



Digital Preservation Workflow Curriculum Development

Findings Briefing and Workshop Outline

Submitted By:



AVPreserve
253 36th St, Suite C302
Brooklyn, NY 11232

Telephone 1.917.475.9630

amy@avpreserve.com
kara@avpreserve.com

December 15, 2016

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Attribute the Digital Preservation Network (DPN) and AVPreserve when citing or reusing this work. <http://www.dpn.org>

Table of Contents

Introduction	3
1. Environmental Scan Findings	3
Digital Preservation Training Comparison	3
Audience.....	4
Topics/Focus	5
Documentation Review	6
Guidebooks/Curriculum	6
Publications	7
2. Interview Findings	7
Interviews conducted	8
Issues discussed by interviewees	8
Workflows	8
Selection.....	8
File Organization	9
Awareness.....	9
Tracking.....	10
Metadata.....	10
Change	10
Potential audience for workshop	10
Content-specific challenges	11
Potential workshop topics / existing gaps in training landscape	11
Other suggestions	12
3. Recommendations	12
Themes / Focus	12
Structure and Delivery	13
Audience	13
Modules: Outline	13
1. Overview / Setting the Stage	14
2. Selection for Submission	15
3. Preparing Content for Submission.....	15
4. Submission	16
5. After Submission	16
Case Study Problem Solving Structure	16
Prerequisites for workshop	17

Introduction

This briefing is the result of research into current digital preservation training and offerings in North America and the UK, as well as digital preservation practice at institutions throughout the US. It (a) summarizes AVPreserve's research in preparation for the development of a modular workshop curriculum that will be aimed at practitioners involved in life-cycle management of valuable digital resources, and (b) provides recommendations and an outline for a workshop curriculum for DPN. In order to determine appropriate scope and content for the workshop, AVPreserve conducted four interviews with digital preservation practitioners and trainers, surveyed the existing digital preservation training landscape, and analyzed the results. Our aim was to identify gaps in the current training landscape, the needs of practitioners, and challenges unique to specific content types.

This document contains three sections that present the work that has been done to this point and which establishes a path forward for the development of the workshop curriculum:

1. Environmental scan findings
2. Interview findings
3. Workshop curriculum recommendations

1. Environmental Scan Findings

The environmental scan involved a comparison of existing digital preservation training offerings (Appendix A), as well as a review of recent publications (e.g., blog posts, guidelines and handbooks, survey results) about training. The focus of this research was on materials in English, since the DPN workshops will be presented to a primarily English-speaking audience.

Digital Preservation Training Comparison

In order to better understand the existing digital preservation professional development landscape, we analyzed several workshops, tutorials, and online courses. This study was focused on short-term, English-language trainings aimed at working professions, and excluded undergraduate and graduate-level semester-long courses. The training options reviewed did not include one-off opportunities, but those that are offered repeatedly or on an ongoing basis, as with some online courses and tutorials.

A summary of findings is presented here. Detailed analysis is available in **Appendix A: Professional Development Comparison**.

The following trainings were reviewed:

Training Name	Host Institution
A Beginner's Guide to the OAIS Reference Model	University of London Computer Centre
Archivematica Camp	Artefactual
DigCCurr Professional Institute: Curation Practices for the Digital Object Lifecycle	UNC Chapel Hill SILS
Digital Curation Workflow and Tools	Lyrisis
Digital Curation: Fundamentals for Success	SAA
Digital Directions: Fundamentals Of Creating And Managing Digital Collections	NEDCC
Digital Preservation Management	MIT Libraries
Digital Preservation of Audio and Video: Fundamentals	SAA
Digital Preservation Training Programme (DPTP) Online 2016	University of London Computer Centre
Digital Stewardship Fundamentals	Lyrisis
DPOE Train-the-Trainer Workshops	Library of Congress
DPTP: An Introduction to Digital Preservation	University of London Computer Centre
DPTP: The Practice of Digital Preservation	University of London Computer Centre
From Theory to Action: A Pragmatic Approach to Digital Preservation Tools & Strategies (Digital POWRR)	Northern Illinois University
Intro to Digital Preservation	Preserve This
Introduction to Preservation Metadata	Lyrisis
Practical Digital Preservation One-Day Workshop	Archives and Records Association with Preservica
Preserving Digital Archives	SAA
Research Data Management and Sharing	UNC Chapel Hill SILS; Digital Curation Centre

