

7-1-2018

Horseshoes, Hand Grenades, and Digital Preservation: When Close Is Good Enough

Nat Wilson

Carleton College, nwilson@carleton.edu

Follow this and additional works at: <https://lib.dr.iastate.edu/macnewsletter>



Part of the [Archival Science Commons](#)

Recommended Citation

Wilson, Nat (2018) "Horseshoes, Hand Grenades, and Digital Preservation: When Close Is Good Enough," *MAC Newsletter*: Vol. 46 : No. 1 , Article 13.

Available at: <https://lib.dr.iastate.edu/macnewsletter/vol46/iss1/13>

This Electronic Currents is brought to you for free and open access by the Journals at Iowa State University Digital Repository. It has been accepted for inclusion in MAC Newsletter by an authorized editor of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Assistant Editor: Sarah Dorpinghaus, University of Kentucky.

Contact Sarah at sarah.dorpinghaus@uky.edu if you would like to guest author a column or have a good idea to share.

Horseshoes, Hand Grenades, and Digital Preservation: When Close Is Good Enough

By Nat Wilson, Carleton College

A Framework for Digital Preservation

When I became the first digital archivist at Carleton College, there was no digital archives program in place. There were a few digitized collections from the college archives, but no plan for ingest, preservation, or access for digital records existed. This was a perfect opportunity to start a program from scratch, and in 2012, we began work on a conceptual framework for digital preservation. We worked with the assumptions that 1) not all records in archives are of equal value, and 2) the care given to a record should match the value we placed on it during appraisal. After assessing the value of a set of records, taking into account the cost and complexity of their preservation needs, we would place the records into one of three tiers. Tier 1 would be our most important records, Tier 2 records those of medium importance, and Tier 3 records of the least important. We hoped this work would make

it easier to allocate our limited resources such as storage space, technical capabilities, staff, and time.

The first requirement of the framework was that it should be widely applicable. It should be general enough to work for all kinds of digital resources from simple sets of images to research data and computer scripts. We were inspired by the developers of the Open Archives Information System (OAIS) to make a framework not too specific with tactical details but instead focused on the higher-level strategic goals of a preservation program. Different tools and specific methods could change over time as long as they accomplished the goals set out in our framework.

The second requirement was the framework should be scalable. At the time, we were dealing with relatively small sets of digital files that were easy to manage. However, growth of the digital archives at Carleton was only bound to increase. The ingestion of new digital accessions would

never slow down, it would never stop—it would only get faster and faster. If we were to be successful, we needed a system that could keep up with this ever-increasing rate of growth.

Prioritizing Records

A number of factors tended to elevate a record in the tiered system. One was the value of the material to Carleton's institutional record, memory, and cultural heritage. Examples included documents created by the board of trustees,



Salvaging Macintosh Plus hard drive in Carleton College Archives, 2014. We leapfrogged from Macintosh Plus to PowerBook 520c in SCSI target mode to compact flash to modern day iMac.

(Continued on page 28)

(Continued from page 27)

annual reports from various divisions across campus, and our historic photograph collections. Another factor was the cost of replacing lost electronic records. Some digitized items would be expensive to rescan, particularly video and audio on old media such as magnetic tape. Other records were extraordinarily difficult to salvage or process when first accessioned due to obsolete hardware and software dependencies. In these cases, we would want to avoid repeating that work. Last, we might need to keep these records for legal reasons.

A number of factors might decrease a record's position in our tiers, many of them having to do with the cost of our preservation efforts. How difficult is it to preserve certain electronic records? For instance, is the collection a set of fairly straightforward TIFF images, or is it a PHP website with a MySQL database? On a similar note, hardware and software dependencies on older electronic records would make us less inclined to preserve a record in its original form. Over the years, Carleton students have created a number of computer programs, including electronic music and computer dating programs in the mid-1960s. All of these resources would be difficult, if not impossible, to maintain in their original forms and they might fall into Tier 2 because of this limitation. Last, are there security or privacy concerns with these records that we would not be able to address without additional training or infrastructure? All of these issues would increase the cost of preservation and would need to be considered during appraisal.

We defined the three tiers as follows:

- Tier 1: Critically important to the institutional record and cultural heritage of the college. In some cases, there may be legal requirements to keep materials indefinitely. Loss would constitute a major blow to future understanding of the history of the college. Every effort would be made and significant costs incurred to recover lost or corrupted files.
- Tier 2: Important to the institutional record and cultural heritage of the college. Loss would compromise our future understanding of the history of the college, but not as significantly as the loss of Tier 1 records. Significant, but not heroic, efforts would be made to recover lost or corrupted files.
- Tier 3: Records useful to understanding the history of the college. Intent is to retain permanently. Loss or corruption would be unfortunate, but not a matter necessitating significant effort or expense to recover.

