or not, may also present challenges. Governance issues, as the Blue Ribbon Task Force indicates, are "broadly speaking, issues of responsibility, authority, accountability, and trust."[57] The public broadcasting system is characterized by decentralized and diverse decision-making structures. It is currently comprised of over 350 member stations including regional stations and uplinks, all of which operate independently and are subject to different governing authorities.

There are many different governance models that exist within public broadcasting that can inform any repository plans. Stations like WGBH and WNET, for example, are community-owned and are responsible to their respective Boards of Directors. Other stations are owned or licensed by colleges or universities, university systems, or by state governments. Meanwhile, the Public Radio Satellite System is set up as a co-op, in which membership is open to all public radio stations and all members have equal access to basic services and decision-making. Planners of a future repository for public television will need to consider questions of responsibility, authority, accountability, and trust with this decentralized system in mind.

Structure and Staffing

A trustworthy repository should have an appropriate number of staff with adequate expertise and skills that are regularly renewed through professional development.[58] Employing skilled staff will be a key issue in building a digital repository for public television. The existing technical expertise at public television stations tend to center on broadcasting, production, IT operation, and web development. If public television chooses to create a new repository, it will need to identify the areas where it needs expertise, find skilled candidates to fill those positions, and secure a financial commitment to maintain that staff.

---

55 Ray Edmondson, “Audiovisual Archiving: Philosophy and Principles.” (Paris: UNESCO, 2004), 4. http://unesdoc.unesco.org/images/0013/001364/136477e.pdf (accessed January 6, 2010).

56 Corporation for Public Broadcasting, "CPB's Goals and Objectives," http://www.cpb.org/aboutcpb/goals/goalsandobjectives/ (accessed January 6, 2010).

57 Blue Ribbon Task Force, “Sustaining the Digital Investment,” 24.

58 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories,” 11.

It is possible that repository services for public television may be performed by an external third-party, such as New York University, the Library of Congress, or a commercial service, rather than within the system. Regardless of who the provider is, though, it is important to keep in mind that preservation standards and practices need to be employed throughout the life cycle on the digital object, not just at the archival stage. As discussed elsewhere in this report, individuals involved in production and other aspects of file creation also have a key role to play in supporting content that will eventually be managed by the repository.

Stated Policies and Procedures

A trustworthy repository should clearly document its requirements, decisions, development, and actions in order to assure stakeholders that it is fulfilling its role. Fortunately, as a first step, the process of developing transparent and accountable procedures and policies has been tested and documented through the PDPTV project. The design of the prototype repository and its services, for example, was arrived at through meetings with designated producer and user communities. Meetings between NYU and the producing station partners about their preservation and access needs produced clearly defined technical requirements for submissions and types of materials for archival storage. Meanwhile, focus groups with the user community resulted in a set of recommendations for a future selection policy. Through its research findings and experiences with the test repository at NYU, the work of the PDPTV project can serve as a useful resource toward the establishment of procedures and policies for a long-term permanent repository for public television materials.

Legal Issues and Rights

The importance of accessibility and its attendant technical requirements have been discussed at length in this report. Providing access to identified user communities, however, also requires careful consideration of rights issues. The legal issues surrounding access can present a steep challenge, as public television programs can be encumbered by a mountain of rights that can make access by anyone other than the copyright holder difficult. These include intellectual property rights, privacy rights, publicity rights, performance rights, and agreements with collective bargaining unions.[59]

Today, rights are further complicated by users' expectations for online access to content from all around the world. Untangling these rights is certain to be complex and potentially costly for a great deal of material. Comprehensive rights documentation will be an essential prerequisite for submission to a preservation repository, so that legal agreements about the use of content can be explicit. However, while rights issues can greatly affect the costs and complexities of preservation, they are not insurmountable and should not be used as an argument against preservation or access.

According to TRAC, a repository should have clear and explicit contracts, licenses, and liabilities with content owners, service providers, and other relevant parties.[60] And, while it may not have any rights to the materials itself, the repository should have a way to manage the rights and restrictions on use of content as defined by those agreements, such as a "policy statement that defines and specifies the repository's requirements and process for managing intellectual property rights."[61]

This does NOT mean that the repository must be responsible for maintaining any of the rights related to the content it is holding, or that it must take on the effort and costs for clearing or negotiating rights

---

59 The PDPTV project has sought to identify and clarify the rights issues involved in making public television content available for access and reuse. In our forthcoming report, PDPTV examines the many contract, intellectual property, and copyright stipulations entailed in the preservation and use of public television programs beyond their initial broadcast window.

60 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories,” 18.

61 Ibid.

agreements on behalf of submitters or users, unless, of course, these are services that the repository has agreed to provide.

It does mean that submitters will need to provide relevant rights and use information to the repository so that it can perform the agreed upon services and create appropriate access policies. The repository will have to tailor the level of access and usage permissions according to the rights status of the content, because without such information, it could be liable for any misuse of copyrighted material in its collections.

Section 108 of the U.S. Copyright Act provides exceptions to the Copyright Act for libraries and archives by allowing them to make copies for preservation purposes, in order to serve the public and to ensure the availability of works over time.[62] Even if the repository's preservation activities are legitimized, though, many initiatives are underway to revise copyright laws in response to the radical transformation in preservation practices dictated by the changes in technology. A repository will have to stay current on such alterations to ensure that "policies are in place to address liability and challenges to those rights."[63]

Criteria 4: Reliable Funding

Undoubtedly, the financial soundness of the repository is central to its sustainability. TRAC affirms that a trustworthy repository should be able to show that it is financially sustainable by *having a business plan* and by adhering *to good business practices that are transparent and compliant with accounting standards and legal requirements*.[64] The business plan should describe how the repository will generate income to meet its costs, and how it will deal with contingencies. There should also be procedures in place to regularly review and adjust business plans as needed. Sound business practice includes good analysis and reporting on risk, benefit, investment, and expenditure, as well as good monitoring for and bridging gaps in funding. Public broadcasting will have to meet these financial requirements if its preservation efforts are to be successful.

* * *

Clearly, there is more to sustainable preservation than just good storage. It also relies on other factors:

- A stable infrastructure that maintains and audits multiple copies of material and protects against disasters is one critical component.
- Mechanisms that guarantee the long-term accessibility of that content must be in place. A user must be able to find, retrieve, and playback files in a timely and efficient manner, or the whole exercise of saving the bits is essentially useless.
- There must be an organization with efficient governance, staffing, and legal policies in place to successfully care for digital content.
- Finally, the finances must be adequate and reliable enough to maintain the repository activities and functions over time.

Without all of these interrelated parts, digital preservation is not sustainable. To accomplish this, the need for changes in the life cycle management of digital public television files is obvious. As stations implement digital workflows from production to broadcast, they are beginning to discover the need for standardized metadata, open and well-supported file formats, unique identifiers, interoperable systems, and criteria for selection and retention of their materials all along the production chain. The changes

62 The terms of this provision are discussed in detail in the aforementioned PDPTV report on intellectual property.

63 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories,” 19.

64 RLG-National Archives and Records Administration Digital Repository Certification Task Force, “Trustworthy Repositories,” 16-17.

necessary to support preservation should be part of a planned shift to digital production and distribution. Actually preserving the programs, however, will take a more concerted effort.

But where will the money come from? Who should pay? How much will it cost? What are the benefits? Because access to adequate funding looms large in the system, the following sections in this report will address solely the issue of generating income, looking at the economic issues that public television faces, potential business models, and the benefits of preservation, as well as the costs.

* * * * *

Chapter 4

Economics of Sustainable Preservation

As we have discussed, digital media are fragile and highly dependent on changing technologies and systems to remain accessible and usable. Maintaining a trustworthy repository with the appropriate technical and organizational infrastructure to support these digital objects requires ongoing financial resources. This point seems quite clear but, as Brian Lavoie indicates, there is more to the issue of sustainability than the simple availability of money. He warns against falling into a "trap" of thinking that "if only more resources could be piped into digital preservation activities, and sustained on an ongoing basis, the economic aspects of the digital preservation problem could be solved"[65].

On the contrary, sustainability relies on a number of economic factors besides just the availability of funds. As we have already pointed out, it also requires establishing sound digital preservation infrastructure and practices. Before resources can be applied to these activities, economic considerations such as defining costs and benefits, and developing business models through which income can be generated, also need to be explored.

Public television's inexperience with financing digital preservation may make the tasks of planning and operating a repository seem a little daunting. The economics are somewhat different and unfamiliar. Part of the aim of the PDPTV project is to demystify these aspects of the process and look specifically at some business models for supporting digital preservation that may be useful and instructive.

The Difficulties of Analyzing The Costs for Digital Preservation

A key economic aspect of digital sustainability is knowing the short- and long-term costs of preserving

65 Brian Lavoie, "The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation," D-Lib Magazine 14, no. 3/4 (2008), http://www.dlib.org/dlib/march08/lavoie/03lavoie.html (accessed January 6, 2010).

and managing digital content. While public television may be able to estimate basic costs such as analog digitization per hour or storage per terabyte, the long-term overall costs to maintain and sustain access are extremely difficult to project. Estimating these kinds of costs in the planning process is challenging because the total cost of ownership is dependent on many factors that are particular to a repository's specific organizational context, architecture and implementation. Some of these innumerable factors include the repository's existing facilities, staff, and hardware and software infrastructure; the complexity and number of file formats to be stored and transformed; the volume of content; electricity costs; metadata requirements; type of storage; timing of activities such as migration; and access requirements. Just because storage costs are going down doesn't mean preservation has become any cheaper; simultaneously, the amount of content produced is rising exponentially, meaning the labor costs involved in managing all the data is going up. Finally, the technical and organizational resources of the repository's parent institution are also variables that will have significant impact on costs.

In recent years, researchers have proposed models to help organizations project the costs of digital preservation.[66] However, these models primarily address text-based files, spreadsheets, or small datasets, and are not always applicable to the unique circumstances of digital video. Digital video file formats used in public television, such as DV and MPEG-2, are much larger and more complex than text files. They therefore place greater demands on storage, bandwidth, and infrastructure. Because of their complex nature, preserving digital video files also involves solving issues related to metadata, codecs, and file wrappers. In addition, public television often produces multiple versions of a program with slight changes in content. Over the course of production and transmission, it creates copies of programs that vary in format and quality. The presence of all of these versions adds to the cost of preservation both in terms of storage and time put into selection, metadata collection, and other maintenance tasks.

The lack of metrics to measure and assess the costs of preserving digital video hampers business modeling and planning. There are also little or no data on the full costs of preserving digital video. Figures for storage or other discrete functions have been published, but for the reasons stated above, they cannot be relied upon to be applicable to situations other than their own.

As a simple exercise, a comparison of the cost to store a terabyte of digital content using various commercial repository service providers shows that expenses can vary by more than an order of magnitude, which reflects the specific configuration of the storage system and the types of services offered:

Sample Costs for Storage Systems		
Service / System	**Description**	**List Cost**
Off-the-shelf RAID system tower	Unmanaged, limited risk protection	~$280 / TB
Google Shared Storage	Online storage service for Google members, for email and photos only	$1250 / TB / year
Amazon S3	Online storage service (fee does not include file transfer or access)	$1800 / TB / year
San Diego Supercomputing Center	Archival tape storage	$500 / TB / year
OCLC Digital Archive	Service designed to support long-term management of data	$7500 / TB / year

Table 1. A sample comparison of the cost to store a terabyte of data using various repository systems/service providers in 2009.

Taking a closer look, SDSC's cost of $500/TB/year reflects the cost to maintain a single offline copy.

66 For examples of cost models, see the Netherlands Archives' Digital Preservation Testbed (http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=286&lang=en), the Life Cycle Information for E-Literature (LIFE) Project (http://www.life.ac.uk/), and the Keeping Research Data Safe study (http://www.jisc.ac.uk/publications/publications/keepingresearchdatasafe.aspx).

Amazon S3, for its price, offers users the convenience of storing and retrieving their data at any time over the web, but makes no guarantees about security or storage permanence.[67] Meanwhile, OCLC Digital Archive provides content migration, six back-up copies, disaster recovery, monitoring and reports, among other services.

As these pricing differences demonstrate, the cost of a repository will vary greatly according to what it offers and whom it serves. A preservation repository will have vastly different needs than a repository meant for high-speed access, which is more concerned with performance than long-term retention. A preservation repository for public television content will have its own requirements where costs will reflect its specific preservation and access needs.

Despite the numerous variables to take into account, defining and determining costs is an important step in planning for a long-term repository for public television content. On the most basic level, it is necessary to know whether there is or will be enough money to perform the tasks laid out. As a first step, the PDPTV project has attempted to document the costs and staffing associated with implementing its prototype repository at NYU. While the actual figures cited by the PDPTV project are unique to the NYU repository, the research nonetheless breaks the process down into cost categories and identifies the factors that influence cost. These costs and cost categories are summarized in **Chapter 5 - Case Study: Preserving Digital Public Television – Costing the Prototype Repository** and in Appendix B**.**

A Variety of Economic Models

Besides understanding costs and benefits, economically sustainable digital preservation depends on the development of economic and business models to guide the repository's activities. This section of the report addresses economic models, which make up the foundation upon which business and cost models are built. An economic model is a simplified representation of an economic situation, which can be expressed in the form of diagrams, graphs, or words. An economic model can be used to gain insight into how and why stakeholders make certain economic choices. It provides a framework for understanding the relationships between economic stakeholders and establishes cause and effect relationships.

An economic model for digital preservation should provide a simple, abstracted view of the digital preservation process that we can use to identify the interests, motivations, and roles of stakeholders involved, and how they are related to each other. An economic model for preserving digital public television would, for example, help to predict how an entity, such as a program producer or a repository, might choose to allocate its resources under given economic constraints, and indicate what actions are required by other entities, such as users or public funders, to ensure that preservation occurs. This knowledge allows us to then devise strategies, or business models, to shape the given economic reality to achieve our preservation ends.

Economic modeling is an important precursor to business planning. The Blue Ribbon Task Force on Sustainable Digital Preservation and Access uses the economic model as its "primary tool" in its analysis of mechanisms and policy instruments to sustain digital preservation activities over time.[68] Describing economic models, and their relationship to business models, the Task Force indicates:

> The development of a good business model relies upon a well conceived economic model. The economic model 'describes' how economic reality works and the business models

67 Peter Murray, "Long-term Preservation Storage: OCLC Digital Archive versus Amazon S3," Disruptive Library Technology Jester Blog, posted May 16, 2008, http://dltj.org/article/oclc-digital-archive-vs-amazon-s3/ (accessed January 6, 2010).

68 Blue Ribbon Task Force, “Sustaining the Digital Investment,” 29.

> provide 'templates' for acting within that reality. With a good economic model, organizations can assess which types of business models will work best for them, given their missions, the environments in which they operate, and the types of products, services, and customers they face.[69]

The development of economic models is a complex exercise. In order to develop the specific business models to effectively sustain preservation, public television will need to apply an economic model based on a defined organizational structure and practical conditions. The ways that a repository could be set up within or outside the public television system will have a bearing on the distribution of economic roles and responsibilities in the preservation process.

Determining an Appropriate Operating Framework

There are two intersecting dynamics in public broadcasting that will influence an economic model for preservation:

- **Stakeholder Roles** – Who has the primary interests in controlling, managing and using the content? Who will benefit directly from participating in a repository? How will these groups interact with and relate to each other?

- **Repository Operation** – How will the repository function, how will tasks be carried out, and what will be the institutional framework? What entities will be involved, and what roles will each have?

Balancing the Economic Interests of Different Stakeholders

First, it is necessary to examine the economic interests that would be reflected in a repository. In *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*, Brian Lavoie suggests that we understand digital preservation as something like an "aftermarket" service; that is to say, digital preservation participates in "the market for goods and services which maintain and/ or enhance the ability of a durable good to provide a stream of benefits over an extended period of time."[70]

The "durable good" in this case is digital information, which will deteriorate and depreciate in value in the absence of these services. Meanwhile, the "aftermarket" itself is made up of stakeholders who occupy critical roles in the preservation process. In Lavoie's simplified economic model, there are three key economic decision-making roles among stakeholders in digital preservation:

- **Rights Holder**: The entity that owns the rights to a digital resource and has the authority to dispose of those rights. The "Rights Holder" of a particular piece of content could be, for example, a copyright holder, a program producer, or a program distributor.

- **Archive**: The entity that implements and manages the preservation process. The "Archive" role could potentially be fulfilled by a producing station, a third-party such as a public institution, or a new entity within the public television system.

- **Beneficiary:** The entity that benefits from the long-term retention of the digital resource. The "Beneficiary" might be any of the stakeholders who would make use of the preserved content.

In a hypothetical preservation scenario for public television, the roles of Rights Holder, Archive, and Beneficiary could be filled by new or existing stakeholders. As mentioned previously, potential

---

69 Ibid., 33.

70 Brian Lavoie, “The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making.” (Dublin, OH: OCLC, 2003), 21.

stakeholders in preserving public television content include copyright holders, program producers, filmmakers, program distributors, educators, educational resource producers, students, academics in higher education, researchers, journalists, and the general public.

