

## CHAPTER 3

### DIGITAL PRESERVATION POLICY AND STRATEGY

#### WHERE DO I START?

**Christine Madsen and Megan Hurst**

As organizations of all types increasingly create, collect, and disseminate collections of digital data and information, preservation policies and strategies are more important than ever. While information and data preservation have historically been, and will continue to be, the domains of libraries and museums, the centrality of digital information and data as organizational and intellectual property means that publishers and nearly all contemporary organizations should have digital preservation strategies and policies in place.

The phrase *digital preservation* can be used to encompass a wide range of activities, from best practices in file-naming and directory structures to fixity-checking, digital forensics, and format migration. It is therefore not surprising that when organizations set out to create digital preservation policies, the process can turn into a never-ending and increasingly detailed exercise in documenting all digital activities. This can be overwhelming and self-defeating.

Digital preservation policies should therefore be designed in such a way that they will actually be used and referred to, actively enabling the work of preservation, and they should align with the organization's overall digital strategy. This chapter presents a simple framework for getting started (or restarted) on digital preservation in any organization. The goal of this framework is not to provide hard-and-fast rules for making decisions about what to preserve, but instead to provide a heuristic foundation for structuring organizational thinking and for gaining a

better understanding of how to embed digital preservation practices and mindset into an organization.

Beginning with the *why* and *what* of digital preservation, this chapter clearly distinguishes between preservation policy, strategy, and operations documents. It provides guidance on content, frequency of review, and participation from staff. By breaking up the often-daunting task of creating a digital preservation policy into three distinct documents, readers will be able to move forward more quickly and easily toward embedding digital preservation activities and support throughout their organization, in a practicable, usable way.

These recommendations, while certainly informed by the literature on digital preservation, are more importantly rooted in real experience. The authors have worked in three large research libraries and one commercial publisher/aggregator over the last fifteen years, designing, shaping, and refining both the digital infrastructure and digital policies of those organizations. During the course of this experience, the authors have recognized emergent patterns that are not being addressed clearly in the academic or trade literature.

## **WHY**

There is a lot of pressure on libraries these days to collect, preserve, and provide access to a rapidly growing number of digital assets. Simultaneously, in non-library organizations, there is increasing pain and frustration associated with the *absence* of library-quality approaches to managing their information for the access, findability, rediscovery, and preservation of these assets. And there is, as we all know and experience in our daily lives, a larger and larger amount of digital data out there to be concerned with. Not only are the *quantities* of digital information growing at an astonishing rate, but there is a larger *variety* of digital assets and formats to look after, and there is mounting pressure to keep and preserve these objects in perpetuity. New

material formats in libraries and archives have never replaced older ones, but have simply been added to the long list of assets and formats that need to be looked after and kept usable. It is this last point—*keeping collections usable*—that makes digital preservation different from the issues that preservationists and conservationists faced in the ages before digital content. The specific challenges of digital preservation are covered elsewhere in this volume, so here we focus on some of the similarities between digital and non-digital preservation. At the policy and planning level, the *digital* is no different than the analog (that is, the physical, or the non-digital). Organizations need to think about what to collect, and how best to store it, describe it, and organize it. And at some point, these organizations will also need to think about what to deaccession.

So, *how* do organizations make these decisions about what to keep and what to deaccession? This question is at the heart of any library's or museum's collection strategy, and so it should be with the digital preservation policy of any organization. It can be tempting to try to provide hard-and-fast rules for making these decisions, but every organization is different—with varying goals, users, and needs. Therefore, the guidance here provides a framework to understand *how to make decisions* that will benefit any organization. The answers about what to preserve and what not to preserve are going to be different for different organizations. The *principles* of how to make these decisions, however, are the same.

## **WHAT**

One of the most common traps in digital preservation today is trying to find a solution before the actual problem has been identified. It is very enticing to say “we need a digital repository” or “we need to create a digital archive” because it sounds like a solution. But what does that actually mean? How will the repository be built? What will it contain? How will it be accessed?

