FP7-ICT-2007-3-231161

Deliverable D6.3.1

Financial Models and Calculation Mechanisms

Netherlands Institute for Sound and Vision

31/01/2011

Document administrative table

Document Identifier	PP_WP6_D6.3.1_FM_calculation_R0_v1.01	Release	0
Filename	PP_WP6_D6.3.1_FM_calculation_R0_v1.01.pdf		
Workpackage and Task(s)	WP6 Audiovisual preservation Competence Centre Task 3 Economic Viability and Business Modelling		
Authors (company)	Marius Snyders, Hans Westerhof, Jeff Ubois (B&G)		
Contributors (company)			
Internal Reviewers (company)	Matthew Addis (IT Innovation)		
Date	31/01/2011		
Status	Release		
Type	Deliverable		
Deliverable Nature	R - Report		
Dissemination Level	PU – Public		
Planned Deliv. Date	31/08/2010		
Actual Deliv. Date	31/01/2011		

Abstract

This Deliverable describes and analyses cost models and reported data for digitisation and preservation. It provides a market analysis for use by archives and by vendors, and suggests what further research is needed to better understand the economics of digitisation, preservation, and archives.

DOCUMENT HISTORY

Version	Date	Reason of change	Status	Distribution
v0.01	09/2011	First Draft	Outline	Confidential
v0.5	3/11/2010	Final draft under review by M.J. Addis ,	Final Draft	Confidential
v1.00	23/01/2011	Final draft v1.00	Pre-release	Restricted
v1.01	31/01/2011	Published	Delivered	Public

Table of contents

Scope

This document is intended to aid in financial planning by archives, libraries, museums, and other custodial institutions, as well as funding agencies and vendors that are concerned with mass digitisation of AV materials. It summarizes and compares of number of cost frameworks, as well reports from various European projects.

It also offers some estimates about the market for digitisation services based on a mix of data from AV archives, financial reports from public companies, and other published sources. For vendors and service providers, the market size estimates are intended to help them understand the archival market, to adjust their product and service offerings accordingly, and to justify investments in product and service development.

This Deliverable overlaps to some extent with work done in the first year of the PrestoPRIME project, notably D2.1.1 Preservation Strategies, which included cost and business model work. The value of D6.3.1. is that it updates and in some areas extends this initial work – which ensures the PrestoPRIME partners and community have an up-to-date resource.

Though PrestoPRIME partners initially discussed splitting the deliverable into two separate documents, one on costs and budgets for archives, and another on market size, primarily for vendors, this was determined to be inappropriate. Both archives and service providers need to understand the same benchmark data - archivists so they can budget effectively, vendors, so they can compete effectively. While some publications focused on the needs of vendors must be a part of the future PrestoCentre's output, the understanding of archival operations and budgeting is useful for them as well.

The original Description of Works notes this Deliverable is focussed on ‘digitisation’ i.e. the act of creating digital content from existing analogue material, and that it will be of interest to those looking to do digitisation and those looking to develop new tools for digitisation. Digitisation was one of the main subject areas of PrestoSpace and much work was done on the costs and tools for digitisation. Therefore, the scope of D6.3.1 has evolved to be somewhat broader and include more on what happens to materials after they have become digital.

This deliverable now functions as both an update to PrestoPRIME’s knowledge base and pointers to information on cost and business models, and a gap analysis (see Section 7) on where further work needs to be done. As such, this report is also valuable input to the future work of PrestoCentre.

Executive summary

Archives in the early stages of mass digitization projects face a enormous number of uncertainties about likely costs. These uncertainties may be addressed through the application of general principles, rough techniques of estimation, comparison with similar projects, and the use of cost models, some of which have reached an impressive level of sophistication.

Each of these approaches represents a different lens on the problem, yet none alone is fully satisfactory. For institutions planning digitization and long term operations, it is most useful to apply multiple approaches and to progressively bound the various economic uncertainties.

The authors were most surprised at how hard it can be to be definitive, at how little useful data about digitization projects has been published, and at the complexity of some of the cost models.

But based on an exhaustive literature review, and on informal conversations with managers at more than twenty archival institutions and service providers, the authors selected the concepts, rules of thumb, comparative project data, and cost models that seemed most useful, and we summarize them here as follows.

Conceptual approaches to thinking about costs are grouped here under the headings sustainability; constants, and predictable changes in archive costs; time, risks and dependencies; trade-offs between cost, quality, and risk; and activity models, lifecycle costing, and OAIS.

The approaches to estimation offered here are based largely on findings of the PrestoSpace project. They provide a general sense of likely costs in a very direct way that is useful for project budgets.

Several meta-analyses and bibliographies of cost models relevant for digitization have been published in the last few years. After examining these, we selected five as most relevant for discussion here: The LIFE Project; Keeping Research Data Safe (Beagrie & Associates); The Danish Cost Model For Digital Preservation; Understanding The Costs Of Digitisation (Curtis+Cartwright); Cost Estimation Toolkit (NASA).

Most of these models require users to provide information, and the authors did an extensive search of relevant data. Some of the single data points gathered from digitization projects, vendors, surveys, and other publications. The extremely wide range in costs reported was somewhat surprising, but some of the ratios between different activities ("ingest is typically more than half the cost"; "disk is about three times as expensive as tape for long term storage") provided interesting ways of thinking about costs.

Perhaps most interesting is the discussion around costs for storage in perpetuity based on a multiple of a single year's total cost of ownership (see Section 5, Costs for Storage in Perpetuity).

The data gathered about cost estimation represents the bulk of the document, but the findings regarding the likely size of the market for services and products aimed at archives

and other custodial institutions is important for many of PrestoCentre's future members. Several studies, including those by PrestoSpace, FIAT and Screen Digest, Numeric, and the Primary Research Group, point to a market that is sizeable, yet fragmented.

Ultimately, this report uncovered a need for better data and further research. Suggestions for particular projects and methods compromise the final section, but to sum up, processes based at PrestoCentre may be a promising answer to the questions facing AV archives embarking on financial planning. These include systems for gathering more extensive and consistent cost data from AV archives, service providers, and funders; improved tools for cost modelling, as are being developed in Work Package 2; partnerships with other cost modelling projects such as Numeric, and commercial research and publishing organizations; and surveys of archives and service providers.

1 Introduction

1.1 Background

Over the last decade, dozens of different financial models for digitisation, digital preservation, and access have been developed by a academics, vendors, custodial institutions, funding agencies and foundations, and projects such as PrestoSpace.

But in the end, as noted by JISC in its 2009 report, Budgeting for a Digitisation Project http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/budgeting-for-a-digitisation-project/, "It is not possible to provide detailed information on budgeting for a digitisation project; these costs will depend on the nature, condition and amount of analogue materials to be digitised and the extent and accuracy of existing metadata. However, some general advice which assigns comparative values to different activities within a project may help when planning a digitisation project."

Rather than offer up yet another attempt at a final answer to the question of how budget for digitisation, this document serves as a guide among different approaches to budgeting and cost estimation. It examines critical trade-offs between price and quality, explains some key concepts used in most cost models, and addresses risks, placing them in the context of timeframes in which action can be taken. It also provides some benchmark data, and approaches to rough cost estimation, and describes how empirical cost data can be gathered by PrestoCentre.

Digitisation brings a variety of associated costs that are difficult to measure. For example, it is often bound up with new cataloguing efforts, or with new access infrastructure. While it is beyond the scope of this document, and the cost models discussed here, it's important to note that digitization brings organizational change. The skills associated with IT operations are different from those typical inside most archives. Even in broadcast archives known for engineering excellence, digitization brings new requirements in hiring and staffing.

It is also worth emphasizing that most archives are still in the very early stages of digitization. As a result, there is a paucity of actual data about costs, a general lack of understanding of cost models, and these uncertainties have to be factored into any budget forecast. As one manager in a very large American institution put it, "we won't really understand costs for another five years - today, it's really easy to spend $10 million before you even find out you're wrong about something."

As a result, the appropriate scope for this deliverable has shifted. While original intent was focused on models for digitization, archival operations much extend beyond that to long term preservation. Therefore, the discussion here has been expanded to cover costs beyond those associated with initial digitization. In another way however, the scope has been narrowed to focus much more on cost models than on business models. Discussions within the community of PrestoPRIME members revealed that there is still enormous uncertainty about how to predict future costs. Without a sense of those costs, the requirements for business models – which address sources of income – are impossible to determine.

