

Fundamentals of AV Preservation

CHAPTER 3

PLANNING, PREPARING, AND IMPLEMENTING REFORMATTING PROJECTS

PREFACE

NEDCC is pleased to offer the Fundamentals of AV Preservation textbook for self-study to anyone with internet access. An HTML version is available online at www.nedcc.org/av-textbook, and downloadable PDFs can be found at www.nedcc.org/publications. Covering the core topics in caring for and reformatting audiovisual collections, this resource supports cultural heritage professionals in their efforts to steward audiovisual materials.

- Chapter 1: Care and Handling of Audiovisual Collections
- Chapter 2: Inventory and Assessment
- Chapter 3: Planning, Preparing, and Implementing Reformatting Projects
- Chapter 4: Managing Digital Audiovisual Collections
- Chapter 5: Disaster Preparedness and Response
- Glossary

Credits

The content for each chapter of the Fundamentals of AV Preservation textbook was created in 2017 by staff members of AVPreserve (now AVP) and edited by NEDCC staff members. The textbook served as the foundation of a multi-session, instructor-facilitated online course launched by NEDCC in 2017. The development of the course and textbook were subsidized by the National Endowment for the Humanities, and in 2022, additional NEH funding supported the transformation of the HTML textbook into downloadable PDFs.

AVPreserve — Authors

Chris Lacinak, *President (Ch. 1 & 3)*
 Rebecca Chandler, *Consultant (Ch. 1 & 2)*
 Amy Rudersdorf, *Senior Consultant (Ch. 4)*
 Kara Van Mallsen, *Partner and Senior Consultant (Ch. 5)*

NEDCC — Editors

Ann Marie Willer, *Director of Preservation Services*
 Sean Ferguson, *Preservation Specialist*
 Becky Geller, *Preservation Specialist*
 Frances Harrell, *Senior Preservation Specialist*
 Kim O’Leary, *Technology and Events Coordinator*
 Danielle Spalenka, *Preservation Specialist*



NATIONAL
 ENDOWMENT
 FOR THE
 HUMANITIES

CONTENTS

i	Preface
ii	Contents
1	Chapter 1: Care and Handling of Audiovisual Collections
1	Care and Handling
2	Grooved Media
7	Magnetic Media
12	Film
15	Optical Media
17	Shipping of Media Carriers
18	Additional Resources
21	Chapter 2: Inventory and Assessment
21	Inventory
28	Selection for Digitization
29	Prioritization for Digitization
33	Chapter 3: Planning, Preparing, and Implementing Reformatting Projects
34	The Digitization Signal Path
35	The Request for Proposal (RFP) Process
36	Internal Project Planning
39	Drafting a Statement of Work
58	Post-Digitization
59	Chapter 4: Managing Digital Audiovisual Collections
60	Digital Preservation Concepts
63	Organizational Infrastructure
65	Storage Infrastructure
70	Active Management
74	Metadata
78	Planning for Obsolescence
80	Prioritization and Phasing
83	Chapter 5: Disaster Preparedness and Response
83	Disaster Prevention and Mitigation
87	Disaster Planning
90	First Response Steps
96	Glossary

CHAPTER 3

PLANNING, PREPARING, AND IMPLEMENTING REFORMATTING PROJECTS

by Chris Lacinak, Founder and President, AVPreserve and edited by NEDCC staff

AT THE HEART OF PLANNING AND IMPLEMENTING a successful reformatting project is clear and strong communication. Whether this work is performed in-house or with a vendor, there is very little difference in the steps that should take place. Due to this fact, and for the sake of simplicity, this chapter will refer to “clients” and “vendors” to mean both internal digitization programs and contracted vendor services. The client is the department, unit, or organization with the need to reformat audiovisual media in their holdings. The vendor is the internal department or an external organization that will perform the reformatting.

Solid communication begins with the client clearly articulating their needs and specifications. This goes beyond specifying the input (e.g. Open reel audiotape, VHS) and output (e.g. 24-bit 96 kHz WAVE file, uncompressed Quick-Time file). When only given these two reference points, too many assumptions are left to chance; each one of these assumptions creates an opportunity for disconnects and miscommunication, greatly increasing the likelihood of the vendor delivering an end-product that does not actually meet the client’s needs.

The need for specificity and detailed communication is particularly important with the reformatting of content stored on legacy audiovisual media, such as Digital Betacam, VHS, audiocassette, and audio open reel. Most organizations do not have the equipment or expertise to view the media they hold in their collections and therefore have no point

of reference for the quality of the moving image and sound recorded on their assets. What is typically known is that it is normal and expected for there to be challenges, audiovisual artifacts, and certain aesthetic characteristics when reproducing content recorded on legacy media. This is due to a combination of obsolescence, degradation, and the fact that legacy technologies generally have specifications, which were typically high-performance at the time of their release but are much lower quality when compared to current technology. How does a client know that a poor-looking or -sounding transfer is due to the original recording quality, degradation, or an error on behalf of the vendor? The answer in most cases is that they almost never know, but this does not render the client powerless or subject to blindly trusting the vendor.

Clients can greatly mitigate risk by using specifications to ensure preservation-oriented workflows and practices are utilized and quality assurance and control measures are incorporated. A great place to document these expectations and specifications is in a statement of work (SoW), a document that is typically incorporated into a request for proposal (RFP) and details all of the requirements of a project with regard to standards, practices, protocols, timeline, and technological specifications. Using an SoW will also help clients fairly perform comparative analysis between vendor proposals. When there are few or no specifications included as part of an RFP, there is no way to identify reasons for differences in vendor proposals,

such as wildly varying pricing, or to understand how the differences relate to vendor workflows and practices.

Many organizations do not have the expertise in-house to draft a detailed SoW for a preservation-oriented reformatting project. This chapter will provide an SoW outline and set of discussion points along with sample SoWs and specifications. These should not be used in a “copy-and-paste” fashion. This will not serve the client, the client’s true needs, the vendor, or the spirit of strong and clear communication. They should be used for establishing a terminology and a framework that organizations can use as a foundation for research, conversations with vendors, and ultimately creating their own SoW.

While solid communication begins with the client, it does not end with the client. Vendors have a critical role to play in this regard as well. Vendor communication begins with responding to a SoW, asking questions, offering suggestions, and engaging in a conversation. These points of feedback can be used by the client as the basis of discussions held with other vendors and for performing further research within the client’s own organization. Vendor communication also consists of keeping in touch with the client at logical points throughout the project, although the expectation for communication from the vendor should be made clear within the client’s SoW.

Too many times, the project ends when a vendor sends the final deliverables of a project to a client and no quality control is performed by the client to ensure that there are no audio or video quality issues or that specifications regarding file naming conventions, organizational conventions, and metadata were met by the vendor. In these cases, a question is raised about the value of creating a set of digitization specifications for a vendor if the client is not going to verify that they have been met and that the deliverable meets their expectations and needs. While the vendor should perform their own quality assurance, the client should perform quality control in addition to this. Mistakes happen, even from well-intended, expert vendors. The client shares part of the responsibility and burden to ensure that the final deliverable meets the specifications outlined in the SoW. For every specification provided to a vendor, the client should have a quality control protocol that verifies that the specification is met.

When issues arise, they may or may not be vendor errors. There needs to be communication between the client and vendor regarding identified issues in order to see them through to a mutually agreed upon resolution, and any necessary rework must be performed and put through the same quality control process.

When solid planning and preparation is followed by strong and clear communication between the client and vendor, the risk of failure or disappointment is greatly reduced. There is an old saying that good fences make for good neighbors. Consider a detailed SoW to be a good fence. It leaves little question about the parameters of the agreement. Some may feel that this inserts a formality or rigidity in the client-vendor relationship that disallows a more friendly and convivial relationship. To the contrary, like good fences, clear and explicit statements of work and specifications allow both clients and vendors to relax, knowing what’s expected of them, fostering a more positive and cooperative relationship.

SECTION 1: THE DIGITIZATION SIGNAL PATH

While the exact technologies and workflows vary from project to project, it is helpful to understand the general setup and process that a vendor will use with your AV materials in the reformatting process in order to determine your internal priorities and specifications. The diagram below shows a very generic signal path with associated AV inputs and outputs. This diagram assumes that the original item has already been inspected and any necessary physical repair and stabilization has already been performed. This diagram also only speaks to AV and not to metadata.

This diagram is extremely simplified in order to show the high-level, salient components of a digitization signal path. An original physical item is played back on a machine. This machine must be calibrated and aligned to the original item in order to achieve a faithful reproduction of the original recording. The signal processor inserted in this diagram could be any number of different devices depending on what the original item format is. For audio discs this may be a preamp and equalizer. For video formats this may be

a time base corrector. For audio open reel this may be noise reduction decoders. Principally, signal processors in a preservation reformatting signal path are not about enhancing the signal. They are devices that are necessary in order to reproduce the original recording with integrity. If they were not present in the signal path, a faithful reproduction could not be achieved. In some cases, their absence may mean that no reproduction could even be achieved.

The output of the signal processor(s) is fed into the analog to digital converter (ADC), responsible for transforming the signal from its analog source to a digital approximation. The quality of the ADC is critically important in the ability to create a digital copy with high precision and integrity. There are video ADCs and audio ADCs. The digital stream of bits created by the ADC is routed to an encoder which takes this raw digital stream of data and converts, maps, and packages it into a file format or wrapper. When this is written to the hard drive as a file, the Preservation Master file is produced as a result. In order to create derivative files such as Mezzanine and Access Copy files, a transcoder is needed to decode the Preservation Master file, thereby creating a data stream that can then be used to convert, map, and package the AV into the appropriate specifications for the Mezzanine and Access Copy file formats.

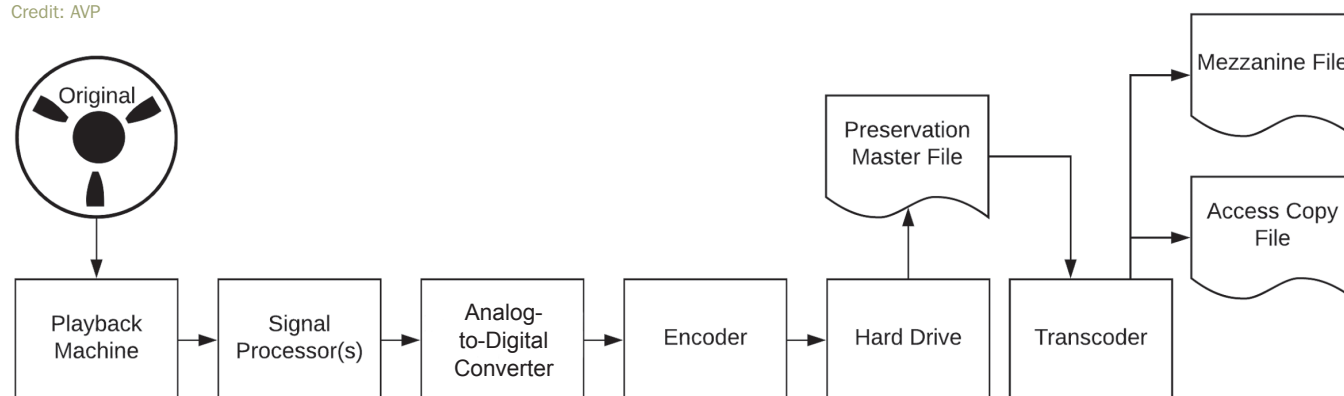
SECTION 2: THE REQUEST FOR PROPOSAL (RFP) PROCESS

For organizations working with either an internal or external vendor, releasing a request for proposals (RFP) is the best way to identify the best fit for the project and establish good communication early. If the RFP process is new to you, thinking of it like hiring an employee can be helpful—many of the steps are similar. There should be a vetting and selection working group established for this purpose, just like a hiring committee would be gathered to evaluate candidates for a job opening. A “job description” should be released and advertised in the form of a written SoW. Vendor proposals should be reviewed, and vendors should be interviewed by the committee, just as you would with a candidate.

