IASA RESEARCH GRANT REPORT: A STUDY ON THE CHANGING PRICES OF AUDIOVISUAL DIGITIZATION, 2006–2015

Rebecca Chandler, AVPreserve, USA

Introduction

Over the past five years in the United States and Europe there has been a decisive decline in the cost of digitization, resulting in historically low prices. This has been due to a variety of reasons, ranging from innovation to marketplace competition. With a shortening window of time in which to act, this economic advantage has been a boon to organizations with holdings of legacy audiovisual media. This is widely recognized and discussed among managers and budget holders within organizations. However, what has not been evaluated effectively is whether or not we are currently in a digitization market bubble.

We hypothesize that obsolescence, degradation, and market factors will begin to place the cost of digitization back on an upward trajectory. What we do not know is how the curve will look. Our experience indicates that organizations are unwilling to believe that prices will increase at the same rate that they decreased, or that they will ever again reach even 50% of what they were just five years ago. However, these discussions are often emotionally-driven and take place without a well-laid foundation of data from which to judge.

With the support of an IASA Research Grant, we set out to design a project that would begin to amass a shared and computable dataset that can help us ask questions and provide projections about our hypothesis. The goal of this project is to quantify and chart in detail the historical pricing trends for digitization over the past 10–15 years. We believe that having this insight and level of detail will provide a solid foundation of understanding. In combination with a strong grasp on key market factors we can assess future costs for digitization of audiovisual materials and generate data-driven predictions that are rooted in reality, and will serve as a risk mitigation tactic to help long-term digitization projects plan more effectively for expenditures and fund-raising.

Aims and Purposes

For two years AVPreserve built and tested a tool that has now been released freely to the public to show raw numbers and visualizations demonstrating the costs of inaction when it comes to caring for audiovisual collections. *COI: The Cost of Inaction Calculator* is an open and free tool that helps organizations analyze the implications of choosing various levels of response with regard to digitizing legacy collections of audiovisual materials. COI is a counter-perspective to the concept of "ROI", or Return on Investment, which is often raised with ill-effect when decision makers analyze digitization and preservation projects. COI recognizes past and present investment in collections in order to add a data point to ROI to help effectively articulate what may be lost in the way of access, intelligence, and finances based on various scenarios.

Building on the concepts behind COI and aiming to bolster an important data point within the calculations, AVPreserve, with this IASA Research Grant, has begun to develop a new dataset to be shared with the field. We have compiled pricing data for digitization going back nine years, plotted it out and analyzed historical trends. This has produced some useful insights and some basis for looking into the future and projecting what price increases in the coming years might look like. Over the next year we will be developing a mechanism to put this data online and to allow others to contribute to the data set so that the insights can be more nuanced and more accurate.

We think this is interesting for a few reasons. One because it has simply never been done. There has been a lot of conjecture but it has never been quantified in any meaningful way. Another because we see that the tendency is for administrators and executives to think that the prices that exist now will persist for the next 15 years; in reality, it is obvious that this will not be true, but it is difficult to say what will be the likely scenario.

During this phase of the project, AVPreserve used existing data, and scrubbed it to maintain complete anonymity on both the service provider side and the service procurement side. AVPreserve extracted this information from and compiled it into one dataset, supplementing the data with contextual information such as size of project, digitization specifications, formats involved, quantities of objects digitized, among others to be determined during the project. In future phases of the project, AVPreserve will reach out to preservation reformatting service providers and ask them to submit their pricing information (again keeping everything confidential and anonymous). AVPreserve will also reach out to people and organizations who have outsourced digitization asking them to provide their specifications and pricing data (again keeping everything confidential and anonymous). This dataset, as it is compiled, will document the nature of digitization service costs, revealing the shifts in costs over time and the discrepancies of pricing based on various factors. Saving our audiovisual heritage requires funding, and we need accurate data to aid our articulation of need.

Methodology

We planned for the overarching project to consist of four activities. This report focuses on activities one and the preliminary results of four. The first activity involved gathering and reporting on data already held by AVPreserve. Our approach was to use the data that was closest to hand in order to develop a stable data model for capture and reporting of the information that will make up this dataset. The second activity will be to gather data from digitization service providers. The third activity will be seeking data from organizations who have procured digitization services and have received quotes/invoices from vendors. This is seen as the most challenging of the datasets due to the variables and lack of controls in place. However, it is still a valuable dataset that is of great interest. Its importance may increase depending on the cooperation sought from vendors. The fourth activity is analysis and reporting, which we perform at the end of each of the first three activities, and is an ongoing product of this research project as the dataset grows and is shared with the field.