2 Concepts

There are a number of concepts common to most discussions of financial models and calculation mechanisms for digital archives. Among the most important of these are sustainability; predictable declines in the cost of storage; project timing for at-risk material; cost/quality/risk trade-offs; and activity models such as OAIS. Familiarity with these concepts makes it possible to compare and interpret different approaches to budgeting.

2.1 Sustainability

Sustainability, the prospect of perpetual continuance, is an important concept for anyone thinking about the economics of archiving. Rather than framing digital preservation as a technical or a policy problem, the concept of sustainability in the archival domain centres on economics in a broad sense.

One approach to understanding sustainability is to create a framework in which all costs can be recorded and allocated or apportioned to specific activities. As a result, some investigations into sustainability have included approaches to estimating future costs, as well as benchmark data from particular projects (see especially the reports from the Blue Ribbon Task Force), and where appropriate we reference these.

Some of of the conceptual thinking from the Preserving Digital Public Television report, funded by the Library of Congress, is summarized in the graphic below.

Another approach to sustainability attempts to include less tangible costs and benefits.

One method of doing so is the balanced scorecard model, which uses both financial and non-financial measurements to provide a sense of the full value of a particular activity. Glasgow University used the balanced scorecard model in their eSPIDA project (http://www.gla.ac.uk/espida/), which helps organizations develop business cases for projects providing intangible benefits. eSPIDA is intended to improve dialogue between funders and project proposers by providing a common understanding of intangibles, the strategic needs of an organization, likely outcomes, and measurements.

Facing Page: A summary of conclusions about sustainability and business models from the Preserving Digital Public Television Project funded by the Library of Congress in the U.S.

MAJOR FINDINGS:

1. Sustainable preservation requires a **sound repository technological infrastructure** plus **creation and distribution practices that support long-term access.**
2. Sustainable preservation requires **ongoing, reliable, sufficient financial support**
3. Proper preservation will **add value to assets** over time
4. The benefits of preservation **will outweigh the costs** if properly managed
5. The costs of long term preservation can be **feasible and manageable** if seen as a long-term investment
6. There are a variety of **repository operating models** that **can maximize efficiency** and allow costs to be shared
7. The needs for preservation are **not in competition with program production**, but are complimentary in an all-digital workflow.

TECHNICAL, PROCEDURAL, AND ORGANIZATIONAL REQUIREMENTS OF A SUSTAINABLE PRESERVATION REPOSITORY (OAIS[1] + TRAC[2])

1 BIT PRESERVATION

requires a repository with...

System Infrastructure
- Replicated, geographically distributed storage
- Auditing of bits and reporting bit loss
- Repair corrupted bits using backups
- Ability to perform mass file migration

Appropriate Technologies
- Software and hardware to maintain access
- Appropriate bandwidth for large file delivery

Security
- Protection against security breeches
- Disaster recovery planning
- Appropriate permissions and access procedures

2 ONGOING ACCESS TO DIGITAL CONTENT

requires new production practices...

Selection
- Identifying appropriate content for long-term preservation
- Retention policies and procedures

Creating Video Content
- High resolution preservation master files
- Open, widely-supported formats and codecs
- No transcoding during production
- Information Life Cycle Management

Metadata
- Quality metadata captured at creation
- System-wide standards (i.e. PBCore)

Technical and Use Policies
- Submission and Dissemination Agreements
- Archival Information Package Creation
- Preservation Planning and Strategies

3 ORGANIZATIONAL INFRASTRUCTURE

will have appropriate...
- Governance and Organizational Viability
- Infrastructure and Staffing
- Procedural Accountability and Policy Framework
- Financial Sustainability and Operating Agreements
- Licenses and Liabilities

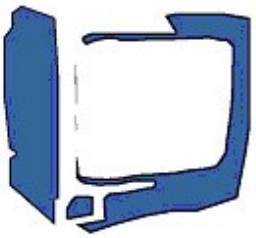
Why is US Public Television Unique?
- Not a centrally-controlled national network
- Over 350 stations produce and broadcast content
- Multiple national program distributors (PBS, APT)
- Varied governance and licensing
- Funding from multiple sources - a large percentage is from voluntary viewer contributions
- Not the dominant content provider in the US

For more information, please visit
http://www.thirteen.org/ptvdigitalarchive/

Contact us!
Nan Rubin, Project Director:
RubinN@thirteen.org
Kara Van Malssen, Senior Researcher:
kvm211@nyu.edu
Karen Cariani:
karen_cariani@wgbh.org

Strategies for Sustainable Preservation of Born Digital Public Television

A Report By: Preserving Digital Public Television, An NDIIPP Project
Kara Van Malssen, New York University
Karen Cariani, WGBH Media Library and Archives
With Nan Rubin (WNET), Yvonne Ng (Witness), Howard Besser (NYU), Irene Taylor (Washington University),Winter Shanck (WNET), Jeff Ubois
Thanks to Joe Pawletko (NYU), David Ackerman (NYU), B Morse (PBS), and Alicia Kubes (NYU)



BUSINESS MODELS TO SUPPORT ONGOING PRESERVATION

Community Model
Rights holders pool resources to financially sustain a repository to provide preservation services

Submitter Fee-for-Service
Rights holders pay a repository to provide preservation services on a fee-for-service basis

Corporate Sponsorship/Advertising
Commercial entity funds the repository directly, or indirectly supports preservation through ads

Public Sponsorship/Philanthropy
Public body or philanthropist funds the repository to provide services that further the public good

User Fee-for-Service
Users pay the rights holder(s) and/or the repository to access preserved digital content at the point of use

Endowment
An endowment is created to fund long-term operations

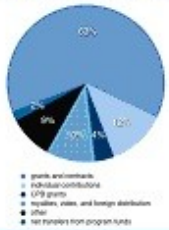
Membership/Subscription
Users and/or rights holders pay dues to the repository in exchange for ongoing access and/or preservation services

Joint Venture Between Rights Holder and Archive
Rights holder(s) and repository enter into a financial agreement in exchange for ongoing access and/or preservation services

Services/Consulting
Services and technical assistance are provided to outside organizations on a straight contract basis to produce revenue to support preservation

WGBH Funding Sources FY08[4]

Public television is experienced in diversifying revenue....



Many Repository Operating Scenarios
- Centralized, repository is part of public TV system
- Centralized, 3rd party repository
- Individual stations maintain their own repositories
- Disaggregated repository shared between public television entities
- Disaggregated repository with 3rd party services
- **Disaggregated repository shared among public television entities, 3rd party administration**

Sustainable projects, "*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).*"
- Ithaka S+R Case Studies in Sustainability, 2009[3]

1 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).
2 RLG-National Archives and Records Administration Digital Repository Certification Task Force. Trustworthy Repositories Audit and Certification: Criteria and Checklist (Chicago and Dublin: Center for Research Libraries/OCLC, 2007).
3 Nancy L. Maron, K. Kirby Smith and Matthew Loy, "Sustaining Digital Resources: An On-the-Ground View of Projects Today," (n.p.: Ithaka S+R, 2009) 21.
4 WGBH, Challenging Times, Critical Impact: WGBH Annual Report 2008-2009 (Boston: WGBH, 2009).

From *Preserving Digital Public Television* Project

2.2 Constants, and predictable changes in archive costs

As in all information technology environments, the costs related to digitisation, storage, and access change over time in fairly predictable ways according to a handful of enduring and persistent patterns that are well known, chief among them, Moore's Law.

Originally formulated in 1965, and applied only to the number of transistors that could be placed on a single chip, the more colloquial and familiar formulation of Moore's Law is that the cost for a given amount of CPU cycles, memory, and storage drop by 50 percent every 12-24 months. Costs for network services also drop, but at somewhat slower rate.

This is easy enough to understand, but when considered over the span of decades, the consequences are not intuitive, in fact, they are quite startling, and their implications are difficult to accept. For example, the Library of Congress conducted a study in the mid-1990s, Television and Video Preservation, which accurately identified many of the problems facing broadcast archives today. But it failed to account for the massive drop in storage costs that has happened in the last decade; the idea that millions of hours of video would be easily available online for free was simply not a part of the future imagined in that report.