In smaller organizations, the SoW is often the entire RFP. In larger organizations, particularly when there are procurement departments, there are often additional administrative, contractual, and legal sections, and the SoW is one part of the RFP. In this case, once the SoW is finalized it is turned over to the people in charge of procurement and typically goes through review with a focus on contracting and legal aspects. In cases in which there are a number of people involved in the process on the client side, it is important

FIGURE 3.1
A Diagram of the Digitization Signal Path

Credit: AVP



before the RFP goes out for all stakeholders to meet, review the SoW, and confirm the timeline to make sure that all parties are on the same page. This should include the primary drafters of the SoW, procurement, legal, IT, and any other stakeholders the project will either depend on or directly impact. Once everyone is in sync and all dates in the timeline are updated and current, the RFP can be released. This may be done by sending the RFP to selected vendors, posting it publicly, or both. In some cases, law or organizational policy may dictate whether an RFP must be posted publicly.

After an RFP is released, it is customary to provide a period in which vendors can ask clarifying questions about the RFP, your organization, or the project. Publishing specific information about how vendors can contact you with questions in the RFP, or even setting up specific conference calls, can keep this process moving quickly.

Once the RFP has been issued, vendor questions have been posed and responded to, and the final proposals are in from vendors, the vetting and selection process begins. Each member of the working group should review each proposal and document comments and questions. Some organizations choose to use a scoring system, rating vendors in specified categories. If this is the case, documenting these scores should be done as part of the review process as well. Once everyone has reviewed the proposals, the working group should convene to discuss. If there are any vendors that are definitely seen as a bad fit, they may be removed from the process. All questions for each vendor that is moving forward in the process should be consolidated and either a meeting should be scheduled with each of the vendors to ask questions or an email should be sent asking for written responses along with a deadline. Once all responses are received, the working group should reconvene to discuss and select a final vendor. Keep in mind that selecting multiple vendors is also a possibility. For very large projects, it might make sense to split between vendors either to spread the load or to work with vendors that are particularly strong with a given set of formats. Keep in mind that the client may need to revisit the vendor candidate pool if the pilot or project goes poorly with a selected vendor. All vendors should be notified of the decision promptly regardless of whether they were selected or not.

SECTION 3: INTERNAL PROJECT PLANNING

Conducting a successful reformatting project requires thoughtful planning that addresses three major questions:

1. What are the human, technology, and financial resources required within the client organization?
2. What are the roles each client staff member will play throughout the project?
3. How will the client organization handle the digital files and original analog materials once they are returned?

Answering these questions before the project gets underway will help the reformatting process run smoothly and avoid issues that may arise from having an incomplete understanding of available funds, team roles, and steps to undertake when receiving digital files. The following subsections provide guidelines to consider when making decisions regarding the reformatting project. Existing literature on effective project planning and management is rich and addresses many of the concerns relevant to AV reformatting projects. Where appropriate, this textbook provides excerpts from important works on project planning.

Committing Staff and Resources

Available staff and funds will invariably play a critical role in determining the scope of the project and the pace at which it is completed. This makes it important to identify available funds and staff hours as soon as possible when planning a reformatting project so that expected goals are realistic and that the projected timeline is manageable. However, identifying these resources is only the first half of the first step of project planning. The second half is allocating resources and delegating roles to staff effectively. A valuable reference that addresses both of these topics is the *Handbook for Digital Projects: A Management Tool for Preservation and Access*.¹⁵ Specifically, a section in Chapter 3, “Project Management: Creating a Plan of Work and Budget,” discusses approaches to committing staff and resources in great depth. And while this reference places a focus on scanning and microfilming, its advice regarding project planning remains true and applicable to audiovisual reformatting.

15 Maxine K. Sitte, ed. *Handbook for Digital Projects: A Management Tool for Preservation and Access*, 1st ed. Andover, MA: NEDCC, 2000.
<https://www.nedcc.org/assets/media/documents/dman.pdf>

Project Team Activities

PROJECT MANAGEMENT

Having a project manager on the client side is critical for a smooth and successful project. The project manager oversees all parts of the project, including, but not limited to: vendor interfacing; managing the client staff; monitoring scheduling and timeline; keeping track of quality control issue resolution and rework status; coordinating with client stakeholders; troubleshooting issues that arise; and proactively working to anticipate and identify prospective issues. Whether this resource is internal to the organization or a third-party provider, it is a critical role that must be filled.

RFP DRAFTING AND PROPOSAL VETTING

The client will be responsible for drafting the RFP, identifying vendor candidates, vetting proposals submitted by vendors, and ultimately selecting the vendor or vendors with which they wish to enter into a contract. This process should be driven by the project manager with the inclusion of other key stakeholders.

CONTRACTING

Once a vendor or vendors have been selected, a formal contract should be written. If the client has contracting, procurement, and/or legal resources on staff they will likely be responsible for this step. If not, the vendor can be asked to provide a draft agreement, but the client should be sure to have a legal advisor review before signing.

THE PILOT PROJECT

A pilot project is a project phase in which a sampling of items is selected to run through the entire process in order to identify and resolve any problem areas or gaps in the processes, protocols, specifications, or understanding. This allows both the client and the vendor to identify and resolve any issues in the process and specifications before launching into full production mode. A pilot phase is highly recommended, especially the first time a project is being performed, the first time a project with a new set of specifications is being performed, or whenever a client and vendor are working together for the first time. Usually it is best to select a sampling of items that will represent

the variances and test as many aspects of the SoW as possible.

There really is no “right” size for a pilot project. Generally, the size is based on a combination of factors including the amount of time it will take the vendor to complete, the number of items needed to represent the variables and diversity of the project, and the number of items seen as sufficient to serve as a proper test. The more complex the specifications and the larger the project, the more important a pilot phase is.

In cases where a pilot project proves very challenging, or in cases of very large projects, it may make sense to perform multiple pilot projects. In the former case this will help ensure that the process is fully ironed out before launching into full production mode. In the latter this may allow a ramping up period where each pilot is successively larger before hitting full throughput.

This is particularly important in the case of large projects where items are being digitized quickly, errors propagate at scale, and much time, energy, and good will can be expended in dealing with identifying and resolving errors. It is much easier to do things right the first time than to have to go back and fix things. This is almost always true, but it’s particularly true with larger-scale projects.

Pilot phases tend to be challenging because they often reveal more issues to be resolved than anticipated and resolving those issues tends to take collaboration between parties, which takes time. Essentially, the pilot process takes time, and it is important to allow for an appropriate amount of time. And as frustrating as it can be to iron out details when everyone wants to get working in full production mode, the up-front hard work pays major dividends throughout the rest of the project. Rework and re-performing quality control on a small number of items at the beginning of the project is far less painful for all parties than doing the same on a large number of items at the end of the project.

PREPARING, PACKING, SHIPPING, AND RECEIVING

Depending on the size of the project and how the original items are stored, this step could take a significant amount of time. Client staff are responsible for carefully packing up all assets to be sent to the vendor(s) and creating

a detailed manifest for each box and for each shipment. Packing and shipping recommendations by carrier type were outlined in Chapter 1, “Section 6: Shipping of Media Carriers.” In some cases, there may be work to do before items can be packed. Some organizations have requirements regarding the minimum amount of processing that must be performed on items before they can be sent for digitization. These requirements are usually focused on registering items with the organization’s system of record, assigning or capturing an identifier, and perhaps capturing one or more additional metadata fields. If these requirements exist, then completing them can take a lot of resources and time. This should be identified and accounted for as a process that precedes packing and shipping. Also remember to account for the fact that you will be receiving and reshelving these materials, too. Additionally, make sure to account for all staff time as well as the costs for materials and supplies for all of these processes.

QUALITY CONTROL

Upon receipt of digitization deliverables from the vendor, client staff must perform quality control. These measures may include testing that ensures that file specifications align with the requested target specifications, checking file naming conventions, making sure all files are accounted for, validating any XML files, and performing qualitative tests such as playing back files to check for audiovisual quality issues. The quantitative quality control may be performed with less expert, but well-trained staff. The qualitative quality control process benefits greatly from using staff that have audio knowledge and are able to identify, interpret, and articulate audio quality issues. The quality control process will take significant staff time and should not be overlooked when preparing timelines and budgets. It is frequently the case that client organizations do not adequately staff this role, resulting in a project bottleneck that can put the success of the project at risk. Aside from human resources for quality control, it is critical that client organizations use the appropriate specialized equipment and environments that quality control requires. Performing quality control using sub-par equipment or while in a poor environment can hinder quality assurance. For AV quality control it is important that clients have, or procure, equipment and use an environment that keeps the following in mind:

- The computer used for quality control must have sufficient processing power, memory, and bandwidth for working with audiovisual files at scale and running the necessary applications and processes.
- The computer used for quality control must have installed, and be able to run, all necessary specialized applications for video, audio, metadata, and file processing.
- The computer used for quality control must have a high bandwidth connection to any storage locations that it will be reading files from or writing files to.
- The audio digital to analog converter used must be of high quality. Audio playback must not be performed using the computer’s internal commodity hardware sound card.
- The audio reproducers (e.g. audio monitors, headphones) must be high quality, capable of reproducing the full audio frequency spectrum with precision, and intended for professional audio use.
- If using headphones for quality control, they should be very comfortable for the user to wear for extended periods. They should be over the ear, high-quality headphones created for professional audio use.
- If using audio monitors for quality control, the room in which quality control is being performed should be acoustically transparent. Rooms may be acoustically treated to achieve this as necessary. The acoustics of the room must not add noise that makes it difficult to perform critical listening (e.g. excessive reflections from close hard surfaces, reverberation).
- The environment that the user is in must be free of noise (e.g. humans, machines, HVAC).
- When performing video quality control, video computer monitors used must be high-quality monitors offering precision video reproduction (e.g. color, contrast) used in professional critical viewing environments.
- CRT video monitors should be used for viewing any files that result from originals that utilize interlace scanning. (This applies to all analog legacy video formats and some more modern digital formats.)

- For critical viewing of video, the user should have total control over the lighting in the room.

