

10-2016

Top Ten Questions and Answers about Digital Preservation for Special Collections and Archives

Jaime Schumacher
Northern Illinois University, jschumacher@niu.edu

Lynne M. Thomas

Follow this and additional works at: <https://huskiecommons.lib.niu.edu/allfacultyother-bookschapters>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Schumacher, Jaime and Thomas, Lynne M., "Top Ten Questions and Answers about Digital Preservation for Special Collections and Archives" (2016). *Faculty Books & Book Chapters*. 13.
<https://huskiecommons.lib.niu.edu/allfacultyother-bookschapters/13>

This Book is brought to you for free and open access by the Faculty Research, Artistry, & Scholarship at Huskie Commons. It has been accepted for inclusion in Faculty Books & Book Chapters by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.

Chapter 14

Top Ten Questions and Answers about Digital Preservation for Special Collections and Archives

Lynne M. Thomas and Jaime L. Schumacher

What Is It?

Digital preservation is, quite simply, the long-term care and maintenance of digital materials (no matter how they were created), with an eye to future access. It is not digitization, although any digitization plan or digital project should also have a preservation component (why put resources into creating a big digital cache without planning to keep it long-term?).

Is This Really *Our* Problem?

In 2009, Whittaker and Thomas argued that “as professionals trained to consider the truly long-term care and storage of materials, our community *must* become part of the digital preservation conversation.”¹ The artifacts and documents that we are creating and collecting right now are increasingly born digital: digital photographs, e-mail, and documents of all kinds. If we don’t find ways to deal with archival materials that takes these creation formats into account, we are not doing our jobs. Literary manuscripts, governmental records, personal papers, and organizational records like newsletters and meeting minutes are likely to have been created digitally before being printed (if, indeed, they were ever printed as opposed to being distributed via e-mail or over the web). In many cases, the digital file is the original artifact. The fonds we are professionally obligated to respect often now exist in the context of a hard drive. Ignoring long-term preservation of digital materials (whether they are born digital or digitized) guarantees that a large part of our contemporary cultural heritage will disappear.

Can't We Just Hire Someone Who Knows Tech and Make Them Do It?

Digital preservation, even when done inexpensively, must be programmatic: a core activity of library/archival life, not something we do as a side project. Ultimately, it is a facet of acquiring and preserving a collection of papers. How many people does that process currently take in your organization? Digital preservation that even minimally meets sustainable standards requires a team of people with diverse skills from across the library or archive, from curatorial acquisitions work to metadata to networking to administrative support, from the traditional to the cutting edge, from technical to advocacy. Assigning a single staffer to solve the problem will fail both your institution and the staffer in question.

Am I Ready to Begin a Digital Preservation Program?

Yes. Many of the skills that cultural heritage professionals already use day-to-day are core to digital preservation, too: selection, metadata creation, advocacy, budgeting, planning, and policy development among them. There are already robust standards, practices, tools and services, and a huge community of practice already in place, at all levels of institutional resources. You won't have to reinvent the wheel. We may, however, need to reevaluate how we approach technical obstacles in our libraries and archives.

Digital preservation options for underresourced institutions in particular may seem impossible on first glance. In 2011, Northern Illinois University, in partnership with Western Illinois University, Illinois State University, Chicago State University, and Illinois Wesleyan University, received a National Leadership Grant from the Institute of Museum and Library Services based on the argument that the Principal Investigators' institutions (and their state consortium) weren't in a position to help address digital preservation challenges based upon current resources.² Our white paper "From Theory to Action," the result of that grant,³ was published in 2011.⁴ It includes case studies for our individual campuses, along with our hands-on testing of several digital preservation tools and services. We found SWOT analysis (Strengths, Weaknesses, Opportunities, Threats/Challenges) very helpful to our process, laying out our technical, money, and organizational options, discussing our short- and long-term goals. This helped us to narrow down our options further based upon what was right for us.

The good news is, no matter what your current resources, some level of digital preservation work is possible. You may already be completing some digital preservation activities at the most basic of levels.⁵ The next step is applying them to a bigger picture and developing an overall plan that meets your institution's needs.

Why Now?

The greatest difference, arguably, between the preservation of digital materials and the preservation of traditional special collection materials is one of time. The speed at which digital materials must undergo processing and archiving and the subsequent frequency with which they must be managed are significantly higher than their paper counterparts.⁶ An acquisition of paper

manuscripts is not at a high risk of physical loss if not processed and shelved within, say, seven years (assuming a climate controlled storage space). Once processed and shelved, these materials can likely survive intact without further active management for decades—this is the beauty of “benign neglect.”

