Chapter 7

Maintaining Digital Materials over Time

This chapter is slightly different from the others because if you have created and consistently implemented workflows for all the other stages of the digital material life cycle, most of the work for ongoing maintenance has been done. When I mention the digital material life cycle, I am referencing the general iterative process of create, acquire, stabilize, appraise, process, and provide access to that I discussed in previous chapters. These actions are all part of the digital preservation process. There is a formalized reference model available from the Digital Curation Centre called the Curation Lifecycle Model, which provides a more comprehensive view of the preservation of digital material, specifically data, from conceptualization through final disposition, where disposition may mean remaining in the preservation cycle or permanent deletion.[1]

In an ideal world, the maintenance of digital materials is automated and your responsibility is to periodically review the system and make upgrades and tweaks as needed to keep the machine running smoothly. In fact, in many institutions this is the current reality, but for others it is an unrealized ideal. In an acknowledgement of this spectrum of resources and abilities, I have structured this chapter based on three of the five functional areas found in the National Digital Stewardship Alliance's (NDSA) "Levels of Digital Preservation."[2] The final two functional areas, content and metadata, have been thoroughly covered in the accession and processing workflows.

The "Levels of Digital Preservation" document was created to be as adaptable as possible to the many different types and sizes of institutions that currently and will in the future curate digital materials. This document focuses on the technological aspects of digital preservation; however, technology is not the only key to maintaining your digital materials. You must also constantly keep your organizational infrastructure and resource allocation at levels consistent with the amount of material you are preserving. The chapter will conclude with discussions on how to maintain both. Depending on your own institutional context, each of these sections may have a lower or higher priority determined by policies, currently allocated resources, and risk assessments. For every one of these sections, documentation is key. It is essential to keep up-to-date records of every aspect of your program in a central location for anyone who is working with digital materials to access. This documentation acts as reference material, advocacy and outreach material, and an audit trail of all you have done to increase the transparency of your actions and policies.

Storage

There are some standard requirements in regard to storage of digital content. Beyond those basic requirements, how robust and diverse your storage is will depend on your institution's resources. If your digital preservation program is entirely standalone, you will make all these decisions. However, most digital preservation efforts are integrated into a wider ecosystem where storage is mostly dictated by the information technology department of your institution. Where that is the case, you will need to have an open line of communication with the information technology staff to come to an understanding of current practice and negotiate possible changes to make the storage of digital materials more robust, as needed. Keep the relationship strong and the conversations on a schedule so you are not caught unaware by any changes or updates to your institution's storage strategy. Also, these conversations should encourage the information technology department to take ownership of its part in the digital preservation program. The more invested it is in the success of the digital preservation program, the stronger a partner it will be.

As a community, digital preservation practitioners encourage that there be at least three complete copies

of all of your digital content. The bare minimum is two copies. Furthermore, these copies must be on separate storage mediums. For example, if you have one complete copy on a computer or your library's server system, the other copy should be on a completely separate type of hardware or in a cloud storage environment. The hardware diversity helps prevent a single point of hardware failure causing data loss. In addition to the number of copies and diversity of storage type, standard practice is to have at least one complete copy of your digital materials in a completely separate geographic location. This could mean having a copy of your backup tapes in a partner institution in at least a different state. Another option is having your cloud storage copy be in a different geographic region than your home institution. The geographic diversity is meant to spread the risk of a single natural or man-made disaster decimating all of your data in one event.

As with all technology, your storage devices will age, so keep a schedule of when you purchased storage, that device's approximate lifetime, and the budget for its replacement. In the case of cloud storage, budget for your subscription every year and leave yourself a little room in the budget for an unexpected increase in needed storage capacity. In some cases, pulling a copy of your materials down from cloud storage incurs a cost, so be prepared for that. Common reasons for needing to retrieve data from the cloud include a local hardware failure that requires you to replace all of your local copies of your data, switching cloud storage services, or needing to retrieve only small amounts of data due to the local copies being irrevocably damaged due to an accident or an act of malfeasance.