Failure to perform good quality control not only stands to waste time and money on the process itself, but also puts the goal of preservation at risk and all of the finances dedicated to the project. If there are concerns about the ability to perform this work, in-house clients may consider hiring a third party to perform quality control for them.

Receiving Digital Copies and Physical Originals

When digitization takes place in a preservation context, it would be remiss for the client organization not to prepare for the eventual return of the digitized files. “Chapter 4: Managing Digital Audiovisual Collections” addresses this topic and provides guidance on the necessary protocols and infrastructure. Organizations should consider this in advance of creating significant quantities of files; they should create plans and be prepared to accommodate both the ingest of media and metadata files into access and preservation systems as well as the long-term storage and management of these files.

MANAGING PHYSICAL ORIGINALS

After the physical assets have been digitized, it may be tempting to assume that they are no longer necessary. However, the physical assets still play a role in the overarching preservation strategy. As long as the original assets are able to be reproduced, they offer risk mitigation against loss of digital files, poorly performed reformatting transfers, and future technological advancements that may bring improved reformatting capabilities that the client may want to keep open as an option. Follow the recommendations outlined in Chapter 1 in order to maintain these backups for as long as possible.

SECTION 4: DRAFTING A STATEMENT OF WORK

An SoW, as defined for the purpose of this text, is a document that provides context to help the reader understand the goals and intent of the project, followed by details of the scope, process, timeline, communication protocols, technical specifications, and a request for pricing. In larger

organizations an SoW may make up one part of a request for proposal (RFP) that also includes legal and contractual information. In smaller organizations it may make up the heart of the RFP with few other components.

The process of drafting an SoW and issuing RFPs provides an opportunity to learn from others, ranging from organizations that have recent experience to share about performing similar work to vendors who are willing to engage and share their expertise and broad experience across a diverse customer base. This section will provide a framework, reference points, and a vocabulary that can be used to prompt conversation with others as you create an SoW. Before issuing your RFP, ask colleagues to review and critique it in order to help confirm and/or shape decisions. If the SoW will be used with a digitization lab that is internal to the client organization, meet with the digitization lab team throughout the SoW drafting process to seek feedback and input. If issuing an RFP for external service providers, the Q&A portion of the process can be a great point of reflection if vendors ask questions that point out unforeseen issues or identify gaps that may have been missed. The author of this chapter worked for a reformatting service provider serving archives between 1999 and 2006, created and taught a graduate level course on this topic for three years, has drafted over 25 SoWs for clients of AVPreserve since 2006, and has managed several of those projects through to completion. Even still, every new SoW and every new implementation brings new insights, lessons learned, and ways to improve and do better the next time. There is always more to learn and ways to continue improving, so use this document as a starting point, but also be sure to use the process as a platform for learning and refining.

There are a number of pieces of information that are important to convey in any SoW, for the sake of both the client and the vendor. The headings below may serve as an outline for an SoW or may simply serve as specific points that are incorporated into an SoW. The thoughts and considerations supporting each heading will provide the information needed to best decide how to incorporate these points into an SoW that is designed to serve a specific organization and set of circumstances.

Before delving into the details below, it may be helpful to review the diagram of the digitization signal path in Section 1 in order to gain perspective on what follows

the vendor selection and pilot project phases of a larger project. This diagram illustrates how many of the items discussed below tie together within the larger process.

Sections of a Statement of Work

ABOUT THE CLIENT AND PROJECT

This section provides the vendor with some context about the client organization and project so they have a basic understanding of who the client is and what their goals are. This serves two important functions. First, it provides frames of reference to ensure understanding and to help avoid misinterpretation based on incorrect assumptions. Second, it allows the vendor to engage as a collaborator more easily because they have the information that will help them to offer suggestions or provide alternative paths to take while reaching the stated goals, objectives, and outcomes.

BRIEF DESCRIPTION

In this section, provide the vendor with the scope of the materials to be digitized. Typically, this would be a breakdown of formats by quantity and duration. If duration of the recorded content is unknown, use the maximum media duration of an asset (e.g. 60 minutes for a 60-minute audiocassette) as a safe outside estimate. If media duration is used, it would be wise to point this out to the vendor so that they know the logic behind these numbers. It is best to provide format information at the variant level. For instance, if the collection contains Betacam, Betacam SP, Betacam SX, and Digital Betacam, do not put all of these formats under “Betacam.” List each of them separately and provide the quantity and total duration for each.

This section points out the importance of performing an inventory of the materials to be digitized prior to drafting the SoW. Depending on the number of items that are candidates for digitization and the size of the client’s budget, there may be the need to prioritize items for selection as well. Read more about performing inventories and selecting for digitization in “Chapter 2: Inventory and Assessment.”

TIMELINE

It is easy to overlook the number of steps that must take place in the contracting and implementation process as well as the number of parties that may need to be involved. Documenting and communicating each step with its associated due dates allows all parties to work more cooperatively and in synchronization. The act of thinking through the process in this level of detail helps to create a realistic timeline that will avoid unpleasant surprises down the road. The timeline below offers an example of what the steps might include. Your timeline may look different depending on the contracting process.

- **RFP distributed:** yyyy-mm-dd
- **Bidder questions:** yyyy-mm-dd
- **Client responses** (Q&A will be provided to all respondents): yyyy-mm-dd
- **Proposals due:** yyyy-mm-dd
- **Award:** yyyy-mm-dd
- **Contracting completed:** yyyy-mm-dd
- **Pilot assets shipped to vendor:** yyyy-mm-dd
- **Pilot phase completed by:** yyyy-mm-dd
- **Media shipments begin:** yyyy-mm-dd
- **Ship files to Client:** To be delivered on a monthly basis for quality control and approval
- **Client QC completed:** To be completed xx days after delivery on a regular basis
- **Rework completed:** Within xx days of completion of project
- **Ship source objects to Client:** Within 30 days of completion of Client approval of rework
- **Project completed by:** yyyy-mm-dd

There are two phases in this process that are most likely to cause timeline issues and problems between clients and vendors. It is important to recognize them as part of the process and to allow sufficient time for them. These are the pilot and rework phases.

The pilot phase is discussed in its own section below. The rework phase, the phase in which vendor mistakes are corrected, is often overlooked entirely. All parties would prefer to think that the project will be smooth sailing and any necessary rework will be negligible at most. Sometimes this is true, but major problems arise when rework hasn't been factored into the timeline at all, especially when everyone is up against a hard deadline, and the vendor delivers right at the end of said deadline. Note that performing a pilot phase will mitigate the amount of rework that is necessary, but it is still important to plan for rework nonetheless.

Aside from performing a pilot phase, another way to avoid rework is to stage the deliverables so that quality control is performed at multiple periods throughout the project instead of all at once at the end of the project. When quality control is performed earlier in the process, it offers opportunities to identify and correct any issues as work proceeds. Also, it allows for more measured staff allocation for both the client and vendor, which is a much better option than performing quality control and rework all at once, usually when there is little time left in the project, a situation which is not conducive to performing good quality control or rework. Spreading quality control and rework out over the course of the project lessens the chance of ending up with unpleasant surprises at the end. One final note on rework is that it quickly leads to the potential for challenging project management if it isn't identified as an auxiliary process to the main digitization effort. It is important to establish a process between the client and vendor on how rework will be managed, tracked, and reported throughout the project.

CLIENT POINTS OF CONTACT

This section should identify points of contact and associated information on the client side for project management, technical matters, administration, contracting, and invoicing. Not all points of contact have to be named in the SoW. Some can be identified as the project evolves, though it is best to provide the information up front if it is known.

COMMUNICATION PROTOCOL

This section defines how the client and vendor will communicate, at what frequencies, and at what point in

the timeline. There are eight points of communication that should be considered and addressed in this section.

1. Shipment of materials from client to vendor

The client should identify the information and means of delivery that will precede and/or accompany the shipment of materials to the vendor so that the vendor can plan and prepare accordingly for the management of this information upon receipt. This typically consists of two documents.

The first document is a shipping manifest that the vendor will reconcile the received materials against. There should be a manifest for each box that is printed and then both included inside and affixed outside of the box. Each box should have a box ID that is documented in a master shipping manifest and that is used for internal tracking purposes within the client organization. The box ID should be clearly placed on each box.

The information on the shipping manifest should include item level identifiers and/or label information that will allow the vendor to quickly identify the item on the manifest when unpacking the boxes. If this proves difficult based on the current labeling and/or identifying mechanisms on the items being shipped, then applying new identifiers to the items should be considered. In addition to including printed copies of the shipping manifests in each box, it is also a good idea to send a digital version of the master shipping manifest containing the total number of boxes, the box ID for each box sent, and a list of all of the items in each box.

The second document is a Source Metadata document, which includes information that will be helpful to the vendor as well as metadata that the client would like to persist through the digitization process and that the vendor should incorporate into the deliverables, potentially to remain as part of an archival package. This might include unique identifiers, title information, rights information, or any type of descriptive or administrative information that makes sense to keep as part of the archival package. The Source Metadata should be provided as a digital file, typically in spreadsheet form or XML. This may serve as an electronic backup to the packing manifest as well.

If the client is responsible for shipping materials to the vendor, this section should also identify how far in

advance the client will send notification of shipment and the Source Metadata document. If the vendor is responsible for transport of the materials, then this section should define when and how the client and vendor will be in communication about pickup to ensure a smooth and coordinated shipment.

Note that this may happen in batches for larger projects as opposed to shipping all items at once for smaller projects.

2. Verification of receipt at the box level by vendor

This should identify how long the vendor has to notify the client that they have received the materials and to offer a box-level reconciliation confirmation, as well as the means of communication and which client point of contact(s) to notify.

3. Verification of receipt at the item level by vendor

This should identify how long the vendor has to notify the client that they have received the materials and offer an item-level reconciliation confirmation. It should also stipulate the expected means of communication and which client point(s) of contact to notify.

4. Routine status updates

The need for status updates throughout the process will be dependent on the size of the project. Projects that will be finished in a few weeks may not need any status updates. Projects that last longer than a month will benefit from them. Projects that last for multiple months require them for successful completion.

This section should identify the frequency and method of status updates. The method may be meetings, emails, updating of shared cloud-based documents, or some combination of these. This section should also lay out the expectation of what information will be provided in status updates. If the method chosen is meetings, this section might also provide a generic agenda for these meetings, typically including a status report on progress from the vendor, reporting on projected progress over the next period, discussion of any outstanding issues, questions from either party, and any administrative or financial matters.

In larger projects, it is often advantageous for the client and vendor to have shared access to a project management spreadsheet (e.g. Google Sheets) or application where each party can see the overall status of the project via summary information in real time, as well as status at the item level where it is possible to track and communicate about each item through the process. For instance, the vendor might document issues encountered in inspection and place them here for the client to have insight into the issue and make a decision about how the vendor should proceed. Or the client might document their quality control findings and the vendor might document their response and/or rework status. This document or application may also specify the unique identifier for the storage media that the generated files are being delivered on in order to make it easier for the client to manage quality control and receipt of final deliverables. If such a system is used, it is important to identify who will be responsible for populating which fields and at which point in the process. Having such a mechanism for project management and immediate transparent access to the project status by all parties promotes tighter coordination and collaboration as well as overall project success.