Looking fifteen years ahead, Cisco's Dave Evans predicted in 2007 "By the year 2025 I'll be able to walk into my local electronics store and buy about 6.3 petabytes of storage for about $100. To give you a sense of the immensity of 6.3 petabytes of storage, consider this: if I had taken a high definition camera and I had set that camera on record at the beginning of the Industrial Revolution, which is about 1700 AD, and let that camera record 365 days a year, seven days a week, and 24 hours a day, I would still have years of storage left in that camera. And that's the amount of storage I'll be able to buy for $100 20 years from now." (see http://newsroom.cisco.com/dlls/2007/hd_060407.html)

By 2029, 11 petabytes of storage will be available for $100—equivalent to 600+ years of continuous, 24-hour-per-day, DVD-quality video (see: Cisco IBSG, 2009, http://www.cisco.com/web/about/ac79/docs/Top_25_Predictions_121409rev.pdf).

Faster processing and larger storage capacities not only allow larger mass digitization projects to be performed for a given budget, they also allow digitization to take place at a higher level of quality.

For archives, all this means that some of the cost / quality tradeoffs that are often so painful for those in the midst of digitization projects will ease over time. For example, film digitization at 4K was impractical 20 years ago, astronomically expensive a little less than ten years ago, and now a standard output from high end film scanners.

2.3 Time, Risks and Dependencies

As institutions, archives differ from most businesses, and therefore most accounting practices, in their very long term focus.

Given that many digitization projects are funded on a one time basis, this translation between annual and one time costs is critical, and at the heart of the cost models discussed throughout this document.

Unfortunately, translating other annual costs into one time payments is more difficult. Common to most cost models is a separation of annual costs from upfront costs.

It is possible to get a sense of the cost for storage in perpetuity for a given amount of data. Deliverable 2.1.1 goes into this in some detail, offering multiples of single year costs today as the cost for storage in perpetuity. In other words, the rapid drop in IT costs makes it possible to translate some annual costs today into one time payments.

Another persistent issue in archive economics is related to the higher costs paid by leading institutions. The early adopters of new technology pay a premium. This not only reflects the needs of early adopters to buy more capacity, but also some realities inherent in workflow engineering: once the bugs have been worked out of a workflow, it can be replicated in another institution for far less than the cost to the first mover.

That dynamic affects scheduling and planning. Perfecting a workflow design may take months or even years. In a five year project to digitize (say) 100,000 hours, it is all but certain that most of the hours will be digitised in the final two years. This is an important point for managers in archives to communicate when reporting on their progress, and it relates to costs because the cost per hour will drop year by year.

Urgency is another major aspect to project timing. How fast are materials degrading? How quickly is equipment necessary for migration becoming unavailable? These considerations impose project constraints that should be included in any project plan.

2.4 Tradeoffs between Cost, Quality, and Risk

PrestoSpace addressed the problem of how archives can best make use of limited resources, and pioneered what is now called the 'cost of quality' approach.

As noted in the PrestoSpace tutorial, Planning Your Preservation Project (http://prestospace-sam.ssl.co.uk/tutorials/T7/T7%2d1%2d1%2d1.html), “Preservation projects are invariably a balancing act between how much the preservation project will cost, what quality can be achieved, how long the project will take and what volumes of material can be preserved within that time…Making trade-offs between these factors [cost, quality, volume and time] is an engineering approach to preservation and aims to find the best compromise given the circumstances.”

Making responsible tradeoffs between cost and quality is difficult. A standard approach in large projects is to test output quality from differently configured digitisation processes, and from different service providers.

Cost / quality tradeoffs also occur in designing quality assurance processes. Sampling the output from large scale digitisation processes is cheaper than having human operators watch every hour of ouput, yet automated quality control mechanisms are not fully mature, especially for film and video.

Tradeoffs between cost and risk are equally problematic. While storing more copies in more places on a wider variety of carriers will certainly decrease the risk of loss, each additional copy means additional costs. Each additional process to ensure file integrity also adds additional costs.

These cost and risk tradeoffs are highly dependent on scale. For smaller archives, making many additional copies may be viable, but for national broadcast archives with petabytes of data, this is not the case.

2.5 Activity Models, Lifecycle Costing, and OAIS

Activity-based costing (ABC) is accounting model that associates costs with particular outcomes, products and services delivered to customers. Its primary benefit is that it can help reduce the unknowns associated with indirect costs and operational overhead. ABC can be combined with activity models specific to archives, such as Information Life-cycle Management (ILM) and Open Archival Information System (OAIS).

ILM is a conceptual scheme that identifies five phases of activity that can be used to analyse the flow of information through a system. These are Creation and Receipt, Distribution, Use, Maintenance, and Disposition.

OAIS is far more detailed, and defined in a reference model maintained by the International Standards Organisation, and developed by the Consultative Committee for Space Data Systems.

The purpose of OAIS is to “establish a system for archiving information, both digitalized and physical, with an organizational scheme composed of people who accept the responsibility to preserve information and make it available to a designated community.” (see http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683).

The reference model describes "archival information preservation functions including ingest, archival storage, data management, access, and dissemination. It also addresses the migration of digital information to new media and forms, the data models used to represent the information, the role of software in information preservation, and the exchange of digital information among archives. It identifies both internal and external interfaces to the archive functions, and it identifies a number of high-level services at these interfaces. It provides various illustrative examples and some "best practice" recommendations. It defines a minimal set of responsibilities for an archive to be called an OAIS, and it also defines a maximal archive to provide a broad set of useful terms and concepts."

OAIS is important to this discussion because it underlies many of the most advanced cost models described later in this document.

3 Making Estimates

Sometimes rough estimates are what is needed to jumpstart a funding proposal, or to make initial decisions about how to apply funds in hand.

Simpler models for cost estimation typically converge on particular measurements (cost per hour, cost cost per item, cost per terabyte), rules of thumb (so many hours to perform a task, such as cleaning a tape), and general frameworks that include most conceivable expenses (equipment and staff costs, for example).

Another approach is to rely on the rough ratios between different cost areas. For example, for mass digitization projects, ingest may account for over one half of total costs, while storage may account for roughly a quarter or less of total costs.

Such simple estimates can go very wrong however, because mass digitization projects commonly run into unexpected costs. For example, a small increase in the number of exceptions in the way a set of materials is being digitised – often due to surprising inconsistencies in material condition, metadata quality – can result in backlogs that are expensive to clear, and which put project schedules at risk. In one national archive, a 10 percent increase in such exceptions increased costs by 25 percent. Ensuring that operators have the training to process exceptions on the spot (perhaps by paying for more highly trained staff) can be more economical than running a process that generates many instances of material that require special handling.

3.1 Guidance from PrestoSpace

The PrestoSpace project documented a number of approaches to cost modelling and estimation, and for most AV archives, the methods outlined in the PrestoSpace deliverables are still among the most useful and relevant approaches to budgeting and planning.

At a somewhat more tactical level, PrestoSpace suggested that there are two approaches to cost modelling. The first is to predict it, the other is to run pilot studies using a representative sample of archival material.

To predict costs, PrestoSpace developed a number of tools and guidelines, all of which could be maintained by PrestoCentre.

For an initial estimate of overall costs, these tools are the most straightforward of any evaluated in this report, and they will become available on the PrestoCentre web site. These tools are also the most suitable of any evaluated in this document for all but the largest AV archives.

The PrestoSpace Storage Calculator (see http://prestospace-sam.ssl.co.uk/hosted/d12.2/calc4.php) helps archivists estimate how much data they will have once they have converted an analogue collection to digital files. Users can choose between shelf length of number of items, set data rates or output types and quality (e.g. VHS, DVD, DV, uncompressed, etc.).

The PrestoSpace Preservation Project Cost Calculator (see http://prestospace-sam.ssl.co.uk/hosted/d13.2/newcalc.php) provides rough estimate of costs for for film, audio, and video digitisation projects. It identifies different levels of material condition (good, difficult, unplayable), and multiplies the amounts of material by cost per hour. It further calculates storage requirements and costs.

Key to making best use of these tools is having accurate information about size and state of collections. Variations in material quality (i.e. the amount of preparation that will be required) and metadata quality (the amount of additional cataloguing that may be required) can have an enormous effect on final cost.

PrestoSpace also developed some other analyses that are useful for archivists engaged in budgeting.

Models and Protocols for PrestoSpace Factory Process (see http://www.prestospace.org/project/deliverables/D3.1_public.pdf) described the models adopted within the PrestoSpace project for the audiovisual Preservation, Restoration and Access workflows. The processes and diagrams here can be useful for modelling processes with outside vendors or for internal operations; by associating a cost with each step in the general workflows described in this document, archivists can get a sense of the full project cost they can expect.

