FROM THE FIELD

Starting up a Digital Preservation (Pilot) Program

Kim Hoffman

ABSTRACT

In 2024, Hamilton College started a project to pilot Archivematica, a digital preservation platform. The author details leading the project as the relatively new Digital Curation and Preservation Librarian and suggests ways that readers could approach such a project, incorporating what she learned along the way.

INTRODUCTION

The Hamilton College Digital Collections team stewards a growing collection of digitized Special Collections and Archives material. I started at Hamilton in January of 2023 as the first Digital Curation and Preservation Librarian, and I quickly realized that robust workflows and systems were needed to jumpstart our digital preservation program. Like many other repositories, we expect that future acquisitions of born-digital, archival content, and digitized A/V material from the archives will expand our capacity needs and requirements for active, preservation-focused maintenance. When digital collections are not merely representations of physical books and paper safely shelved in the archives but instead capture at-risk media on legacy carriers, or are themselves unique, our digital preservation strategy takes on a greater importance. While much of the digital collections are accessible online and backed up to local servers, they have so far been passively stored without the benefit of much, if any, active digital preservation management.

HAMILTON COLLEGE INSTITUTIONAL BACKGROUND

Hamilton College is a mid-size, suburban campus located in the central New York village of Clinton on the ancestral homelands of the Oneida Indian Nation. It was originally established as Hamilton-Oneida Academy in 1793.[1] The college serves around 2,000 undergraduate students, with a library of around 35 staff, plus about 40 members of the college's IT organization, another branch of the same organizational unit.[2] Digital collections are managed by a team of seven full-time employees who generally hold responsibility for materials digitized for viewing on the web. This team includes my position as the Digital Curation and Preservation Librarian. Born-digital archival materials are managed by the Archives department.

Digital preservation encompasses the various managed activities necessary to preserve and maintain digital assets for the long term. Those managed activities include everything from writing policies to confirming the authenticity and integrity of individual files. While storage is a necessary consideration for digital preservation, simply parking files in long-term storage is not sufficient. Many tools exist that can assist with active preservation management, and it is possible to create a custom workflow by chaining several software utilities together that each specialize in one or two preservation tasks. However, many organizations find it useful to adopt a digital preservation software platform that will provide a full suite of managed preservation activities, such as fixity checking, "a method for ensuring the integrity of a file and verifying it has not been

About the Author

Kim Hoffman (khoffman@hamilton.edu) is Digital Curation and Preservation Librarian, Hamilton College.

This contributed column was submitted on 27 May 2025 and published on 16 June 2025.

https://doi.org/10.5860/ital.v44i2.17452

altered or corrupted," and normalization, which migrates files from obsolete file formats to new formats that have been selected for their long-term viability.[3]

While Hamilton had been interested in a comprehensive digital preservation solution for some time preceding my tenure, the project had not yet bubbled to the top of the priorities list for me as I settled into a relatively new role. The nudge we needed to start exploring our options was learning that our digital asset management vendor, the Metropolitan New York Library Council (METRO), was considering developing an integration between Archipelago, the open-source digital objects repository that we use for our digital collections assets, and Archivematica, a prominent open-source software solution for managed digital preservation that integrates with the storage repository of your choice. Our team reasoned that, if successful, integration between these two products could be an opportunity to add a digital preservation solution with a minimum of disruption to our existing digital collection workflows. Additionally, since development was still in the earliest stages, we anticipated being able to influence project direction and new features. We formed a small, informal group to explore what a project to pilot the use of Archivematica would entail, composed of the Associate Director of Digital Initiatives and Technology, who headed up the Digital Collections team at the time, the director of our unit, and myself. When ultimately we were not able to pilot the integration with Archipelago, the project shifted to focus entirely on working with Archivematica as a standalone system.

This project was my first experience piloting a software product, and there were times throughout the process during which I felt somewhat lost. While there is plenty of literature about digital preservation basics, creating workflows, selecting tools, and the other day-to-day concerns of working with content for long-term preservation, I did not find any resources that helped me understand where to start when piloting something on this scale, from scratch. Only with hindsight have I been able to see the structured approach that I wish I had understood from the beginning. With the remainder of this column, I will describe the steps that I took—or wish I had taken!—from start to finish, in the hope that it will provide a template for others embarking on a similar pilot project.

STEP ONE: SELECT THE TOOL(S) TO PILOT

Because the Hamilton team was specifically interested in Archivematica, "a web- and standards-based, open-source application which allows your institution to preserve long-term access to trustworthy, authentic, and reliable digital content," we knew that this was the only tool we were interested in testing for this pilot.[4] Before you make a similar decision, consider questions like your budget and how much technical expertise you have internally to contribute to maintaining the software going forward. Comprehensive digital preservation software falls into one of two categories: commercial or open source. While commercial software carries a greater up-front price tag, open-source software requires more time and support from the institution to implement and maintain and may carry other costs for things like external support. At Hamilton, we are fortunate enough to generally have strong internal support for open-source software solutions, so our team knew that we would be able to support Archivematica if it met our functional needs.