Roles among stakeholders can vary and overlap, and with such a broad group, various interests and demands on the repository can become contradictory. One stakeholder could, for example, be both a Rights Holder and a Beneficiary, or be both a Rights Holder and the Archive. The relationship of one stakeholder to another in a given situation produces what could be called an "organizational scenario". Since an organizational scenario describes the structure of a particular digital preservation activity, it is possible that a single organization can take part in a number of organizational scenarios at once.[71] **Appendix A: Economic Models Based on Market Analysis of Stakeholder Interests** takes a closer look at these organization scenarios and further describes how different scenarios could have a relationship to funding incentives.

The Influence of Repository Operation

The second dynamic comes from the repository design. The distribution of Rights Holder, Archive, and Beneficiary roles among stakeholders and the constraints they face are not simply arbitrary. They represent genuinely different interests in how the repository could function, and the characteristics of the economic situation will depend on determining how the repository is set up, how responsibilities are distributed, and who the repository serves.

But these are not the only considerations that must be taken into account in repository design. Because of the nature of digital content and networked environments, the division of labor and responsibility within a repository can be allocated in many different ways. For instance, the constituent parts of a repository may be carried out holistically within a single institution, or they may be spread across multiple institutions where each specializes in a separate activity. Factors that must be considered include:

- The resources that producers and stations are able to contribute to preservation efforts;
- The amount of material that needs to be preserved;
- Where content currently resides;
- The cost/benefit of sharing responsibilities versus outsourcing to third-parties;
- The need to reuse the content or make it available;
- Financial stability of the station or producer, and resources potentially available;
- The existing expertise in the system;
- The interests and resources of other entities, such as CPB or PBS.

At the same time, in planning for preservation of its digital content, public television must determine which operating design best suits its needs, given its existing structures, including the possibility that it will support a combination of models. There are a wide range of possible options, each of which assumes different economic factors and available resources.

- **Centralized repository for all public television materials within public television system:** A single existing or new organization within the public television system would be responsible for the full range of preservation activities.

- **Centralized repository for all public television materials via third-party:** A trusted third-party organization would be responsible for the full range of preservation activities (i.e. the PDPTV prototype repository at NYU).

---

71 For a detailed discussion of model Organizational Scenarios, please see Appendix A. A preservation repository for public television would do well to examine these scenarios and the various organizational structures implied by each model.

- **Individual producing stations maintain their own repositories:** Following the model of the existing analog videotape libraries at many stations, each producing station maintains its own repository for its own content.

- **Decentralized disaggregated repository shared between public television entities:** A decentralized but networked approach in which responsibilities are shared between multiple existing public television sites.

- **Decentralized repository that relies on third-party services:** A disaggregated but networked approach in which trusted third-parties provide public television with repository services, such as storage or data**.**

- **Decentralized repository shared among public television entities, and administered by a third-party:** A third-party entity is made responsible for administering the repository, with storage or distribution nodes (or other services) at certain sites throughout the system.

As these many examples indicate, the design of the repository can place authority and responsibility for preservation with many different entities, which are not necessarily mutually exclusive. The Blue Ribbon Task Force indicates that there is a key connection between repository design and business planning, and suggests that organizations think strategically about how the distribution of roles and responsibilities can reflect who the stakeholders are and align with the ways they fund preservation. The Task Force asks, "Does the responsibility for ensuring that preservation takes place reside with the appropriate stakeholder or stakeholders, given the particular choice of mechanism for obtaining resources to support a digital preservation activity?" If given fixed options, they indicate that "Sometimes the best method of coordination will involve a transfer of preservation responsibility from one entity to another."[72]

Assessing Cost-Effectiveness

Along with which repository set-up best matches the available avenues for funding, the choice of repository model will also depend on which one is most efficient for the public television system. Public television will have to determine, for example, what repository services are more cost-effective to bundle in a centralized location, which may allow it to take advantage of economics of scale (and scope), and which may be more effective shared across many institutions or even managed by individual institutions.

To date, most research has indicated that disaggregated but networked approaches are more cost efficient than centralized ones.[73] When repository tasks are shared, labor and responsibility can be divided in sensible and effective ways. Brian Lavoie and Lorcan Dempsey of OCLC Research have found that a disaggregated system can allow users to combine services in various ways to suit their needs

---

72 Blue Ribbon Task Force, “Sustaining the Digital Investment,” 22.

73 For example, in its 1996 report “Preserving Digital Information”, the Task Force on Archiving of Digital Information asserted that "the most effective and affordable strategy for developing a system of digital archives is to assume a distributed, rather than centralized, structure for collecting digital information objects, protecting their integrity over the long term, and retaining them for future use” (p. 21, accessed January 6, 2010 from http://www.clir.org/pubs/reports/pub63watersgarrett.pdf). More recently, the Joint Information Systems Committee (JISC) “Digital Moving Images and Sound Archiving Study” indicates that "the way forward as envisioned by many is to disaggregate the tasks undertaken by a digital repository, so that not all repositories need to undertake all tasks" (p. 99, accessed January 6, 2010 from http://www.jisc.ac.uk/media/documents/programmes/preservation/moving%20pictures%20and%20sound%20archiving%20final%20version.doc).

Finally, a position paper from a NSF/JISC Repositories Workshop, April 16, 2007, on “The Need for Formalized Trust in Digital Repository Collaborative Infrastructure” argues that “The increased number and diversity of those concerned with digital preservation—coupled with the current general scarcity of resources for preservation infrastructure—suggests that new collaborative relationships that cross institutional and sector boundaries could provide important and promising ways to deal with the data preservation challenge. These collaborations could potentially help spread the burden of preservation, create economies of scale needed to support it, and mitigate the risks of data loss.” (http://www.sis.pitt.edu/~repwkshop/papers/berman_schottlaender.html).

while enabling service providers to realize economies of scale.[74]

The AVATARm[75] project in the UK, a research and development initiative that is developing solutions to improve the utility of digital file storage for broadcast archives, advocates a approach that combines local storage, grid storage technologies, and third party storage services in order to reduce risk, leverage economies of scale, and improve access times. They make an important point on this issue of cost effectiveness:

> Value chains and business models are also changing and increasingly delivered through multiple service providers or organizations (e.g. outsourced services, federated preservation across organization, etc.). The economies of scale, power, cooling, and staff costs that can be achieved by organizations such as Google mean that as network costs continue to fall, in-house solutions will become increasingly expensive compared with outsourced or federated models.[76]

While potentially more cost-effective and efficient, an important consideration for disaggregated networked repositories is *interoperability.* Interoperability, as defined by the ISO/IEC 2382 Information Technology Vocabulary, is "the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units." A disaggregated repository relies on the interoperable exchange of data. This will likely represent a concern for public television because of the current lack of standards, the variety of formats, and the various software and hardware employed across the system. System-wide life cycle management would need to be implemented in order to address problems with interoperability, and public television would have to build or use tools, interfaces, and infrastructure to allow distributed, heterogeneous sources to be accessed and manipulated. Besides technical interoperability, public television would also need to develop compatible and coordinated work practices and communications. In planning for a repository, these costs should be considered, but also weighed against the cost of inefficiency and lost opportunities resulting from inadequate interoperability.

Given the already decentralized structure of the public broadcasting system and the availability of technology to facilitate a networked approach, a repository with distributed services is a strong option. To take advantage of the disaggregated approach, however, public television will need to determine "a sensible deconstruction of the digital preservation process into a set of more granular services" as well as "the optimal degree of specialization across preserving institutions" given its existing infrastructure.[77] With its limited staff expertise in digital preservation, for example, it may make sense for public television to locate resources for technical planning, support, and training within a single entity which can then provide that service to others in the repository system. This could be an entity that currently exists in the public broadcasting system, a newly formed entity, or a third-party.

The Need for a Trusted Solution

Whether the repository will be operated within the public broadcasting system or by an external third-party, concerns about trust and control are relevant. Each repository model will require stations to collaborate and commit to shared policies and procedures to some degree. In addition, if a third-party is involved, public television will need to be confident in the sustainability of the service provider to ensure that its content is safe and retrievable over time.

---

75 http://www.avatar-m.org.uk/

76 Matthew Addis, et al, “Sustainable Archiving and Storage Management of Audiovisual Digital Assets,” 27.

77 Ibid.

The notion of "trust" can be defined as "the willingness of a party to be vulnerable to the actions of another party, based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party."[78] For public television, it may be a tricky issue. Some repository models will require a level of trust that does not currently exist within the system, or between the system and external third-parties. It is unlikely, for example, that many stations would willingly give custody of their content to another existing entity within the public television system to perform preservation. If public television chooses to build a centralized repository, individual stations would likely only trust a new entity created outside of the auspices of any existing one.

When a climate of trust is required, one way of building it is through mechanisms that produce its correlative element, control. Control, in this context, "refers to the processes that are used to monitor and enforce activities, e.g. through things like governance structures, contracts, or adherence to standards"[79] In the digital preservation community, the TRAC criteria and checklist is an example of a control mechanism. Such mechanisms may be useful if public television chooses to employ a repository model that requires building some degree of trust among stakeholders.

Funding Preservation

Once the public television system has decided on which repository set-up best suits its needs and existing infrastructure, and has developed an understanding of the economic relationships that exist between the stakeholders involved in that arrangement, it can then begin to articulate the specific business models that can be used to support its preservation activities.

The sheer volume of content that a public television repository could offer will automatically position it to become an invaluable resource to a number of user communities. The key to sustainability will be ensuring that the repository remains indispensable to those groups. After closely studying 12 digital projects for it's 2009 study "Sustaining Digital Resources: An On-the-Ground View of Projects Today," (2009) the Ithaka S+R authors concluded that, "A sustainable project covers its operating costs through a combination of revenue sources and cost-management strategies and continues to enhance its value based on the needs of the user community." (Maron, et al, 2009)

Regardless of the organizational model of the repository, whether it is a distributed network or a single entity, public television has an incredible opportunity to provide customers with unique access to potentially the largest source of high quality educational, cultural, and historical programming in the United States. The potential added benefits to the entire public television system of a unified search portal, interfaced with similar repositories around the world, could provide revenue both for rights holders as well as provide reliable ongoing funding for preservation.

What then, are the strategies that public television can use to provide sustained financial support for digital preservation? There are myriad business models currently used to support public television production and distribution. Likewise, numerous models are in use by digital repositories around the world. These business models respect the economic model that governs the context in which the repository operates. Successful revenue models for digital preservation are tailored specifically to the repository's context; this includes organizational setting of the repository, the types and needs of users, and copyright status of content, amongst other factors. Carefully strategized business models allow digital resources to flourish, and become highly valuable to their user communities. Poorly conceived models, or those that aren't continuously re-evaluated, are often doomed to irrelevance.

---

78 Michael Day, "Toward Distributed Infrastructures for Digital Preservation: The Roles of Collaboration and Trust," The International Journal of Digital Curation 1, no. 3 (July 2008), http://www.ijdc.net/index.php/ijdc/article/view/60/39 (accessed January 6, 2010).

79 Ibid.

Like the operation of a public television station, a digital preservation repository will not survive by relying on one revenue source alone. In the economic climate of 2010, those working in the non-profit sector have experienced all too painfully what happens when stock prices tumble and endowments evaporate. Sustainable digital preservation for public television content will depend on the ongoing development of creative revenue strategies that combine multiple business models. Both traditional and new sources of funding should be examined. Options such as adding a line item for preservation in production budgets, as well creating new revenue streams through repository services, will have to be considered.

The following sections of the report explore existing business models in public television and describe others that could also potentially be used to support digital preservation. Building effective business models for preserving digital public television content will involve consideration of the various stakeholders involved and their relative roles. It also requires thinking about other factors such as the public television mission and available resources.

Some of the main challenges of devising mechanisms to support preservation within public broadcasting will be with building relationships between public television entities, ensuring fairness to all stakeholders, and making sure the models are consistent with the public television mission. With its experience in developing diverse and creative business models to support its programming and station operations, public television is well equipped to meet these challenges. It can draw from and adapt its existing business models to provide ongoing support to preservation. It can also learn from the experiences of other organizations.

Current Income Sources of Public Television

Public television relies on a wide range of sources for revenue. This diversity means that public television is not overly dependent on any single source for its survival, which has contributed to the sustainability of the system over time.

However, is quite possible that public television's traditional income streams may not be able to sustain digital preservation on their own. If public television chooses to include preservation within its production budgets, it may need to raise additional funds from its traditional sources of funding to cover the increased costs. The best solution will involve a combination of additional funds from traditional sources and new sources of income that generate revenue specifically for preservation.

Currently, institutional income streams originate from a wide variety of sources.

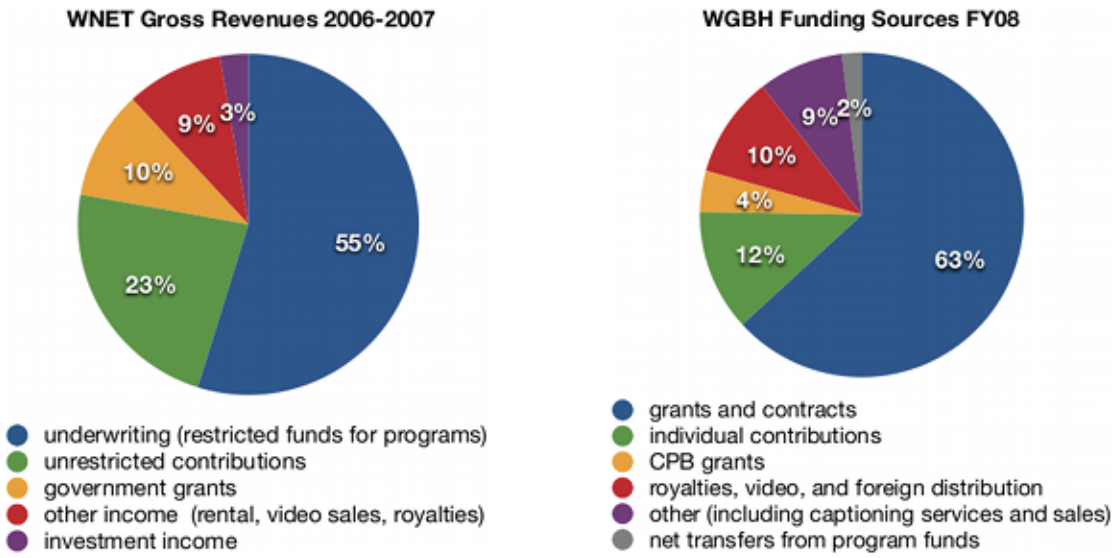

Figure 3. Reported gross revenues and funding sources for WNET and WGBH, respectively.
Sources: Building the House of Ideas: Educational Broadcasting Corporation Report to the Community 2006-2007[80] *and Challenging Times, Critical Impact: WGBH Annual Report 2008-2009.*[81]

Memberships and Large Gifts (Support from Individuals)

Public television relies heavily on memberships and large donations from individuals to generate unrestricted revenue. Viewers become members of local stations by contributing money. Stations provide various avenues for donation, including on-air pledge drives, online forms, and regular automatic deductions. To encourage donations at all levels, stations offer many options for giving. At WNET and WGBH, for example, donors may make one-time contributions or monthly gifts; they may also make larger donations in the form of bequests, endowments or directed gifts.

In 2006-7, 23% of gross revenue at WNET came from unrestricted contributions, part of which is made up of contributions from individuals and businesses. In 2008, approximately 50% of funding for WGBH's local TV and radio services came directly from audience support (not shown in diagram above).[82]

Grants from Public Sources

Public television also relies on a variety of grants from government sources. Local stations and PBS regularly receive grants and contracts from federal, state and local government agencies and from the Corporation for Public Broadcasting (CPB), which receives an allocation from Congress. These grants accounted for about 10% of WNET's revenue in 2006-7 and 12% of WGBH's funding in 2008. About 9% of PBS' total revenue in 2008 came from grants and contributions.[83] Grants often require a matching contribution that must come from the grantee.

---

80 Educational Broadcasting Corporation, "Building the House of Ideas: Educational Broadcasting Corporation Report to the Community 2006-2007," http://www.thirteen.org/home/pdfs/EBCAnnualReport_BHI.pdf (accessed January 6, 2010).

81 WGBH, Challenging Times, "Critical Impact: WGBH Annual Report 2008-2009" (Boston: WGBH, 2009), http://support.wgbh.org/site/PageNavigator/annualreport (accessed January 6, 2010).

82 Ibid.

83 BDO Seldman, LLP, "Public Broadcasting Services and Subsidiaries, Consolidated Financial Statements and Independent Auditors' Report, Years Ended June 30, 2008 and 2007" (Bethesda: BDO Seldman, 2008), http://www.pbs.org/aboutpbs/content/annualreport/2008/2008FullFinancialHighlights.pdf (accessed January 6, 2010).

Grants from Private Foundations

Major funding for programming and other projects comes from foundations and other private sources. Funding from private foundations is frequently for larger amounts and does not require a matching contribution. Grants from foundations accounted for approximately 5% of WGBH's revenue in 2008.