5. Shipment of digital files from vendor to client

This should identify how far in advance and who the vendor will notify within the client organization before shipping the deliverables for quality control. Note that the vendor should maintain possession of the original materials until quality control has been fully completed in case the vendor needs access to the items for inspection or rework.

This section should also identify the information that the client expects the vendor to send in advance of and/or along with the shipment. At minimum this should be a shipping manifest detailing the number of boxes being sent, the box ID for each box being sent, the number of items being sent, and the item ID for each item being sent. This should be included in print form inside and outside of each box, and a master shipping manifest should be sent electronically. The electronic version might also include a manifest of the contents of each item being sent.

The client should identify the turnaround time on notifying the vendor of receipt at the box and item level to ensure closing the loop. Note that this may happen in batches for larger projects as opposed to shipping all items at once for smaller projects.

6. Quality control findings and resolution

This should identify how quality control will be managed and tracked throughout the process. Quality control consists of the client performing quality control, client documentation and reporting of findings to vendor, discussion, and investigation of any identified issues by vendor, rework as necessary, and re-performance of quality control as necessary.

This should identify the method of documenting and tracking this information, who is responsible for documenting which parts, and the turnaround time for each part of the process.

Properly managing this auxiliary process can be confusing and is often difficult in the midst of a digitization project. Establishing the protocols and methods of communication here will go a long way in making this process as smooth as possible. To reiterate, having a shared project management spreadsheet or application that incorporates quality control documentation and management is advised.

7. Shipment of original materials from vendor to client

This should identify how far in advance the vendor will notify the client before shipping the original materials, the means of notification, and who should be notified. It should also identify the information that the client expects the vendor to send in advance of and/or along with the shipment. At minimum this should be a shipping manifest detailing the number of boxes being sent, the box ID for each box being sent, the number of items being sent, and the identifying information for each item being sent within each box. This should be included in print form inside and outside of each box, and a master shipping manifest should be sent electronically.

The client should identify the turnaround time on notifying the vendor of receipt at the box and item level to ensure closing the loop. Note that this may happen in batches for larger projects as opposed to shipping all items at once for smaller projects.

8. Notification of ability to delete information from vendor systems

Before a vendor deletes client information from their systems, there should be explicit agreement from the client. This ensures that all quality control issues are resolved and that the client has time to ensure that all digital information is secured, with multiple copies stored in multiple locations, and any ingest routines are completed. Obviously, this needs to happen in a reasonable amount of time, as it places a burden on the vendor to maintain this data on their systems. Therefore, this section should not only identify the means of communication but also the anticipated turnaround time in order to allow the vendor to plan accordingly.

SHIPPING

The shipping section should be used to discuss any aspects related to the required method, speed, protocols, care and handling standards, material specifications, and insurance requirements related to packing and shipping. This is a good place to identify who is responsible for providing packing materials, performing packing and shipping, and paying for shipping. Note that if the vendor is responsible for any of the labor or costs associated with packing and shipping, be sure to make it clear that they need to include this in their proposal and pricing.

PILOT PROJECT

This section should lay out the parameters of the pilot project, including the size of the sample to be included, the formats that will be selected, the conditions under which a second pilot project may be performed, and any other relevant details. The size of the sample should be based on the size and variety of the formats selected for digitization in the SoW. In general, the pilot project will be most useful if it includes a sampling of each of the formats the client plans to reformat. This will give the client and the vendor the opportunity to test the specifications outlined in the SoW for each type of material. A second pilot project will be warranted if errors due to the reformatting process are identified or if specifications the client outlined in the SoW are found to be unsuitable for the goals of the project.

DEFINITIONS

One of the largest communication issues that tend to cause problems is how different people apply different definitions and assumptions, resulting in miscommunication and misunderstanding. This section should be used to define any terms that are commonly interpreted in different ways or that the client is using in a way that may not be obvious to others.

CARE, HANDLING, AND STORAGE

The care, handling, and storage of materials is a common point of differentiation between vendors that are preservation focused and those that are not. This section provides an opportunity to convey requirements and request proof of adequate care, handling, and storage from vendors.

Relevant aspects to raise here in the context of care, handling, and storage include:

- staff experience;
- company policies, protocols, and practices;
- storage and facility environmental specifications; and
- security.

MEDIA TREATMENT AND PREPARATION

Vendors have differing approaches to media treatment and preparation, such as cleaning, baking, and repair. These differences can have significant cost implications and may represent differences of opinion with the client as well. This section should be used to communicate any specifications that the client feels strongly about so that all vendors are referencing the same specifications in their pricing. In the absence of preferences from the client, this is a good place to ask the vendor to provide information about their standard media treatment and preparation protocols in order to provide more insight into their practices and to help you understand pricing variances between vendors. Note that this is also an area that may separate preservation-oriented vendors from non-preservation oriented vendors.

REFORMATTING

Similar to the section on media preparation and treatment, the reformatting section is a place to document any client preferences and/or to request information on vendor standard protocols. There are four main areas that are generally addressed in this section:

1. Reproduction Setup

Referring to the actions performed to the equipment and signal path immediately before beginning the transfer, reproduction setup is an important cost driver and differentiator for analog sources. Typically, this speaks to calibration and alignment processes performed on the playback equipment in order to help ensure a faithful reproduction of the original recording when playing back the source item. Different media types and formats have different reproduction setup routines, and this section should speak to each specific one. If there are no client preferences, inquiring about the vendor's standard protocols for this will allow comparisons between vendor proposals and pricing.

2. Signal Path

The path that the audio and/or video signal follows from the output of the playback device to the input of the recording device is the signal path. The quality of the devices in the signal path, the cable and interfaces used, and the way in which the signal path is constructed has a significant impact on the quality of the transfer. Stating preferences and/or gaining insights into a vendor's signal path is a useful piece of information for evaluation.

3. Image and Sound Processing

Best practices for preservation have dictated for decades that "flat" transfers be made when creating a preservation master. "Flat" transfers are unaltered digital versions of the content on the analog media. The intent of this approach is clear: not altering the signal in any way that detracts from a faithful reproduction of the original recording. However, the interpretation and application of "flat" differs from format to format and from person to person. It is important to forego the assumption that everyone has the same interpretation of "flat" and instead to be specific about what types of image and sound processing are and are not acceptable.

Inquiring about the vendor's standard protocol and opinion on this matter is appropriate here.

It may be the case that mezzanine or access copies (see section below for more details) have different specifications for image and sound processing. Some organizations want enhanced copies for access while others prefer to enhance only specific pieces of content, on demand and as needed. Any preferences should be stated here.

4. Destination File Format Specifications

Whereas the other subsections of the reformatting section offer some latitude insofar as the client may ask the vendor for input or for an opinion, the specifications for destination file formats does not. This is where the client must document in as much detail as possible the target digital formats that they want to receive. These may differ based on media types and format, and in these cases, it is important to provide details for each distinct set of specifications and to clearly identify which media types and formats they apply to.

There are three common destination file types, each of which has a distinct role to play:

- a. **Preservation Master**—The primary role of the Preservation Master is to provide an authentic reproduction of the original recording that enables a path to sustainability and long-term access.
- b. **Mezzanine**—The primary role of the Mezzanine is to offer a copy that is ready for use in standard production workflows and systems. There is more than one use case which suggest there can be a routine need for a high-quality version (e.g. for editing or broadcast). While a Preservation Master is in a format and resolution that makes it cumbersome to work with, and an access copy may not be of sufficient quality, the Mezzanine-level file fulfills the need. Mezzanine files are not always needed and should be created only when there is a clear and frequent need. An alternative to creating Mezzanine copies of all files is to create them on-demand from a copy of the Preservation Master file. In cases where the

need for Mezzanine level files is infrequent, this is often the best option.

- c. **Access Copy**—The primary role of the Access Copy is to provide a file that can be reviewed using standard commodity hardware and non-specialized software and infrastructure in order to maximize accessibility of the content.

Depending on the media type and format, file format specifications for each of these file types might include: wrapper specifications, codec specifications,¹⁶ bit depth, sample rate, fourcc code, color space, chroma subsampling, compression algorithm, scan type, pixel height and width, aspect ratio, recording standard, number of audio channels, handling of closed caption information, handling of timecode information, bit rate, optimization for streaming vs download, and endianness.

For video sources it is especially important to specify that all channels of audio be digitized as part of the recording. Video recordings may have anywhere from a single channel to multiple channels of audio depending on the format and how it was recorded. Not all playback devices will necessarily play all channels of audio on a videotape. For formats such as Betacam SP, there are decks that can record and playback four channels, and there are decks that can record and playback two channels. The latter deck is less expensive, so it is not uncommon for vendors to have two-channel decks. When a videotape with four channels of audio is played in a deck that supports only two channels of audio, it may not be possible to identify that there are additional channels present on the tape. The same is true when playing a videotape with timecode in a deck that does not support timecode. Unfortunately, lack of awareness of these issues over the past decades has almost certainly resulted in the loss of a great deal of audio and timecode information. Ideally, the presence of all audio channels and timecode should be identified and reproduced in the digitization process.

An aspect of file format specifications that is not obvious is that, even with all of the above-mentioned details, different encoders using the same exact specifications might construct files differently. If a large

16 Joshua Ranger, "A Primer on Codecs for Moving Image and Sound Archives." 2010. <https://www.avpreserve.com/a-primer-on-codecs-for-moving-image-and-sound-archives>

number of files resulting from a digitization effort share the same specifications but are constructed in multiple ways, this inconsistency can lead to challenges in the performance of quality control, variable results in compatibility with software applications, and eventually, challenges with migration and transformation.

One way to mitigate this issue when working with a single vendor is to utilize MediaInfo¹⁷ or a similar application to create a profile for each target file format and to utilize this profile as the lower level specification moving forward to ensure consistency. This can also be useful with multiple vendors as a conformance point that each vendor must meet to help ensure as few inconsistencies as possible.

It is true that any archive that is involved in generating, acquiring, and preserving digital files will have to learn to work with the inherent variability and diversity that comes along in this domain. While this is the reality, it is also true that consistency is a friend to preservation, and where we have the ability to insert our control to gain consistency we should.

HEAD AND TAIL CONTENT

This section speaks to how to handle the head and tail of content, common parlance for the beginning and end of a program or piece of media. A range of challenges may be present, and it is important to communicate how the vendor should handle each of these scenarios.

The first scenario that occurs is that there may be content at the beginning and/or end of program material that may or may not be relevant or related to the content of interest for preservation. For instance, the beginning of a piece of media may have minutes of black or silence before there is any program material. Or 20 minutes of program material might be recorded onto a 60-minute piece of media, leaving 40 minutes of black or silence at the end. Or there may be bars, tone, countdowns, or slates before the program material. In almost all cases, no organizations are interested in a vendor recording extended periods of black or silence. Usually organizations want to maintain bars, tones, slates, and the like for the preservation master. It is important to let the vendor know how they should handle these issues when encountered. Note that it is

common for an SoW to state how many seconds of black or silence a vendor should record at the end of a program before stopping the recording. However, it is important to include in the SoW that the vendor is responsible for ensuring that there is no additional content recorded after a period of black or silence to avoid missing content in the digitization process.