STEP TWO: GATHER STAKEHOLDERS

You will need, at a minimum, group members representing the following roles and responsibilities: someone with control over, or insight into, the budget for the department and where this project fits in; someone representing the IT unit or someone with relevant technical expertise; someone deeply familiar with the collections that are targeted for digital preservation; and last but not least, the person who will be principally responsible for implementing the digital preservation system.

While gathering the right group matters, it is at least as important to secure an ongoing commitment from all involved for the length of the pilot project, including written documentation that describes roles and responsibilities as well as the anticipated timeline and project phases. Ideally, this group would continue to meet until the selected software solution has been implemented at the production level and you have had the opportunity to work through some of the early kinks.

For the Hamilton project, our in-house team needed to work closely with our digital asset management vendor to understand what was possible for a software integration that had not yet been fully tested elsewhere. While vendor representatives are unlikely to form a part of a library's team, it is important to build in time to meet with them if you will require their input. We found that because we had more than one project in the works with the same vendor, it was also useful to anticipate needing extra time for back and forth with our representative, since they were juggling multiple competing priorities with our institution.

STEP THREE: DEVELOP USE CASES

As part of the preparation phase, before you ever see the new software, it can be useful to identify the formal use cases your team envisions for the software you are testing. I have found that this is a particularly good exercise for revealing expectation mismatches between team members with differing roles. If you begin with a long, comprehensive list, the process of winnowing that list down to key use cases can also help to clarify the most important outcomes for the project.

For example, the Hamilton team identified the following as some of the potential use cases for Archivematica:

- Restore a damaged or missing file from a backup
- Migrate obsolete file formats to current preferred formats to maintain usability
- Verify file checksums to ensure fixity
- Upload metadata alongside content, with the ability to update if needed
- Embargo resources (e.g., presidential papers)
- Nice to have: provide some kind of limited mediated access to researchers for embargoed content

STEP FOUR: IDENTIFY QUESTIONS

Ultimately, pilot projects are designed to test software and workflows that are new to you and your organization. While kicking the tires without direction is also a valuable exercise, it is important to take time at the project outset to identify what specifically you need to learn to determine whether this software solution will work for your institution. If you have already done the work of identifying use cases, it is likely that relevant questions have arisen during that process.

For example, the Hamilton team identified the following as some of the questions we needed to answer during the early stages of our pilot:

- Can we securely mingle objects that should be dark (accessed only by staff, or embargoed even from staff) with objects that are part of our public archive?
- Will everything be put into Archivematica?
- What happens if we decide to get out of Archipelago?
- Checksumming—how does Amazon Simple Storage Service (S3) checksumming interact with Archipelago checksumming?
- Who gets access to Archivematica? Are different permission levels available? How are accounts provisioned?
- Do we know what storage configuration options are available or that we are considering? For example, we might be able to use Amazon Glacier for something like this rather than S3.

The Hamilton team also discussed whether there were any questions we specifically did not need to answer during the pilot, or at least during the initial phases, to curtail scope creep.