Underwriting for Programs and Station Operations

Underwriting for programs is another primary source of revenue for public television producers. Underwriters are typically commercial businesses that provide support for program production or acquisition. Public broadcasters are legally prohibited from airing advertisements, but they can accept business donations and, if they do, they must air announcements to acknowledge the donation based on specific guidelines established by the Federal Communications Commission. More than half of WNET's gross revenues in 2006-7 took the form of restricted funds to underwrite national television programs and educational projects.

Public television also receives underwriting support for station operations from local foundations and corporations. This support funds general broadcasting activities, as opposed to funding specific programs. Underwriters for station operations receive on-air recognition and local recognition, such as at an event.

Royalties, Video Sales, Foreign Distribution

Producers and other rights holders can receive revenue from video sales, licensing and royalties. WGBH, for example, sells programming via home video and to schools and libraries. It also sells stock footage from WGBH collections to external commercial producers. In addition, WGBH operates WGBH International, which participates in co-productions and distribution, as well as licensing for WBGH, PBS, and independently produced programs. In 2008, royalties, video, and foreign distribution accounted for 10% of the station's funding.

Services for Hire

WGBH offers services-for-hire in a range of television-related areas of expertise, including captioning and video description, production services, and scenics. The WGBH Archives also offers rights clearance and internal research services. Service revenue accounted for approximately 4% of the station's funding in 2008.

Endowment Income

Some stations and other institutions in public broadcasting have been successful raising large enough sums to create an endowment. The interest generated by the endowment is used as unrestricted income for the institution.

Benefit Events

Virtually all public broadcasting organizations produce benefit events as fundraisers, as well as providing an opportunity to meet supporters in person. The events can range from modest gatherings to high price gala black-tie affairs. Approximately 5% of WGBH's funding came from these types of "Other" sources in 2008, which includes special events such as the WGBH Auction.

Current Funding that Supports Preservation

Examining the way that preservation activities are now funded in public television may provide an indication of the current state of affairs, and how public television can best proceed in its planning for a future initiative. As this is being written, very little investment has been made in preservation system-wide, and overall, it is fair to say that current funding for preservation in public television is grossly inadequate to sustain a long-term digital preservation program.

Most preservation has been done locally, by individual stations or producers, solely to manage their own materials. Two examples of stations with funded archives are PDPTV project partners WGBH and WNET. The WGBH Media Library and Archives is supported primarily through discretionary funds from the station, but also operates with diverse funding streams. Falling within Administrative and Institutional Resources, the Library and Archives receives part of the money set aside for overhead costs. Its role is considered to be protecting WGBH's institutional assets.

The WGBH Media Library and Archives writes grants to federal sources and foundations to fund specific projects, such as putting its Vietnam collection archives online, and also sells footage to commercial producers (while providing WGBH producers with material free of charge). WGBH accepts donations via its online Open Vault portal to support preservation and access, although it has not taken full advantage of this avenue to date by promoting it. Researchers who wish to use legacy materials on obsolete formats must pay the cost of transfer to a current format and the production of an access copy.

The Thirteen/WNET Tape Archives is also maintained by the station. For preservation activities, the Tape Archives relies primarily on user fees. When researchers request materials, they are responsible for paying the cost to transfer the content, creating a preservation copy for the Archives, and creating an access copy for their own use. The Tape Archives has been successful in obtaining a small number of grants for special projects, but is not supported with on-line donations or similar income.

Ongoing support for preservation activities, then, has generally been funded through station operating costs because the items to be preserved are considered institutional assets. Only a few stations such as WGBH and WNET actually do this, however, and even so, their limited resources do not allow them to fully exploit those preserved assets. In most cases, operating costs cover only storage and some management, technical and organizational activities. Generally, in order to process or make accessible special collections, the Archives must raise dedicated funding for the initiative. At the station level, this is often seen as competing with operating or production funds, which leaves the Archives with fewer opportunities to raise funds or increase support on its own.

While the existing archives in the public television system have done an admirable job in safeguarding a vast amount of content with only limited resources, the current methods of funding preservation will not suffice in an all-digital environment. Covering storage costs with institutional operating funds may be enough to maintain videotapes in the medium-term, but simple storage cannot sustain digital content. This avenue, which is only a stop-gap measure for analog materials, cannot begin to meet ongoing management needs of digital objects, which, as outlined here, are many. New revenue streams, and new approaches, are needed if fast-growing collections of digital materials are to be preserved.

Exploring New Business Models to Support Preservation

Outside of traditional sources of funding, it is possible for public television to develop business models to generate new income to sustain digital preservation. Existing revenue models for public television are quite varied, and what is in place can offer direction to a new repository. Ithaka S+R concluded that a significant factor in the success of a digital project was the development of diverse revenue sources. Sustainable projects, they found, "experiment with different revenue models to find the ones that are the best fit for the project; show willingness to try new models; cultivate the ability to identify and communicate the value of the resource to the target audience (of customers, authors, subscribers and so forth)."[84]

This section describes a number of business models that are being used by digital repositories worldwide. These illustrate some of the different possible models that public television might adopt to

[84] Nancy L. Maron, K. Kirby Smith, Matthew Loy, "Sustaining Digital Resources: An On-the-Ground View of Projects Today," (New York: Ithaka S+R, 2009), 21. http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability/report/SCA_Ithaka_SustainingDigitalResources_Report.pdf (accessed January 6, 2010), 6.

sustain its own digital preservation activities. Many of the examples involve higher-learning educational institutions, although other public broadcasting, non-profit and commercial organizations are also represented. While each revenue model is discussed in turn in this report, clearly two or more of these strategies can be used in combination in order to support ongoing digital preservation activities.

The following business models are reviewed here:

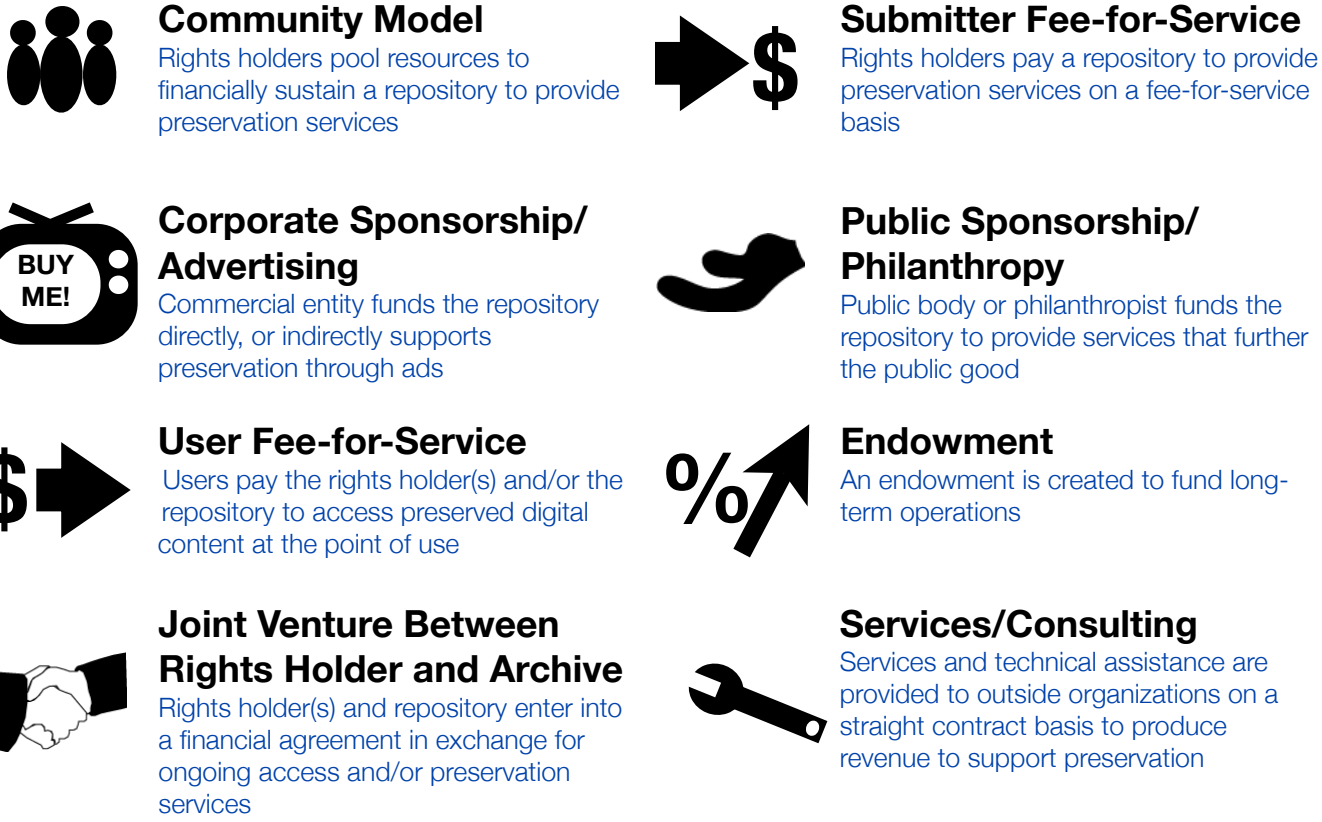

Figure 4. The revenue models discussed in this report.

Community Model

In the community model, rights holders pool their resources to financially sustain a repository to provide preservation services from which they benefit directly. All members of the community who benefit from preservation share in supporting and making decisions about the functioning of the repository.

An example of the community model in the digital preservation domain is *HathiTrust,*[85] a non-profit shared repository for academic research libraries in the US. Originally conceived as a collaboration between the 13 universities in the Committee on Institutional Cooperation and the University of California, the partnership is open to all interested research libraries for a one-time start-up fee (based on volume added to the repository) and an annual fee for ongoing curation. Its primary designated community is comprised of the members of the institutions represented by the participating libraries.

While the primary function of HathiTrust is to provide long-term preservation rather than access, all public domain materials are accessible to researchers, and all content is discoverable through Google

85 HathiTrust, "HathiTrust: A Shared Digital Repository," HathiTrust, http://www.hathitrust.org (accessed January 6, 2010).

and other online search technologies. HathiTrust is governed by an Executive Committee made up of library deans and CIO's from the founding institutions. Its budget is maintained as a separate budget within the University of Michigan budget system.[86]

Submitter Fee-for-Service

In this business model, rights holders pay a repository to provide preservation on a fee-per-service basis. Unlike the community model where costs are shared, this model allows individual rights holders with heterogeneous needs to pay just for the services they require. To be effective, the repository needs to institute a correct pricing scheme that adequately recovers the recurring and future costs of maintaining the resource.

There are many examples of the submitter-fee-for-service model in the digital preservation domain. The *Archaeology Data Service (ADS*)[87] is an example of a repository that charges a one-time fee for service. ADS collects, catalogs, manages, preserves, and encourages re-use of digital resources created by and of interest to archaeologists in the UK. The ADS charging policy stipulates that costs shall be recovered from the body funding the archaeological research through a one-time payment collected at the time of deposit. Submitters therefore pay a fee that covers the total future costs of preservation. This model allows ADS to adhere to its central tenet that resources be free on the user end, in keeping with professional ethics that requires primary data be openly available. It should be noted that the success of the ADS model is also tied to a mechanism that provides an additional incentive for rights holding submitters: deposit to the ADS is a requirement for funding from the Arts and Humanities Research Council.

The *Harvard University Library Digital Repository Service (DRS*)[88] is an example of a repository that charges submitters an ongoing fee for service. DRS is a preservation and access repository that provides managed services to the Harvard community to ensure the viability and usability of its stored content over time. While DRS' infrastructure is supported by the university, it is partially financed on a cost recovery basis by fees from depositors. There is no initial charge to deposit objects, but object owners or collection managers must pay for ongoing storage and transformation. DRS also offers optional fee-based services such as custom databases, virtual collections catalogs, and custom programming. All objects stored in the DRS are considered to be Harvard University resources, and a version of all deposited content is accessible to the Harvard community.

Public Sponsorship / Philanthropists & Foundations

In this model, a public body or philanthropist funds the repository to perform digital preservation to further the public good. As with the other models, this model has strengths and weaknesses. As Albert W. Darimont explains in *Environmental Scan of Pricing Models for Online Content: Report II Business Models for Object Repositories*,[89] relying on an external public body or philanthropist for support may not always allow for the true cost of the service to be recovered. There is also the constant threat that funding may be reduced due to competing interests beyond the repository's control. On the other hand, public sponsorship or philanthropy generally allows it time to develop its services. The model also allows the repository to preserve a greater breadth of works since it is less economically obliged to meet

---

86 HathiTrust is currently undergoing a TRAC audit by the Center for Research Libraries. More information can be found at http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-portico-and-hathitrust (accessed January 6, 2010)

87 Archaeology Data Service, "About the ADS," ADS, http://ads.ahds.ac.uk/project/about.html (accessed January 6, 2010).

88 Harvard University Library Office for Information Systems, "DRS Policy Guide," President and Fellows of Harvard College, http://hul.harvard.edu/ois/systems/drs/policyGuide/wwhelp/wwhimpl/js/html/wwhelp.htm (accessed January 10, 2010).

89 Albert W. Darimont, “Environmental Scan of Pricing Models for Online Content: Report II on Business Models for Object Repositories” (n.p.: OnDisC, 2002).

market demands.

Quite a few examples of public sponsorship and philanthropy are found in the digital preservation domain. One is *The Depot* in the UK.[90] The Depot is a short-to-medium term repository for UK academics whose institutions do not yet have a repository to ensure that important research is safeguarded and openly accessible in the meantime. The Depot is the prototype service of EDINA / SHERPA, and was designated as part of a £15 million commitment towards Digital Repositories and Preservation activities by the Joint Information Systems Committee (JISC), a body that supports UK government objectives.

Public television is very familiar with the public sponsorship and philanthropy model, which it uses to fund programming and other initiatives that serve the public good. Staff and infrastructure, such as the PBS Foundation and departments within WNET and WGBH, are already in place to solicit and maintain this kind of support.

User Fee-for-Service

In this model, users pay the rights holder and/or the repository to access preserved digital content at the point-of-use. As Guthrie, Griffiths, and Maron indicate in their 2008 report "Sustainability and Revenue Models for Online Academic Resources," a pay-per-use model can effectively accommodate a large casual user base with occasional or unpredictable needs, and users that require specific content in a time sensitive manner.[91] This kind of model requires good search indexing and a correct pricing scheme.

This model works very well for *L'Institut national de l'audiovisual (Ina)* which uses such fees to provide professional users with high quality content. The business model of Ina offers an example of a repository that uses diverse revenue streams, but where most of the revenue is generated from rights licensing through user fees. Their public website, ina.fr, covers its own operating and staff costs through a combination sources, including advertising revenue, revenue from downloads and video clip rentals, and a DVD on Demand program.[92] Though this site does not produce additional revenues to support preservation, it meets Ina's mission to provide content to the public, while at the same time generating awareness of the collection and reaching wide audiences at its own site and at partner sites, including many popular French newspapers. This awareness in turn helps steer professional customers to Inamédiapro, where they can search for, discover, order, and download high resolution audio, video, and metadata, using state of the art web tools. The rights licensing side of this site generated €14.5 million in 2008. After paying its personnel and non-personnel costs (including € 3.5 million in royalty payments to rights holders), this revenue stream resulted in almost € 6.5 million to support the organization, including its preservation activities.[93]

Public television has experience with using the user fee-for-service model to fund preservation. At both WNET and WGBH, researchers who want to access legacy materials must cover certain costs. Those include the cost of researching and clearing rights, locating copies, making transfers, the cost of their research copy, and the cost of providing the Archive with a preservation copy. Public television also has extensive experience with videotape sales, which operates similarly to a user fee-for-service model.

---

90 Joint Information Systems Committee, "The Depot," HEFCE, on behalf of JISC, http://www.jisc.ac.uk/whatwedo/programmes/reppres/depot.aspx (accessed on January 6, 2010).

91 Kevin Guthrie, Rebecca Griffiths, and Nancy Maron, "Sustainability and Revenue Models for Online Academic Resources: An Ithaka Report (n.p., Strategic Content Alliance / Ithaka 2008), 31-32. http://www.ithaka.org/ithaka-s-r/strategy/sca_ithaka_sustainability_report-final.pdf (accessed January 6, 2010).

92 Nancy L. Maron, K. Kirby Smith and Matthew Loy, "L'Institut national de l'audiovisuel: Free Content and Rights Licensing as Complementary Strategies," (New York: Ithaka S+R, 2009), 3.