Additionally, there are a variety digital preservation workshops held around conferences each year on rotating topics, including those held at the iPRES, NDSA, PASIG, and others. Typically these are one-off workshops, developed specifically to be presented at a single conference and not repeated. These were not included in this review.

Audience

Audiences for training offerings were reviewed as part of the analysis. In most workshop descriptions, several professional roles are indicated as the target audience, capturing a broad range of job titles in the library and archive professions. Each role listed below was specifically mentioned in the training descriptions:

Role	Mentions
Archivist	8
Librarian	5

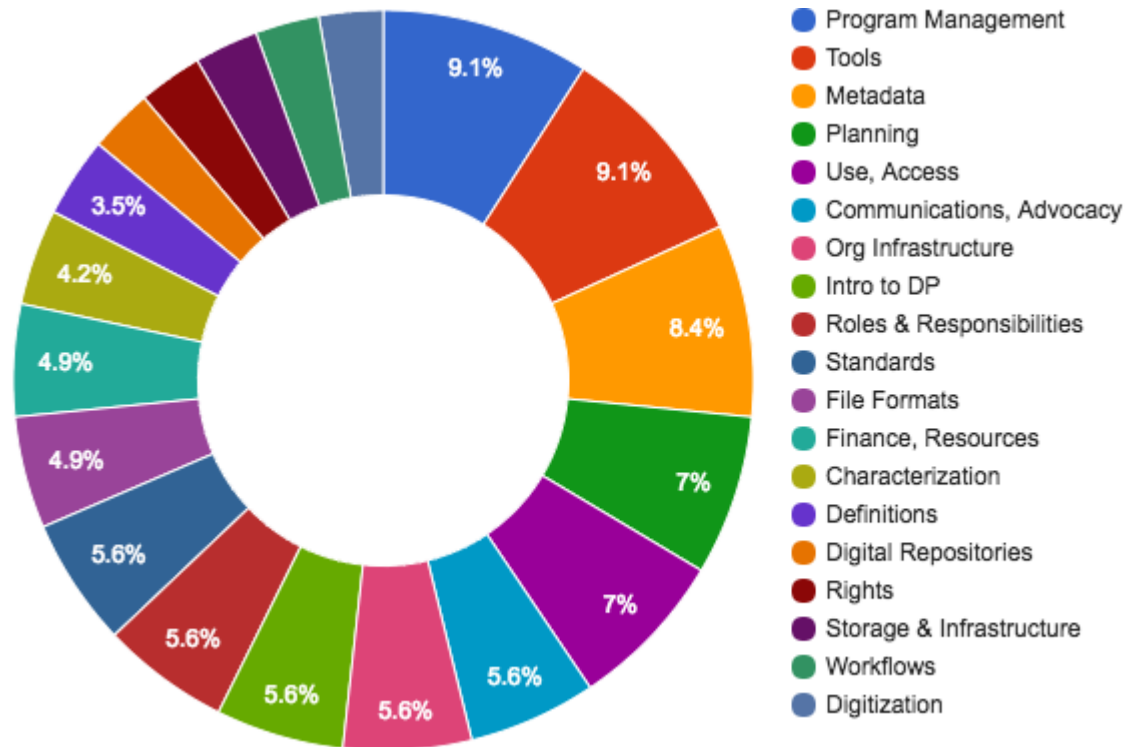
Records and Information Manager	4
Collection Manager	3
Digital Curation/Preservation Practitioner	3
Information Professional	3
Curators	2
Practitioner, general	2
Administrator	1
Conservators	1
Developer/System Administrator	1
Digital Librarian	1
Digital Preservation Manager	1
Library Director	1
Managers, general	1
Metadata Technician	1
Museum Professional	1
Registrar	1
Visual Resource Curator	1

While descriptions do not always explicitly state the skill levels to which the training opportunities are directed, they can be summarized as:

Level	Occurrence
Intermediate	5
Beginner	1
Beginner/Intermediate	3

Topics/Focus

Professional development trainings are focused on a wide variety of topics. We analyzed the topic of each module within the trainings reviewed, and found the following distribution of topics. (It should be noted that some modules were assigned more than one topic.) The chart below includes topics that were identified in three or more modules.