STEP FIVE: DESIGN WORKFLOWS

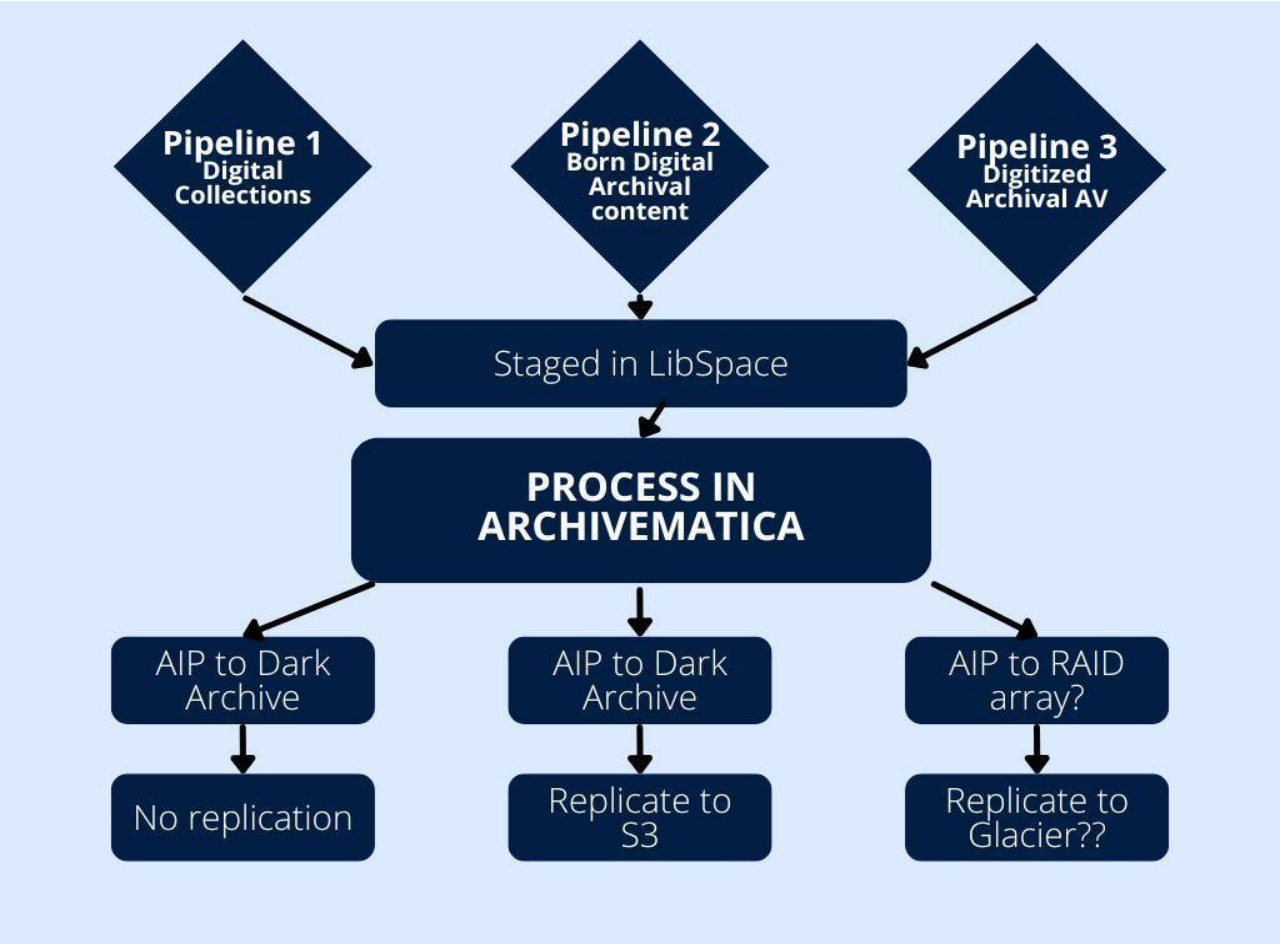
Think through the specific outcomes you hope to achieve with digital preservation software; again, use cases should be helpful here. For at least a few of these outcomes, design the workflows you will need to follow to accomplish these outcomes. Specify where the files will come from: are they accessioned? If not, do you already have the files? Where will they be staged? Who will do the work to prepare and ingest the files? Does your software solution specify a file structure for ingest? Finally, if Archival Information Packages (AIPs) will be used, determine where they will be stored. You will likely not know the answers to all these questions from the beginning; instead, the act of designing workflows will help shape your inquiry.

For the Hamilton project, I created a visual (figure 1) to convey our early thoughts about the most important workflows for our software. Even such a simple, provisional visual aid made conversations about workflows more concrete and easier to facilitate.

STEP 6: UNDERSTAND THE MINIMUM VIABLE PRODUCT

In the case of the Hamilton project, one of the desired features that initially drew us to Archivematica—a potential integration with our digital asset management system—was removed from the scope of the pilot. This feature was newly under development at the time we proposed our project and implementation exceeded our and the vendor's capacity. This need to scale down the pilot helped us to identify our true metrics for success. While the product integration was intriguing and may be something we are able to explore at another time, what mattered most was implementing an active digital preservation software solution. This realization allowed us to move ahead with the pilot and realign with department priorities. With the benefit of hindsight, knowing that testing this integration was not required for success would have allowed us to move forward more quickly.

Figure 1. Three early proposed workflows for moving digital content through Archivematica at Hamilton College.



CONCLUSION AND NEXT STEPS

While the project developed in an unexpected direction, ultimately, we are making progress toward department goals for growing our digital preservation program. At this point, the Hamilton team could consider extending the pilot phase to compare capabilities of additional products; however, we have been able to determine that Archivematica meets our functional requirements, even in the absence of the product integration we hoped to implement. For now, our plan is to begin transitioning to Archivematica for those digital assets prioritized for long-term preservation, fine-tuning the ingest process as we go along.

This was my first experience testing out software through conducting a pilot. If I could start again from the beginning and apply what I know now to navigating the process, I am confident that our project would have proceeded more quickly and more smoothly. We embarked on the pilot as a fairly informal effort, with a loose team structure for early meetings and an expectation that as the person primarily responsible for implementation, I would be able to access more help if and when it became necessary. While the informal nature of our efforts was in keeping with our institutional culture and felt comfortable to us, it also meant that we lacked ongoing commitment and check-ins that might have helped overcome obstacles throughout the process. We also lacked written documentation of the expected timeline and project steps, which might have helped us to set expectations early on and more quickly highlight when the project was languishing. Applying even

this lightweight structure to the project might have seemed to us at the time like unnecessary overhead but would instead have provided useful support. The flexibility we sought through informality could instead have come through proceeding iteratively through the process outlined in this column, returning to earlier steps—such as by revisiting our use cases, or even by considering a new software solution to test—as needed.

ENDNOTES

[1] "College History," Hamilton College, accessed April 30, 2025, https://www.hamilton.edu/about/history/full.

[2] "Facts," Hamilton College, accessed April 30, 2025, https://www.hamilton.edu/about/facts.

[3] "Glossary," Digital Preservation Handbook. Digital Preservation Coalition, accessed April 30, 2025, https://www.dpconline.org/handbook/glossary#D.

[4] "Home," Archivematica, accessed April 30, 2025, https://www.archivematica.org/en/.