93 Ibid., 8.

Membership / Subscription

In this model, users and/or rights holders pay dues to the repository in exchange for ongoing access and/or preservation services. Guthrie, Griffiths, and Maron indicate that this model requires core beneficiaries of significant size who can afford and are willing to pay. This implies that the content must either be unique, authentic, or otherwise valuable, or that the repository services add value that is difficult to replicate (for instance, if the repository can deliver content in a timely manner). In employing this model, the repository must assume some financial risk, as it must make some upfront investment in infrastructure and operations before subscribers can be sought out. Although access can be limited to members according to their subscription level, this model in effect limits usage and the potential impact of resources compared to models that encourage open access.[94]

An example of the subscription model in the digital preservation domain is *Portico.* Portico is a non-profit long-term repository service whose mission is to preserve scholarly literature published in electronic form and to ensure that these materials remain accessible to future scholars, researchers, and students. Its main beneficiaries are publishers and academic institutions, which constitute Portico's main source of funding. Member publishers pay annual contributions that are tiered according to journal revenue, and provide electronic source files in a timely manner to the archive. In exchange, Portico performs conversion to archival file formats, ongoing migration, and other archiving activities. Meanwhile, member libraries also pay an annual contribution. In return, they receive the assurance that content will remain accessible in the future in the event that a publisher stops operation or ceases publishing a title.[95]

Public television already utilizes several versions of this model for generating revenue, as individuals and corporations are willing to pay to ensure continued access to public television's high-quality content. Local stations rely on memberships for a significant portion of their income, and infrastructure is in place to support different types and tiers of membership on the local and the national level. In keeping with its educational and access-oriented mission, public television does not limit benefits to paying members; non-members can still freely watch public television shows. Instead, stations offer nominal, but exclusive, thank-you gifts to members. In this way, public television does not limit the potential usage and impact of its programming.

This, too, is a common membership model of support already existing within the public broadcasting system that could be applied to preservation services. For example, PBS is a membership organization that provides programming and distribution services to public television stations, with several tiers of membership dues and services.

Joint Venture Between Rights Holder and Archive

In this model, the two entities enter into an agreement that is mutually beneficial, allowing the digital resources to be preserved.

The *CLOCKSS (Controlled Lots of Copies Keeps Stuff Safe)* initiative is an example of a joint venture between Rights Holders and Archive/Beneficiaries for digital preservation. CLOCKSS is a partnership between scholarly publishers and research libraries whose mission is "to build a sustainable, geographically distributed dark archive with which to ensure the long-term survival of Web-based scholarly publications for the benefit of the greater global research community."[96] Participating publishers give CLOCKSS librarians permission to collect, preserve and provide access to their content using LOCKSS (Lots of Copies Keeps Stuff Safe) software. Like Portico, CLOCKSS libraries only make content accessible in the event that the publisher stops operation or ceases to publish a title. CLOCKSS is currently sustained by fees from both participating publishers and libraries, but it has a five-year goal

---

94 Guthrie et al., “Sustainability and Revenue Models,” 27-31.

95 Portico, "Portico," Ithaka Harbors, Inc., http://www.portico.org/index.html (accessed January 6, 2010).

96 CLOCKSS, "CLOCKSS: A Trusted Community-Owned Archive," CLOCKSS, http://www.clockss.org/clockss/Home (accessed January 6, 2010).

of establishing an endowment from publisher and library contributions. (See below for *Endowments*.)

Services / Consulting

Digital repositories have found that their expertise, resources, and technologies enable them to generate supplemental revenue by offering services or consultations to similar institutions. These services might include digitization, cataloging, customizing web tools, and advising on preservation projects. Ina, for example, provides digitization services and licenses Ina software. These services must be carefully priced so that they are revenue-generating, and not a drain on an organization's resources.

Parallels exist in public television in WGBH's services-for-hire, in which independent producers can pay a fee for particular expert services by station staff, or WNET, renting out its editing suites to outside producers when they are not being used in-house.

Corporate Sponsorship / Advertising

In this model, a third-party sponsor funds or provides in-kind resources to the repository to provide preservation services from which the Rights Holder benefits. In exchange, the Rights Holder trades on its intangible assets (e.g. its reputation) by allowing the corporate sponsor to be associated with it.

An example of the Corporate Sponsorship Model in the digital preservation domain is *DSpace@MIT.*[97] Dspace@MIT was developed as a joint research project between MIT Libraries and Hewlett-Packard (HP). MIT Libraries worked with development partners at HP to plan the repository, as well as a hardware expert at HP and MIT's Information Systems to scale the DSpace system. The hardware for the system was donated by HP. MIT Libraries sustains the repository through corporate sponsorship, combined with institutional support, expertise from institutional collaborators in the DSpace Federation, and some fees for extraordinary services. MIT Libraries' goal was to make the system freely accessible to both submitters and users.

Corporate sponsorship can also come in the form of online advertising on public access portals. This model has been used by Ina in France to support its public website.[98]

The corporate sponsorship model for digital preservation is somewhat similar to the underwriting model that public television uses to support programming or station operations. In public television, corporations and businesses fund programs or operations in exchange for on-air recognition and other acknowledgment from the broadcasting station.

Endowments

While the endowment model's "permanent" support and free access to users is appealing, establishing a sufficient endowment is quite challenging. As Guthrie, Griffiths, and Maron indicate, an organization must accumulate enough capital to support the repository's activities through its income from investments and interest. A host institution must therefore be able to raise about 20 times the amount of the repository's annual operating budget in order to sustain the service. This requires the institution to build infrastructure and employ staff in development, donor relations, and investment management.[99]

The task of building an endowment is enormous and can take many years, even in a good economic climate. But, as Maron, et al. found recently, "The dire economic environment of 2009 has highlighted the risk associated with reliance on investment income. Project leaders using this model are watching

---

97 Mary R. Barton and Julie Harford Walker, "Building a Business Plan for DSpace, MIT Libraries' Digital Institutional Repository" (n.p.: Texas A&M University, 2003), http://dspace.mit.edu/bitstream/handle/1721.1/26700/Barton_2003_Building.pdf?sequence=1 (accessed on January 6, 2010).

98 http://www.ina.fr/

99 Guthrie et al., "Sustainability and Revenue Models," 46-47.

the values of their endowments drop, suggesting the importance of diversification of revenue streams."[100]

Public television is quite familiar with this model of revenue generation. Many local stations such as WNET have long-established endowment efforts that solicit gifts from members and others. WGBH successfully built an endowment to offset additional operating costs of its new facility in 2007 as part of its capital campaign. PBS has also established a separate supporting foundation that focuses explicitly on soliciting large gifts to build an endowment.

Developing an Appropriate Mix of Fundraising Strategies

As this report makes evident, an important part of sustaining a long-term repository for digital public television content is ongoing financial support. This support can come from existing sources that traditionally fund programming and station operations, or from new avenues developed specifically to generate revenue for digital preservation. Public television will certainly have to rely on a mix of strategies in combination to sustain preservation, just as it has done in the past to cover its other essential programming and operation costs.

Most critical though, public television should ensure that the strategies it selects match the economic situation -- the stakeholders involved, their economic constraints, the relationships between them, and the incentives and disincentives they face -- by assessing the options against a relevant economic model. Public television will also need to evaluate how well potential strategies align with its mission and values, existing staff capabilities and strengths, investment resources, an acceptable level of risk, and whether they provide other ancillary benefits.[101]

Many factors will need to be weighed when selecting the best strategies to support a repository for public television, and in addition to the familiar sources listed above, there are many as-yet unexplored possibilities for generating income. For example, if public television chooses to begin incorporating preservation costs within production budgets, it will need to demonstrate to its traditional funders that preservation is not an add-on activity, but rather integral to the future of television production. Some preservation activities will entail additional costs, but others may simply represent changes in the way some steps in production are performed.

Public television can also show traditional funders that investing in preservation can result in long-term cost savings and the creation of new opportunities for older material. Building sound life cycle management in the digital production workflow will not only facilitate sustainable preservation but also make production more efficient and organized. The technical and organizational aspects of sustainable preservation outlined earlier in this report provide support for these arguments.

Meanwhile, if public television chooses to find new ways of supporting preservation instead of, or in addition to, covering costs through existing budgets, it will need to develop new business models. As this report indicates, many creative business models are currently in use by organizations to support their digital preservation activities. Public television can draw from these, as well as from its own vast experience with diverse business models, to devise appropriate approaches to funding the ongoing costs necessary to sustain the digital preservation of its content.

* * * * *

100 Maron, et al., "Sustaining Digital Resources," 25.

101 Guthrie et al., "Sustainability and Revenue Models," 24.

Chapter 5

Case Study: PRESERVING DIGITAL PUBLIC TELEVISION
Costing the Prototype Repository

As discussed in **Chapter 4: Economics of Sustainable Preservation**, projecting "ballpark" costs for a repository with unspecified parameters is meaningless because of the many interrelated variables that influence costs. It is possible, however, to identify and describe the basic expense categories for operating an actual digital preservation repository. Once specific operating parameters are applied, a useful framework can be constructed that may provide guidance for a more general costing model.

What follows is an analysis of the costs associated with operating the prototype digital repository built by NYU Digital Library Technology Services for the Preserving Digital Public Television project. The costs are summarized below, and outlined in greater detail in the three spreadsheets that are attached as appendices.

The purpose of this case study is to demonstrate the range of skill sets that were required, the balance of staff time, and the other expenses of the project within the organizational framework of New York University. Although most costs will be highly specific to the organizational context and cannot be generalized, we hope this case study can be used as a reference by a future repository as it calculates its personnel and operational costs.

Building a "Model Repository"

The initial goal of the PDPTV NDIIPP grant from the Library of Congress was to develop a small model preservation repository that could provide guidance for the wider public television system. The PDPTV repository was therefore developed to comply with the ISO standard OAIS Reference Model for the design of a trustworthy digital repository. Here, it provides a vocabulary for discussing digital preservation repository functional activities, and is the source for many of the terms used in this case study.

The PDPTV prototype repository has only ever existed in a start-up phase and not as an ongoing, operational facility. Thus, the costs associated with longer term preservation activities, such as forward migration of content to new file formats and refreshing of storage media, could not be calculated for this case study. Furthermore, content in the prototype repository is only accessible to the submitting entities (i.e. the public television project partners). Since the repository does not support search and retrieval by a wider user community, no access interface was created, and no cost for this was calculated. In a future repository, development and/or technologies would be needed in these areas, and would contribute significantly to overall costs.

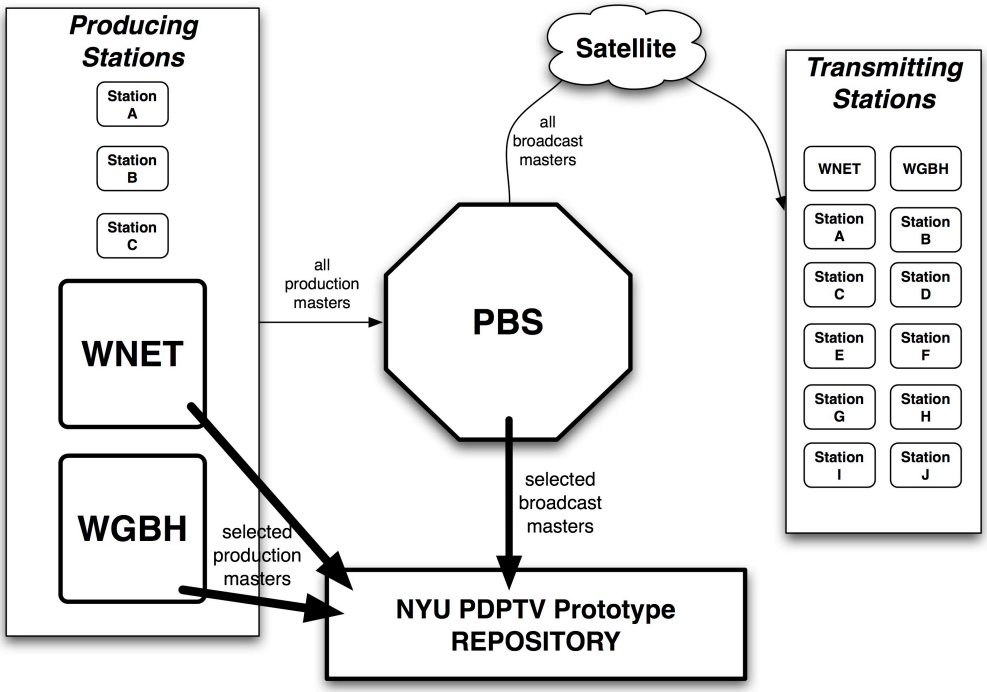

Figure 5. Sources of program files for the Model Repository

Based on the parameters of this specific repository, which are described in more detail below, the PDPTV project has attempted to identify and describe the relevant costs associated with its construction and operations. For this reason, the costs stated in this model can only be said to reflect the specific implementation of the repository at NYU, and not necessarily the costs that any future public television repository may incur operating under different circumstances.

The PDPTV Repository Implementation at NYU

Most preservation repositories, including the prototype repository at NYU, are implemented in accordance with the ISO Reference Model for an Open Archival Information System. OAIS defines the various functions of a repository within the following entities:

- **Common Services**: not a distinct functional entity, but rather describes services that support distributed computing applications overall.
- **Ingest:** the entity that receives the submission, performs quality assurance, and generates the

archival package to be transferred to storage.

- **Archival Storage:** the entity that transfers the archival package to storage, checks for errors, manages storage, and performs back-ups.
- **Data Management:** the entity that administers the database that contains information about the repository's holdings and system information, generates reports, and performs access queries.
- **Administration:** the entity that negotiates submission agreements, monitors and audits the functionality of the repository system, establishes standards and policies, and provides physical access control (i.e. locks, security guards, etc) for the repository.
- **Preservation Planning:** the entity that develops preservation strategies and standards, monitors the designated community, monitors technology, and develops package designs and migration plans.
- **Access:** the entity that coordinates access activities, generates the dissemination package, and delivers the response to the consumer.

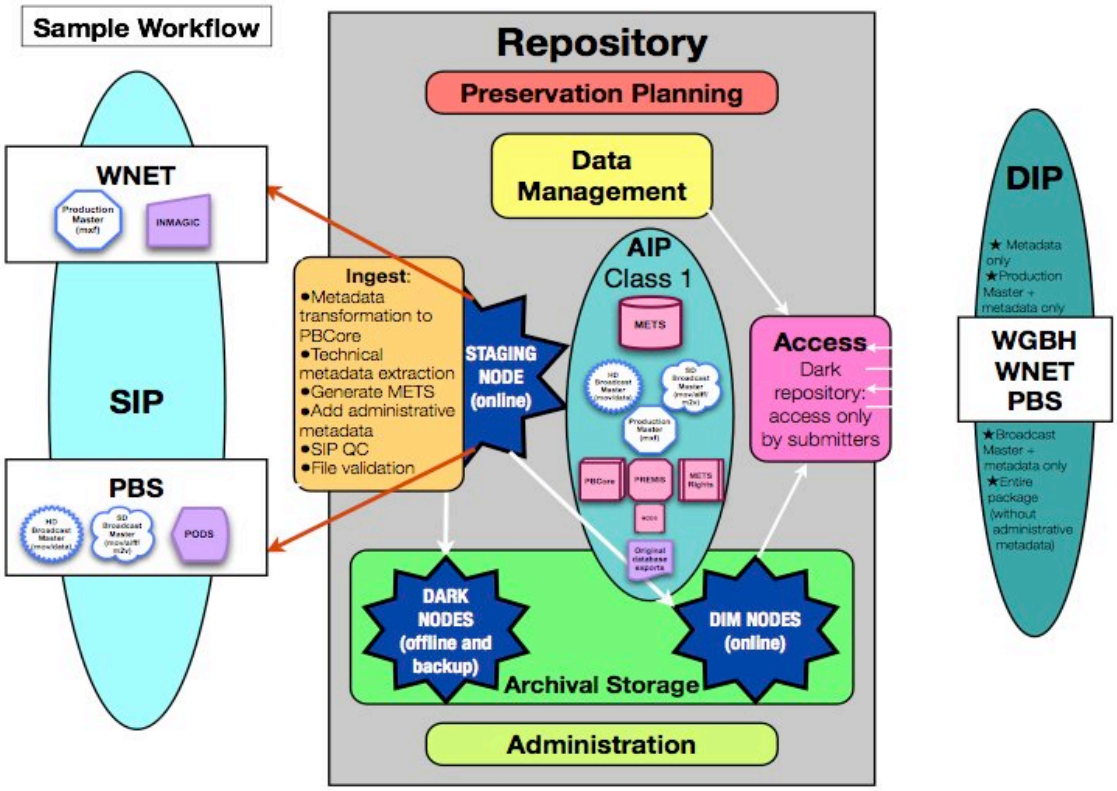

Figure 6. The OAIS Functional Entities in the PDPTV repository

The prototype repository for the PDPTV project is managed and maintained by New York University's Digital Library Technology Services, a joint effort by NYU Libraries and Information Technology Services (ITS). ITS is comprised of the Office of the Chief Information Technology Officer and five service departments, and provides computer, network, and Internet services for NYU, one of the largest private universities in the United States. The cost of leases and licenses for the prototype repository are significantly influenced by NYU's large size, which gives it bargaining power and the ability to spread costs across the institution to advantage of economies of scale.