Once records were categorized into these tiers, we defined a series of preservation policies for each, including

1. Fixity checks
2. Preservation metadata
3. Backup cycles
4. Backup media
5. Migration practices
6. Disposition of the original
7. Recovery tests
8. Format support

The full description of tiers and preservation activities is available at wiki.carleton.edu/x/JNq6.

Assessments and Compromises

As we apply the framework to digital records in the archives, we have found our efforts to be only half as successful as we had hoped. The goals set for Tier 1 items have mostly not been met, often because they are beyond our control or capabilities. Many of the actions for Tier 3 records ended up increasing our workload instead of decreasing it, which largely defeated the purpose of having a low priority tier. Tier 2 is the only area where we have been able to achieve most of our goals.

Some of these shortcomings were due to factors beyond our control. For instance, our policies on backup cycles called for Tier 1 items to be in a LOCKSS network, ensuring our data was stored as multiple copies in distinctly different geographical areas. However, the Information Technology Services Department at Carleton was not willing to support a LOCKSS system, the Archives Department did not have the funds to support its own LOCKSS system, and we had no authority to make the Information Technology Services Department change its policies or priorities. Luckily, the backup methods available to us through Carleton have been acceptable, even if they fall short of our goals laid out in the framework.

Our own lack of skill, time, staff, or technical capabilities caused other failures of the framework. For example, the framework recommended that we verify the checksums for our data about once per year. With Bagit, the software we currently use to run these fixity checks, we can process about one terabyte of data every 25 hours. With our current holdings of nine terabytes, and the fact that the Bagit application causes a noticeable drain on processing power, that operation would require a machine dedicated to this task for nearly 10 days. We currently do not have

enough computers to perform fixity checks at this level.

While we may have had only partial success when applying our framework, it is important for us to do as much as we can to preserve our electronic records. To continue to make progress in our work, we have had to make a number of compromises—with our colleagues, our managers, and the pressures of reality. For our backup practices, we currently accept the services offered by our Information Technology Services Department. It does not offer everything we had hoped for, but service has been good, and technicians work with us to develop effective but affordable solutions. If we see serious problems with its methods, we will address them at that point. However, in the meantime, we will continue relying on its servers for storage and backups.

For checksum validation, I mentioned that we do not have enough computers to run checks on all our files every year. Instead, we have developed a method to validate the checksums of 10 percent of our digital files, selected at random, every one to two years. This method is not

perfect, but surveys like this can alert us to failing storage media and unintended alteration of our files.

Conclusion

We cannot be so concerned with the perfection of our digital preservation that we fail to act, as is the case of numerous institutions. Many archivists have told me about their desire to tackle the digital records in their archives, but also their dismay because they lack sufficient resources or technical capability. Who can blame them? Even archives with strong institutional support, technical expertise, and staff would have a difficult time satisfying all the requirements of professional standards for digital preservation. However, some action is better than no action; every good practice we adopt helps reduce the risk to our collections. As professionals, we will always try to push our practice to higher levels of quality, but in some cases, we have to accept some imperfection to keep moving forward.

WHY DO EMPLOYERS WANT TO HIRE CERTIFIED ARCHIVISTS?

“The CA designation informs employers that an applicant has a clear level of competence for any archival position. When you receive dozens of resumes, the CA is extremely helpful in evaluating a candidate’s essential qualifications.”

-- Marcus Robyns, CA
University Archivist
Northern Michigan University

Employers who have recently advertised for Certified Archivists:

Wayne State University, MI	Temple University, PA	City of Albuquerque, NM
University of Arkansas, AR	University of Virginia, VA	Winthrop Group, NY
IU Health Methodist Hospital, IN	Rutherford County Archives, TN	Atlanta University Center, GA
Missouri Secretary of State, MO	The Daughters of Charity, MD	Jewish Institute of Religion, OH
Texas A&M University, TX	University of Georgia, GA	East Carolina University, NC
Union College, NY	City of Austin, TX	City of Newton, MA
	St. Louis University, MO	20th Century Fox Archives, CA

Create more opportunities
for yourself - register for
the 2019 ACA exam today!

Academy of Certified Archivists

www.certifiedarchivists.org • 518-694-8471 • aca@caphill.com