Who will look after it? Simply having something called a “repository” or an “archive” certainly does not ensure its long-term preservation. Similarly, people jump into methodological or detailed technical discussions about approaches to digital preservation. What is the best file format for digitized collections? How should we structure the descriptive metadata? What is the minimum amount of technical metadata? What is the *right* number of copies to keep? But again, these questions alone—even when answered—do not help an organization establish a successful digital preservation program. It may seem like it is simply deferring the inevitable, but organizations will save time and money, as well as increase their chances of long-term success, if they think *first* about their local *strategy and policies* before trying to decide on any digital preservation *solutions*.

As with any other strategic endeavor, it is best to start with a basic plan that focuses first on understanding the context:

1. What does your organization want to accomplish?
2. What is the mission of your organization?
3. What are the goals of your organization?

In other words, begin with a strategy, not with a policy. Clearly separating these two tasks of strategy and policy into two documents is the first step towards creating something meaningful and usable.

A strategy is defined as “a plan, scheme, or course of action designed to achieve a particular objective, esp. a long-term or overall aim.”<sup>1</sup> In order to create a strategy, one needs to know one’s objective. A digital preservation strategy is therefore about supporting the institution in its goals and mission through the preservation of digital data and artifacts. So, before writing a digital preservation strategy, one needs to think about the overall mission and goals of one’s

organization. A strategy should provide a clear connection between the goals of the organization and the digital preservation activities in which it is engaging.

So, if a strategy is setting out *why* the organization is doing what it is doing, then what is a digital preservation *policy*? Policy, as the word implies with its close connection to “politics,” is necessarily about people. A policy (“a principle or course of action adopted or proposed as desirable, advantageous, or expedient”)<sup>2</sup> is there to help staff adhere to and/or contribute to the strategy. A policy should:

1. Provide the people on the ground, doing the work, with straightforward instructions about what to collect, what not to collect, and what to deaccession.
2. It should provide a cost-effective way for people to make these decisions.
3. It should provide them with the confidence to make decisions on their own without having to check with numerous stakeholders.
4. And it should provide some clarity for end users about why the organization is making the decisions it is making (that is, it should provide a clear and substantial answer to *why* something may or may not be kept).

Making this distinction between policy and strategy is, first, a means of breaking the task into smaller, achievable, usable deliverables. Trying to create a single, combined strategy and policy is a common practice, and it is commonly where organizations fail. In 2008, and again in 2013, two junior fellows at the United States’ Library of Congress published a review of existing preservation strategies and policies. The analysis was of thirty-three digital preservation policies/strategies from around the world.<sup>3</sup> The primary output in both of these exercises was a taxonomy of areas covered in these documents. The list of topics from the 2013 analysis included:

- Access and Use
- Accessioning and Ingest
- Audit
- Bibliography
- Collaboration
- Content Scope
- Glossary/Terminology
- Mandates
- Metadata or Documentation
- Policy/Strategy Review
- Preservation Model/Strategy
- Preservation Planning
- Rights and Restriction Management
- Roles and Responsibilities
- Security Management
- Selection/Appraisal
- Staff Training/Education
- Storage, Duplication, and Backup
- Sustainability Planning

This is an incredibly useful taxonomy that is both comprehensive and specific. Trying to create a single document that incorporates all of these topics, however, is trying to tackle too much at once. It conflates three different things, which makes the task of writing a digital preservation

policy overwhelming—and it likely makes the final product too long, too cumbersome, and in all likelihood unusable.

The taxonomy given above is far more useful if it is broken into three distinct parts. The first is strategy. As emphasized above, strategy should be intricately linked with, and drawn from, institutional strategy. Importantly, digital preservation strategy should not try to encompass something larger in scope than the institutional strategy. The parts of the Library of Congress taxonomy that should be included in strategy are:

- Collaboration
- Content Scope
- Mandates
- Policy/Strategy Review
- Sustainability Planning

Policy, as mentioned above, should be separate from strategy and in some ways can be thought of as the “connective tissue” between the strategy and the operations. Looking again at the Library of Congress taxonomy, the topics that should be included in a digital preservation policy are:

- Access and Use
- Metadata or Documentation
- Preservation Model/Strategy
- Preservation Planning
- Rights and Restriction Management
- Selection/Appraisal
- Staff Training/Education