PrestoSpace Deliverable D13.5, Service Level Agreements for Storage: Report and sample documents (see http://www.prestospace.org/project/deliverables/D13-5.pdf) can help archives negotiate with storage service providers. Using the information in this document, it is possible to negotiate more effectively for cloud and other service based approaches to storage. While the level of enthusiasm and hype regarding cloud storage has only increased since this document was published, the authors note "There are various reasons for using a service provider for storage, including cost, expertise, security. The general idea is to give the job to someone who is highly professional and experienced, and offers a good price – and let them worry about it. Life should then be simpler. As shown by the documents presented, it isn't all that much simpler. Using a service provider has a technology and complexity all its own."

PrestoSpace also developed more detailed approaches to costing that account for the cost of risk when digitising materials, as well as cost/quality trade offs.

For those writing funding proposals, PrestoSpace also developed a document called The Case for Investment in Digital Archives (see http://prestospace-sam.ssl.co.uk/asset_arena/text/es/PS_D12.7_Investment_in_Digital_Archives.pdf). This is intended as "a resource for those who are making a policy case for digital archives, for digitisation programmes or for projects designed to exploit the corresponding digital content." The document makes "the case for more investment in these areas, covering the nature of the case to be made, the concepts of market failure, contingent valuation, public value and the internal business case. Summaries are given of the main policy developments in the area at an EU level and of some prominent case studies."

3.2 Guidance from Other PrestoPRIME Deliverables

Other PrestoPRIME deliverables directly address the issue of costs.

In particular, the cost calculators now under development and set for deployment on the future PrestoCentre web site, and the cost modelling discussions in Deliverable 2.1.2, Tools For Modelling And Simulating The Use Of Resources By Migration-Based Preservation, are directly relevant. Other discussions of costs occur in Deliverable 2.1.1 (on file format migration strategies and cost models) and in the risks from using IT systems to store and manage files (ID3.2.1).

The new PrestoPRIME cost modelling tools are more advanced than previous budgeting tools because they help users relate strategic and operational choices to their budgets. D2.1.1. summarized the functions of the tools as follows:

- The long-term planning tool is designed to support decision making on what storage strategy to use, for example how many copies to make of files in archive, what storage technologies to use to hold them, and what measures to take to maximise the long-term integrity of these files.

- The interactive simulation tool is designed to support more operational levels of decision making, e.g. how to allocate resources to tasks such as ingest, access and maintenance, and how to react to unforeseen events, e.g. failures in storage systems, peaks in load.

For a more in-depth discussion of these tools, please see D.2.1.1.

4 Detailed Cost Models

Over the last decade, dozens of different cost models for digitisation, digital preservation, and access have been developed by academics, vendors, custodial institutions, funding agencies and foundations, and projects such as PrestoSpace.

These models vary widely in their purpose, focus, completeness, flexibility, level of abstraction, time horizon, amount empirical data included, and other factors. Some have had significant investments -- more than 500,000 Euros has been invested into the Life Project, close to 150,000 Euros has gone to Keeping Research Data Safe, and NASA has supported two full time people for seven years to develop its CET model.

Some have been developed into spreadsheets and even software applications that aid in the estimation of storage requirements, and ultimately, long term costs. These models have also evolved significantly over time, and several comparisons between them [Mageto, Zeller] have been published.

A few have attempted to take an empirical approach, relying the cost data from particular projects, but that data is sparse: even with many mass digitisation projects under way in the EU, the reality is that most digitization projects are so heterogeneous that making generalizations from them about costs is impractical. As a result, building a comprehensive model for AV digitisation based on lots of data hasn't yet been done. Gathering real data about costs, responses to tender offers, and actual budget data from member institutions will become a key activity for PrestoCentre. This effort is still in its early phases however; some of budget data collected so far is discussed in Section 5.

Two approaches to categorizing costs run through most of the models discussed in this section. One approach is grounded in traditional accounting practices, and breaks archival costs down into capital costs (equipment and infrastructure) and operating costs (staffing, services, utilities, etc.). The other approach is based on an activity model, often OAIS. This approach may be used within the traditional accounting model to analyse operating costs, or to address operating costs inside a traditional accounting model.

As of 2010, several of the most sophisticated of models [Beagrie, Curtis+Cartwright] have converged on the OAIS model as a way of distinguishing between different phases of archival operations, and segmenting projects into smaller pieces that are can be more accurately estimated. Other long term models have been developed as a result of studies in sustainability [BRTF, PDPTV], or for specific parts of the archival process, such as migration.

Rather than duplicate this work, this section focuses on making the work that has been done slightly more intelligible and useful for future PrestoCentre members and users. The different models described below were developed to meet different needs. Factors such as project scale, type of funding, type of media, staffing, and relative project maturity can all influence the choice of model, and of course, many archives may want to combine elements from different models.

The formal cost models discussed below require considerable effort to implement. For managers of large scale projects, it may be appropriate to combine approaches found in different models.

4.1 Life Project

Developed in the U.K. and funded by JISC, the LIFE Project has developed a general model based on a lifecycle approach to costing. More than 500,000 GBP have been spent in development of the LIFE costing models, and as a result, they are among the most detailed and complete available.

The LIFE model assigns costs to six main categories, each with a more detailed set of elements, and concludes the full cost for a digital collection is represented by the equation:

LT=Aq+IT+MT+AcT+ST+PT L is the complete lifecycle cost over time 0 to T,

where Aq is acquisition, I is ingest, M is metadata, Ac is access, S is storage, and P is preservation. Each of these components is further broken down into subcategories. This is presented in a somewhat alternative form in the figure below.

Creation or Purchase	Acquisition	Ingest	Bit-stream Preserv.	Content Preserv.	Access
Creation	Selection	Quality Assurance	Repository Admin.	Preserv. Watch	Access Provision
	Submission Agreement	Metadata	Storage Provision	Preserv. Planning	Access Control
	IPR & Licensing	Deposit	Refresh	Preserv. Action	User Support
	Ordering & Invoicing	Holdings Update	Backup	Re-ingest	
	Obtaining	Reference Linking	Inspection	Disposal	
	Check-in				

From *LIFE* Project

The LIFE project expanded the portion of its “content preservation” model to build the LIFE Generic Preservation Model (GPM)

The LIFE project has also undertaken very broad surveys of the state of the art in cost modelling, and has applied its model in a number of settings. These uses of the model have been written up into case studies, which are mainly focused on the preservation of books and journals, are none the less useful for archivists in the AV domain.

4.2 Keeping Research Data Safe (Beagrie & Associates)

These studies, funded with more than 100,000 GBP from JISC, are focused primarily on research data, but they address the same set of needs, and is grounded in discussions with practitioners. They are economically sophisticated, and aimed mainly at large institutions. Though neither KRDS1 or KRDS2 focuses specifically on the needs of media archives, these models are abstract enough to offer helpful guidance for AV archivists.

The first study, KRDS1 (see http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf), provides a cost model for preserving research data. It identifies significant variables, such level of description, access models, and levels of user support. It accounts for issues such as the "first mover" cost (the institution that develops a new approach will pay more than institutions that later adopt that approach). It also provides an activity model for archiving based on LIFE, NASA CET, OAIS, and TRAC, and helps planners associate costs with each activity.

The second study, KRDS2 (see http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx), applies data from actual projects to KRDS2. The data provided validates the model developed in KRDS1, and can be used by project planners to estimate their own costs. It offers a simplified as well as a complex version of the activity model, and direct access to survey responses from a number of large European data preservation projects.

4.3 The Danish Cost Model for Digital Preservation (CMDP)

Funded by the Danish Ministry of Culture, and created by the Royal Danish Library and the Danish National Archives, the Cost Model for Digital Preservation (CMDP) (see http://www.costmodelfordigitalpreservation.dk/), is intended to offer "a generic model for the calculation of costs related to preservation of digital materials at ALM institutions."

Based on the OAIS standard, the CMDP takes an activity based approach that breaks down functions described in the OAIS model into components that can be estimated and related to each other.

As part of the project, researchers did a very extensive investigation into the LIFE models and their effectiveness. In addition, the model has been tested against data from actual projects, so users can get a sense of CMDP's likely accuracy.

The project is currently focusing on creating a more detailed model of costs related to ingest, and may be extended to look at storage costs as well.