Given the way that the PDPTV repository is set up within NYU, it was often impossible to isolate that fraction of total infrastructure resources consumed by the prototype repository apart from other initiatives and services at the University. Many of the costs reported here, therefore, are only estimates based on the relative size of the PDPTV project within the entire NYU Digital Library structure.

For this specific project, staff salaries were the most significant cost, followed by hardware, backup and recovery service fees, network services, software and electricity/utilities. Some of these, especially those related to repository design and software development, are non-recurring one-time costs. Others are ongoing expenses necessary to maintain repository operations once it has been established.

Repository Architecture

Rather than build separate applications for each of its various projects, the NYU Digital Library has built a single DSpace-based Preservation Repository to preserve and provide access to content for a number of its projects, including PDPTV. DSpace is an open source digital repository platform developed by MIT Libraries and Hewlett-Packard to manage digital assets.[102] It is widely adopted among research and higher education institutions, cultural heritage organizations, and state libraries and archives.

The Preservation Repository (PR) is composed of a set of loosely coupled components communicating over stable interfaces. This approach allows components to be optimized, scaled, and even re-implemented with minimal impact on other PR components. The PR is designed to be "content-neutral" so that vastly different digital objects can be stored and preserved by the system. OAIS functional entities are implemented using project-independent code when possible, and project-specific code when required. Within the Ingest functional entity, for example, PDPTV-specific scripts are used to generate the Archival Information Package (AIP) whereas quality assurance and file validation are handled by project-independent code. Meanwhile, the Archival Storage and Data Management functional entities are implemented using DSpace software, backed by a Postgres open-source relational database instance. Dissemination is handled by project-specific code and the Common Services are provided by project-independent code.[103]

Content Stored in the Prototype Repository

To design a repository specifically around the needs of public television, the PDPTV project collected a variety of sample content. In the prototype repository, the technical issues and repository functions explored were primarily focused on solving the problems of how best to manage and organize test files and their associated metadata to create Submission Information Packages (SIPs) and Archival Information Packages (AIPs) as described by the OAIS Reference Model.

More than 35 hours of completed programming, totaling more than 78 hours of program files (including production master and broadcast master files) were ingested into the prototype repository, along with 45 minutes of source footage. The test files originated from three sources – program masters were submitted by producing stations WNET and WGBH, and compressed distribution versions of the same programs were submitted by PBS. These included standard definition video files drawn from national series *Nature, Frontline* and *Religion and Ethics Newsweekly*, *Antiques Roadshow*, and *American Experience*, plus a selection from the local program *New York Voices*. High-definition broadcast files were collected for *Nature*, as well as HD production masters and source content from the new WNET nightly news show *WorldFocus*. This provided a mix of both high- and low-resolution program file formats from several sources, accompanied by a wide range of metadata.

---

102 More information about DSpace is available at http://www.dspace.org.

103 The design of the repository is discussed in detail in a forthcoming PDPTV project report, “Repository Design Report I with Attached Metadata Plan.” which will be available by June 2010 at http://www.thirteen.org/ptvdigitalarchive/

The costs associated with running a repository are greatly influenced by the volume of material that the repository must ingest, store, and manage. It affects, for example, the amount and type of hardware required, electricity and utility usage, and computer processing and staff time, among other costs. The types of content held in the repository also influences costs, such as staff costs for software development and preservation planning. The costs identified in this case study therefore reflect the fact that the PDPTV prototype repository only holds a minimal amount of content, but in a variety of forms that reflects the types of materials that might be submitted, stored, and retrieved by a fully functioning repository for digital public television content.

The degree of standardization in file formats, metadata, and processes also greatly influences cost. The Initiation and Ingest phases of the PDPTV repository were costly and time-consuming because of the array of formats and metadata that the repository received from submitting stations. If, for example, a standardized metadata schema and dictionary such as PBCore were widely used throughout the public television system, there would be great cost savings for a future repository.

Storage

The Preservation Repository, which holds the PDPTV prototype repository content, uses a subset of NYU's 21 Sun Fire X4500 servers, which in all comprise over 400 terabytes of raw storage. The prototype repository content is stored using a combination of the Storage Resource Broker (SRB) Data Management System, an ORACLE relational database management system (RDBMS) instance, and Sun's Zetabyte File System (ZFS). SRB is a storage software system developed at the University of California - San Diego that supports shared collections that can be distributed across multiple organizations and heterogeneous storage systems (i.e. grid storage). SRB manages and controls access to three of the X4500 ("thumper") storage servers at NYU. The SRB Metadata Catalog (MCAT) is backed by the ORACLE database instance. Additional thumpers running ZFS are used for Archival Storage. Back-up and recovery services are provided by an external vendor.

NYU leases its storage and computing hardware at a cost of approximately $1 million over 4 years. Its university-wide ORACLE site license is approximately $2.5 million over 5 years. The cost for networking and bandwidth is significant, but a figure could not be provided at the time of this report. NYU contracts back-up and recovery of all its records, including email, to an external vendor at a cost of approximately $500,000 a year. The costs for hardware, software, bandwidth, and back-up associated with the PDPTV project cannot be actually separated out from these overall figures, but are instead estimates based on the proportion of resources likely consumed by the Digital Library.

Staffing

Several people are involved in implementing and managing the prototype PDPTV repository. The number of staff and their job descriptions are reflective of the needs of the PDPTV repository, but also of the existing infrastructure at NYU, Digital Library, and ITS. In mapping out the costs of the prototype repository, PDPTV identified the primary staff roles and attempted to accurately quantify the staff-time spent on operating the prototype repository.

The types of staff involved in implementing and managing the PDPTV prototype repository are:

- **Project Manager:** Responsible for overseeing the initiation, planning, implementation, and closing of the project.

- **System Architect:** Develops system architectures to meet functional requirements specifications, ensures that system architectures integrate with existing systems, pinpoints optimization and migration opportunities, performs optimization/migration cost-benefit analyses, generates resource requirement forecasts, monitors applications and services for compliance with Service Level Agreements and develops corrective action plans if necessary.

- **Preservationist:** Monitors and communicates with Designated Communities (i.e. in this case,

public television partners) to determine selection criteria / policy, SIP / AIP / DIP requirements, service policies and agreements, and format, documentation, and procedural standards. Works closely with repository staff, leading the implementation of metadata standards, performing analysis of submitted content, assisting in preservation planning, determining workflow and repository requirements. Supervises research and monitors the community for technological changes. Represents the project at conferences and symposia.

- **Programmer/Analyst:** Creates a functional framework for repository, and tests the functionality of the repository. A programmer/analyst for this project would have a background in computer or information science. He or she would also have experience developing applications using Java; experience with relational database management systems (RDBMS) and digital repositories (i.e. DSpace); experience with digitization projects involving library and digital library metadata and standards (e.g. METS, MODS, Dublin Core, PREMIS); knowledge of Apache, Tomcat, Java, SQL, XML, XSLT, a scripting language (Perl, Ruby, etc); experience with data grid software such as SRB or iRODS; experience working in a Unix/Linux environment; and familiarity with video file formats and packages (MXF, Quicktime), video encoding standards (DVC Pro, MPEG4), and video metadata standards (i.e. PBCore).

- **Systems Administrator:** Oversees smooth operation of hardware and software on computing system; performs installation, testing, maintenance, upgrades and administration of operating system and application software; fine-tunes system configuration for reliability and performance; develops monitoring and testing tools; troubleshoots problems; ensures system remains operational; monitors systems for performance and security; implements system policies to adhere with departmental policies and standards; and researches and recommends configurations for new systems based on vendor and industry trends and contacts.

- **Database Administrator:** Maintains and upgrades database systems to ensure optimal database system performance, security, and reliability.

- **Support Staff:** Provides user support; assists repository staff with day-to-day operations.

As mentioned above, staff salaries represent the most significant cost for the repository. While the PDPTV prototype repository only exists in a "start-up" phase, it is possible to roughly distinguish the staff time spent on developing the repository into two categories:

- time spent on tasks that only need to be performed once to initiate operations;
- time spent on tasks that must be performed repeatedly or on an ongoing basis if the repository is to exist into the future.

The purpose of dividing costs up in this way is to give public television a better sense of the short- and long-term costs of running a repository. Staff demands will be high at start-up, but are reduced significantly once the repository becomes operational. The following figures are estimates and projections based on the experience of members of the PDPTV team:

Repository Staff - Prototype		
Staff Type	**Full-Time Equivalents - One-Time (Start-Up) Only**	**Full-Time Equivalents - Would be Ongoing Staff**
Project Manager	0	0.1
System Architect	0.75	0.4
Preservationist	1.0	1.1
Programmer/Analyst	4.8	0.31
Systems Administrator	0.1	1.2
Database Administrator	0	0.05

Support Staff	0	0.1
TOTAL	**6.65 FTE**	**3.26 FTE**

Table 2. Total estimated staff time to develop the PDPTV prototype repository, divided into costs that are anticipated to be only one-time, and costs that are anticipated to be recurring, if the repository were to continue to exist.

Summary of Key Costs by 'Functional Entity'[104]

The costs associated with the prototype repository were delineated and described according to OAIS functional entities (Common Services, Ingest, Archival Storage, Data Management, Administration, Preservation Planning, and Access), described above. Within each functional entity, the PDPTV team identified individual functions or responsibilities that involved some cost for the prototype repository. These costs are quantified where possible; described in terms of type (i.e. Staffing, Hardware, Software, Bandwidth, Electricity/Utilities or Service Fees); as one-time or ongoing costs; and as increasing or decreasing costs. As discussed above, it is very difficult to provide precise cost figures. Estimates stated in the summary sections below and in the appended spreadsheets are based on the PDPTV team's project experience and through discussions with NYU administrators.

Setting Up the Repository

This phase of preservation activity is not actually part of the OAIS model, which describes an already-functioning repository. Given the fact that start-up costs are quite significant, however, the PDPTV team included this phase as a discrete entity in its costing model. An Initiation phase is similarly included in the cost model proposed by Neil Beagrie, Julia Chruszcz and Brian Lavoie in their JISC-sponsored report on *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities*,[105] a key study in the costing of digital preservation. For the PDPTV project, this phase includes the functions:

- Project design
- Data management plan
- Repository architecture design and installation
- Network set up, testing, monitoring, maintenance
- Creation of customized interfaces and applications

The costs for this stage are primarily for staff. Because the start-up phase will only be initiated once, these staff costs are anticipated to be one-time only costs:

Initiation/Start Up Costs	
Staff Type	**One-Time (Start-Up) Full-Time Equivalents**
System Architect	0.75
Programmer/Analyst	2.75
Preservationist	1.0
System Administrator	0.1

Table 3. The estimated staff costs involved in the Initiation phase, stated in terms of Full-Time Equivalents. All staff costs in the Initiation phase are one-time costs.

Common Services

Common Services support and pervade all of the functions of the repository. According to the OAIS

104 Appendix B details the cost points for the NYU prototype preservation repository according to type and function.

105 Neil Beagrie, Julia Chruszcz and Brian Lavoie, "Keeping Research Data Safe: A Cost Model and Guidance for UK Universities" (n.p.: Higher Education Funding Council for England, 2008), http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf (accessed January 6, 2010.

model, Common Services include operating system services, network services, and security services.[106] For the purposes of describing costs, the PDPTV costing model includes the hardware and software that make up the technological infrastructure supported by Common Services within the entity. The cost model also includes the cost of electricity and utilities within Common Services.

At NYU, support for the large-scale technological infrastructure is centralized in ITS. Because ITS provides computer, network, telephone, and Internet services to the entire University, it was difficult for the PDPTV team to isolate and quantify the Common Services associated specifically with our relatively small project. In general, however, these costs include staff, hardware, software licensing, bandwidth, and utilities. These costs are primarily ongoing costs that will likely increase over time with added volumes of material.

Common Services		
Cost Center	**Cost Type**	**Cost Description**
Computing Hardware	Staffing	Ongoing, increasing cost for Sys Admin. Depends on traffic volume, frequency of periodic hardware refreshes.
	Hardware	Ongoing, possibly constant or increasing cost. NYU currently has a 4-year lease that includes computing and storage hardware, and maintenance. Generally, computing hardware costs depend on transaction volume.
Software Licenses	Staffing	Ongoing, increasing cost for Sys Admin.
	Software	Ongoing, constant cost. An estimate of the cost for individual ORACLE license. All other software used in repository is open source or developed in-house.
Operating System Services	Staffing	Ongoing, increasing cost for Sys Admin.
	Software	Ongoing, constant cost, since it is bundled with computing hardware lease
Network Services	Staffing	Ongoing, increasing cost for Sys Admin Depends on traffic volume.
	Hardware	Ongoing, possibly constant or increasing cost. Depends on repository traffic volume.
	Bandwidth	Ongoing, possibly constant or increasing cost. Depends on repository traffic volume.
Security Services	Staffing	Ongoing, increasing cost for Sys Admin.
Electricity and Utilities	Electricity/ Utilities	Ongoing, possibly constant or increasing cost. Depends on cost model for electricity and utilities.

Table 4. Description of the costs associated with the Common Services functions within the PDPTV prototype repository. Costs associated with Common Services are ongoing costs that are likely to increase over time and with volume of material.

The total cost of staff salaries for Common Services for the PDPTV prototype repository are estimated to be equal to 1.0 System Administrator FTE on an ongoing basis. All "Computing hardware" costs, including maintenance, are part of a four-year $1 million lease, which also includes storage hardware. As mentioned above, NYU's computing infrastructure includes 21 X4500 servers, with over 400 TB of storage. It is impossible for PDPTV to determine its "share" of the overall hardware and maintenance

106 Consultative Committee for Space Data Systems, “OAIS Reference Model,” 4-3 - 4-4.

costs; the total cost incurred by NYU is given here to provide public television with a general sense of the cost to lease this volume of computing power and storage.

It was not possible to obtain NYU's cost for network hardware or bandwidth at the time of this report; it perhaps suffices to say, however, that these are significant costs. Meanwhile, for the software licensing cost, it may be more instructive for public television to consider a different configuration. For example, the list cost of an individual ORACLE license is estimated between $12,500 to $50,000 rather than NYU's $2.5 million/5 year university-wide license. Finally, the cost of electricity and utilities reflects usage of 18 kW/year, which is an estimate of the amount of power to run one rack of hardware, equaling approximately $25,000/year. All of these costs are based on facilities and hardware in use or available in mid-2009. Undoubtedly, they will not reflect the costs for options that will be available in 2010 or later, when an actual repository might be planned.

Ingest

The Ingest entity provides the services to accept Submission Information Packages (SIPs) and prepares them for archival storage and management. The costs incurred within this functional entity are primarily non-recurring one-time costs for staff to develop software used for ingest. The PDPTV team estimates that 2.0 Programmer/Analyst FTE were required to develop software for the Ingest process. Hardware and bandwidth used in Ingest are part of Common Services costs, discussed above. Software costs are minimal, as most of the software is open-source or developed in-house by staff Programmer/Analysts.

Ingest		
Cost Center	**Cost Type**	**Cost Detail**
Receive SIP	Hardware (Common Services)	Ongoing cost. Cost/TB will decrease, but total volume may increase over time.
	Bandwidth (Common Services)	Ongoing cost. Depends on transaction volume.
	Software (Common Services)	Ongoing, constant cost. Software that is not part of Common Services is open source or developed by staff in-house.
Quality assurance/ file validation	**Staffing**	**One-time cost for Programmer/Analyst**
	Hardware (Common Services)	Ongoing, possibly constant or increasing cost. Depends on processing time on servers, related to hardware maintenance.
	Electricity/ Utilities (Common Services)	Ongoing cost, possibly constant or increasing cost. Depends on processing time on servers and electricity/ utilities cost model.
Generate AIP	**Staffing**	**One-time cost for Programmer/Analyst**
	Hardware (Common Services)	Ongoing, possibly constant or increasing cost. Depends on processing time on servers, related to hardware maintenance.
	Software (Common Services)	Ongoing, constant cost. Software that is not part of Common Services is open source or developed by staff in-house.
	Electricity/ Utilities (Common Services)	Ongoing cost, possibly constant or increasing cost. Depends on processing time on servers and electricity/ utilities cost model.
Transfer AIP	Hardware (Common Services)	Ongoing cost. Depends on transaction volume.
	Bandwidth (Common Services)	Ongoing cost. Depends on transaction volume.

	Software (Common Services)	Ongoing, constant cost. Software that is not part of Common Services is open source or developed by staff in-house.

Table 5. Description of the costs associated with the Ingest functions within the PDPTV prototype repository. Ongoing hardware and bandwidth costs are accounted for within Common Services (see above). Software costs are minimal as most software is developed in-house or open-source. The primary cost for Ingest is for staff, estimated to be a one-time start-up cost of 2.0 Programmer/Analyst FTE.

Archival Storage

The Archival Storage entity provides the services to store, maintain, and retrieve Archival Information Packages (AIPs). The costs in this entity are ongoing costs for staff, hardware, and utilities related to the processing time on the servers. PDPTV estimates that the ongoing staffing requirement for Archival Storage will be 0.1 System Administrator FTE (to manage storage) and 0.1 Programmer/Analyst FTE (to build validation and verification tools). Hardware and utilities costs are accounted for within the Common Services entity discussed above. The annual cost of off-site back-up and recovery services is also quite significant, estimated at approximately $50,000 a year for the NYU Digital Library as a whole.