The above speaks to how to handle the creation of the preservation master. Specifications used for the preservation master may not apply to mezzanine or access copies. Depending on how these lower-resolution target formats will be used and by whom, the client may choose to eliminate anything that is not content of interest (e.g. bars, tone, slate).

REFERENCE FILES

As discussed in the introduction, one common and significant challenge when working on digitization of legacy media is the issue of not having a reference of quality for the original recording due to lack proper equipment, expertise, and labor resources. Therefore, when a digital file is returned from a vendor it is difficult to assess the quality of the work performed. There may be indicators of possible issues, but it is rare that qualitative review yields evidence of obvious errors on behalf of the vendor. Qualitative indicators that are identified in quality control often lead to an exchange with the vendor and further investigation that may confirm whether or not an issue is a vendor error, part of the original recording, or perhaps due to degradation. However, this confirmation is based on the review and feedback of the vendor. The client has little control to make their own assessment and judgment on the matter. This is simply a reality of doing this work, and because clients have little control when it comes to evaluating the output, it is necessary to have control over the processes and practices. This is the reason for the great level of detail and consideration given to these matters throughout an SoW.

Another way to mitigate the risks associated with not having a reference for quality of the original recording is to provide a known reference signal (e.g. bars, tone, black, silence) and have the vendor input it to the same system(s) that will be used to digitize client materials. The vendor is asked to replace the output signal of the playback device

17 <https://mediaarea.net/en/MediaInfo>

used to reproduce client content with the output of a signal generator that outputs the specified reference signals. The specified reference signals are then routed through the exact same signal path and captured in the exact same way that the client's audio or video signal will be routed and captured. This provides some insight to the client on the quality of the signal path and systems that are being used by the vendor. The client can utilize test and measurement software such as waveform monitors, vectorscopes, oscilloscope, audio level meters, and other tools to assess the quality of the recorded signal. Because the input signal is a known reference, variances between the input signal and the recorded output signal can be identified. In essence, it is letting the client know what impact the vendor system is having on the signal, which speaks directly to the quality of the signal path and system, as well as the degree to which it is in good operating condition and calibrated.

This is not a catch-all by any means. This is something that is done periodically, likely under the careful attention of the vendor, and may not be representative of the many tens, hundreds, or thousands of hours that this same system may be put to use for digitizing original materials. It also only speaks to the part of the system that follows the playback device, which means that it is not speaking to the quality and condition of the playback device, or the quality of the setup, calibration, and alignment performed by the operator. However, while it is not perfect, it is still far better than having no insights or reference points. This section of the SoW should provide detailed specifications for the reference signals to be used, the period of time that each should be played, and the precise order in which they should occur. It should also identify how frequently and under what circumstances to create reference files.

Typically, a client will ask for a reference signal per setup, or distinct signal path, being used for digitization. For instance, if the vendor is using four Betacam SP decks and digitization stations and two VHS decks and digitization stations, then this would yield six reference files. A distinct signal path may also be defined by the equipment that is used in between the playback deck and capture station.

When the reference file is digitized, the specifications used should be the same as the specifications for the

preservation master. It is also recommended that the client request that mezzanine and access copy files be made using the same method, hardware, software, and protocols that will be used with the client's original materials. This provides an opportunity to assess both qualitative and quantitative aspects of the derivative creation process as well, utilizing a known reference signal.

For small projects, it may be the case that only one set of reference files per setup needs to be created for the entire project. For longer term projects, it is advisable to identify the frequency with which reference files should be generated. This may be at the batch level if the project is being split into batches, but it could also be weekly or monthly. Whatever the case, it is wise to have periodic snapshots of the vendor's system, as opposed to just one, to ensure that the equipment is in good condition and that the system is properly calibrated and aligned for the duration of the project.

The above approach works well for audio and video, but film is somewhat trickier. For one, there are no widely adopted test films to fulfill this purpose. While test films with different test patterns and signals have been produced from time to time, there is no standard test film that has emerged in the way that test recordings became the norm for audio and video, or even for still image digitization, which has widely-adopted test targets. At this point there is no common test protocol that is used in film digitization projects that is similar to what is discussed above. Creating test films for use on film scanners would be a resource-intensive endeavor, which is certainly a contributing factor for why they are not in wide use.

On the plus side for film, there is often much greater insight into the quality of the original film. Because a client can see the images on a film and can see if a film is deformed or scratched, the disconnect that exists for audio and video is diminished for film.

Finally, in addition to the items mentioned above regarding the contents of this section of the SoW, the client should specify organizational and file naming conventions for the reference files and any associated embedded or external metadata to be delivered along with them.

DIRECTORY STRUCTURE AND FILE NAMING

Digitization projects produce files of all types (e.g. media files, metadata files), and the client must provide specifications for the vendor on how to name and organize them. When establishing file naming conventions for a collection, most people think in terms of newly-derived files reformatted from other sources. In reality, there will be more and more born digital content that already has filenames deposited with archives. In some cases, this content can be renamed to fit the archive's naming structure with no loss of information, but in other cases, such as with MXF files, the inherited naming structure refers to complex file and directory structures that must be maintained in order to preserve the entire content. Naming structures should be flexible enough to recreate any necessary naming conventions.

There are multiple questions that should be answered before drafting organizational and file naming specifications.

- **Are there existing organizational and naming conventions?**

Many organizations already have organizational and naming conventions that can be adopted or adapted. Some organizations are unhappy with their current organizational and naming conventions, so this may be an opportunity to address that. If you choose to depart from existing conventions, be sure to maintain awareness of any internal systems and processes that may have a dependency on the existing organizational or naming conventions and update accordingly. If there are no current conventions, then the opportunity exists to start from a blank slate and create a new convention without concern for inconsistency or incompatibility with internal systems and processes. For tips on file naming, see below.

- **What is the most persistent and pervasive identifier in use in the client organization?**

Naming conventions for directories and files are often based on the identifiers for an object. It is also common for organizations to have multiple identifiers for items. When selecting one to use, give thought to which of these is the most persistent and pervasive, both historically and moving forward. Are any systems or processes dependent on any particular identifier? Sometimes organizations have no identifiers or find that existing identifiers are unsuitable. In this case it may

be appropriate to establish a new identifier scheme and assign new identifiers to items; however, if there are existing identifiers, check to see if they will, in fact, work. Do not rush into creating a new identifier scheme without first diligently thinking it through, as the implications are far reaching, and ending up with too many identifiers can be a problem in and of itself.

- **Will this be the SIP or AIP, or will this organization and file naming be an interim specification?**

SIP and AIP are terms from the Open Archival Information System (OAIS) that stand for “submission information package” and “archival information package” respectively. Although these terms have specific meaning within the context of OAIS, a SIP can roughly be thought of as the components (e.g. metadata and media files) that make up the submission to the archive for longer term storage. These components may undergo further processing as part of the ingest routine before establishing an AIP. An AIP can roughly be thought of as the components that make up the collection of items that the archive will manage over the long term.

Some organizations create specifications for deliverables of a digitization project that equate to a SIP. Once the deliverables are received from a vendor, the client will put them through their own internal processes to generate an AIP. These processes may end up altering the organization and naming of the directories and files delivered from the vendor. In this case it is important that the directories and files are organized and named in way that is aligned with internal systems and processes and that makes the ingest and processing as efficient as possible.

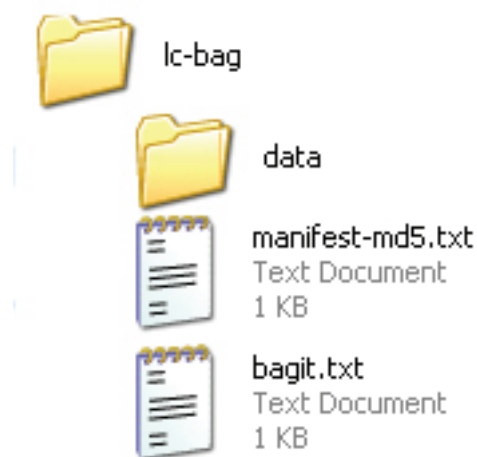
Other organizations create specifications for deliverables of a digitization project that equate to an AIP. They ask the vendor to organize and name their directories and files in such a way that that no further reorganization or naming is necessary to create the AIP. This approach requires consideration of more than just organization and naming of directories and filenames; it requires consideration of the metadata deliverables and parsing through which fields will be produced by the vendor and also of which fields must be provided by the client. (See information about the Source Metadata document under Communication Protocol, above.)

- **To bag or not to bag?**

BagIt is a file packaging format standard developed by the Library of Congress and the California Digital Library and maintained by the Internet Engineering Task Force. BagIt was originally developed as a means of sending and receiving files in a “bag” that would provide built-in data integrity verification for the recipient of the data. In other words, it allows someone who is receiving files to verify that all of the files sent to them are present and have not been altered since the sender created the bag. Another use case that exploits features of the BagIt specification is longer term storage in archives. Routine file attendance and data integrity verification are activities of a digital archive, and some organizations choose to use bags as an encapsulated AIP.

FIGURE 3.1
Bag File Structure

Credit: <https://blogs.loc.gov/thesignal/files/2015/03/bags.jpg>



Whether the vendor’s deliverables are being treated as a SIP or an AIP by the client, bags may prove useful. If being treated as a SIP, a bag can be used to verify file attendance and data integrity by incorporating a bag validation tool into the ingest routine. If being treated as an AIP then file attendance and data integrity can be validated upon receipt, and on an ongoing basis.

At its core, the BagIt specification provides file attendance and data integrity validation through the creation of checksums for every file and a manifest that documents all files in a bag. If the client chooses not to use bags in the deliverables, then it will be important to provide

an alternate specification for the provision of manifests and checksums from the vendor so that file attendance and data integrity can be validated upon receipt by the client.

- **Film**

As is the case throughout this document, film proves to be somewhat of a different animal. Digitized audio and video produce a single file from a single original item, with some exceptions for multiple sides of an item or very long recordings. With film, the most widely adopted preservation master formats produce a file for each frame of the film. This results in many thousands of files from a single film, often stored in a directory that essentially serves as the “wrapper.” Each of the frame-level files are named using a common base name with sequential modifiers appended. These names cannot be modified because players are dependent on the file names in order to play them back properly. The directory is typically named with only the base file name. Some organizations choose to specify that a bag be created from the parent directory so that there is some encapsulation of the directory that provides an automated way to validate file attendance and data integrity at the level of the film.

In addition to being an organizational and file naming consideration, the deliverable of the preservation master for film has logistical implications. Because the structure is different, it may take additional planning when determining how the files will be managed and used and whether or not the difference in structure will have implications.