4.4 Understanding the Costs of Digitisation (Curtis+Cartwright)

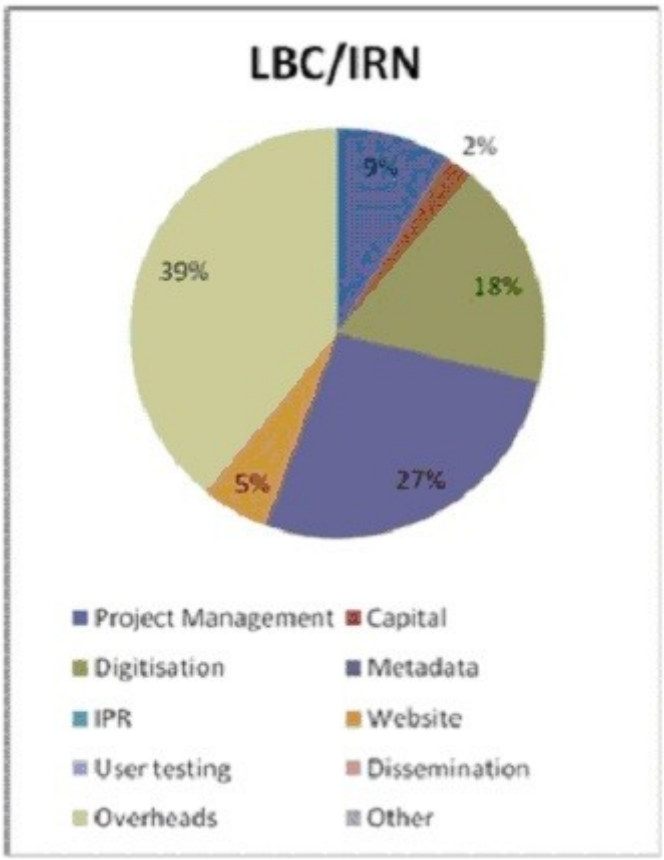
Developed on behalf of the Joint information Systems Committee (JISC) by the UK organization Curtis + Cartwright Consulting Ltd, Understanding the costs of digitisation, detail report (27 October, 2009) by Max Hammond and Claire Davies asserts that "It is not possible to provide a formula (or even approximate figures) to cost a generic project."

Rather than using the OAIS model to organize a model budget, Curtis + Cartwright took a case based approach, interviewing the managers of a number of projects, and then organizing findings around tasks, such as planning, content selection, procurement and subcontracting, project management, quality assurance, metadata, and rights management. The report identifies management issues that are particularly difficult, such as recruiting qualified staff, especially given the limited term of most projects; rights clearance; preparation for ingest; and service delivery.

The discussion of decisions and options at each stage in the digitisation process is then discussed in terms of likely budgetary implications. For example, how much conservation will be required prior to digitisation, how heterogeneous are the materials in the collection, how well catalogued the collection is, what tendering process must be used, clearance requirements, satisfying partners and stakeholders, and web design.

The case studies section shows the rough allocation between different phases of the projects studied, ie the percentage of costs in each project devoted to project management, digitisation, metadata, web site design, and so on.

For example, Curtis + Cartwright analyse an archive of radio recordings at Bournemouth University, and report the costs are apportioned as shown in the graphic below.



From *Curtis + Cartwright*

Overall, while this report does not provide much in the way of exact amounts, it does provide a solid discussion of the issues facing digitisation managers, and should be used as a check-list against digitisation budgets.

4.5 Cost Estimation Toolkit (NASA)

Developed to support multi-year budgeting and planning for complete lifecycle costs related to large scientific data sets, the U.S. National Aeronautics and Space Administration's (NASA) Cost Estimation Toolkit (CET) (see http://opensource.gsfc.nasa.gov/projects/CET/) is very applicable to the needs of AV archives.

The CET has been under development since 2002; as a result, it is one of the most evolved models available. It is based on a large collection of projects that allow cost estimation based on known costs. As an aside, it's worth noting that NASA was deeply involved in the creation of the OAIS model; the CET is another reflection of NASA's interest in data archiving.

The toolkit breaks costs down into functional areas of activity (see figure TK) (this is not OAIS?), and recognizes staff and non staff costs. It includes an Excel spreadsheet listing categories and personnel, and extensive documentation.

The output from CET shows estimated costs year by year in table and graphical form.

For AV archives, its weakness is in it lack of data specific to digitisation, and its strength is in data management, and its exhaustive list of activities to plan and decisions to take.

Ingest	Receipt and capture of incoming data, verifying data integrity, possibly including a format conversion.
Processing	Generation of derived products from input data, either operationally (scheduled or on-demand) or on an ad hoc basis.
Documentation	Ensuring documentation to standards of products generated or received, cataloged, archived, and distributed.
Archive	Storage and preservation of data and product holdings, ranging from long term to temporary working storage.
Access and Distribution	Distributing data to users by network or media, perhaps including fully automated system-to-system access.
User Support	Providing assistance to users in understanding data, products, services, etc., in selecting and using data and products.

The functional areas covered by NASA's *Cost Estimation Toolkit (CET)* .

5 Benchmark Data: Prices and Reported Project Costs

The greatest weakness of all the cost models discussed in Section 4 above is the lack of meaningful cost data. Therefore, one of the planned outputs from PrestoCentre will be cost information that will allow member archives, funding agencies, and others to budget more effectively.

The costs for different digitisation services and systems vary widely. In part, this reflects the nature of the field, which is still characterized by one-off projects and artisanal approaches to digitisation. While mass digitisation has truly arrived for books and documents, the same is not the case for audio, video, and film. Though certain IT costs, such as storage, might seem easier to arrive it, there are in fact considerable differences between large scale systems in their performance.

There have been a few attempts to systematically gather cost data. In that regard, one of the most helpful sources of information is the Numeric study (see Section 5.5). Numeric found extreme variation between expected costs and reported costs. Expected costs ranged from 0.60 to 150 Euros per hour for sound, and from €5 to €3113 per hour for film and video. Median reported costs varied significantly by institution, from €6.42 per hour for audio in higher education libraries, to €58.42 at national libraries, and €78.84 at broadcasting institutes. Videotape digitization ranged from €5.46 per hour for audio in higher education libraries, to €100 at broadcasting institutes, and €120 at special or other libraries. Film ranged from €8 per hour for audio in higher education libraries, to €678 at broadcasting institutes, and €1040 special or other libraries.

5.1 PrestoCentre and Cost Monitoring

For archives, libraries, museums, funding agencies and others, PrestoCentre could work on cost monitoring to provide access to “reality based” cost information, drawn from tenders, project reports, informal discussions with PrestoCentre’s future members, and partnerships with other collectors of cost information.

To date, this data collection activity has been largely informal, with a mix of reports and tender offers provided by member archives. It is planned to become available on the PrestoCentre site, and accrete there over time.

To accelerate collection of this data, PrestoCentre could also engage in active discussions with several nationally- and EU-funded projects that gather cost information, including Numeric. PrestoCentre could be pursuing cost data sharing arrangements with commercial publishers.

5.2 Other Reported Costs for Digitisation

In the meantime, it’s helpful to examine some of the high and low figures for digitisation, preservation, storage, and access reported in other surveys, and in other projects.

5.2.1 Audio

The price for mass digitisation of audio has dropped considerably over the last ten years. There has also been a move from price per minute or price per hour to price per reel. One recent large project was contracted out to a service provider for 7 Euros per two hour DAT tape – fair below the 60 Euros per hour paid by one PrestoPrime member in the early part of the 2000s.

5.2.2 Videotape

Video tape digitisation is typically priced by the hour and by the tape. Recent prices quoted in discussions on the Association of Moving Image Archivists list range from as little as $15 per hour for VHS video tape to up to $200 per hour for 2 inch Ampex tape.

5.2.3 Film

As noted in the Numeric study, prices for film digitisation vary widely, as do the types of film being digitised, and the final output format.

One of the larger recent tenders for film scanning at Beeld en Geluid allowed up to 500 Euros per hour for material that was in need of additional cataloguing and restoration.

Others have reported figures in the 100 Euros per hour range for HD and 2K quality scans of 16mm news footage.

5.3 Costs for Storage

Storage costs have fallen continuously for the last 60 years to the point where the price for a terabyte of disk storage has dropped well under 100 Euros.