- Activity 1: Existing data compilation and entering: includes coming up with the data model for documentation and reporting
- Activity 2: Vendor data coordination, compilation, entering: more coordination, more data, more unknowns
- Activity 3: Client data coordination, compilation, entering: more people to organize, more follow through, more data and more unknowns
- Activity 4: Analysis, Visualizations, Reporting: this activity is repeatable as the dataset grows

Approach

The first step we needed to take in beginning to gather data was to decide on an appropriate data model for the project. As we were unsure at this point of exactly what analysis we were going to be able to perform at the end of Activity 1, we decided to record as much data as we could, in order to avoid needing to go back to retrieve more. The fields we selected are as follows:

Key
Description/Rationale: Each entry, based on format type, requires a unique identifier in order to differentiate between them.
Data type: 16 digit sequential number
Validation: n/a
Vocabulary: n/a

Date
Description/Rationale: This date corresponds to the date the vendor sent the proposal to the client.
Data type: date
Validation: ISO 8601
Vocabulary: n/a

Project ID
Description/Rationale: In order to link projects with multiple vendor responses, we assigned each project a Project ID number. A project is defined as a single client with a single RFP— a project may have multiple vendor responses.
Data type: 6 digit sequential number
Validation: n/a
Vocabulary: n/a

Batch ID
Description/Rationale: In order to differentiate between multiple vendor responses to a single project and to link multiple formats within a vendor response, we assigned each vendor response a Batch ID number. A batch is defined as a vendor's response to an RFP— a batch may have multiple format entries.
Data type: 6 digit sequential number
Validation: n/a
Vocabulary: n/a

Vendor country
Description/Rationale: The country in which a vendor is located.
Data type: free text
Validation: n/a
Vocabulary: n/a

Vendor state/province
Description/Rationale: The state or province in which a vendor is located.
Data type: free text
Validation: n/a
Vocabulary: n/a

Client country
Description/Rationale: The country in which a client is located.
Data type: free text
Validation: n/a
Vocabulary: n/a

Client state/province
Description/Rationale: The state or province in which a client is located.
Data type: free text
Validation: n/a
Vocabulary: n/a

Resource type
Description/Rationale: A generic description of the format type, e.g., is it a sound recording or is it a moving image object?
Data type: list
Validation: resource type list
Vocabulary: sound recording; moving image

Quantity of format in project
Description/Rationale: The number of items of a particular format the client requested a quote for in their RFP. Vendors may offer discounts on bulk projects.
Data type: integer
Validation: n/a
Vocabulary: n/a

Media
Description/Rationale: The method with which sound or visual elements are affixed/stored on the media type.
Data type: list
Validation: media list
Vocabulary: magnetic; optical; mechanical

Media type
Description/Rationale: A mid-level description of the format type in order to support generalizations based on geometry and physical structure.
Data type: list
Validation: media type list
Vocabulary: open reel; cassette; disc; cylinder; wire

Source format
Description/Rationale: A specific description of the format.
Data type: list
Validation: source format list
Vocabulary: audio cassette; 1/4 inch audio tape; DAT; lacquer disc; metal disc; pressed 78RPM disc; pressed LP disc; pressed 45RPM disc; CD; mini disc; wire; cylinder; VHS; U-Matic; betacam; digital betacam; DV; 8MM; betamax; HDCam; laserdisc; XDCam; 2 inch videotape; 1 inch videotape; 1/2 inch videotape; D2; D3; DVD; 2 inch audio tape; 1/2 inch audio tape; DTRS; IMX MPEG; 1/2 inch digital audio tape; D1; 1/2 inch digital audio tape; 1/4 inch digital audio tape; MII; digital compact cassette; VHS audio; instantaneous disc; transcription disc; D5; ADAT; CVC; 1 inch audio tape; 8-track; Dictaphone; flexidisc

Format type
Description/Rationale: Any additional specifications needed to differentiate one format from another.
Data type: free text
Validation: n/a
Vocabulary: n/a

Average estimated run time
Description/Rationale: The estimated runtime of an item. Vendors may charge differently based on length, making this an important variable when determining and comparing costs per minute.
Data type: integer in minutes
Validation: n/a
Vocabulary: n/a

Digitization type
Description/Rationale: Different types of digitization will result in different costs. High throughput workflows, or one operator performing multiple transfers simultaneously, will result in lower costs than one-to-one workflows, or one operator performing one transfer. It would not make sense to lump these two workflows together during analysis.
Data type: list
Validation: digitization type list
Vocabulary: high throughput; one-to-one; unknown

Preservation master target
Description/Rationale: Costs may vary based on how a client requests the preservation master be delivered - as a physical carrier or as a file.
Data type: list
Validation: preservation master target list
Vocabulary: carrier; file; na