Archival Storage		
Cost Center	**Cost Type**	**Cost Detail**
Receive data	Hardware (Common Services)	Ongoing cost. Cost/TB will decrease, but volume may increase over time.
	Bandwidth (Common Services)	Ongoing cost. Depends on transaction volume.
	Software (Common Services)	Ongoing, constant cost. Software that is not part of Common Services is open source or developed by staff in-house.
Storage systems (including administration and operation)	**Staffing**	**Ongoing, increasing cost for Sys Admin**
	Hardware (Common Services)	Ongoing cost. Depends on volume. Cost/TB will decrease, but total volume may increase over time.
	Software (Common Services)	Ongoing, constant cost. Software that is not part of Common Services is open source or developed by staff in-house.
File validation	**Staffing**	**One-time cost for Programmer/ Analyst**
	Hardware (Common Services)	Ongoing, possibly constant or increasing cost. Depends on processing time on servers, related to hardware maintenance
	Electricity/Utilities (Common Services)	Ongoing, possibly constant or increasing cost. Depends on processing time on servers, electricity/ utilities cost model
Offsite storage	**Staffing**	**Ongoing, increasing cost for Sys Admin**
	Service Fees	**Contract with external vendor for back-up and recovery services.**

Table 6. Description of costs associated with the Archival Storage functions within the PDPTV prototype repository. Estimated ongoing staff costs are 0.1 System Administrator FTE and 0.1 Programmer/Analyst FTE per year. Hardware and electricity/utilities costs for archival storage are included as part of the Common Services entity. NYU contracts back up and recovery services to an external vendor for an annual fee.

Data Management

The Data Management entity provides the services to populate, maintain, and access descriptive

information about the archive holdings and the administrative data to manage the archive. Functions normally include administering and updating the database, performing queries, and generating reports. For the PDPTV prototype repository, data management functions presented only nominal costs, as most costs depend on volume, and relate to hardware maintenance and electricity for running the servers. The main Data Management costs for the prototype repository were for staff salaries:

Data Management	
Staff Type	**Ongoing FTE**
Programmer/Analyst	0.2
Database Administrator	0.05

Table 7. Costs associated with the Data Management functional entity within the PDPTV prototype repository were primarily staff salaries. They are anticipated to be ongoing costs.

Administration

The Administration functional entity provides services for the overall operation of the archive system. The functions within the Administration entity that involved costs for prototype repository include:

- General management
- Determine selection criteria / policy
- Negotiate submission agreements
- Manage storage configuration
- Depositor support
- Implement and maintain standards and policies

The costs to provide these functions are primarily ongoing staff salaries:

Administration	
Staff Type	**Ongoing FTE**
System Architect	0.10
Project Manager	0.10
Preservationist	0.60
System Administrator	0.10

Table 8. Costs associated with the Administration functional entity within the PDPTV prototype repository were primarily staff salaries. They are anticipated to be ongoing costs

Preservation Planning

The Preservation Planning entity monitors the system environment and provides recommendations to ensure the long-term accessibility of the repository's holdings to the Designated Community. For the PDPTV prototype repository, the following functions were performed as part of Preservation Planning:

- Monitor designated community
- Monitor technology
- Develop preservation strategies and standards

As with the Administration entity, the costs associated with the functions in this entity are mainly ongoing staff salaries:

Preservation Planning	
Staff Type	**Ongoing FTE**
System Architect	0.30
Preservationist	0.50

Table 9. Costs associated with the Preservation Planning functional entity within the PDPTV prototype repository were primarily staff salaries. They are anticipated to be ongoing costs.

Access

The Access entity provides services that support users in finding, requesting, and receiving information stored in the repository. Access functions are quite limited for the PDPTV prototype repository, which is primarily a dark archive whose contents are accessible only to submitters/owners and to the Library of Congress. For the prototype repository, these minimal functions include:

- Generate DIP
- File validation
- Deliver response
- User support / training

The primary costs associated with the Access entity in the PDPTV prototype repository are for staff salaries, which are quite minimal (0.01 ongoing Programmer/ Analyst FTE; 0.1 ongoing Support Staff FTE). Software is developed by staff in-house, and other costs, for hardware and electricity/utilities, are accounted for under Common Services.

Access		
Cost Center	**Cost Type**	**Cost Detail**
Generate DIP	**Staffing**	**One-time cost for Programmer/ Analyst**
	Hardware (Common Services)	Ongoing cost. Depends on processing time on servers, related to hardware maintenance
	Electricity/Utilities (Common Services)	Ongoing cost. Depends on processing time on servers, electricity/ utilities cost model
File validation	**Staffing**	Ongoing cost for Programmer/ Analyst
Deliver response	**Staffing**	**Ongoing cost for Programmer/ Analyst**
	Hardware (Common Services)	Ongoing cost. Depends on processing time on servers, related to hardware maintenance
	Bandwidth (Common Services)	Ongoing cost. Depends on transaction volume.
	Electricity/Utilities (Common Services)	Ongoing cost. Depends on processing time on servers, electricity/ utilities cost model
User support/ training	**Staffing**	**Ongoing cost for Support Staff**

Table 10. Access functions for the PDPTV prototype repository are very limited. Costs associated with the Access entity within the prototype repository were mainly one-time (start-up) and ongoing staff salaries.

In planning a repository, all these functions need to be performed, whether by the repository itself or by outside entities. Even without specific costs assigned to each task, we believe this to be useful for planning. Based on this extensive breakdown, actual costs for building and operating a modest preservation repository for public television, including staffing and other expenses, could be projected along several different lines. Start up activities require a major investment, but operating costs could be projected out over an extended period to be level or declining, based on the size of the operations and volume of content.

* * * * *

Chapter 6

CONCLUSION: Public Television has the Capacity to Support Digital Preservation

Since the PDPTV project began, all television broadcasters have switched their analog transmitters to digital signals, and the partner stations have begun to implement end-to-end file-based program production workflows, a development that will soon be practiced at local broadcast stations throughout the system. Along with swapping videotape machines for computer servers, these changes have also led to adopting new approaches to editing, sharing and maintaining file-based program content. It is evident that new approaches for long-term preservation must also be incorporated into these workflows if we want programs to exist long-term in this new digital environment.

This report has examined the problem of digital sustainability as a broad range of related issues, overturning the popular notion that preservation is limited to storage costs and operating income. Instead, we set out to demonstrate that sustainable preservation and access can only occur when there are trustworthy digital repositories -- entities with the technical and organizational infrastructure capable of preserving bits and managing digital content, so that it is findable and usable. In this context, sustainability encompasses not only the resources needed to sustain the materials being preserved, but also the systems, activities and staffing needed to sustain the organization or services doing the preservation.

We have outlined how the size, scope and activity of any long-term digital repository must rest on many variable and interrelated factors, starting with identifying the underlying goals and extending to reflect:

- The mission for preserving specific materials;
- The stakeholders who have an interest in preserving such materials;
- The scale necessary to care for the materials;

- The services and operations that might be required;
- The structures for governance and decision-making;
- The range of income sources and revenue streams that can be exploited.

Once an organizational and technical framework has been described, the requirements can be broken down and described, operating models can be considered, and responsibilities can be assigned across multiple constituent groups. That makes it possible to project reasonable costs for repository functions such as archival storage, access, data management, and administrative services.

It is worth noting that the problems addressed here are not unique to collections in the US, but are common to institutions and organizations with valuable digital collections around the world. A number of international initiatives are building and operating repositories that are trying to solve the technical and functional challenges of preserving digital files, and compared to Europe and Australia, the US lags behind in finding the solutions and establishing the infrastructure capable of dealing with its national digital video dilemma. Fortunately, we can learn from the experiences of these television and video repositories abroad, which are beginning to have enough of a track record that they can provide extremely helpful guidance to us in understanding issues of long-term sustainability.

Within such a global context, this project has tried 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, and as such, it must be an important component in the complete life cycle management of digital productions. As this report has discussed, the adoption of preservation standards and practices are necessary not only for accessibility of the content in the long run, but for the continued usability of the files in the near- and medium-term future. For this reason alone, preservation should not be in competition with production. In fact, digital preservation practices serve the new needs of both ongoing and long-term access to digital content in a cost-effective way.

U.S. public broadcasting already relies on relatively diverse funding sources to support its current operations. This diversity has contributed to the durability of public television over time, but even without many funds to spare, there are many familiar cooperative and collaborative business models within the system that meet the interests of different sets of stakeholders. Together with an understanding of the needs of digital preservation, public television is capable of adapting and developing new ways of generating income to support its preservation activities.

It is exactly this potential that gives digital preservation its intrinsic value -- it allows the ability to give old content new life, to meet the expanding demands by viewers, and to create new sources of revenue from archived materials, all while fulfilling the public television mission to educate, inform, and entertain.

This means that the challenges of preservation should not be as daunting as was once believed -- as long as they are a thoughtfully planned component of public broadcasting's goals and resources, the costs can represent a manageable long-term investment that can be shared among multiple stakeholders and supported by diverse revenue streams.

The Promise of the *American Archive*

This project has reiterated that digital preservation requires appropriate infrastructure, sound policies and practices, and reliable and sufficient financial support. The sustainability of digital repositories as the central tool in digital preservation, is a critical issue in today's uncertain economy. Historically in public broadcasting, audiovisual preservation has had no dedicated funding but has been supported by “soft money” that comes and goes according to the availability of project funds and the efforts of grant writers.

But in the digital era, this model will not suffice to ensure the survival of our cultural heritage. Instead, we

believe the support for program preservation and long-term access must be a system-wide commitment – building on and expanding the investments in existing infrastructure and production that have already been made, so that not only the costs but also the benefits of preservation can be shared across all the stakeholders throughout public broadcasting.

For the very first time, the Corporation for Public Broadcasting is investing in preservation activities by allocating funding for *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." [107]

The American Archive is being planned as a totally new entity, independent of existing public broadcasting institutions, in part as a response to growing momentum within public broadcasting from projects like ours. It will take several years to shape, outline how it will operate, design a governing structure, and build sustainable funding streams. Nonetheless, public broadcasting stations across the country have endorsed the project and as this is being written, development of *The American Archive* is moving along rapidly. We hope this report will contribute towards the planning that goes into making its activities and services sustainable.

* * *

Viewers keep reminding us that public television programming is precious and has made an indelible imprint on American culture. However, public television is just now beginning to recognize and reap the benefits of preserving and maintaining its programming assets. The ability to find and re-use material for production translates to immediate and long-term cost savings, and the capacity to re-market and re-package content that has been saved and kept viable means that public television can reach new audiences in new ways. In the end, the savings and revenues recovered from exploiting renewed assets, not to mention other intangible values, may well outweigh the costs.

The public broadcasting community is ready to express its desire to keep our television content vibrant and useful. The technical conditions necessary to operate a 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 to be done now is to continue building commitment across the entire system to make the required investment to sustain digital preservation over the long-term. Combining local efforts with national initiatives like the American Archive can insure that the responsibility for saving this American media legacy will be maintained over time, not solely to serve the public broadcasting system, but because it will benefit the American public as a whole.

* * * * *

[107] Speech to the Association of Public Television Stations, February 2007.

APPENDIX A: Economic Models Based on Market Analysis of Stakeholder Interests

Defining the 'Marketplace'

In *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*, Brian Lavoie suggests that we understand digital preservation as something like an "aftermarket" service -- that is, digital preservation participates in "the market for goods and services which maintain and/ or enhance the ability of a durable good to provide a stream of benefits over an extended period of time."[108] The "durable good" in this case is digital information, which will deteriorate and depreciate in value in the absence of these services.

Meanwhile, the "aftermarket" is made up of stakeholders who occupy critical roles in the preservation process. In Lavoie's simplified economic model, there are three key economic decision-making roles among stakeholders in digital preservation as discussed in Chapter 3:

- **Rights Holder**: Owns the rights to a digital resource and has the authority to dispose of those rights. This entity will have the most direct interest and control over the uses of the content, and potentially direct recipient of financial benefits. The "Rights Holder" of a particular piece of content could be, for example, a copyright holder, a program producer, a program distributor, or all of them in combination.

- **Archive**: As the entity that implements and manages the preservation process, the "Archive" role must be able to operate and sustain itself. It could be a producing station, a third-party such as a public institution, or a new entity within the public television system.

- **Beneficiary:** There are a number of entities that can realize potential benefits from the long-term retention of the digital resource, including all of the stakeholders who would make use of the preserved content.

In a hypothetical preservation scenario for public television, the roles of Rights Holder, Archive, and Beneficiary could be filled by new or existing stakeholders. As mentioned previously, potential stakeholders in preserving public television content include copyright holders, program producers, filmmakers, program distributors, educators, educational resource producers, students, academics in higher education, researchers, journalists, and the general public.

Roles among stakeholders can vary and overlap, and with such a broad group, various interests and demands on the repository can become contradictory. One stakeholder could, for example, be both a Rights Holder and a Beneficiary, or be both the Rights Holder and the Archive. The relationship of one stakeholder to another in a given situation produces what could be called an "organizational scenario". Since an organizational scenario describes the structure of a particular digital preservation activity, it is possible that a single organization can take part in a number of organizational scenarios at once.

108 Brian Lavoie, "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making" (Dublin, OH: OCLC, 2003), 21.

Five Economic Scenarios

Based on such flexible roles, Lavoie describes five possible scenarios under which digital preservation can be organized (see Figure 1). Each of these possible combinations assumes a set of benefits dictated by the economic interests of the relationships:

- **Demand Side Model**: One stakeholder is the Rights Holder and Beneficiary; another stakeholder is the Archive. The Rights Holder would organize archive activities primarily for its own benefit.
- **Centrifugal Model**: One stakeholder is the Rights Holder, another stakeholder is the Archive, and yet a third stakeholder is the Beneficiary. Here, the economic interests are roughly balanced equally among the stakeholders.
- **Centripetal Model**: The Rights Holder, Archive, and Beneficiary are the same stakeholder. In this scenario, all benefits would be captured by this single entity.
- **Supply Side Model**: One stakeholder is the Rights Holder and Archive; another stakeholder is the Beneficiary. As such, the Rights Holder and Archive must both serve the Beneficiary.
- **Consolidated Model**: One stakeholder is the Rights Holder; another stakeholder is the Archive and Beneficiary. The interests of the Rights Holder are in contrast to the other two entities.

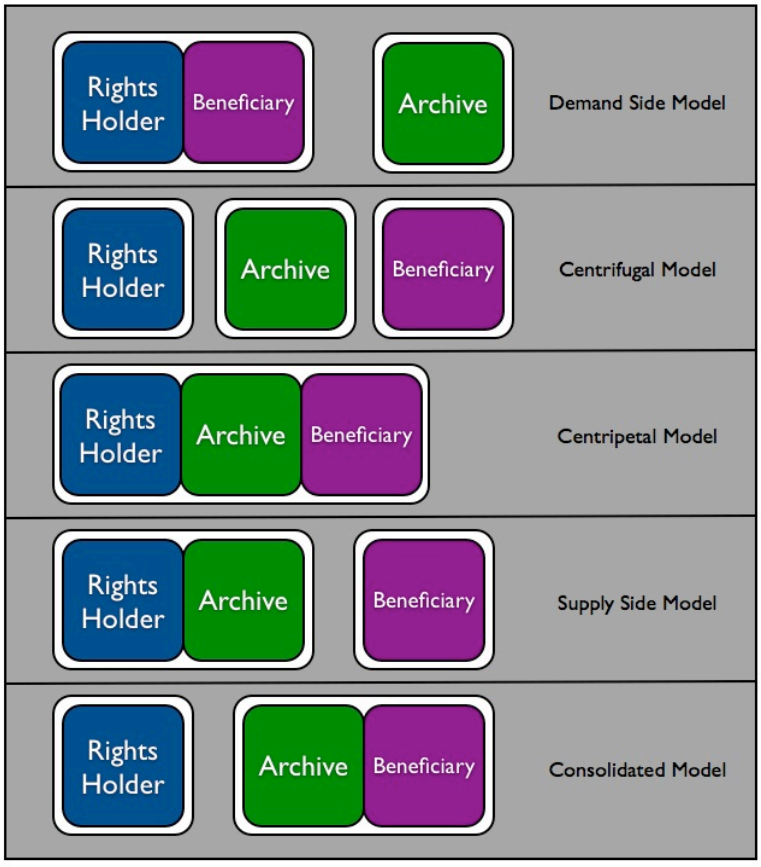

Figure 7. The five "organizational scenarios" in which a digital preservation activity can take place, as defined by Brian Lavoie in The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making.

The Need to Correct Incentive Gaps and Market Failures

As the figure illustrates, each scenario presents a model where the benefits for preservation can be shared among the entities, and by association, whereby the costs and risks of operating a repository would also be shared among the entities.