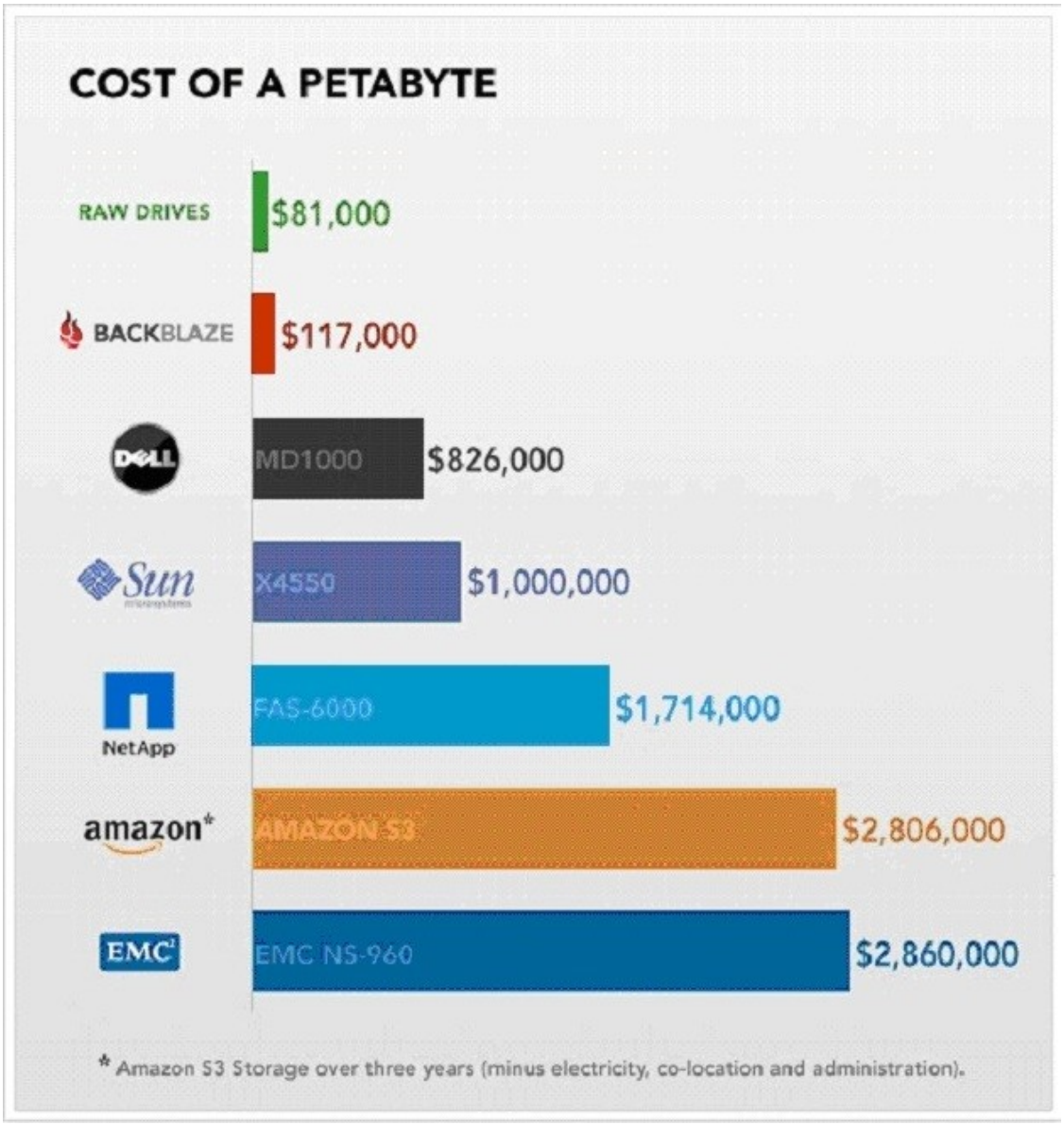
Unfortunately, the requirements of AV archives make the cost calculation for a petabyte more complex than simply multiplying 100 Euros by the number of terabytes needed (thought at least one project, described below, attempts to do just that). Redundancy, performance, and storage management requirements, essential for large scale operations, drive the costs up by an order of magnitude or more, and at petabyte scale, the operational costs of power become significant, as do the drive failure rates.

5.3.1 Costs for disk storage

To set a lower bound for storage costs on disk, it is useful to look at some published reports on that problem.

BackBlaze is a California-based online back service that has documented an approach to storage on hard disks that is intended to cut costs to the barest minimum. (see http://blog.backblaze.com/wp-content/uploads/2009/08/cost-of-a-petabyte-chart.jpg).

The low costs reflected in the BackBlaze numbers are not realistic for PrestoPrime members, and the figures shown for the commercial offerings are more in line with what large AV archives can expect to pay as of 2010.



From *BackBlaze*

5.3.2 Tape versus Disk (SDSC)

Researchers at the San Diego Supercomputer Center, University of California San Diego; La Jolla, CA, USA have done extensive work on overall costs for mass storage, and on the economics of tape compared to disk storage.

In Disk and Tape Storage Cost Models (2007), Richard L. Moore, Jim D'Aoust, Robert H. McDonald and David Minor state that "current estimates of the total 'bit preservation' cost of storage are ~$1500/TB/yr for SATA disk and $500/TB/yr for enterprise-class tape archives; thus the current difference between tape and disk costs is a factor of about three."

Media cost is a relatively small fraction of the total cost of ownership. They report that "the raw media cost (spindles/arrays/controllers for disk, cartridges for archives) is only a modest percentage of the total cost of delivering sustainable storage - ~36% for disk and ~20% for tape."

The researchers add that “the differential between the cost of delivering disk and tape storage is likely to diminish in the foreseeable future,” and discuss the importance of usage models: systems designed for frequent access are more expensive. This is shown in costs for services: “For example, Amazon S3 offers storage for ~$1850/TB/yr with a transmission (access) charge of $205/TB. For 'write-once- read-rarely' storage, this is cost-effective storage; for data which are frequently accessed the cost can become quite high (e.g. $4200/TB/year for once/month access).”

5.3.3 Costs for Storage in Perpetuity

PrestoPrime Deliverable D2.1.1 Audiovisual Preservation Strategies, Data Models and Value-chains, provides an extensive discussion of storage costs. For purposes of rough estimation involving large volumes of storage, the D2.1.1 suggests archivists:

1. Take today’s cost for commodity storage media (SATA hard disk drives or LTO4 tapes) and multiply by 10 to get the total annual cost of storage using that media.
2. Repeat step one if you want multiple copies, e.g. disk and tape.
3. Multiply the result of step two by 4 to give the TCO for a lifetime of storage.

Given different assumptions about the rate of cost declines, D2.1.1 provides the following table to allow archivists to estimate the cost for perpetual storage based on a multiple of current one year costs:

Number of years for annual cost to halve	Total cost as a multiplier of year 1 cost
1	2
2	3.3
3	4.9
5	7.7
10	14.9

From *PrestoPrime Deliverable D2.1.1*

A similar approach to developing what the authors call “Pay Once, Store Forever” and “Write Once, Read Forever” models was developed in 2010 by Serge J. Goldstein Mark Ratliff at Princeton University. In this paper, the authors identify three variables that can be used to calculate total storage costs. These are

- C = the initial cost of the physical storage required to preserve a file.
- D = the rate (as a fraction) at which the cost of storage decreases, on a yearly basis.
- R = the average number of years that elapse before the storage device must be replaced.

The total cost of storage, forever (T), is then calculated as T=C x 1/1-(1-D)^R.

Perhaps more remarkably (and usefully), the authors suggest the total cost of storage in perpetuity is now US$6/GB.

6 Estimating market size

6.1 Introduction

This section is intended to help vendors understand the overall shape and size of the market, and the typical concerns of AV archives.

Spending by European AV archives is difficult to track: funding and budget information is difficult to gather, services to archives are of very different kinds so markets are disconnected, and preservation budgets may be buried within broadcasters, libraries and universities.

In addition, the market itself is not always distinct. For example, storage systems built for corporate IT may find their way into AV archives. This isn't necessarily bad: it means that archives can benefit from the competition in larger markets, and from development costs underwritten by larger organizations.

That said, there are several ways to obtain rough estimates of the AV archives market, and several studies that have gathered significant amounts of market data. We outline these in the following sections.

One approach to estimation is to run rough calculations based on data such as the number of hours of material held by national archives. So for example, given 100mm hours of material on hand (a number taken from PrestoSpace), and a typical rate of 100 Euros per hour for digitisation, the upper limit for digitisation services is roughly 10 billion Euros.