Documentation Review

Guidebooks/Curriculum

Our scan of existing documentation revealed that there are several online handbooks, curricula, and guidelines openly available, including, amongst others:

- “Digital Preservation Handbook” by the Digital Preservation Coalition¹
- “Curriculum Framework for Digital Curation” by DigCurv²
- “Digital Preservation: Sustaining Media Art” by Matters in Media Art³
- “Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems” by MIT Libraries and Cornell University⁴
- “Curation Reference Manual” by DCC⁵
- “Strengthening and widening the European Infrastructure for Social Science Data Archives” by the Consortium of European Social Science Data Archives⁶
- PERICLES Modular Training Package⁷

¹ <http://handbook.dpconline.org/>

² <http://www.digcurv.gla.ac.uk/>

³ <http://mattersinmediaart.org/sustaining-your-collection.html>

⁴ http://www.dpworkshop.org/dpm-eng/eng_index.html

⁵ <http://www.dcc.ac.uk/resources/curation-reference-manual>

⁶ <http://cessda.net/CESSDA-Services/Projects/CESSDA-SaW>

⁷ <http://pericles-project.eu/training-module/>

Publications

One of the most informative publications we reviewed was the summary of the 2015 Digital Archiving & Preservation Training Needs Survey, conducted by the University of London Computing Center along with the Digital Curation Centre (DCC) and the Digital Preservation Coalition (DCC) in the UK.⁸ 216 respondents provided input on what they would like to see in future digital preservation training offerings. The blog post on the findings summarized the takeaways as follows:

- People want to **learn about strategy and planning**, not exclusively Digital Preservation theory, not exclusively technology.
- People are clear that Digital Preservation training will bring them benefits directly related to their job / organization / collections.
- People want to **learn by doing**; they want to **see case studies** of how others have done it.
- People would **rather attend one-day courses** than try and get a week off work.
- The price of a course is less of a barrier to attendance than one might think.
- An online offering would be welcome, but there's no strong evidence that it would be popular.
- Everyone wants to know more (there's an appetite for learning), and everyone **wants to feel confident** about Digital Preservation.

Some of these conclusions match what we heard during our interviews, as described below.

2. Interview Findings

This section summarizes what was expressed by individuals interviewed for this project. Each interviewee was asked to comment on the following topics:

- Experience with day-to-day challenges of digital preservation processes, including selection, metadata production, packaging for ingest, ingest activities, and ongoing management
- Format-specific challenges to digital preservation that they encountered, including text, still images, and audiovisual content
- Whether their institution is a part of a digital preservation consortium (e.g., DPN, APTrust, MetaArchive) — *if applicable*
- Whether they have submitted content to DPN — *if applicable*
- If the interviewee or their staff has attended any professional development programs focused on digital preservation, and if so which
- Interviewee's perspective on gaps in existing training and needs of practitioners

⁸ <https://dart.blogs.ulcc.ac.uk/2016/02/10/2015-digital-preservation-training-needs-survey-what-we-learned-so-far/>

Interviews conducted

The following interviews were conducted specifically for this project:

Name	Position	Institution	Date
Rosalyn Metz	Director of Library Technology and Digital Strategies	Emory University	2016-11-14
Sam Meister	Preservation Communities Manager	Educopia	2016-11-14
Daniel Johnson	Digital Preservation Librarian	University of Iowa	2016-11-14
Nancy McGovern	Digital Preservation Program Manager	MIT Libraries	2016-11-17