However, these benefits and risks are not necessarily shared equally. All these organizational scenarios, except the Centripetal Model, creates "incentive gaps" and "market failures." These gaps and failures, discussed in turn below, work as disincentives that discourage stakeholders from wanting to participate in the preservation process. It is important to know where and how disincentives are produced in an economic chain, in order to devise strategies to correct them. Without deliberate intervention in the right place, this model predicts that digital preservation activity would come to a halt.

While identifying incentive gaps and market failures is not too difficult, choosing the right market or non-market mechanism to address those disincentives is the challenge. Understanding how the roles, responsibilities, and benefits of digital preservation are allocated among the stakeholders involved in any given scenario is a fundamental first step in this process. As the Blue Ribbon Task Force indicates, "any analysis of the economic mechanisms appropriate for sustaining digital preservation activities over the long-term should be preceded by an examination of the *economic context* in which these models are expected to operate."[109]

Incentive Gaps

The distribution of roles among stakeholders often results in a misalignment between those who have the authority to undertake preservation (the Rights Holder), those who have the ability to do it (the Archive), and those who would benefit from it (the Beneficiary). As Lavoie explains, the division of Rights Holder, Archive, and Beneficiary roles among multiple stakeholders produces incentive gaps which can block preservation from taking place.

The incentive gaps revealed by this economic model may be applied to anticipated scenarios in preserving digital public television, giving us a sense of what needs to be done to ensure that preservation takes place. For example, if a local producing station wants to preserve a digital program that it owns so that it can be re-aired in the future, the station is both the Rights Holder and the Beneficiary of preservation. If, however, the station does not have the equipment or staff to do the preservation work, an external entity like a third-party repository would be required to serve as the Archive.

This scenario is an example of a Demand Side Model in Lavoie's scheme, in which the Archive is not the Beneficiary, and which thereby creates an incentive gap for the Archive. Why should the Archive repository supply preservation services for the local television station? The economic model tells us that some non-market or market mechanism -- e.g. a legal mandate to provide preservation services; a mission to serve public television; payment from the station, etc. -- must be used to correct the lack of incentive and motivate the Archive to perform preservation. Each of the organizational scenarios, except one, produces its own inherent incentive gaps like this one.

The exception is the Centripetal Model, in which the Rights Holder, Archive, and Beneficiary are all the same entity. This scenario does not require external incentives to motivate activity, as a single organization controls all risks and benefits from the repository. An existing example of this scenario within public television would be WNET or WGBH and its respective station-based archives. As rights holders, each station is self-motivated to maintain its own internal archive, which performs preservation work for the benefit of the station itself.

---

109 Blue Ribbon Task Force, “Sustaining the Digital Investment,” 35.

In the context of public broadcasting, these scenarios can be used to identify various specific stakeholders within the system and determine successful motivations and strategies to overcome the incentive gaps inherent in the relationships between participating entities.

Market Failures

Like incentive gaps, market failures revealed by the economic model can also help to explain how a stakeholder might act in a given situation. Market failures are "imperfections associated with the workings of the market mechanism,"[110] which discourage stakeholders from participating. Lavoie argues that the market failures that occur in his economic model for digital preservation result from the characteristics of digital resources and the preservation process, and lend to problems normally faced by public goods, like the free-rider/positive externalities problem.

One market failure occurs as a result of a scenario in which the preservation of a digital resource benefits stakeholders *other* than those with the authority to preserve them. The preservation of an important groundbreaking television program, for example, may be of great cultural value to society as a whole. If the rights-holding station does not directly benefit, however -- say, because it cannot re-air the show due to underlying rights issues -- it may be difficult for the station to justify bearing the cost of preservation. This is an imperfect situation in which the overall benefit of preserving the resource may be great, but the benefit to be reaped by the Rights Holder is insufficient to motivate it to act.

This is a very common situation in the public television system, which keeps the majority of programs out of re-distribution and inaccessible to the public largely because of the risk and costs associated with program productions dense with complex layers of rights.

Other market failures can occur because digital preservation resembles a public good. As Lavoie explains, digital preservation is non-rivalrous (i.e., consumption by one does not diminish potential supply for others) and non-excludable (i.e., the cost of excluding non-paying consumers is high). In other words, once someone pays for the service, it does not cost more to allow others to benefit from it. Moreover, it is difficult to prevent others from benefiting. This creates an imperfection in the market in which parties can "free-ride," off others who are willing to pay.

If all parties know that they can free-ride off someone else who is willing to pay, it follows no one will want to be the one who pays. If a public television station preserves the broadcast master of an acquired program, for example, the independent producer of said program may choose not to pay for its own preservation masters, knowing that someone else has already safeguarded it for the future. Similarly, if a public television station knows that another program is being preserved elsewhere, it may be less motivated to select or prioritize it for preservation.

Ultimately, this may produce a situation in which each potential stakeholder holds off on paying for preservation, hoping that another stakeholder will do it first. Within public broadcasting, this can mean clarifying who holds responsibilities for preservation, as well as who has resources to contribute towards this activity.

Finally, another market failure occurs in this model when end-users (i.e. Beneficiaries) have heterogeneous, and even conflicting, preservation and access demands. In a public television context, this may resemble a situation in which a local station wants a full-quality production master of a program, while a high-school class only cares about being able to view a copy of the program online. As Lavoie explains, the Archive could supply different levels of service in order to take advantage of the entire market.

This presents a problem similar to the "free-rider," however, in that low-end users can benefit from the

110 William J. Baumol and Alan S. Blinder, *Economics: Principles and Policy*, 7th ed. (Orlando: The Dryden Press, 1997), 323.

high-end preservation paid for by high-end users. The high-school may only be interested in paying for the ability to see an online video today, but years from now when it wants to access the program again, it will benefit from the fact that the local station has paid for the ongoing storage and migration of the high-quality file in the intervening time. Lavoie calls these "spillover benefits."

As with incentive gaps, market failures can be corrected with appropriate strategies or business models. A mechanism that spreads costs across all potential beneficiaries evenly, for example, would eliminate the problem of free-riders. A pricing scheme that forces users with heterogeneous demands to pay for the level of service they require, at tiered prices that reflect the overall cost of preservation at that level, could address the problem of spillover benefits. These kinds of strategies need more exploration to devise appropriate and affordable prices for services.

Summary: Roles, Organizational Scenarios, Incentive Gaps, and Market Failures

In Lavoie's economic model for preservation, the arrangement of economic stakeholder roles, or the organizational scenario, can create disincentives for participating in the digital preservation process. Effective business models or other non-market solutions would work to address the incentive gaps and market failures presented by a given scenario, thereby encouraging preservation to take place.

Organizational Scenario	Incentive Gaps	Potential Market Failures
Demand Side Model (Rights Holder and the Beneficiary are the same entity; the Archive is separate)	no incentive for Archive	• free-ridership if a digital resource has multiple Rights Holder/ Beneficiaries • spillover benefits and market segmentation if there is heterogeneous demand
Centrifugal Model (Rights Holder, Archive, and Beneficiary are all separate entities)	no incentives for either Rights Holder or Archive	• positive externality, benefits accruing mostly to those who cannot initiate preservation • spillover benefits and market segmentation if there is heterogeneous demand
Centripetal Model (Rights Holder, Archive, and Beneficiary are all the same entity)	no incentive gaps	
Supply Side Model (Rights Holder and the Archive are the same entity; the Beneficiary is separate)	no incentive for the Rights Holder / Archive	• positive externality • spillover benefits and market segmentation if there is heterogeneous demand
Consolidated Model (Rights Holder is a separate entity; the Archive and Beneficiary are the same entity)	no incentive for the Rights Holder	• positive externality

Table 11. A summary of Lavoie's organizational scenarios, their inherent incentive gaps, and their potential market failures.

The challenge for public broadcasting is to understand these dynamics and respond to them when designing sustainable preservation services by devising appropriate incentives that encourage each stakeholder group to participate.

* * * * *

Appendix B: NYU PDPTV Repository Costing Case Study

The tables on the following pages provide additional detail about the costs associated with the prototype repository developed at New York University for the Preserving Digital Public Television Project. These tables identify the cost points associated with planning, developing, and operation of the repository. They should be seen as very specific to New York University. As Chapter 5 demonstrated, many of the costs for this project were difficult to break out from the cost of larger university-wide or library-wide systems which are purchased or licensed by the university for a variety of departments and functions. However, our hope is that these tables will help others with repository planning, and can be used as a guideline for identifying staffing, bandwidth, hardware, software, and other requirements.

- **Table B-1: Costs By Type / Function** breaks down costs associated with the repository by type and OAIS functional entity (Ingest, Archival Storage, etc.) and process within that entity, and identifies the type of staff needed for each function.

- **Table B-2: Staff By Type / Function** details the various types of staff involved in the repository's planning, development, and ongoing operation. For each type, FTE (full time equivalent) staff that were needed for each OAIS functional entity are identified, and the staffing needs for that type / function on a start-up (one time) and ongoing basis.

- **Table B-3: Start Up vs Ongoing Costs** takes a closer look at the cost types (staffing, hardware, software, etc) for each OAIS functional entity and notes whether that was a one-time start up cost or an ongoing cost.

Combined, these tables should provide a picture of the different types of costs, the staff roles and time, and the different needs at start-up versus ongoing operation for the PDPTV repository.

Appendix B-1: Costs by Type / Function

FUNCTIONAL ENTITY	FUNCTION/COST POINT		STAFF	HARDWARE	SOFTWARE	BANDWIDTH	ELECTRICITY AND UTILITIES	SERVICE FEES	OTHER
Initiation	Project design		Architect Programmer/Analyst						
	Data management plan		Architect Preservationist Programmer/Analyst						
	Repository architecture design and installation		Architect Sys Admin (at both repository and contributing nodes)						
	Network Set Up, testing, monitoring, maintenance		Sys Admin (at both repository and contributing nodes)						
	Creation of customized interfaces and applications		Programmer/Analyst						
Common Services	Computing hardware	Servers, desktops, and maintenance	Sys Admin Staffing depends on traffic volume, frequency of periodic hardware refreshes. NYU Digital Library administers its own desktops. Maintenance cost is included in computing hardware lease.	$250,000 (i.e. NYU lease is $1 million over 4 years. Includes computing and storage hardware, and maintenance). Generally, the cost of computing hardware depends on traffic volume.					
	Software licenses		Sys Admin		$12,500-$50,000 Estimate of cost for individual ORACLE license. All other software used in repository is open source or developed in-house.				
	Operating system services	Core services needed to operate and administer the application platform, and provide an interface between application software and the platform	Sys Admin		Included in computing hardware cost.				
	Network services	Capabilities and mechanisms to support distributed applications requiring data access and applications interoperability	Sys Admin Depends on traffic volume.	A significant cost. Depends on repository traffic volume.		A significant cost. Depends on repository traffic volume.			

Appendix B-1: Costs by Type / Function

FUNCTIONAL ENTITY	FUNCTION/COST POINT		STAFF	HARDWARE	SOFTWARE	BANDWIDTH	ELECTRICITY AND UTILITIES	SERVICE FEES	OTHER
	Security services	Access, authentication, confidentiality, integrity, and non-repudiation controls and management of communications between senders and receivers of information	Sys Admin						
	Electricity / Utilities						$25,000 Estimate of NYU's cost to run and cool a rack of servers and other hardware (~18kW).		
Ingest	Receive SIP	File transfer, staging storage		Included in Common Services' computing hardware costs	Some portion of ORACLE site-license (SRB runs using ORACLE). Included in Common Services' software licenses.	Included in Common Services' bandwidth costs			
	Quality Assurance / File Validation	Scripts to validate SIP; generate checksums	Programmer/Analyst	Depends on processing time on servers. Included in Common Services costs.	$0 Open source or developed in-house		Depends on processing time on servers. Included in Common Services costs.		
	Generate AIP	Normalize filenames, generate AIP directory, metadata extraction, checksum and metadata generation	Programmer/Analyst	Depends on processing time on servers. Included in Common Services costs.	Some portion of ORACLE site-license (SRB runs using ORACLE). Included in Common Services' software licenses.		Depends on processing time on servers. Included in Common Services costs.		
	Transfer AIP	Transfer to storage, update database		Included in Common Services' computing hardware costs	Some portion of ORACLE site-license (SRB runs using ORACLE). Included in Common Services' software licenses.	Included in Common Services' bandwidth costs			
Archival Storage	Receive data	Transfer to storage volume, update database		Included in Common Services' computing hardware costs	Some portion of ORACLE site-license (SRB runs using ORACLE). Included in Common Services' software licenses.	Included in Common Services' bandwidth costs			
	Storage systems	Storage systems administration / operation; manage storage hierarchy	Sys Admin	Included in Common Services' computing hardware costs	Some portion of ORACLE site-license. Included in Common Services' software licenses.				
	File validation	Scripts to validate and verify AIP, error correction	Programmer/Analyst	Depends on processing time on servers. Included in Common Services costs.	$0 Open source or developed in-house		Depends on processing time on servers. Included in Common Services costs.		

Appendix B-1: Costs by Type / Function

FUNCTIONAL ENTITY	FUNCTION/COST POINT		STAFF	HARDWARE	SOFTWARE	BANDWIDTH	ELECTRICITY AND UTILITIES	SERVICE FEES	OTHER
	Offsite storage	Storage and recovery services; data transfer	Sys Admin Depends on volume (cost per TB will decrease as volume increases)					Back up and recovery services from Sungard $500,000/ year for NYU records. Back-up hardware, media, support are included. An estimate of the Library's share of this is approx. $50000	
Data Management	Administer database	Maintain schema and definitions; update database	Database Admin Programmer/Analyst	Depends on processing time on servers. Included in Common Services costs, but nominal for prototype repository.			Depends on processing time on servers. Included in Common Services costs, but nominal for prototype repository.		
Preservation Planning	Monitor designated community	Monitor producers' product technologies and service requirements	Preservationist						
	Monitor technology		Architect						
	Develop preservation strategies and standards	Develop format, documentation, and procedural standards, AIP/SIP designs based on reports from designated communities and technology watch	Architect Preservationist						
Administration	General management	Operational and strategic planning, staff management, staff professional development	Project Manager Architect Preservationist						Travel, conferences, tuition remission
	Determine selection criteria / policy		Preservationist						
	Negotiate submission agreements	Determine SIP/AIP/ DIP requirements, submission schedule, service level agreements	Preservationist Architect						
	Manage storage configuration	Monitor system, audit system operations	Sys Admin						
	Depositor support		Preservationist						
	Implement and maintain standards and policies	Implement format, documentation, and procedural standards; migration goals and schedule; storage management policies	Architect Preservationist Sys Admin						
Access	Generate DIP	File and metadata transformation, staging storage	Programmer/Analyst	Depends on volume and processing time on servers. Included in Common Services costs.	$0 Open source or developed in-house		Depends on processing time on servers. Included in Common Services costs.		