METADATA

Metadata generation is a byproduct of any digitization project. Whether and how the generated metadata is captured and delivered is up to the client and must be specified. Metadata specifications for projects generally come in two types: external and embedded.

External

External metadata is delivered as a sidecar file, usually either in XML, CSV (spreadsheet format), or a combination of the two. If using a spreadsheet format, CSV is advisable over any proprietary format because it is a transparent

and widely adopted format for data exchange that can be opened in any text editor or spreadsheet application.¹⁸

In addition to specifying the format for delivery, the client must specify the fields, vocabularies, and structure of the metadata being delivered. The types of information encompassed in external metadata deliverables include technical, administrative, preservation, and descriptive metadata.¹⁹ Utilization of standards helps avoid reinventing the wheel and saves some time and effort. Sometimes organizations will adopt a standard as a whole, and other times organizations will choose to use standards in part, selecting

fields or sections along with their associated vocabularies from one or more standards. Organizations often have their own fields and vocabularies that they want included in the metadata deliverable as well. Some organizations may request a single master metadata deliverable while others may ask for multiple metadata file deliverables. There is no single right way to come up with a metadata specification. The right way will be whatever proves to be functional within the client organization. Here are some considerations that will help guide an organization to an answer:

FIVE TIPS FOR FILE NAMING

1. Beware of filename dependency

Filenames are not actually part of the file but rather are part of the file system and are therefore not dependably persistent over time and across systems. Instead, the Unique ID (UID) assigned to the object should be the constant identifier used to track and maintain the provenance of the file. The UID may end up being the same as the filename, but regardless, be sure to embed the UID inside the file in an appropriate and documented place.

2. Do not overthink filenames

Whether the filename is a randomly generated value or not, be systematic. Think, “Is this logical? Can I spell out the rules easily enough to do batch renaming?” In trying to create the perfectly contained and expressed filename or UID structure, there is a strong temptation to overthink to the point that they become non-systematic or too idiosyncratic to be logically parsed. If a naming structure is not systematic enough to have a piece of software perform a series of logical renaming steps, then manual renaming and retyping of filenames may be required at some point in the future.

3. Do not use filenames as database records

Filenames are not the place to cram in a bunch of descriptive and structural information. That’s what databases are for! All we require from a filename and ID is that they act as a link to the database record for that unique object. Trying to cram excessive descriptive information into a filename creates unwieldy names and is often futile because of how often conditions or conventions change and new scenarios come up over time. Having filenames that are tied too closely to specific scenarios creates inflexible structures that require non-systematic revision when situations change, which creates the predicament described in tip #2.

4. Keep filenames machine-readable

There is often an urge to make a file naming structure decodable by humans, but it also needs to be decodable by computers. Avoid characters that are not URL compatible, that require escape characters, or that are reserved by operating systems. Limit options to numbers, letters, periods, and underscores.

5. Consider existing filenames

As noted above, the inherited filenames structure of born-digital content may refer to complex file and directory structures that must be maintained in order to preserve the entire content.

18 See “Managing Data in Spreadsheets” for a tutorial on how to avoid accidentally corrupting data when managing it within spreadsheets. <https://www.avpreserve.com/papers-and-presentations/managing-data-in-spreadsheets-tutorial>

19 “Chapter 4: Managing Digital Audiovisual Collections” goes into these types of metadata further.

How will the metadata deliverables be used?

Considering the questions and points posed below will help determine:

- The format(s) to specify in the SoW
- The fields, vocabularies, and structure to specify in the SoW
- The number and types of external metadata deliverables to specify in the SoW

Questions and Considerations:

- Will deliverables be used as the AIP and kept in packages for long-term storage as-is upon receipt from the vendor?
- Will metadata be ingested or imported into systems? If so, what are the specifications for import into those systems?
- Is the client able to transform the metadata deliverables to meet any system import specifications as necessary, or do they need the vendor's deliverables to conform with the system import specifications?
- Is it easier to send the vendor pre-existing metadata to have them incorporate into their deliverables along with the metadata they are generating, so either (a) the metadata can be ingested in a streamlined way all at once instead of combining things together after the deliverables are received, or (b) the vendor is delivering a package that meets the client's AIP specification? The client may even want to ask for external metadata deliverables that are included as part of an AIP in addition to metadata deliverables that are delivered outside of an AIP that can be used for ingest into systems. If planning on delivering metadata to the vendor for incorporation into their deliverable, the client should specify this as part of the SoW and include it as part of the Source Metadata document(s) sent to the vendor.

What information is desired for capture?

Considering the questions and points posed below will help determine:

- Who will populate which fields?
- When will those fields be populated?

Questions and Considerations:

- In order to put some shape around an answer to this question, standards and metadata used in existing internal systems can serve as a guide. This might also be a good time to start capturing fields that are needed but have not been historically captured. Some standards that are typically used or referenced include:
 - PBCore (<http://pbcore.org>)
 - reVTMD (<https://www.avpreserve.com/products/avi-metaedit-revtmd>)
 - AES 57 (<http://www.aes.org/publications/standards/search.cfm?docID=84>)
 - PREMIS (<https://www.loc.gov/standards/premis>)
- Once there is a list of fields and associated vocabularies, think about which fields it makes sense for the vendor to capture. Of the fields that remain, which metadata already exists in some form? If the balance of fields contains required fields, there will need to be a determination of where and how those will be populated.
- When deciding which fields it makes sense for the vendor to populate, a typical approach is to have them capture any identified fields that would be permanently lost if they weren't captured at the time of digitization (e.g. playback devices, settings used for the digitization process, transfer operator, reproduction issues) and any fields that would be more efficiently captured by the vendor than by the client (e.g. original item specifications, duration).
- Many organizations inquire about whether or not they should ask vendors to generate descriptive metadata, such as describing the content of a recording during the transfer process. Most of the time this does not make sense for two reasons: 1) Vendors are not subject matter experts in the content being digitized and are likely only able to provide the most generic descriptive metadata, and 2) digitization workflows have shifted dramatically toward high throughput workflows that do not lend themselves to systematically capturing descriptive metadata. This question is usually based on the perception that because the vendor is already watching the material, they can easily document what they are seeing. In this way it appears to be an easy add-on or a potential byproduct of the digitization process; however, this is

not usually the case. Descriptive metadata generation should be considered as a separate service with its own set of specifications and special systems and expertise.

When and how will metadata deliverables be validated?

Considering the points posed below will help determine:

- The number and types of external metadata deliverables to specify
- When and how quality control will be performed on these deliverables

Validation is the process of vetting external metadata deliverables to make sure that they conform to the defined fields, vocabularies, and structure that make up the specification. One advantage to using standards in whole and having metadata delivered as independent files is that there is often an XML Schema Definition, or XSD, for a given standard. An XSD provides the rules against which an instance of a document claiming to be in conformance with a given standard can be validated. This may be useful not only to determine that the metadata received meets the specified standards, but also for potential future use cases in which a standard for data exchange, migration, import, or transformation to an organization or system is required or useful.

If such potential future use cases are less of a priority, then validation may be performed in another way. The client might choose to create a local XSD file to validate against as part of quality control or ingest, or the systems importing the data might perform validation as part of the import process.

Embedded

Embedded metadata can be most simply defined as metadata that is stored inside the same file, or container, that stores the audiovisual signal to which that metadata refers.

In many ways, one can think of embedded metadata as the file-based domain's equivalent of the physical domain's labels, annotations, and written documentation stored inside of material housing, or even as "in-program" annotations such as audio and video slates at the head of a recording.²⁰

Every file format has distinct embedded metadata specifications and fields. For instance, there are different options for embedding metadata in WAVE²¹ files than there are in MP3 files.²² Embedded metadata is what enables the display of information such as artist and title in applications that play back media files. The primary goal of embedding metadata for the purpose of preservation should be to identify:

- the object in an instance where it is dissociated from its external metadata
- the holding organization
- the data source that holds information about the object
- the copyright status

The Federal Agencies Digitization Guidelines Initiative (FADGI) published guidelines for the use of embedded metadata in WAVE files,²³ which may be adapted to other formats accordingly.

There are a few additional considerations to keep in mind when it comes to creating embedded metadata specifications.

- Digital files that are acquired by an organization, rather than created through digitization, likely will have existing embedded metadata that was generated by people, software, and/or hardware prior to acquisition. In the interest of maintaining the authenticity of the original object, these files should undergo a different process with regard to embedded metadata in order to maintain the authenticity of the original object.

20 Joshua, Ranger. "Embedded Metadata in WAVE files: a look inside issues and tools." 2014. <http://www.avpreserve.com/papers-and-presentations/embedded-metadata-in-wave-files-a-look-inside-issues-and-tools>

21 Federal Agencies Audio Visual Digitization Working Group. *Task 5.4: Assess Options for Embedding Metadata in WAVE Files and Plan the Audio Metadata File Header Tool Development Project: Assessment Report and Initial Recommendations*. http://www.digitizationguidelines.gov/audio-visual/documents/AVPS_Audio_Metadata_Overview_090612.pdf

22 "ID3 tags are the audio file data standard for MP3 files in active use by software and hardware developers around the world." <http://id3.org>

23 Federal Agencies Audio-Visual Working Group. *Embedding Metadata in Digital Audio Files: Guideline for Federal Agency Use of Broadcast WAVE Files*. http://www.digitizationguidelines.gov/audio-visual/documents/Embed_Guideline_20120423.pdf

- While some file formats/wrappers have robust options and toolsets for reading and writing embedded metadata, others are lacking in this regard. For instance, MXF and DPX files, common target formats for video and film, present significant challenges. In cases like these, side-car files with the embedded metadata equivalent may be created as an interim solution, or the choice may be made not to have an embedded metadata deliverable for these files.
- When creating an embedded metadata specification, consider how most applications manifest and present this information. The choice of where to store the metadata within the file may be influenced by how accessible it is to users across an array of applications and tools.
- Embedded metadata can be fragile and may be accidentally erased and/or augmented if not handled in a considered way.²⁴

QUALITY ASSURANCE AND CONTROL

Vendor Quality Assurance

Quality assurance is actually addressed at several points preceding this section in the SoW, but this is the place to bring about clarity on two fronts. The first is to request details about the vendor's quality assurance protocols in their proposal. The second is to be explicit about protocols and practices required to perform the statement of work.

This is particularly important because differences in quality assurance protocols and practices are often the biggest driver of variation in pricing offered by vendors. It is important both to clearly understand the details of their quality assurance plan to gain insights into differences that may manifest in pricing and to make sure that each vendor provides a proposal that meets the core quality assurance requirements. This enables more of an apples-to-apples comparison.

Client Quality Control

In the same way that it doesn't make sense to create a rule that can't be enforced, there is little sense in creating

a specification that is not checked for compliance. The client and vendor are partners in ensuring success, and it is the obligation of the client to perform quality control in order to live up to their side of the partnership. There will be quality issues in every project; this is simply a reality. Having a reasonable number of errors does not mean that the vendor did a bad job. On the flip side of that statement, if the vendor did fail to perform quality work or there was a misunderstanding and misapplication of a specification and subsequently quality control is not performed, then the client is complicit in the poor outcome. It does not matter if a client and vendor have been working together for many years using the same set of specifications; quality control should be performed diligently and routinely for all work performed. With this in mind, it is important for the client to have a documented quality control protocol accompanying each SoW to ensure that it is performed comprehensively and consistently. At a high level, this protocol should provide a check for each specification and requirement in the SoW.