Document all current storage mediums, their physical locations, and the backup schedule. Keep track of how much data you have stored and be aware when you are about to reach the limits of your current storage capacity. Careful documentation will help you plan for gradual increases in storage capacity and slowly build the necessary increases into your operating budget. However, there may come a time when an accession could exponentially exceed your current storage capacity. If this were to occur, there are three avenues open to you: requesting the donor to provide financially for the storage of the material, advocating for additional storage resources on an expedited time line, or making the difficult decision to recommend another institution for the donor to work with.

Integrity

This section focuses on maintaining the integrity of the digital materials in your custody. In this context, integrity is ensuring that you can prove that the digital content received from the original creator is the same content that a researcher eventually has access to. This is much easier to do with physical materials such as papers and books than it is with digital content due to the very nature of digital material. Digital content is meant to be copied, shared, and modified easily, so maintaining a static version of the materials requires establishing the state of the material when it arrives in your institution and then maintaining that state throughout the life of the digital material while it is under your care.

If you have established an accessioning and stabilization workflow, the first part of this has been done. You have checked all incoming material for viruses and documented the thumbprint of each file by recording the materials' checksums. As part of your digital preservation maintenance workflow, you will periodically regenerate checksums and compare them to ensure no changes have been made to the material. If an undocumented change has been made, delete the local copy and pull down a new version from your preservation masters. Many digital asset management systems automate this process and create an audit log that you can monitor. However, if you do not have one of these systems, there are several tools available, one of which is Fixity, that you can use to at least automate the checksum comparison on a schedule and that will provide you with a report after every pass that you can use to determine if further action needs to be taken.[3] The audit logs of these comparisons are just as critical as the performance of the comparisons themselves because the audit logs provide transparency about the state of your materials and any changes made to them.

There will be instances where you will deliberately change the nature, format, or content of the material in your care. This will cause a change in the checksum and throw a flag during the integrity checking process. Again, thorough documentation of all actions taken during processing will help explain the flags raised during the first round of integrity checks after the changes have been completed. Part of guaranteeing integrity is doing these checksum comparisons after every transfer of materials and keeping the audit logs of the integrity checks in a secure location.

Control

This section is all about security, who has access to content, who is authorized to modify and delete content, and how you keep track of those permissions. The single greatest enemy of digital material, besides time, is human interaction. The most common reason, at my institution, that we have had to depend upon backups of our digital content was accidental modification by a user. It is imperative that your institution

have a strong permissions structure that automatically denies unauthorized users access to content.

System administrators, usually housed in the information technology department for local network systems, are responsible for setting these permissions, but this not an automatic process. To be able to set these permissions, system administrators need to know who, what, and how—who has access, to what content, and how far that access extends. This information is most easily communicated in a table where you list each user, what they have permission to, and how far that permission extends (see table 7.1). Annually, at the very least, this table should be reviewed and permissions changed as needed. Also, after an event such as a staff member leaving or a new person being hired, this table should be updated and immediately communicated to the system administrators. This security plan works just as well for digital asset management systems and other tools and services you use to carry out digital preservation workflows, so be clear what each table is for and keep each document up to date.

Organizational Infrastructure

Digital preservation relies on a stable budget of financial resources and personnel time. Organizational infrastructure plays a vital role in maintaining that stability. From the top administrator down to the newest employee, knowing who is in your organization, what their responsibilities are, and what their priorities are will help you to successfully advocate for your digital preservation program and find willing partners in performing your responsibilities.

Another vital tool in your knowledge bank is a thorough understanding of the policies and procedures that guide your institution. Those policies will determine who you should approach for help on specific projects and when to advocate for more resources in the cycle of fiscal allocations. For example, your institution may require that all software be managed by the information technology department. In this scenario, any time you want to test or implement a new piece of software, you must get approval from your information technology department and work with it to have the software downloaded on your personal work machine. This even includes software that is open source and therefore does not require financial resources. Building and maintaining a strong relationship with your information technology department personnel will make this process much less frustrating.