Mezzanine target
Description/Rationale: Costs may vary based on how a client requests the mezzanine copy be delivered - as a physical carrier, as a file, or not at all.
Data type: list
Validation: mezzanine target list
Vocabulary: carrier; file; na

Access target
Description/Rationale: Costs may vary based on how a client requests the access copy be delivered - as a physical carrier, as a file, or not at all.
Data type: list
Validation: access target list
Vocabulary: carrier; file; na

Lump sum cost
Description/Rationale: Used when a vendor provides a lump sum cost estimate for digitization. Not used when the Preservation master generation cost, Mezzanine generation cost, or Access generation cost fields are used.
Data type: integer
Validation: n/a
Vocabulary: n/a

Preservation master generation cost
Description/Rationale: Used when a vendor does not provide a lump sum cost estimate for digitization, but rather breaks down the cost into separate actions. Not used when the Lump sum cost field is used. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Mezzanine generation cost
Description/Rationale: Used when a vendor does not provide a lump sum cost estimate for digitization, but rather breaks down the cost into separate actions. Not used when the Lump sum cost field is used. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Access generation cost
Description/Rationale: Used when a vendor does not provide a lump sum cost estimate for digitization, but rather breaks down the cost into separate actions. Not used when the Lump sum cost field is used. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Inspection & prep
Description/Rationale: A vendor inspects an item for physical issues that may inhibit digitization and then prepares the item for digitization. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Quality control
Description/Rationale: The cost for a vendor to inspect an item for physical issues that may inhibit digitization and then prepare the item for digitization. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Metadata cost
Description/Rationale: The cost for a vendor to generate and format technical metadata - either as an XML document, spreadsheet, or embedded into the file - for delivery to the client. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Cleaning cost
Description/Rationale: The cost for a vendor to clean the item in preparation for digitization. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Dehydration cost
Description/Rationale: The cost for a vendor to bake or dehydrate an item, if it exhibits signs of sticky shed syndrome, in preparation for digitization. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Lubrication cost
Description/Rationale: The cost for a vendor to lubricate an item, if it exhibits signs of soft binder syndrome, in preparation for digitization. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Humidification cost
Description/Rationale: The cost for a vendor to rehydrate an item, if it exhibits signs of extreme brittleness, in preparation for digitization. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Mold abatement cost
Description/Rationale: The cost for a vendor to remove mold from and thoroughly clean an item. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Other cost
Description/Rationale: Other costs itemized in the vendor's response that do not fit in one of the cost categories above. 0 is used when the action is included in another price, null is used when it is not included or it is unknown.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Total cost per object
Description/Rationale: The sum of Lump sum cost, Preservation master generation cost, Mezzanine generation cost, Access generation cost, Inspection and prep, Quality control, Metadata cost, Cleaning cost, Dehydration cost, Lubrication cost, Humidification cost, Mold abatement cost, and Other cost for an item.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Total cost per minute
Description/Rationale: The Total cost per object divided by the Average estimated run time for the format.
Data type: integer
Validation: n/a
Vocabulary: integer; 0; null

Dataset

This project is ongoing and we do not yet have analysis and results to share. In the coming year we will run statistical tests against the dataset that we have created in order to identify trends and make assertions about the changing costs in digitization services. For the purposes of this report, we will provide a summary of the dataset and a few calculations.

The dataset that we have aggregated to-date consists of 25 projects spanning the years 2006–2015 across two countries (US and Canada) and eight different content-holder locations (New York, Maryland, Texas, Massachusetts, California, New Jersey, Iowa, and Ontario). Projects represent unique organizations who have bid out for digitization services. In response to these 25 projects, the dataset includes 59 responses from digitization service providers. The current dataset only includes data from service providers in the US. A benefit to this dataset would be the addition of project data and batch data from organizations and service providers across the globe.

Within these 59 responses from digitization service providers, we have accumulated 1,765 format-specific price points. Because these price points are for multiples of specific formats and not for individual objects, this actually represents price points for 2,682,044 audiovisual carriers.

Moving Image Format	Quantity of Carriers in Sample	Average Cost per Carrier 2006–2015
1/2 inch videotape	1951	$263.17
2 inch videotape	1612	$252.22
8MM	48786	$230.27
betamax	54831	$167.32
D2	4527	$126.87
1 inch videotape	50409	$119.55
betacam	1120604	$102.48
VHS	159596	$88.08
U-Matic	228767	$84.87
DV	364996	$76.18
D5	111	$71.17
laserdisc	234	$70.00
XDCam	44282	$64.23
D1	6521	$64.19
digital betacam	156508	$63.17
D3	2898	$60.12
DVD	73399	$57.59
MII	1163	$51.83
HDCam	45353	$51.56
IMX MPEG	12092	$38.72

Table 1. Quantity and total average cost of moving image formats represented in the dataset for this project. Average represents samples from 2006–2015.