Issues discussed by interviewees

Interviewees brought up a range of current, day-to-day digital preservation challenges, and made several suggestions for improvement, as summarized below:

Workflows

- Organizations don't have just one workflow. Workflows are highly variable depending on the *content source* (digitization, born-digital archives, IR/self-deposit, media generated during events, etc.). The resulting packages have different content, metadata and structure. Content format (text, image, AV) has much less impact on digital preservation workflows than the source of that content.
- People struggle to understand that some aspects of collections management is not directly digital preservation, but still in service of long-term accessibility.
- "Doers" (practitioners) aren't authorized to make decisions, and managers aren't always confident to make decisions. People need to make the connection between decision points and workflows to understand how things weave together.

Selection

- Decision-making about what goes into a service like DPN is often shared between collection owners and digital library staff, who have different perspectives (e.g., cost vs. content). It is sometimes difficult for people to come to an agreement, especially when space is limited (Emory). In other cases, the content owners trust the digital preservation librarian to guide decision-making (Iowa).

- It is difficult to get content into a state that is “done” enough for submission into a repository. For example: a manuscript that is missing pages, ETDs that aren’t in PDF version (only Word). People are uncomfortable ingesting these things that “might not be ready” and having to worry about versioning the content later. Versioning metadata is less of a concern (Emory). In other cases, they don’t digitize things until they are well organized and described (Iowa). In all cases, readiness plays a major role in decision-making.
- People are getting stuck trying to make selection decisions for content they may not understand well.
- At the University of Iowa, selection for inclusion in DPN is determined by:
 - Whether content is well organized. Older projects that don’t meet current standards will likely be passed over for now.
 - Materials that are at risk, and cannot be reformatted again (e.g., AV)
 - Rare and unique content

File Organization

- Working with collection managers on file organization is a big hurdle. Everyone does things differently, especially depending on the source of the files and type of files (e.g., born-digital archives, digitized, etc.).
- SIP completeness varies greatly between institutions interviewed. Emory doesn’t want to ingest content into a repository that isn’t complete with preservation metadata. Iowa doesn’t submit metadata into DPN other than the “space” title in DuraCloud. Instead, they relate the files back to metadata held locally.
- Collection management systems often dictate organizational principles. This can be frustrating as organizations start to outgrow these systems (Iowa). They can also result in very different, incompatible organizational approaches for different content types (Emory).

Awareness

- Not everyone understands the difference between digital preservation and digital libraries. It can be hard to communicate why there is a distinction and why it is important.
- “People need help creating a plan, sticking with it, and making sure there is buy in.”
- “People have different opinions on what digital preservation should be and what is acceptable.”

Tracking

- It is hard to understand what steps have been taken on an object. Because each system fulfills a different function, it is hard to get an end-to-end view of an asset. For instance: what is in DPN already? If no one knows that something is in there, it essentially isn't. What has gone through Archivematica? What hasn't?
- Preservation packages themselves are probably the best indicators of what has/has not happened.
- Complete inventories of digital collections largely do not exist.

Metadata

- Creating preservation metadata seems to be a big hurdle.
- People are struggling with knowing what metadata to capture when in the workflow, how and how much. How do you know if the content needs to be run through additional tools or not?

Change

- People need to become comfortable with the work being a moving target, a sense that what is being built right now (e.g., workflows, tools) is subject to rapid change.
- Quality and standards are also constantly changing. One interviewee noted that early digitization projects are not being selected for preservation because they don't meet today's quality standards.
- People need help with moving from project to production (as opposed to moving from theory to practice).

Potential audience for workshop

Interviewees made the following suggestions for a potential workshop audience:

- Collection owners, such as archivists, special collections librarians, and collections staff
- Digital asset creators, e.g., digitization team
- Digital archivists
- Digital preservationists — particularly in cases where they do the prep work to go into the preservation environment on behalf of others
- People who will be creating workflows
- People who will be implementing workflows

There was also a suggestion that collection owners are not the right target for a digital preservation training.