Everything else should fall under operations. These are the actual procedures that should be taken to ensure the proper ingest, management, and preservation of digital content. From the Library of Congress taxonomy these are:

- Accessioning and Ingest
- Audit
- Roles and Responsibilities
- Security Management
- Storage, Duplication, and Backup

As mentioned earlier, one of the reasons for breaking this taxonomy into three parts is to make it more manageable and usable; that is, to make these documents into something that will actually be used. Three short documents rather than one long one are more likely to be read, understood, put into practice, referred to on an ongoing basis, and updated periodically.

## **WHO**

Another important aspect of digital preservation policy and planning is to understand *who* should be involved in what levels of strategic and policy planning and who is responsible for implementation. Another risk of creating a single large, comprehensive policy is creating a mismatch between the *what* and the *who*—that is, a misunderstanding of who is actually responsible for creating, updating, and adhering to a single digital preservation policy.

Conversely, another benefit of the three-tiered structure outlined above is that different people (or different roles) at different levels of the organizational structure can be both involved in, and assigned responsibility for, writing each of the three different documents.

In order to ensure alignment between the preservation strategy and the institution's strategy, the writing of the digital preservation strategy should be led by senior leadership (or

have some overlap with the group responsible for the institutional strategy). The digital preservation policy should be created by the organization’s managers, and ideally, by those managers who have oversight and/or contact with end users. And finally, the crafting of the operations policy should be led by the “techies”—those who are at the front line of the digital infrastructure.

In accordance with their separate intentions, owners, and audiences, each of these documents should have a separate shelf life (see table 3.1). Operations documents should be reviewed annually and revised as needed. Revision of policies and associated documents can happen every 3–5 years. And the review and revision of the strategy document should align with the timing of the review of the institutional strategy.

**TABLE 3.1**

**CONTENTS OF EACH PART OF THE THREE KEY DIGITAL PRESERVATION DOCUMENTS**

<b>Document</b>	<b>Content</b>	<b>Owner</b>	<b>Frequency of Update</b>
<i>Strategy</i>	<ul style="list-style-type: none"> <li>• Collaboration</li> <li>• Content Scope</li> <li>• Mandates</li> <li>• Policy/Strategy Review</li> <li>• Sustainability Planning</li> </ul>	Senior Administrator	Aligns with organizational strategy update
<i>Policy</i>	<ul style="list-style-type: none"> <li>• Access and Use</li> <li>• Metadata or Documentation</li> <li>• Preservation Model/ Strategy</li> <li>• Preservation Planning</li> <li>• Rights and Restriction</li> </ul>	Manager	3-5 years

	Management <ul style="list-style-type: none"> <li>• Selection/Appraisal</li> <li>• Staff Training/Education</li> </ul>		
<i>Operations</i>	<ul style="list-style-type: none"> <li>• Accessioning and Ingest</li> <li>• Audit</li> <li>• Roles and Responsibilities</li> <li>• Security Management</li> <li>• Storage, Duplication, and Backup</li> </ul>	Technical Lead	Annual