Audio Format	Quantity of Carriers in Sample	Average Cost per Carrier – 2006–2015
2 inch audio tape	8799	$255.31
1/4 inch digital audio tape	1	$250.00
DTRS	17169	$97.84
1/2 inch audio tape	2168	$96.68
Dictaphone	204	$92.50
VHS audio	285	$83.67
wire	32	$79.60
pressed LP disc	57418	$75.86
1/2 inch digital audio tape	66	$74.87
transcription disc	816	$74.33
instantaneous disc	4598	$74.22
flexidisc	1102	$67.50
lacquer disc	168	$64.28
digital compact cassette	63	$62.16
pressed 45RPM disc	8	$50.00
1/4 inch audio tape	73908	$49.08
DAT	34406	$48.64
ADAT	9393	$47.95
CVC	2	$45.00
8-track	22	$42.50
pressed 78RPM disc	54	$39.33
audio cassette	54120	$31.95
CD	38501	$26.78
mini disc	101	$14.40

Table 2. Quantity and total average cost for digitization of audio formats represented in the dataset for this project. Average represents samples from 2006–2015.

Source Format	2010 avg/ item	2011 avg/ item	2012 avg/ item	2013 avg/ item	2014 avg/ item	2015 avg/ item
1/2 inch videotape	$331.11		$188.57		$190.00	$165.27
2 inch videotape	$292.82	$292.24		$206.75	$250.00	$229.41
8MM	$368.35		$60.00	$40.30	$61.50	$43.30
betamax	$244.91		$40.00		$51.00	$27.54
D2	$330.80		$40.00	$41.49		$71.17
1 inch videotape	$206.53	$167.24	$205.31	$76.36	$128.50	$55.95
betacam	$262.08	$82.24	$38.13	$34.44	$23.25	$23.35
VHS	$221.80		$49.40	$35.47	$38.38	$27.47
U-Matic	$204.70	$59.74	$39.06	$33.88	$27.79	$30.88
DV	$187.30	$54.74	$50.83	$36.94	$39.94	$24.28
D5						$71.17
laserdisc						$70.00
XDCam	$140.42	$84.74		$44.53	$27.38	$23.93
D1			$50.00			$87.84
digital betacam	$158.36		$26.67	$30.71	$22.38	$24.57
D3		$72.24		$41.00		$71.17
DVD	$137.41			$29.27	$18.86	$14.02
MII				$20.33		$57.38
HDCam		$92.24		$50.67		$28.93
IMX MPEG				$38.72		
2 inch audio tape	$333.33	$282.35			$234.00	$137.50
1/4 inch digital audio tape	$250.00					
DTRS		$223.60		$46.67		$65.19
1/2 inch audio tape		$168.60			$71.67	$86.86
Dictaphone						$92.50

Source Format	2010 avg/ item	2011 avg/ item	2012 avg/ item	2013 avg/ item	2014 avg/ item	2015 avg/ item
VHS audio						$83.67
wire						$79.60
pressed LP disc	$114.46	$138.25			$70.00	$60.73
1/2 inch digital audio tape	$63.34			$75.83		
transcription disc						$74.33
instantaneous disc						$64.50
flexidisc						$67.50
lacquer disc	$56.63	$87.25				
digital compact cassette				$46.67		$71.46
pressed 45RPM disc					$50.00	
1/4 inch audio tape	$51.32	$78.59	$60.00	$36.31	$92.43	$33.88
DAT	$89.82	$143.60		$39.79	$35.67	$48.90
ADAT						$47.95
CVC						$45.00
8-track						$42.50
pressed 78RPM disc						$39.33
audio cassette	$94.74	$80.77	$35.00	$30.00	$27.73	$17.89
CD	$63.85	$118.60		$28.67	$19.57	$12.44
mini disc						$14.40

Table 3. Average digitization cost per item per format per year for 2010–2015. Values represent data contained in the research data set currently.

Much work remains to be done with this dataset. We hope these brief numbers provide a summary of the dataset and an understanding of the strengths and weaknesses of the data. We welcome any additions to the dataset so that the questions asked of it reveal answers with richer nuance and accuracy.

The author wishes to thank Bertram Lyons for his support in framing and executing Activity One of this project. Special thanks also to the IASA Board for providing AVPreserve the opportunity to explore this topic in detail.