Quality Control Check Categories

The checks to be performed can be categorized as quantitative or qualitative and automated or manual. The categories that a given check falls into will determine the best method and approach for implementing it.

- **Automated Quantitative:** These are checks that are performed and reported with reliability based on logic applied by a software application. For example, checking a client-specified embedded metadata field in a file may be performed using an application such as MDQC.²⁵ Checking that a particular copyright statement is present in an embedded copyright field or checking that the sample rate of all audio preservation masters is 96 kHz are quantitative checks that can be performed in an automated way with a definitively reliable result.
- **Automated Qualitative:** These are checks that use algorithms to perform the job in an automated fashion and that require subjective judgment calls by humans to increase the reliability of reporting. For example, automated tools that identify and report on audio and

24 Association of Recorded Sound Collections (ARSC) Technical Committee. *A Study of Embedded Metadata Support in Audio Recording Software: Summary of Findings and Conclusions*. <https://www.avpreserve.com/tools/a-study-of-embedded-metadata-support-in-audio-recording-software-summary-of-findings-and-conclusions>

25 Metadata Quality Control (MDQC), <https://www.avpreserve.com/tools/mdqc>

video errors fall into this category. These types of systems may be useful for creating more efficient workflows, but they will over- and under-report issues and require human review.

- **Manual Quantitative:** These are quantitative checks for which either (a) there is no logic that can be implemented in an automated way or (b) the resources required to develop an automated routine are not justified by the level of manual effort required to do the same task. For instance, checking that the title field contains the correct title based on the original item or reviewing the file naming and organizational conventions of deliverables from a vendor fall into this category. Depending on the circumstances, these checks may or may not be able to be automated, or they may be able to be automated but there are so many conditions in the logic that it would be overly complex and resource-intensive to build the appropriate tool.
- **Manual Qualitative:** These are checks that require the subjective judgment of a human. These can be used in combination with automated qualitative checks or used alone. An example of the former is the human review of a report generated by an automated qualitative tool that identifies issues in the audio and video. An example of the latter is ensuring that the qualitative differences between the preservation master, mezzanine copy, and access copy for a given original item are within acceptable limits of variable quality.

Quality Control Resource Planning

Depending on the specifications used, the available data, and the tool set at hand, a quality control protocol consisting of a combination of these checks should be documented. Once the appropriate categories and associated methods are identified for each check, it is necessary to analyze the allocation of human and machine resources needed to make sure that the time it takes to implement the quality control protocol aligns with available resources.

Some checks will be computer- and labor-intensive, while others will require very few resources and can be performed quickly. For those that are quick and require few resources, it makes sense to perform them 100% of the time. On the other hand, resource-intensive checks will

require scaling down the percentage of files checked based on the available resources.

Consider the scenario where a staff member has 20 hours per week dedicated to quality control for deliverables coming in at a pace of 250 digitized items per week. Let us say that performing the audio and video quality checks takes five minutes per original item. At 250 items per week, this activity alone will require just over 20 hours. With all of the other quality control checks that must be performed, as well as the handling of media and other administrative tasks, it is not feasible to spend all 20 hours solely on the audio and video quality check. A solution to this is to perform the check on a sampling of materials. Say that after adding up the time that all of the other quality control checks and administrative tasks will take, there are only four hours that can be allocated toward checking of audio and video quality. At five minutes per original item, this means that there is enough time to perform this check on 48 items, or 19%.

Sampling is a widely adopted and perfectly acceptable approach, but it is important to keep a couple of things in mind. The first is that if the error rate of the sample set is high and/or there are consistent errors found, the sample size should be increased to track down the severity and extent of the issue. The second is that the higher the sample size, the greater the confidence and the lower the margin of error. While scaling and sampling can be used within reason, it does have its limits, and it is important for a successful outcome to provide sufficient staff for the performance of quality control. It is also important that the staff performing quality control has the appropriate expertise necessary to make subjective judgment calls.

The above text mentions human resources, and it is equally necessary to analyze and plan for the compute resources that are required to implement the quality control protocol. Like the human resource calculations in the previous example, the approach here is to identify the length of time it will take for computers to perform their job. This will make sure that there are no bottlenecks that may become barriers to the performance of tasks by staff or slow down the required throughput to meet the demand. For computer-intensive processes, the options may be to utilize sampling for certain processes or to increase the compute resources to scale accordingly.

Quality control documentation and rework

Documenting and resolving quality control issues can be confusing and challenging. It is advisable to have a shared system (e.g. Google Sheets or a custom software application) that allows both the client and the vendor to comment on quality control issues in real time, where there is no likelihood of version control issues, and where all current data is immediately available to all parties at the item level. Having an explicit quality control status is useful so all parties can see if an issue is pending resolution, requires rework, or has been resolved. Be sure that any rework goes through the quality control process again; do not make the assumption that rework must be correct. Also, consider updating data tape, such as LTO tapes, or hard drives that resulted in quality control issues and the need to perform rework.

Reference Files

There is a preceding section of the SoW that specifies reference files that the vendor will record through their system and that will give the client insight into the performance of the vendor systems. In order to check these reference files, it is necessary to procure software that will analyze these files. This likely includes software with video waveform and vectorscope displays and audio level meter and frequency analysis displays. Using this software to determine the extent to which the reference files meet standard targets will provide the client with the information they need to either identify potential issues or confirm the health of the vendor's systems.

Financial Planning for Quality Control

Implicit in the above discussion of human and compute resources is the need to plan and budget for resources within the client organization. Budgeting the appropriate amount of time for properly-skilled staff, as well as for specialized equipment and software, to perform quality control tasks is critical to ensuring the success of a digitization project. These should be planned for and included in the project budget from the outset.

DELIVERY

This section of the SoW defines several aspects of the delivery of media from the vendor to the client. Note that the file naming and organizational conventions

are integrally tied to this section. Specifications of the relationship between the vendor deliverable and the AIP and whether to bag or not should be incorporated here.

Media

The first item to specify here is the media and formatting that the vendor should use when sending the deliverables to the client. This is usually hard drives and/or data tape. Be sure to specify lower level requirements regarding file systems (e.g. NTFS, HFS+, LTF5) to ensure that data is able to be accessed and worked with upon receipt. Otherwise, specifications for delivery on hard drive is straightforward.

Delivery on data tape requires consideration of the following factors:

- *Hardware and software*

Assuming that the client is venturing into data tape for the first time or they are going to upgrade data tape formats, the client should first look within their own organization to identify the existing hardware and software infrastructure. If it uses current technology that makes good sense from a preservation perspective, determine if it is possible to share these resources and take advantage of the organization's investment. Sometimes there may be an overlap in technology but a big disconnect in policy, or there may be no additional capacity or bandwidth. For instance, it is often the case that an organization's IT department uses data tape for backup purposes. However, these solutions often utilize proprietary software (see next bullet point), have limited retention policies that don't align with preservation, and generally manage the data using very different workflows and practices when compared to a preservation environment. This is not always the case though, and sometimes there can be resource sharing that utilizes separate policies, practices, and management.

If there is no opportunity to take advantage of existing infrastructure within the client organization, or it does not make sense, then it is necessary to budget for the appropriate hardware and software to be able to access the data on the data tapes. The client must be sure to source this equipment and budget for it from the outset.

- *Proprietary versus open*

Writing and reading data to data tape requires an intermediate application. Until recently, these intermediate applications were all proprietary applications that utilized their own methodology for writing and reading data. This reliance on a proprietary third-party application to access data represents an increased risk to the preservation of that data. Within the past few years, LTFS (Linear Tape File System) was introduced for the LTO (Linear Tape Open) data tape format. LTO and LTFS are both open standards resulting from a cooperative effort across manufacturers to agree on and publish a set of standards for both a data tape format and the methodology for writing and reading data to the data tape format. This mitigates the increased risk of using a proprietary third-party methodology for accessing data on data tape. This may or may not be reason enough to select LTO and LTFS as the specification, but this general point should be thoroughly thought through as part of the specification process.

- *Implications to quality control protocol*

As opposed to hard drives which are instantly accessible, data tape, despite the claims of manufacturers, simply takes longer to work with. It often requires copying data from tape to a local drive in order to perform the quality control checks. This can add significant machine and labor time. In some cases, the bottlenecks that this creates may dictate the details of the quality control protocol and require more sampling than may be considered ideal. An alternative approach to scaling down the amount of quality control is to request that some items be delivered on both hard drive and LTO to enable faster quality control and ingest of materials.

- *Library system versus storage on shelves*

Data tape devices range from single tape drives to tape library systems that can hold many tapes. The former is fully manual and results in storing data tapes on shelves. This has some advantages, but it falls short of being able to implement a digital storage solution that meets the requirements of a digital preservation environment. The latter better meets the demands of a digital preservation environment because the automated nature of a tape library enables working across a large set of tapes, and this approximates the functionality found when working with disk-based servers and network

attached storage devices. The downside is that the cost and complexity is much greater compared to a single drive approach.

- *Migration*

Data tape formats have a limited lifespan. Some have roadmaps that define how frequently updated versions will come out and how much backward compatibility will be supported. Regardless of delivery media and storage choice, however, migration is a reality, and the greater amount of effort required with data tape requires additional planning. Thought should be given to when the migration should occur, the level of effort involved, and the costs to procure new hardware and staff or to outsource the effort.

Shipment guidelines

In addition to specifying the delivery media, this section should specify how the media is packed, shipped, and identified at the box level and media level. Specifications may include conventions for identifiers and where and how they should be applied. Furthermore, there should be specifications for the metadata that accompanies the media delivery, identifying the contents of the delivery media at the item level. This may be included in a shipment manifest, in the project management documentation, or both.

This section of the SoW should also specify when and how the original items should be shipped back to the client. Typically, originals are not shipped back until all quality control and rework is complete for all of those items. They are also sent separately from the delivery media to mitigate the risk of loss during the delivery process.

This is also a good place to reiterate who is responsible for performing and paying for the shipment and any shipment protocols or requirements.

Protocol for deletion

The client must be explicit about expectations for when and how the vendor will delete the client's data from their systems. There is a balance to be aware of here between the burden placed on the vendor who must store large quantities of data on their systems and the need of the client to thoroughly work through the quality control process and make sure that the data is safely on client systems before the vendor deletes their data. The turnaround time for client quality control and vendor rework is defined in

the timeline section of the SoW, but this is a good place to reiterate that timeline and to add explicit instructions that the vendor should not delete any client materials until given written approval to do so.