There’s another wrinkle, too, in terms of how much time we have, and the rate at which tools and services change or become obsolete. If you have ever experienced the selection process for, say, a library management system/online catalog, a rather typical library approach will sound deeply familiar: extensive research, followed by calls for proposals, endless meetings, and a lot of money spent. The result is a series of practices, policies, and tools that we rely upon for a decade or more.

Digital preservation tools and services change far too quickly for that model to be viable.

We need instead to plan for three- to five-year technology cycles, with choices that allow us graceful entry and exits from the tools we are using at any given time. This is especially important in organizations that have budgeting issues, whether cyclical or systemic. What is affordable and sustainable now may not be so in three to five years. The goal is to make sure that at the end of a given period, there are ways to get your data back and options for what to do next.

Sustainable digital preservation programs incorporate an iterative process. The speed with which tools and services are created, tested, adopted, and superseded means that we can’t just make a single, “big” choice and assume we are all sorted out for the next decade with minor maintenance. You will set up your policies and practices and find that they don’t actually work, or change your mind about your tool selection. *That’s okay*. Trial and error are the order of the day. In most cases, one system/solution does not fit all, organizationally or technically.

All of These Standards Look Complicated and Expensive. Should I Be Worried about Becoming a Trusted Digital Repository or Obtaining TRAC?

If you are just learning about digital preservation for the first time, there are lots of options to bring yourself up to speed.⁷ You will need to understand the basic vocabulary and processes of digital preservation but not necessarily all of the nuts and bolts at an examining-code level. The advantage of not being on the bleeding edge is that the vast majority of the digital preservation tools and services any given organization will consider have already incorporated relevant standards (such as Open Archival Information Systems) into their development. TRAC (Trustworthy Repositories Audit and Certification) and TDR are laudable and important goals but may not be right for every organization. They require significant staffing and resource levels which may be out of reach of smaller or less-well-resourced organizations. Digital preservation is an iterative process; it’s not just making a single choice and then sitting back, but making a series of ongoing choices as the next set of options becomes available, based upon previous choices.

The available systems designed to incorporate large numbers of functions may assume a level of preparation of files that your organization may not have completed. Many institutions will need to perform a level of “triage” on their digital collections to prepare them for ingest into more comprehensive

systems. Some tools and services may assume that you have a full PREMIS (Preservation Metadata: Implementation Strategies) or MODS (Metadata Object Description Schema) record ready to go at the point of ingest. Most tools and services assume that files have already been removed from temporary carriers such as CDs or flash drives. File formats may need to be standardized, metadata may need to be generated, and an inventory of the amount of extant material will be necessary to determine storage needs and collection growth. All of these are activities that can be taken on before a specific solution is selected, during the policy planning and budgeting process.⁸

How Do I Know What Tools and Services Are Best?

You have numerous options, with new developments almost daily. You will need to think in terms of your organization's goals and needs. Think of it as building a tool belt. It is unlikely that a single tool or service will solve all of your problems, but you may be able to find a limited group of tools that do most of the job at hand.

Here are some things to consider:

- Do you need a comprehensive system for managing your digital files, tools that only perform a couple of functions, or just secure, inexpensive storage that isn't solely on your local server?
- What file formats are your priorities? PDFs? Image files? E-mail?
- Do you plan to host software at your institution, or do you need a hosted elsewhere solution?
- Do you have sufficient support to maintain open source software, or would a vendor-based solution (either for-profit or nonprofit) work better for your organization?
- Which digital preservation tools may work the best with the systems you already have in place? Your selections will be based in part on which systems for creating metadata and/or access you are already using (e.g. ARCHON, ArchivesSpace, DSpace, Fedora, and ContentDM).
- Do your goals include emulation and migration, or merely saving materials at the bit level?
- Would a consortial or group solution be more appropriate for your setting? If so, does the solution on offer meet most of your needs?

To research and compare-and-contrast individual tools, consider the Community Owned Preservation Tool Registry.⁹ The Digital POWRR white paper¹⁰ includes our discussions of the software and services that we tested in-depth, along with our proposed individual institutional choices, which were all completely different from one another.

How Do I Choose What to Save?