Appendix B-1: Costs by Type / Function

FUNCTIONAL ENTITY	FUNCTION/COST POINT		STAFF	HARDWARE	SOFTWARE	BANDWIDTH	ELECTRICITY AND UTILITIES	SERVICE FEES	OTHER
	File validation	Scripts to validate DIP	Programmer/Analyst	Depends on processing time on servers. Included in Common Services costs.	$0 Open source or developed in-house		Depends on processing time on servers. Included in Common Services costs.		
	Deliver response	Receive request, deliver response, transmit DIP	Programmer/Analyst	Depends on processing time on servers. Included in Common Services costs.	$0 Open source or developed in-house	Included in Common Services' costs.	Depends on processing time on servers. Included in Common Services costs.		
	User support / training		Support Staff						

Appendix B-2: Staff by Type / Function

	PROJECT MANAGER		SYSTEM ARCHITECT		PRESERVATIONIST		PROGRAMMER/ ANALYST		SYSTEMS ADMINISTRATOR		DATABASE ADMINISTRATOR		SUPPORT STAFF		TOTALS	
	One-Time	Ongoing	One-Time	Ongoing	One-Time	Ongoing	One-Time	Ongoing	One-Time	Ongoing	One-Time	Ongoing	One-Time	Ongoing	ONE-TIME (START-	ONGOING FTE
Initiation	0	0	0.75	0	1	0	2.75	0	0.1	0	0	0	0	0	4.6	0
Common Services	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Ingest	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0
Archival Storage	0	0	0	0	0	0	0	0.1	0	0.1	0	0	0	0	0	0.2
Data Management	0	0	0	0	0	0	0	0.2	0	0	0	0.05	0	0	0	0.25
Preservation Planning	0	0	0	0.3	0	0.6	0	0	0	0	0	0	0	0	0	0.9
Administration	0	0.1	0	0.1	0	0.5	0	0	0	0.1	0	0	0	0	0	0.8
Access	0	0	0	0	0	0	0.05	0.01	0	0	0	0	0	0.1	0.05	0.11
TOTALS	0	0.1	0.75	0.4	1	1.1	4.8	0.31	0.1	1.2	0	0.05	0	0.1	6.65	3.26

Appendix B-3: Start Up vs Ongoing Costs

FUNCTIONAL ENTITY	FUNCTION/COST POINT		COST TYPE	ONE TIME (START-UP) COST	ONGOING COST	*NOTES*
Initiation	Project design		Staffing	✔		
	Data management plan		Staffing	✔		
	Repository architecture design and installation		Staffing	✔		
	Network Set Up, testing, monitoring, maintenance		Staffing	✔		
	Creation of customized interfaces and applications		Staffing	✔		
Common Services	Computing hardware	Servers, desktops, and maintenance	Staffing		✔	
			Hardware		✔	*Depends on volume*
	Software licenses		Staffing		✔	
			Software		✔	
	Operating system services	Core services needed to operate and administer the application platform, and provide an interface between application software and the platform	Staffing		✔	
			Software		✔	
	Network services	Capabilities and mechanisms to support distributed applications requiring data access and applications interoperability	Staffing		✔	
			Hardware		✔	*Depends on transaction volume*
			Bandwidth		✔	*Depends on transaction volume*
	Security services	Access, authentication, confidentiality, integrity, and non-repudiation controls and management of communications between senders and receivers of information	Staffing		✔	

Appendix B-3: Start Up vs Ongoing Costs

FUNCTIONAL ENTITY	FUNCTION/COST POINT		COST TYPE	ONE TIME (START-UP) COST	ONGOING COST	*NOTES*
	Electricity / Utilities		Electricity/ Utilities		✔	*Depends on electricity/utilities cost model*
Ingest	Receive SIP	File transfer, staging storage	Hardware		✔	*Assuming that cost/ TB will decrease, but that volume may increase over time*
			Bandwidth		✔	*Depends on transaction volume*
			Software		✔	
	Quality Assurance / File Validation	Scripts to validate SIP; generate checksums	Staffing	✔		
			Hardware		✔	*Depends on processing time on servers, related to hardware maintenance*
			Electricity/ Utilities		✔	*Depends on processing time on servers electricity/ utilities cost model*
	Generate AIP	Normalize filenames, generate AIP directory, metadata extraction, checksum and metadata generation	Staffing	✔		
			Hardware		✔	*Depends on processing time on servers, related to hardware maintenance*
			Software		✔	
			Electricity/ Utilities		✔	*Depends on processing time on servers electricity/ utilities cost model*
	Transfer AIP	Transfer to storage, update database	Hardware		✔	*Depends on transaction volume*

Appendix B-3: Start Up vs Ongoing Costs

FUNCTIONAL ENTITY	FUNCTION/COST POINT		COST TYPE	ONE TIME (START-UP) COST	ONGOING COST	*NOTES*
			Bandwidth		✔	*Depends on transaction volume*
			Software		✔	
Archival Storage	Receive data	Transfer to storage volume, update database	Hardware		✔	*Depends on volume. Cost/TB will decrease, but volume may increase over time.*
			Bandwidth		✔	*Depends on transaction volume*
			Software		✔	
	Storage systems	Storage systems administration / operation; manage storage hierarchy	Staffing		✔	
			Hardware		✔	*Depends on volume. Cost/TB will decrease, but total volume may increase over time.*
			Software		✔	
	File validation	Scripts to validate and verify AIP; error correction	Staffing	✔		
			Hardware		✔	*Depends on processing time on servers, related to hardware maintenance*
			Electricity/ Utilities		✔	*Depends on processing time on servers, electricity/ utilities cost model*
	Offsite storage	Storage and recovery services; data transfer	Staffing		✔	*Decreasing cost/TB (salary distributed over more volume), but total TB may increase over time*
			Service Fees		✔	*Depends on vendor's fees*

Appendix B-3: Start Up vs Ongoing Costs

FUNCTIONAL ENTITY	FUNCTION/COST POINT		COST TYPE	ONE TIME (START-UP) COST	ONGOING COST	*NOTES*
Data Management	Administer database	Maintain schema and definitions; update database	Staffing	✔	✔	*Primarily an ongoing cost, but some that will only be one-time cost*
Preservation Planning	Monitor designated community	Monitor producers' product technologies and service requirements	Staffing		✔	
	Monitor technology		Staffing		✔	
	Develop preservation strategies and standards	Develop format, documentation, and procedural standards, AIP/ SIP designs based on reports from designated communities and technology watch	Staffing		✔	
Administration	General management	Operational and strategic planning, staff management, staff professional development	Staffing		✔	
	Determine selection criteria / policy		Staffing		✔	
	Negotiate submission agreements	Determine SIP/AIP/DIP requirements, submission schedule, service level agreements	Staffing		✔	
	Manage storage configuration	monitor system, audit system operations	Staffing		✔	
	Depositor support		Staffing		✔	
	Implement and maintain standards and policies	Implement format, documentation, and procedural standards; migration goals and schedule; and management policies	Staffing		✔	
Access	Generate DIP	File and metadata transformation, staging storage	Staffing	✔		

Appendix B-3: Start Up vs Ongoing Costs

FUNCTIONAL ENTITY	FUNCTION/COST POINT		COST TYPE	ONE TIME (START-UP) COST	ONGOING COST	*NOTES*
			Hardware		✔	*Depends on processing time on servers, related to hardware maintenance*
			Electricity/ Utilities		✔	*Depends on processing time on servers, electricity/ utilities cost model*
	File validation	Scripts to validate DIP	Staffing	✔		
	Deliver response	Receive request, deliver response, transmit DIP	Staffing		✔	
			Hardware		✔	*Depends on processing time on servers, related to hardware maintenance*
			Bandwidth		✔	*Depends on transaction volume*
			Electricity/Utilities		✔	*Depends on processing time on servers, electricity/ utilities cost model*
	User support / training		Staffing		✔	

Works Cited

Addis, Matthew, Richard Beales, Richard Lowe, Lee Middleton, Charlotte Norlund, Zlatko Zlatev, "Sustainable Archiving and Storage Management of Audiovisual Digital Assets," *SMPTE Motion Imaging Journal* Vol. 118, No. 8 (November/December 2009) p. 27-32.

Amazon Web Services. "Amazon Simple Storage Service." Amazon. http://aws.amazon.com/s3/#pricing (accessed January 6, 2010).

Archaeology Data Service, "About the ADS." ADS, http://ads.ahds.ac.uk/project/about.html (accessed January 6, 2010).

Barton, Mary R. and Julie Harford Walker. "Building a Business Plan for DSpace." MIT Libraries' Digital Institutional Repository. N.p.: Texas A&M University, 2003. http://dspace.mit.edu/bitstream/handle/1721.1/26700/Barton_2003_Building.pdf?sequence=1 (accessed on January 6, 2010).

Baumol, William J., and Alan S. Blinder. *Economics: Principles and Policy*. 7th ed. Orlando: The Dryden Press, 1997.

BDO Seldman, LLP. "Public Broadcasting Services and Subsidiaries, Consolidated Financial Statements and Independent Auditors' Report, Years Ended June 30, 2008 and 2007." (Bethesda: BDO Seldman, 2008). http://www.pbs.org/aboutpbs/content/annualreport/2008/2008FullFinancialHighlights.pdf (accessed January 6, 2010).

Beagrie, Neil, Julia Chruszcz and Brian Lavoie. "Keeping Research Data Safe: A Cost Model and Guidance for UK Universities." N.p.: Higher Education Funding Council for England, 2008. http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf (accessed January 6, 2010).

Blue Ribbon Task Force on Sustainable Digital Preservation and Access. "Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation, Interim Report." N.p.: Blue Ribbon Task Force on Sustainable Digital Preservation and Access, 2008. http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf (accessed January 6, 2010).

Bradley, Kevin. "Defining Digital Sustainability," *Library Trends* 56, no. 1 (Summer 2007), http://muse.jhu.edu/journals/library_trends/v056/56.1bradley.html (accessed January 6, 2010).

CLOCKSS. "CLOCKSS: A Trusted Community-Owned Archive." CLOCKSS, http://www.clockss.org/clockss/Home (accessed January 6, 2010).

Consultative Committee for Space Data Systems. "Reference Model for an Open Archival Information System (OAIS)," CCSDS 650.0-B-1, Blue Book, Issue 1. Washington, DC: CCSDS Secretariat, 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf (accessed on January 6, 2010).

Corporation for Public Broadcasting. "CPB's Goals and Objectives." Corporation for Public Broadcasting. http://www.cpb.org/aboutcpb/goals/goalsandobjectives/ (accessed January 6, 2010).

Darimont, Albert W. "Environmental Scan of Pricing Models for Online Content: Report II on Business Models for Object Repositories." N.p.: OnDisC, 2002.

Day, Michael. "Toward Distributed Infrastructures for Digital Preservation: The Roles of Collaboration and Trust." *The International Journal of Digital Curation* 1, no. 3 (July 2008), http://www.ijdc.net/index.php/ijdc/article/view/60/39 (accessed January 6, 2010).

Deegan, Marilyn and Simon Tanner, "Key Issues in Digital Preservation." In *Digital Preservation*, edited by Marilyn Deegan and Simon Tanner. Digital Futures Series. London: Facet, 2006.

Edmondson, Ray. "Audiovisual Archiving: Philosophy and Principles," CI/2004/WS/2. Paris: UNESCO, 2004. http://unesdoc.unesco.org/images/0013/001364/136477e.pdf (accessed January 6, 2010).

Educational Broadcasting Corporation, "Building the House of Ideas: Educational Broadcasting Corporation Report to the Community 2006-2007," http://www.thirteen.org/home/pdfs/EBCAnnualReport_BHI.pdf (accessed January 6, 2010).

Gantz, John F., Christopher Chute, John McArthur, Alex Manfrediz, Stephen Minton, David Reinsel, Wolfgang Schlicting, and Anna Toncheva, Irida Xheneti. *The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010.* (Framingham, MA: IDC, 2007): http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf (Accessed January 6, 2010).

Gantz, John F., Christopher Chute, Alex Manfrediz, Stephen Minton, David Reinsel, Wolfgang Schlicting, and Anna Toncheva. *The Diverse and Exploding Universe: An Updated Forecast of Worldwide Information Growth Through 2011*. Framingham, MA: IDC, 2008. http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf (accessed January 6, 2010).

Getty Images, "gettyimages," Getty Images, http://www.gettyimages.com/ (accessed January 6, 2010).

Google. "Purchase Additional Storage." Google. https://www.google.com/accounts/PurchaseStorage (accessed January 6, 2010).

Guthrie, Kevin, Rebecca Griffiths, and Nancy Maron. "Sustainability and Revenue Models for Online Academic Resources: An Ithaka Report." N.p., Strategic Content Alliance/ Ithaka, 2008. http://www.jisc.ac.uk/media/documents/themes/eresources/sca_ithaka_sustainability_report-final.pdf (accessed January 6, 2010).

Harvard University Library Office for Information Systems. "DRS Policy Guide." President and Fellows of Harvard College. http://hul.harvard.edu/ois/systems/drs/policyGuide/wwhelp/wwhimpl/js/html/wwhelp.htm (accessed January 6, 2010).

HathiTrust, "HathiTrust: A Shared Digital Repository," HathiTrust, http://www.hathitrust.org (accessed January 6, 2010).

"It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation, Final Report," Workshop on Research Challenges in Digital Archiving and Long-Term Preservation, April 12-13, 2002," sponsored by the National Science Foundation, Digital Government Program and Digital Libraries Program, Directorate for Computing and Information Sciences and Engineering, and the Library of Congress, National Digital Information Infrastructure and Preservation Program, 2003. http://www.digitalpreservation.gov/library/resources/pubs/docs/about_time2003.pdf (accessed January 6, 2010).

Joint Information Systems Committee, "The Depot," HEFCE, on behalf of JISC, http://www.jisc.ac.uk/whatwedo/programmes/reppres/depot.aspx (accessed on January 6, 2010).

Kelly, Kevin. "Better than Free," The Technium blog, posted on January 31, 2008, http://www.kk.org/

thetechnium/archives/2008/01/better_than_fre.php (accessed January 6, 2010).

LaCie. "LaCie Biggest S2S: SATA II - 5-Disk RAID System Tower." LaCie. http://www.lacie.com/us/products/product.htm?pid=11211 (accessed on January 6, 2010).

Lavoie, Brian. “The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making.” Dublin, OH: OCLC, 2003.

Lavoie, Brian F. "Of Mice and Memory: Economically Sustainable Preservation for the Twenty-first Century," in *Access in the Future Tense*. Washington, DC: Council on Library and Information Resources, 2004. http://www.clir.org/pubs/reports/pub126/lavoie.html (accessed on January 6, 2010).

Lavoie, Brian F. "The Fifth Blackbird: Some Thoughts on Economically Sustainable Digital Preservation." *D-Lib Magazine* 14, no. 3/4 (2008), http://www.dlib.org/dlib/march08/lavoie/03lavoie.html (accessed January 6, 2010).

Lavoie, Brian and Lorcan Dempsey. "Thirteen Ways of Looking at... Digital Preservation." *D-Lib Magazine* 10, no. 7/8 (July/August 2004), http://www.dlib.org/dlib/july04/lavoie/07lavoie.html (accessed January 6, 2010).

Maron, Nancy L., K. Kirby Smith, and Matthew Loy. “Sustaining Digital Resources: An On-the-Ground View of Projects Today.” New York: Ithaka S+R, 2009), http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability/report/SCA_Ithaka_SustainingDigitalResources_Report.pdf (accessed January 6, 2010)

Moore, Richard L., Jim D’Aoust, Robert H. McDonald and David Minor. "Disk and Tape Storage Cost Models." San Diego: San Diego Supercomputer Center, University of California, 2007. http://users.sdsc.edu/~mcdonald/content/papers/dt_cost.pdf (accessed January 6, 2010).

Murray, Peter. "Long-term Preservation Storage: OCLC Digital Archive versus Amazon S3." Disruptive Library Technology Jester Blog, posted May 16, 2008, http://dltj.org/article/oclc-digital-archive-vs-amazon-s3/ (accessed January 6, 2010).

Palm, Jonas. “The Digital Black Hole.” N.p: Training for Audiovisual Preservation in Europe, 2008. http://www.tape-online.net/docs/Palm_Black_Hole.pdf (accessed January 6, 2010)

PrestoSpace. “Tutorial: Media Migration and Obsolescence in Mass Storage: Obsolescence Issues in IT Storage Media.” PrestoSpace: Digital Preservation of Audio/Visual Material. http://digitalpreservation.ssl.co.uk/storage/T2/T2-1-2.html (accessed January 6. 2010).

PrestoSpace. “Metadata and Cataloging.” PrestoSpace: Digital Preservation of Audio/Visual Material. http://digitalpreservation.ssl.co.uk/metadata/ (accessed on January 6, 2010).

Portico. "Portico." Ithaka Harbors, Inc., http://www.portico.org/index.html (accessed January 6, 2010).

Quint, Barbara. "OCLC Introduces High-Priced Digital Archive Service." Information Today, May 2, 2008, http://newsbreaks.infotoday.com/nbReader.asp?ArticleId=49018 (accessed January 6, 2010).

RLG-National Archives and Records Administration Digital Repository Certification Task Force. “Trustworthy Repositories Audit and Certification: Criteria and Checklist.” Chicago and Dublin, OH: Center for Research Libraries/ OCLC, 2007. http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf (accessed January 6, 2010).

Rosenthal, David S. H. "Bit Preservation: A Solved Problem?" Paper presented at iPRES2008, London, UK, September 2008. http://www.bl.uk/ipres2008/presentations_day2/43_Rosenthal.pdf (accessed on January 6, 2010).

Santayana, George. *The Life of Reason: or the Phases of Human Progress*. vol. 1, chapter 12 (1905).

SEO Economic Research. “Outline of Benefits: Results of 'Indicators For Cost-Benefit Analysis Images for the Future'.” Amsterdam: SEO, 2006. http://www.beeldenvoordetoekomst.nl/assets/documents/imagesforthefuture_outlineofbenefits_2006.pdf (accessed January 6, 2010).

Smith, Abby. "Issues in Sustainability." *First Monday* 8, no. 5 (2003), http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1055/975 (accessed January 6, 2010).

Smith, Abby. "Valuing Preservation." *Library Trends* 56, no. 1 (2007), http://muse.jhu.edu/journals/library_trends/v056/56.1smith.html (accessed January 6, 2010).

Thibodeau, Ken. "Preserving Records, Saving Appearances: Perspectives on Digital Preservation." in *Preserving Authentic Electronic Records: Preliminary Research Findings*. Proceedings from an International Symposium, February 17, 2001, University of British Columbia, edited by Luigi Sarno, 92-102. Vancouver: UBC/ Italian Government Cultural Office, 2001. http://www.interpares.org/display_file.cfm?doc=ip1_symposium_2001.pdf (accessed January 6, 2010).

U.S. Congress. Public Broadcasting Act of 1967, US Code Title 47, 1967, S. 396.

WGBH. “Challenging Times, Critical Impact: WGBH Annual Report 2008-2009.” Boston: WGBH, 2009. http://support.wgbh.org/site/PageNavigator/annualreport (accessed January 6, 2010).