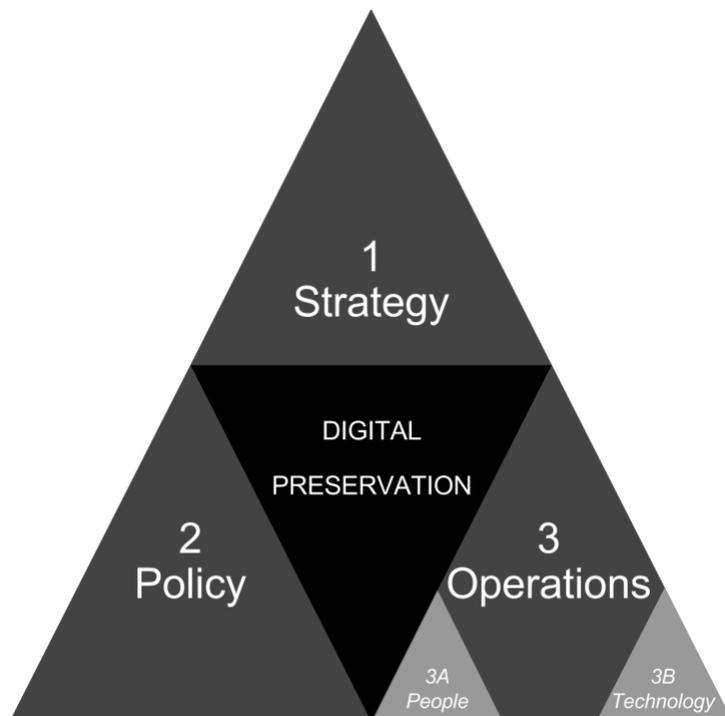
### ESSENTIAL COMPONENTS OF SUCCESSFUL DIGITAL PRESERVATION PROGRAMS

The Library of Congress taxonomy of topics, when structured into these three categories, provides a good framework for setting about writing or revising a strategy, policy, and operations documents supporting digital preservation. While these are good headings to help structure the process, it is also important to remember that digital preservation needs three things to be effective:

1. Technology
2. Organizational Support
3. Resources

Technology will inevitably be a part of digital preservation, but a common mistake is to *start* with or *focus primarily* on technology. As discussed above, digital preservation solutions should not be purchased or built without a firm sense of the organization’s goals. Technology should be the end point, not the beginning. Furthermore, an organization needs to believe that digital preservation is critical to its mission and goals. A strategy which makes that connection explicit is important. Thirdly, to state the obvious, digital preservation is not free. Resources—both

human and financial—will need to be allocated to create a successful digital preservation program. A team of committed people at all three of these levels is needed: strategy, policy, and operations (see figure 3.1). There are some libraries that devote more financial resources to physical boxes to hold physical materials than they do to digital preservation. The invisibility or lack of tangibility of “the digital” often leads to a misperception that digital preservation is a low- or zero-cost endeavor. This is, of course, not true. And this means that advocacy is also required at each of these levels. One could argue, as with data and information security, that the investment in digital preservation, and the strategies and policies supporting it, should be proportional to the potential risks, and negative impact, of total loss of the data or information under consideration.



**FIGURE 3.1**

**THE THREE-PART APPROACH TO DIGITAL PRESERVATION**

Thus far, we have provided an understanding of how and why organizations should approach digital preservation at these three levels in order to create a local strategy and policy that suits their needs. If digital preservation is planned and implemented in this three-tiered method, it cannot only inform the technical approach, but it can also make digital data and information collections *more preservable*.

Let us take an example that looks in detail at a common question addressed by many institutions when facing digital preservation planning:

“If we start to collect and preserve datasets and research data, do we also need to start preserving the software that was used to create and manipulate the dataset?”

There is no right or wrong answer to this question. The answer is simply “It depends.” If one has a clear understanding of the organization’s digital preservation strategy and how it ties into the organization’s mission, and from that emerges a clear policy about what the organization will preserve, then the question can be approached differently.

“What are the ‘significant properties’ of the research data (or any given digital asset)? And how do they fit with what we as an organization are trying to achieve?”

Breaking down this question with this in mind, one would then ask:

- Is this a case where a tool has been created to facilitate the outputs of the research?
- Is there something inherently new about the methods and the software?
- Or, conversely, is it simply the research data itself that is important?

These questions should then be answered within one’s overall institutional context. For example, does the organization collect the raw materials of research? That is, does the organization have as part of its mission to preserve the process of scholarship or research (archives, notes, manuscripts, etc.), or does it only collect the outputs of research (publications)? If the answer is the latter, why would it change its approach simply because the outputs are digital?

Employing this heuristic, and being clear about the organization's goals and intentions from the beginning, gives staff throughout the organization an opportunity to communicate with stakeholders and end users (in this example, university faculty, researchers, etc.) about how to make datasets more preservable. That is, it provides an opportunity to engage stakeholders in a conversation about the significant properties of their digital outputs. It also provides an opportunity to instruct them on the benefits of how separating the data from the software or interface in which it is created or presented can have a significant impact on the preservability (and therefore longevity) of their research data.