Interviewees suggest making it clear whom the workshop's target audience is. They note that with many existing workshops the exact intended audience is not clear, and a wide variety of people end up going, who may not be the ones who need it most.

Content-specific challenges

Interviewees did not feel strongly about challenges of different formats, other than AV and web crawls. Instead, nearly all noted that challenges are specific to the *source* of content creation, for example:

- Born-digital vs. digitized
- Digitized in-house vs. outsourced
- Born-digital submission to a repository vs. removable media in physical boxes

Potential workshop topics / existing gaps in training landscape

Interviewees had many suggestions for topics and areas of focus for the new workshop, including:

- Practical SIP creation: What does ingest look like (SIP → AIP creation for different types of content)
- Metadata: What needs to be captured and when in order to make things accessible over the long term
- Steps you can take to meet NDSA or ISO criteria
- Moving past the current state: identifying organizational needs → developing requirements → building out workflows → filling in with tools (mapping needs to tools)
- How to get your content ready to go into anything: What is the bare minimum you need?
- SLAs: how to make them equitable for you
- Very specific end-to-end workflows that are diagram out or walked through, with examples/case studies from organizations
- Roles and responsibilities
- Advocacy

They also suggested that the workshops should help participants with:

- Drilling down into implementation
- Focusing on why things need to be done the way they do
- Knowing if you are doing things right or wrong
- Translating concepts into day-to-day practice

- Understanding the need for continual assessment and reality of constant change—testing and experimentation
- Understanding the importance of establishing a “good enough” practice (e.g., start with secure storage), and gradually working toward a more robust digital preservation system
- Letting things settle and play out over a year or two, not jumping at new tools (i.e., maintenance vs. innovation)

Other suggestions

Interviewees made the following additional suggestions:

- The SAA publication, “Trends in Archive Practice, Becoming a TDR” by Steve Marks (2015)⁹ was referenced due to the detail in which it goes into practical implementation, not just abstract concepts.
- Don’t reinvent existing trainings. The DPN workshop should look at potential collaboration with Digital Preservation Management Workshop (for high-level concepts) and Digital POWRR (for tools).
- Partnering or mentorship program between DPN members might be useful.

3. Recommendations

Following the interviews and environmental scan and our analysis of both, we have reached several conclusions and make the following suggestions for the workshop.

Themes / Focus

We recommend that rather than provide a broad-brush overview of digital preservation concepts and practices, the curriculum should be oriented around problem solving. To that end, we suggest that the modules center on concrete actions that can be taken to move forward with digital preservation, including the need to periodically revisit strategies and shift as changes in technology and staff occur, or a new type or source of content is acquired. Additionally, we recommend that the curriculum take a “problem tree” approach, starting with an end goal (secure, long-term preservation storage) and moving back through the process to identify the individual steps (ingest, pre-ingest, selection) that are required to meet that goal, and removing roadblocks along the path toward the goal.

For each topic, the curriculum may address:

- Assessment of the current state
- Identification of gaps and root cause analysis
- Decision-making and prioritization
- Establishing practice

⁹ <http://saa.archivists.org/store/module-8-becoming-a-trusted-digital-repository/4679/>

Upon completing the workshop, participants should come away with the confidence to develop on current practice to ultimately submit content into digital preservation storage systems (such as DPN).

Structure and Delivery

The structure of the curriculum will be modular and will include a combination of the following delivery methods:

- Case studies
- Large group problem solving/discussion
- Small group activities
- Peer problem solving (Come up with recommendations for a peer institution, before looking inward.)
- Hands-on activities
- Presentations

Audience

We recommend that at least two people from each institution attend the workshop. This would include an individual responsible for digital preservation activities, as well as one collection manager/digital archivist responsible for acquisition and collection-based decision making. Because selection and preparation for long-term preservation requires decision making that ultimately affects staff outside of the digital preservation unit (especially those with “ownership” over collections), communicating what is involved, how it is performed, and why certain actions are taken is key to ensuring that collections receive the care and handling they need.