Many an archivist will be faced with the donor who wants to know why you don't wish to accept all 80,000 photographs in the donor's Flickr or

Instagram account, or on the 2 TB hard drive on offer. Selection/assessment and prioritization are even more important skills in digital preservation, when the assumption is that cheap digital storage means “just saving everything.” Selection for posterity is (in part) what cultural heritage workers *do*. There is nothing stopping us from treating different types of digital materials differently from one another, in the same way that libraries and archives often have different levels of access or preservation choices (or lack thereof) for different kinds of analog materials.¹¹ Please see Chapter 8, “Digital Acquisitions and Appraisal,” by Sarah Barsness and Anjanette Schussler in this volume for additional details on this process.

Digital preservation is more than just keeping bits safe: provenance and contextualization are important, and what we do best! Creating and maintaining context is another part of our professional practice easily applied to digital materials. Once processed and archived, digital materials require ongoing, active care so as to ensure that file formats do not become obsolete and the setting in which they are archived remains viable.

How Do I Find the Resources to Do This?

The short answer is advocacy, planning, and budgeting. One of the key components of a robust digital preservation program is administrative buy-in, particularly at a campus/organizational affiliation level. You are working toward solving an organizational problem, not just a library-specific problem. Consistent, long-term communication across organizational levels is best done with a team of people. This is a marathon, not a sprint. Developing a new program in any organization requires laying groundwork, lots of communication, long-term planning and budgeting, and leveraging what is already in place in your organization wherever possible.

The Digital POWRR project developed a series of one-page handouts to aid in this process, aimed at communicating with different organizational roles, particularly outside of the library.¹² One of the most important lessons we learned from our process is that of quantifying risks.

What Is the Cost of Doing Nothing?

The answer to this question, especially with specific examples, is what will prompt momentum out of organizational inertia. Potential stakeholders may not yet know that digital preservation affects their ability to do their jobs. Explaining to your alumni association or development office that if nothing is done, there will be no photographs 25 years from now of, say, your football team’s first trip to a major college bowl game is likelier to get their attention. State institutions may be affected by university records disappearing through lack of long-term preservation that could cause legal or compliance issues. Initiating these conversations and pointing out that the library or archive has expertise to help other units solve problems that they didn’t realize that they had may drive your proposed solutions forward.

What are you waiting for?

Notes

1. Beth M. Whittaker and Lynne M. Thomas, *Special Collections 2.0* (Santa Barbara: Libraries Unlimited, 2009), 100. Their chapter “The Elephant in the Room”

opens with a quote from Paul Conway's 1996 "Preservation in the Digital World," Council on Library and Information Resources, Washington, D.C., March 1996, accessed November 19, 2015, <http://www.clir.org/pubs/reports/conway2/index.html>.

2. Over the course of the period of the study, the Open Access to Research Articles Act passed in Illinois. This law requires that research created at state-supported institutions be made publicly available and *preserved long term*. Our campuses now also have a legal mandate to somehow create digital preservation programs. We're still working on it due to it being an unfunded mandate. The act can be found at <http://www.ilga.gov/legislation/publicacts/fulltext.asp?Name=098-0295>, accessed November 19, 2015.

3. Grant number IMLS award LG-05-11-0156-11.

4. "From Theory to Action: Good Enough Digital Preservation for Under-Resourced Cultural Heritage Institutions," accessed January 6, 2016, <http://commons.lib.niu.edu/handle/10843/13610>.

5. See the National Digital Stewardship Alliance levels of preservation, which is an easy assessment rubric: <http://www.digitalpreservation.gov/ndsa/activities/levels.html>, accessed December 4, 2015.

6. Audiovisual preservation is a whole other discussion. A good starting point is the Library of Congress's page on the topic: <http://www.loc.gov/preservation/care/record.html>, accessed November 19, 2015.

7. The Digital POWRR site has a Digital Preservation 101 page with a collection of links, for example; <http://digitalpowrr.niu.edu/digital-preservation-101/>.

8. We particularly recommend Ricky Erway's "You've Got to Walk before You Can Run"; <http://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf>.

9. Community Owned digital Preservation Tool Registry, accessed January 6, 2016, http://coptr.digipres.org/Main_Page.

10. Jaime Schumacher et al., "From Theory to Action: Good Enough Digital Preservation for Under-Resourced Cultural Heritage Institutions," August 2014, Institute of Museum and Library Services, accessed January 6, 2016, <http://commons.lib.niu.edu/handle/10843/13610>.

11. The University of Utah has created a handy flowchart that illustrates this point: http://digitalpowrr.niu.edu/wp-content/uploads/2014/04/UnivUtah_DigPresDecFlowchart.pdf, accessed November 19, 2015.

12. These can be found at http://powrr-wiki.lib.niu.edu/index.php/One_Pagers_tailored_to_educate_different_professionals, accessed December 10, 2015.