One of the easiest examples to illustrate this concept is to look at a technology that has been around since at least the fifth century—the book. The book is a technology that stores and delivers text to a reader. It is agnostic about the content, which may be graphic or textual, in any language, and in a large number of mediums.

To extend this to the proposed model for digital preservation, though, is to ask, “What is important or significant about a book?” As with the question about research data and software above, the answer is “It depends.”

Does the organization collect or study bindings, papermaking, and bookmaking? Then the very format of the book may be significant. If only the *content* of the text matters, then it is not. The answers to these two questions will have significant implications for an organization's collection and retention policies. It may make the decision between purchasing books and e-books clear, or it may help an organization decide to invest more in preservation. As these questions help structure *physical* collection policies, so too should they help structure digital collection and preservation policies. This is the conversation—about what is essential and what

is not—that should be had with any individual seeking digital preservation services from an organization.

Banks in the United States have understood this issue for many decades. For years, banks returned canceled checks to their customers, who held onto them in case a payment was disputed. In this case, the physical check itself *seems* like a valuable and essential object. Banks, however, recognized that keeping a copy was sufficient, and so for many years they would microfilm checks before sending them back to customers.<sup>4</sup> When storage began to become an issue—both for consumers and for the banks—the banking system moved toward routinely destroying checks after making the copies. As the copies moved from analog to digital, further efficiencies were gained by sending digital images of checks rather than physical copies. These efficiencies were eventually codified in law.<sup>5</sup> The original check is no longer an essential object.

In an example from the university context, it is increasingly common for faculty members to approach their library about preserving their digital projects. Maintaining outdated database or web-based software can be difficult and expensive, so many libraries are faced with having to say “no” to these faculty members simply for economic reasons. However, approaching this problem with the question of essentiality, as outlined above, may lead to a favorable outcome for both parties. The relevant questions in the case of a faculty member might be:

- What is important or unique about this research project?
- Is the software itself important? That is, does it represent something innovative or unique?
- If not, then the technical architecture of the project can be implemented in a way that will facilitate the extraction and preservation of the data.

This heuristic and collaborative approach also provides an opportunity for an organization to have these conversations early in the life cycle of a digital project or endeavor. Having the conversation early will increase the likelihood of creating something that is infinitely more likely to be preserved in the long term. It is likely that the technologies used in any given digital project will be out of date in less than twenty-four months. But the data should last. It should be designed to last. Wherever possible, an organization should separate the data from the software and interface. A clear, usable preservation strategy, policy, and operations guide will elicit, guide, and inform these conversations and decisions. It will get people thinking from “Day 1” about choosing a long-term, preservation-friendly approach.

Using the three-tiered approach of separating strategy, policy, and operations will encourage and facilitate conversations that will help to separate the essential from the optional. In this way, a digital preservation policy moves from being a largely theoretical document that is seldom read (and likely not adhered to) to a heuristic framework that helps to make day-to-day decisions and to build a robust, preservable digital archive.

## NOTES

1. “Strategy, n.,” *OED Online*, Oxford University Press, June 2017, [www.oed.com/view/Entry/191319](http://www.oed.com/view/Entry/191319).
2. “Policy, n.1.,” *OED Online*, Oxford University Press, June 2017, [www.oed.com/view/Entry/146842](http://www.oed.com/view/Entry/146842).
3. Bill LeFurgy, “Analysis of Current Digital Preservation Policies: Archives, Libraries and Museums,” August 2013, Library of Congress, *The Signal*, <https://blogs.loc.gov/thesignal/2013/08/analysis-of-current-digital-preservation-policies-archives-libraries-and-museums/>.
4. For further information on this shift, see Thomas Watterson, “Deciding Whether to Let Your Bank Keep Those Canceled Checks,” *Christian Science Monitor*, September 25, 1985, <https://www.csmonitor.com/1985/0925/wise2.html>.

5. Robert Moïse, “Where’s My Cancelled Check? A Review of the Basics,” The Tax Advisor, July 1, 2012, <https://www.thetaxadviser.com/issues/2012/jul/tpp-july-story-02.html>