The workshop should be presented at an intermediate level, recognizing that some attendees will have a better foundation than others (e.g., digital preservation managers vs. collection managers). If the digital preservation manager can play the role of expert and support the growth of the collection manager, this will help to build confidence in the digital preservation manager’s abilities from both perspectives. The collection manager will be asked to prepare for the training with pre-workshop readings and tutorials (such as the “Digital Preservation Management” online tutorial from MIT).

Workshop capacity: 24-30 max.

Modules: Outline

We feel strongly that the workshop curriculum should focus on areas that are currently not receiving much attention in existing trainings, with an emphasis on getting content into digital preservation environments, and managing that content over time. Suggested modules include:

- Results will show immediately where people are getting stuck and provide a way to spotlight the problems that people are having and focus the conversation around those: Why are you getting stuck? Who else shares those challenges? What are the internal/external organizational factors that cause challenges or have the potential to move you forward?¹⁰

2. Selection for Submission

- Presentation Topics:
 - Developing criteria for selection of existing digital collections
 - Content source/stream considerations (digitized, born-digital, etc.) - with case studies for text, images, AV
 - Prioritization
 - “Done-ness” challenge
 - Legal issues
 - Roles and responsibilities
 - Inventory tracking
- Exercise: Case study problem solving (see structure below)

3. Preparing Content for Submission

- Presentation Topics 1:
 - Content source/stream considerations (digitized, born-digital, etc.)
 - Deciding how and when to capture metadata
 - Determining what is “good enough” for preservation
- Exercise 1: Case study problem solving (see structure below)
- Presentation Topics 2:
 - Understanding and conforming to repository requirements
 - Identification
 - Packaging
 - Sample packages for specific content types (text, image, AV)
 - Fixity creation
 - Local file management (until you move content into long-term storage), i.e., get content into secure storage and don’t store on removeable media
 - Inventory tracking 1 (includes an inventory template “handout”)
- Exercise 2 (Hands on): Packaging according to storage environment requirements (e.g., creating bags according to DPN and APTrust specs)

¹⁰ Using a tool such as Survey Monkey or Poll Everywhere (<https://www.polleverywhere.com/>)

4. Submission

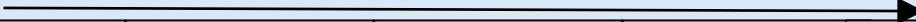
- Presentation Topics:
 - Upload/Transfer of packages
 - Verification, communication, feedback
- Exercise (Hands on): Submission of packages to DPN and using Exactly

5. After Submission

- Presentation Topics:
 - Ongoing responsibilities of the service vs. the collection owner/manager
 - Ongoing monitoring and decision making
 - Inventory tracking 2 (includes an inventory template “handout”)
- Exercise: Case study problem solving (see structure below)

Case Study Problem Solving Structure

- For each step / module
 - First: Present the goal / provide an overview
 - Second: Perform case study problem solving, which includes:
 - Assessing current state (entire group discussion)
 - Review the case study
 - Identify gaps and cause (root cause analysis): What is preventing us from reaching the goal?
 - Determining possible solutions (presentation)
 - Best practice / “good enough” / “how do I know I am doing it right”
 - Consider resource constraints
 - Decision-making and prioritization (small group discussion)
 - Solutions will be detailed, step-by-step, and based on a provided template (see example below)
 - Solution will answer who, what, when, and identify impact on success, potential problems, and mitigation strategy

1. Goals 2.			1. Stakeholders 2.	
<i>Step-by-step process</i> 				
Step	Step	Step	Step	Step N
Who is				
1. Challenges 2.			1. Solutions 2.	

- Put what we've learned into practice (entire group discussion)
 - Discuss relative merits of each proposal

Prerequisites for workshop

- The level of knowledge expected for participants will differ depending on the attendee's role in their organization. Those with less experience with digital preservation will be required to do more up-front preparation than those with more practical knowledge.
 - Prerequisites for digital preservation staff: Assumed previous training and baseline working knowledge of digital preservation.
 - Preparation methods for other staff: May include readings and online trainings and tutorials about digital preservation.