Another approach is to look at production figures. The study How Much Information (Lyman 2003) estimates 68 petabytes of television broadcasts were created in 2002, and reported that "World radio stations produce 320 million hours of radio broadcasting, which would require 16,000 terabytes to store; we estimate 70 million hours are original programming, which would require an annual storage requirement of about 3,500 terabytes. World television stations produce about 123 million hours total programming; we estimate about 31 million hours are original programming, requiring about 70,000 terabytes of storage."

Yet another approach is to distinguish between hardware vendors, software vendors, and service providers, and to then look at very specific problems within the AV domain. For example, repairs and replacement parts for old equipment is a small market, but arguably underserved.

Other rough estimates may be done by looking at funding programs and events, and how much money is going from public institutions towards digitisation programs. For example, JISC maintains a list of funding organizations concerned with digitisation (http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/potential-sources-of-funding-for-digitisation-projects). Of course, this does not include the market for commercial broadcasters and archives.

While these approaches may allow one to arrive at a single large number for single large market, the reality is that the market is extremely fragmented. It's simply not that useful to lump film restoration services, enterprise disks storage manufacturers, custom software

developers, and Internet Service Providers into a single category. To address this problem, the PrestoSpace documents referenced below, broke the market up according to phases of archival operations.

6.2 PrestoSpace findings

The PrestoSpace project conducted surveys of film archives, broadcast archives, and service providers in an attempt to understand the market for services and systems for archives. PrestoSpace also compared its results with some user survey data from the TAPE project. Much of the survey data is from before 2005, but for vendors selling into archives, the framing of the questions, and some of the priorities expressed useful.

One of the more interesting findings from this is that outsourcing a variety of services is a critical component of future plans at most archives, but there are some points of resistance. Deliverable 2.1 User Requirements Final Report noted that "71% of broadcast archives and all the film archives are willing to outsource part of their collection migration or subcontract part of their planned migration. Reasons why archives choose not to outsource are varied, but include: to save money; to have more control over data file and metadata management; belief they either have the expertise or want to develop it in-house."

Also interesting was the willingness to work with remote service providers. D20.1 SET1 : Survey on Target Users noted "It was almost evenly split between respondents willing or unwilling to use a service provider outside of their national borders."

Another point from Deliverable 2.1 User Requirements Final Report is that the range of services desired by archives is broader than the range offered by most service providers. " It is interesting to note that two of the three “must have” services (bulk preservation and cleaning/physical repair) and the top 5 “nice to have” services are currently offered by only half or less of the service providers. Service Providers however do express a willingness to offer the services in the future."

6.3 Screen Digest, FIAT/IFTA, and FOCAL

In 2010, Screen Digest, FIAT/IFTA, and Focal released The Global Trade in Audio-visual Archives, which they describe as "the first report to assess the business of archive content and provides the first estimates on the global volume and value of trade in archive content."

The report estimates there are 43m hours of content held in the world's archives, and that "audio-visual archive content generates Euro 364m in revenue a year and has grown in value at a compound annual growth rate of nearly eight per cent over the past five years."

The report notes several barriers to increased sales of archival material, noting that "Finding and accessing that content can be even more problematic, with issues surrounding digital asset management, metadata and customer access portals all at the fore."

Of the 43 million hours on hand, 21 percent is cleared, though 87% of archive content has been catalogued and 61% has been made available on-line. Over the last five years, archival holdings have grown by 45 percent, bu the value per hour has fallen by 15 percent to about 10 Euros per hour.

News footage accounts for the bulk of revenue, followed by stock footage, with 55 percent of sales coming from television program producers, followed by corporate users, advertisers, educators, and movie makers.

The report also states that "72% of content held is on tape, of which 40% is digital tape. Just 20 per cent is held on film with a further 10 per cent on another form of HD source." This suggests there are tens of millions of hours of non-digitised material on hand, but it is unclear what the market for digitisation services might be.

More about the report is available at http://www.screendigest.com/reports/201074c/10_08_the_global_trade_in_audio_visual_archives/view.html.

6.4 Primary Research Group

The International Survey of Library & Museum Digitisation Projects (see http://www.primaryresearch.com/release-200811052.html) is based on responses from over one hundred digitisation programs at libraries and museums in the U.K., Germany, Canada, Australia, and the U.S.

Though it is not specific to AV, the overall patterns it reports are suggestive. Nearly half of these projects rely on outsourcing for some portion of their digitisation projects, but less than ten percent had outsourced a project entirely. Just over half (52%) had developed some form of digital asset management software in house, and most reported significant investment in cataloguing: “The mean percentage of labor time required for digitisation projects that is accounted for by cataloging and metadata tasks is about 37%, with a range of zero to 85%.”

6.5 Numeric

The responses from the Numeric study http://www.numeric.ws/ can be used to limn the overall size of the European market.

Numeric collected survey responses from 788 of 5752 European cultural institutions identified as relevant from a much larger pool of potential stakeholders. While it is unclear in economic terms what fraction of the total spending those 788 institutions represent, Numeric reported that “Based upon exchange rates in September 2008, the unadjusted total value of the digitisation budgets for responders was €80 million, representing 1.1% of the total reported institutional budgets (€7 billion).”

It also found a significant fraction of work was undertaken by outside contractors: "The majority of digitisation work (63%) appears to be undertaken in-house, by the institutions’ own staff. A further 6% of the work was completed by lead partners working on behalf of a group of other institutions; notably among archives, film institutes and national libraries.

External specialist contractors appear to be handling the significant remaining volume (31%)."

In addition, Numeric found that the vast proportion of holdings in European memory institutions is yet to be digitized – on the order of 87 percent for 17 million hours of AV materials in responding institutions (see Table 17 in Deliverable 8, Study Report). And while many institutions do not see digitization of their collections as critical (this was especially true in the library sector), nearly ¼ of respondents do.

The organizers of the Numeric study suggest it will evolve into a time series. As that happens, a detailed picture of digitization markets should emerge over the next years.

7 Conclusion and Recommendations for Future Work

Given the rapid changes now happening in digestion projects, this Deliverable opens a number of areas for future work by both PrestoPRIME and PrestoCentre.

Clearly, there is a need for better cost data, not only to improve the complex cost models, but to help funders, and small and medium sized archives and digitisation projects create rough estimates. As noted by several informants, there is almost complete lack of consistent data about costs for digitization and operations. A systematic approach to collecting this data on an ongoing basis – possibly by PrestoCentre - would have tremendous value.

This data might also be collected through a survey of PrestoPRIME partners on existing costs and business models for digitisation, access, storage, and preservation.

The abundance of complex cost models stands in contrast to the relative paucity of meaningful data based on the operating experience of AV archives. Yet most of the cost models developed so far have problems; they may require assumptions that are difficult to justify, or data that archives don't yet have They may involve a level of complexity that makes them unwieldy for any sort of practical budgeting, or in the case of models developed for research data or more generic archival settings, lack understanding of the special requirements associated with AV. The shortcomings in the existing cost models make it difficult for archives to begin thinking clearly about business models.

There is also a need to account for the wide variability in cost data that has been reported, and to develop clearer conversions between annualized costs and costs in perpetuity.

The unmet need for better cost modelling is being addressed in other PrestoPRIME deliverables. Based on the assessment here, and in informal conversations, it appears the new cost calculators maybe be the most useful tools for AV archives.

Not only archival institutions need data on real costs. The relationship between archive budgets, and markets for archival products and services may also be usefully explored. The Numeric project is doing this, but insight into the market that AV archives represent is still lacking. Vendors need real numbers on how much of a market there could be for new products and services, whether the market is ready for these products/services and how much archives might be willing to pay for them.
Projects to gather data of interest to vendors, and to incent them to build new products and services, were carried out under PrestoSpace and TAPE (see http://www.tape-online.net/docs/tracking_the_reel_world.pdf), but the data from those projects is now several years old.

Building relationships between PrestoCentre and projects such as Numeric, and with commercial organizations such as Screen Digest, could be an excellent start on such data gathering.

Similarly, there is a need for continuous monitoring of emerging business and operational models for digital archives. This would allow convergence on best practices, make it possible for custodial institutions to negotiate new types of collaborative agreements, and allow PrestoCentre future members to learn from the successes – and mistakes – of other

institutions. In particular, greater attention to outsourced or federated digital preservation or archive hosting – in particular role of so called ‘cloud providers’ in addition to incumbent service providers (e.g. facilities houses) - would be useful.