Importance of an Internal Ingest Protocol

Once quality control is complete, rework performed, and final deliverables received for a given set of items, there will be an ingest process. The details of the ingest process will depend on the given organization and may range from simply placing LTO tapes onto shelves and documenting the location to copying the data off and running it through extensive processes in order to prepare and deposit the data into digital preservation repositories and populate access systems. Whatever the case, think through the details of the ingest process. Consider the specifications of the deliverables and the delivery media and think through the handling of the media, the processing of any data, and the updating and population of systems. Document these details and analyze and calculate the human and compute resources required and the associated timeline. There may need to be meetings with others in the client organization or third party storage/system providers to figure out the details of ingest and to resolve any prospective issues. The quality control and ingest protocols go hand in hand and are a critical part of internal documentation.

Staffing

Aside from quality control and ingest of delivered items, there will be effort associated with receiving, reconciling, and reshelving the original items once they are received. Be sure to plan for this and staff the effort adequately.

RFP RESPONSE

In this section, the client communicates to prospective bidders how they should respond to the RFP. The SoW has many specifications and requests for specific pieces of information from the vendor. Providing a checklist for the vendor along with a general comment about how they should respond and what should be included in their proposal is useful for all parties.

PRICING INFORMATION

If the client does not provide specific direction on how bidders should provide pricing information, it becomes nearly

impossible to perform comparative analysis. It is also the case that there are unknowns on the client side regarding the quantity of items that will be digitized (this is usually dependent on the pricing that comes back from the vendors) and the program duration of the content on the media being digitized; therefore, it is valuable for the client to provide a structure for pricing information that is consistent across vendors and that builds-in variables that provide a better understanding of pricing under different scenarios. This may be accomplished by building a range of durations and quantities into the client-specified reporting structure for pricing. Any specified reporting structure should include costs for media, supplies, shipping, and any supplementary services that are needed.

QUESTIONS

The SoW is a detailed document with many parts that have had a great deal of consideration put into them. It is in the best interest of all parties to allow the vendor the opportunity to ask questions about the SoW. This will help them better understand the SoW and put together a representative proposal, or it may expose flaws in the SoW that need to be corrected. The number and quality of the questions asked also provides an indication of how thoroughly the vendor has read the SoW.

This section should reiterate the date by which any questions from vendors are due, who they should be delivered to, how they should be delivered, and the timeline and method for responses.

Many organizations choose to share all questions (anonymized) and all answers with all participating bidders. Others choose to simply respond privately to the inquiring bidder. Regardless of approach, it should be documented here so the vendors will know who will see their questions and the associated responses.

Examples of Statements of Work

The following guides offer examples of much of what is discussed above, and they can serve as references for developing an SoW. Note that the title of these documents uses the phrase “Request for Proposal.” In the context of this chapter we would refer to these as Statements of Work.

- Guide to Developing a Request for Proposal for the Digitization of Audio. <https://www.avpreserve.com/papers-and-presentations/guide-to-developing-a-request-for-proposal-for-the-digitization-of-audio>
- Guide to Developing a Request for Proposal for the Digitization of Video (and More). <https://www.avpreserve.com/papers-and-presentations/guide-to-developing-a-request-for-proposal-for-the-digitization-of-video-and-more>
- Digitizing Motion Picture Film: Exploration of the Issues and Sample SOW. http://digitizationguidelines.gov/guidelines/FilmScan_PWS-SOW_20160418.pdf

SECTION 5: POST-DIGITIZATION

The text above discusses the importance of creating an ingest protocol and coordinating with other stakeholders in the client organization (and perhaps partner organizations) who are depended upon or directly impacted. Also discussed as part of the ingest protocol was walking through the details of populating client systems with media and metadata. All of these are critical elements of properly planning for post-digitization. One aspect not yet covered is how to calculate storage requirements for each of the target systems, in total and over time. This will be an important piece of data for IT infrastructure planning, for budgeting for the delivery media in the project, and for logistical planning for quality control, ingest, and longer-term storage.

The basis for calculating storage requirements will be an inventory of items selected for digitization. Populate a spreadsheet with the following columns:

- Format
- Quantity
- Estimated average duration (use media duration if program duration is unknown)
- Estimated total duration (quantity x estimated average duration)
- Preservation master file size (GB/TB per hour/min x estimated total duration)
- Mezzanine copy file size (GB/TB per hour/min x estimated total duration)
- Access copy file size (GB/TB per hour/min x estimated total duration)

- Total file size (PM file size + Mezz copy file size + Access copy file size)

Summing across all formats will provide the total required storage capacity. If the capacity of the delivery media is known, then the quantity of media to purchase can be calculated, and the client can plan for what will be received.

Knowing total storage capacity is helpful, but a lower level piece of information that is more useful when planning for IT storage infrastructure is how that data will be produced over time. This is much more true for large projects than small projects where everything will come in at once.

To calculate the storage capacity growth over time, the client organization will need to have either specified or have a sense of the frequency and quantity of batches being delivered from the vendor. Note that this same information is also useful for planning for quality control and ingest staffing. Once the frequency and quantities (and possibly formats) of deliveries are known, the spreadsheet created to calculate total storage capacity can be used to calculate storage capacity growth over time. Furthermore, if it is known which target formats will populate which target systems, this calculation can be performed at the target system level.

CONCLUSION

Managing a successful reformatting project requires effective communication, a basic knowledge of AV digitization techniques, and a deep understanding of how to work with a vendor. Creating a Request for Proposal and a Statement of Work informed by a detailed inventory and guided by the walkthrough detailed in this chapter will help to ensure that project goals are understood by all and that files are properly delivered to the institution. Additionally, a process for quality control, rework, and ingest should be in place to make sure that returned files are vetted and effectively stored in the repository.

It will be the responsibility of the institution to establish a sustainable preservation program to ensure that the delivered files are well cared for long into the future. This requires planning, policies, a storage infrastructure, active management, and metadata. These topics as well as basic concepts in digital preservation are described in “Chapter 4: Managing Digital Audiovisual Collections.”

GLOSSARY

Active management: The performance of consistent and ongoing digital preservation activities (e.g. fixity and validation) to ensure a digital file's continued access for as long as necessary.

Artifact: Anomalies during visual or aural representations of recordings.

Audit trail: The information associated with a digital file that tracks the transactional history of it from the point of capture or ingest to know whether it has been managed without change to the bits that make it up and according to relevant policies and standards.

Authenticity: The quality of being genuine and free from tampering and is typically inferred from internal and external evidence, including its physical characteristics, structure, content, and context.⁵⁰ Trustworthiness.

Back coat: Layer added to some magnetic tape to help support the magnetic recording layer. The back coat reduces tape friction, dissipates static charge, and reduces tape distortion.

Binder system: System through which magnetic particles are held by a binder to a substrate layer.

Bit rot: The corruption, loss, or decay of bits, the building blocks of digital files.

Carrier type: Refers to the physical carrier of the AV material. Examples of carrier type include reels and cassettes.

Checksums: Alphanumeric strings that reflect the uniqueness of every digital file.

Curation: The activities that are performed on a digital file throughout its lifecycle, including selection and appraisal, description, ongoing care and management, long-term access, and/or deaccessioning/disposal.

Degradation: The process in which the quality or integrity of an object is destroyed over time.

Delamination: In disc media, the process that causes layers to separate from the support base.

Digital preservation: The active management of digital content over time to ensure ongoing access.⁵¹ It is an integral part of curation (see definition above).

Digitization: The representation of an object, image, sound, moving image, or document by generating a series of numbers that describe a discrete set of its points.

File attendance: Ensuring that there are no missing or unexpectedly present files in a given location.

Fixity: File fixity refers to the property of a digital file being fixed, or unchanged. Fixity checking is the process of verifying that a digital object has not been altered or corrupted.⁵²

Governance: In the informational sense, governance is the set of structures, policies, procedures, processes, and controls implemented to manage information at an enterprise level, supporting an organization's immediate and future regulatory, legal, risk, environmental and operational requirements.⁵³

50 "Authenticity." Glossary of Archival and Records Terminology. Society of American Archivists. <http://www2.archivists.org/glossary/terms/a/authenticity>

51 "About." Digital Preservation. Library of Congress. <http://www.digitalpreservation.gov/about>

52 Wikipedia. File Fixity. https://en.wikipedia.org/wiki/File_Fixity. Accessed December 29, 2016.

53 Wikipedia. Information governance. https://en.wikipedia.org/wiki/Information_governance

Ingest: The process by which digital files and their associated metadata (called a Submission Information Package, or SIP) is deposited or submitted into a digital repository.

Latency: In computer networking, latency is the time interval between the request for information, such as a digital file, and the retrieval or display of that file to the user by the system.

Machine transport: Playback equipment.

Mandrel: A cylindrical rod placed through a cylinder and used to rotate it for playback.

Media type: AV materials are classified as audio, video, or film during the cataloging and inventory processes.

Metal evaporated tape process: Process in which magnetic particles are vaporized from a solid and deposited onto a substrate layer.

Migration: Converting from one format to another format considered to be of greater stability.

Obsolescence: The state of being which occurs when an object or practice is no longer wanted or used. Usually occurs when a new technology supersedes the old.

Preservation planning: A process by which the general and specific needs for the care of collections are determined, priorities are established, and resources for implementation are identified.

Refreshing: Copying information content from one storage media to the same storage media.⁵⁴

Reproduction method: Method in which a recorded signal is played back from a physical media object.

Risk management: The systematic control of losses or damages, including the analysis of threats, implementation of measures to minimize such risks, and implementing recovery programs.⁵⁵

RPM: Rotations per minute. Used to indicate recording speed for discs and cylinders.

Sidecar file: A file that is stored next to the AV file in the same directory.

Signal path: The route that an audio signal travels from source to output. This may be within a single device (CD to speaker within a stereo system) or within a workflow (original audio recording to reformatted digital file).

Slipping: Tape pack problem in which either single strands or groups of strands are misaligned and migrate to rest against the edge of the flange. May cause edge damage to the tape or film.

Splice: When two ends of a tape or film are joined together using specially formulated splicing tape.

Sticky shed syndrome: A condition resulting from the deterioration of the binder in magnetic tape that results in gummy residues on tape heads during playback.⁵⁶

Storage architecture: The computing and network infrastructure required to store digital files.

Storage capacity: The amount of data a storage device can hold, often measured in gigabytes (GB), terabytes (TB), and petabytes (PB).

Storage media: Devices on which data is stored. These include computer hard disks, optical disk drives, USB drives and other external hard drives, DVDs, and magnetic data storage tapes.

Stylus: A hard point following a groove in a phonograph record and transmitting the recorded sound for reproduction.

Substrate: The backing film needed to support the magnetic recording layer of a magnetic tape.

Tails out: A method for winding tape onto a reel where the end of the tape is on the outside.

54 Digital Preservation Coalition Digital Preservation Handbook Glossary <https://dpconline.org/handbook/glossary#R>

55 "Risk Management." Glossary of Archival and Records Terminology. Society of American Archivists. <https://www2.archivists.org/glossary/terms/r/risk-management>

56 "Sticky Shed Syndrome." Glossary of Archival and Records Terminology. Society of American Archivists. <https://www2.archivists.org/glossary/terms/s/sticky-shed-syndrome>