Let us consider another scenario, one where all purchases must be requested and approved through a specified process before the end of the fiscal year. In this situation, you need to be aware not only of the structure of your institution's fiscal year—for example, July 1 to June 30 or January 1 to December 31—but also the rhythm of that fiscal year. Historically, has there been a purchasing freeze at least a month before the fiscal year ends? On average, how long does it take requests to work through the process? Are certain types of requests more likely to be approved than others? All of this information will help you format your requests and craft the most successful argument for your request and the ideal time in which to place it.

In the final scenario, I would like to address changing organizational infrastructure. It is the nature of any institution that people will come and go and leadership priorities will change. You need to build out your digital preservation program and all of its workflows to be sustainable amid change. It is critical that your workflows be able to survive and persist through these changes. This is where having extensive, up-to-date documentation of how processes are done and *why* processes are done is vital to the longevity of your program. That documentation should include what *positions* are responsible for each step in the process. That way, if a person leaves your organization, you can better determine who should do the work in the interim and what skills should be included on the job description of the new person being hired.

Table 7.1
An example permissions document that includes the user, what they have permission to, and how far that permission extends

User	Username	Role	Dark Archive			Working Files			Use Copies		
			Read	Write	Delete	Read	Write	Delete	Read	Write	Delete
Suzy Q	suzy.q	student worker	no	no	no	no	no	no	yes	no	no
Kirk McCoy	kirk.mccoy	digital archivist	yes	yes	yes	yes	yes	yes	yes	yes	yes
Jane Doe	jane.doe	oral historian	yes	yes	no	yes	yes	yes	yes	yes	yes

When leadership changes, this documentation can be used to help advocate for the continuation of your current program or for an increase in resources to expand your digital preservation efforts. Change is inevitable. Strong workflows and documentation can help you weather changes much more easily and efficiently.

Resource Allocation

When considering resource allocation, I am encouraging you to consider how you budget your financial resources, staff time, knowledge, and technological resources devoted to the ongoing maintenance and access of your digital materials. With digital materials, it is easy to become hyper-focused on financial resources dedicated to ongoing storage costs, subscription costs for digital asset management systems or other software used to preserve and provide access to your digital content, and the costs of equipment. However, another vital resource that needs to be managed is personnel time and knowledge.

At many institutions, staff are being asked to learn new skills and perform new tasks while maintaining their existing workload. In other scenarios there is one expert who is expected to be responsible for almost every aspect of digital content management, with some help from information technology departments. Perhaps you have a situation that is both: a staff member has been asked to become the expert and take over the responsibilities of digital preservation on top of their current responsibilities. This sets an institution up for gaps in preservation management when the expert leaves or when staff have to reprioritize their workload and digital preservation is put at the bottom of the list. So, just as you budget money, you need to budget staff time and expertise.

Spread out the responsibilities and expertise among multiple staff members as much as possible so that the pieces of digital preservation each person is asked to do fit well with their existing knowledge base or job responsibilities. Dedicate resources to training and continuing education for staff because the standards and practices of digital preservation are constantly changing simply due to the nature of the rapid evolution of digital materials. Encourage team building and communication between the archival and curatorial experts and the information technology experts. They have to work together to preserve digital content. It is easy for information technology experts to become overburdened by requests for expertise and resources specific to digital preservation in addition to the information technology department's existing workloads, so be deliberate about how digital preservation is added to that. Remember, the more content you add to your collections, the more resources you will need to allocate to the digital preservation effort. If you cannot sustain your digital preservation program on your current level of resources—financial, technological, and personnel—prioritize preserving the materials that you currently have and do not accept new accessions.

Notes

1. Sarah Higgins, "The DCC Curation Lifecycle Model," *International Journal of Digital Curation* 3, no. 1 (2008): 134–40, https://doi.org/10.2218/ijdc.v3i1.48.
2. "Levels of Digital Preservation," National Digital Stewardship Alliance, Digital Library Federation, http://ndsa.org//publications/levels-of-digital-preservation/.
3. "Fixity Pro," AVP, https://www.weareavp.com/products/fixity-pro/.