Finally, a survey of Service Providers to find out what they would like to know from archives to inform their product roadmaps, and where there are currently big areas of uncertainty or barriers to take-up that prevent the market from being accessed (e.g. trust in use of cloud storage / remote content hosting), would be useful. A similar project was a part of PrestoSpace, but it has not been updated since.

References

Ascent Media (2010). FORM 10-K - March 12, 2010. http://www.faqs.org/sec-filings/100312/Ascent-Media-CORP_10-K/#102

Ascent Media (2010). Online Content and Media Content Marketplace - An Evolution. http://www.ascentmedia.com/buzz/industry-insight/Online-Media-Content-Marketplace.aspx. Accessed September 6, 2010.

Backblaze (2009). Cost of PB graphic http://blog.backblaze.com/wp-content/uploads/2009/08/cost-of-a-petabyte-chart.jpg. Accessed August 24, 2010.

Beagrie, Chruszcz, and Lavoie (2008). Keeping Research Data Safe: A Cost Model and Guidance for UK Universities. London: JISC.
- http://www.jisc.ac.uk/publications/reports/2010/keepingresearchdatasafe2.aspx Accessed August 24, 2010.
- http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf Accessed August 25, 2010.

Blue Ribbon Task Force (2008). Interim Report. http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf
A Selective Literature Review on Digital Preservation Sustainability.
http://brtf.sdsc.edu/biblio/Cost_Literature_Review.pdf

Curtis+Cartwright (2009). Understanding the Costs of Digitisation.
http://www.jisc.ac.uk/media/documents/programmes/digitisation/digitisation-costs-full.pdf accessed September 6, 2010.

ELECTRONIC RESOURCE PRESERVATION AND ACCESS NETWORK (2003).
http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf [simple]

Goldstein, S.J., Ratliff, M. (2010). DataSpace: A Funding and Operational Model for Long-Term Preservation and Sharing of Research Data. Office of Information Technology, Princeton University.
http://dspace.princeton.edu/jspui/bitstream/88435/dsp01w6634361k/1/DataSpaceFundingModel_20100827.pdf accessed December 20, 2010.

IBM (2008). Manage exponential information growth and costs with archiving and retention solutions from IBM ftp://ftp.software.ibm.com/common/ssi/pm/br/n/tsb03002usen/TSB03002USEN.PDF

JISC (2009). Budgeting for a Digitisation Project.
http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/budgeting-for-a-digitisation-project/

Kejser, Ulla Bøgvad; Nielsen, Anders Bo; Thirifays, Alex (2009). Cost Model for Digital Curation: Cost of Digital Migration. http://escholarship.org/uc/item/4d09c0bb. Accessed August 23, 2010.

LIFE (Life Cycle Information for E-Literature). http://www.life.ac.uk/. Accessed September 3, 2010. (see also Ayris, P., Mcleod, R., & Wheatley, P. (2006). Lifecycle Information for E-literature: Full Report from the LIFE Project. London, U.K.: University College London and the British Library and LIFE: costing the digital preservation lifecycle (2007). Wheatley, Paul et al. http://eprints.ucl.ac.uk/4914/1/4914.pdf)

Lyman, Peter and Hal R. Varian, "How Much Information", 2003. http://www.sims.berkeley.edu/how-much-info-2003 accessed September 11, 2010.

Mageto, Doreen Kerubo (2009). Cost Factors in Digital Preservation. http://www.longrec.com/Intranet/ResearchResults/RecommendedPractices/CostFactorsForDigitalPreservation.pdf.

Moore, R. L.; D'Aoust, J.; McDonald, R. H.; and Minor, D. (2007). Disk and tape storage cost models. Archiving 2007. http://chronopolis.sdsc.edu/assets/docs/dt_cost.pdf, accessed September 15, 2010.

Nationaal Archief of the Netherlands. http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=185&categorie=6

Numeric project Study Report (2009). Chartered Institute of Public Finance and Accountancy (CIPFA). http://www.numeric.ws/, accessed September 15, 2010.

The Preserving Digital Public Television Project (2010). Strategies for Sustainable Preservation of Born Digital Public Television. Accessed at http://www.thirteen.org/ptvdigitalarchive/files/2009/10/PDPTV_SustainabilityStrategies.pdf accessed September 2, 2010.

Palm, J. (2006). The Digital Black Hole. National Archives, Stockholm, Sweden. http://www.tape-online.net/docs/Palm_Black_Hole.pdf. Accessed August 23, 2010.

Peterson, M., Zasman, G., Mojica, P., & Porter, J. (2007). 100 Year Archive Requirements. Survey: Storage Networking Industry Association (SNIA). http://www.snia.org/forums/dmf/knowledge/100YrATF_Archive-Requirements-Survey_20070619.pdf. Accessed August 23, 2010.

Primary Research Group (2008). The International Survey of Library & Museum Digitisation Projects. http://www.primaryresearch.com/release-200811052.html. Accessed September 7, 2010.

PrestoSpace (2005-2008). D20.1 SET1 : Survey on Target Users. Available at http://prestospace.org/project/deliverables/D20.1_public.pdf and Deliverable 2.1 User Requirements Final Report PrestoSpace Questionnaire 1 Preservation and Digitisation Plans: Overview and Analysis. Available at http://prestospace.org/project/deliverables/D2-1_User_Requirements_Final_Report.pdf both accessed September 8, 2010.

PrestoSpace (2004-2008). http://wiki.prestospace.org/pmwiki.php?n=Main.PS. Help http://wiki.prestospace.org/, http://prestospace.org/project/public.en.html.

PrestoSpace (2004-2008). Making a Preservation Plan – and Budget. http://www.hatii.arts.gla.ac.uk/confweb/tape/richard_preservation_strategy_handout_3.pdf

Science and Technology Council. (2007). The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials: Academy of Motion Picture Arts and Sciences (A.M.P.A.S.).

Screen Digest, FIAT/IFTA, and FOCAL (2010). The Global Trade in Audio-visual Archives http://www.screendigest.com/reports/201074c/10_08_the_global_trade_in_audio_visual_archives/view.html. Accessed September 10, 2010.

Sun Microsystems (2007). Archiving Movies in a Digital World. http://sun.com/storagetek/disk_systems/enterprise/5800/ArchivingMoviesinaDigitalWorld.pdf. Accessed September 2, 2010.

Training for Audiovisual Preservation in Europe (2008). Tracking the Reel World. http://www.tape-online.net/docs/tracking_the_reel_world.pdf.

Zeller, Jean-Daniel (2010). Cost of digital archiving : Is there an universal model? Hôpitaux Universitaires de Genève. http://www.bar.admin.ch/eca2010/00732/00869/index.html. Accessed September 2, 2010.

Annex : Example of a PrestoCentre Activity

Sharing Experiences: Survey of Digitisation Project Budgets and Other Cost Information

To assist libraries, museums, and archives with budgeting, PrestoCentre will be collecting financial data from its members regarding digitisation, storage, preservation, and access. Because costs and accounting procedures vary so widely between different projects, we collect only the most basic data in the online form here, and for those institutions that are willing, follow up with a short personal interview.

Institution

Your name

Your email address

Project name

Project dates

Total project budget €
- One time
- Annual

Total collection size (items or hours)

Types of media being digitized
- Film (hours, items)
- Videotape (hours, items)
- Audiotape (hours, items)
- LP records (hours, items)
- Paper records (pages)

Total expected collection size (terabytes) in
2011
2012
2013
2014

Number of staff engaged in project

Is digitization being performed by an outside contractor or partner? If so, please name them here, and briefly describe the approach to arriving at a price (e.g. cost per hour, cost per item).

If you were to break the project cost up by the following phases, what approximate percentage would you assign to each phase?

Acquisition (including inventory, selection, appraisal)
Ingest (including material preparation, metadata, scanning, quality assurance)
Storage (including system architecture, equipment)
Long term operations (including transcoding, migration, operation)
Access (including web development, right clearance)

If you were to break the project cost up by the following functions, what approximate percentage would you assign to each function?

Personnel
Capital equipment costs
Other operating costs

To what extent does the work involve restoration? Are these costs accounted for separately?

To what extent does the work involve cataloging? Are these costs accounted for separately?

To what extent does the work involve software development? Are these costs accounted for separately?

Are there any plans for cost recovery, e.g. by the sale of rights to the material?

How and by whom do you expect the material will be used (e.g. by broadcasters, educators, the public, film makers, scholars)?

Are there any supporting documents or project plans you can provide?