Alternatives for Long-Term Storage Of Digital Information.

Chris L. Erickson
Brigham Young University
HBLL 2217, Provo, UT 84602
1-801-422-1851
chris_erickson@byu.edu

Barry M. Lunt
Brigham Young University
265 CTB, Provo, UT 84602
1-801-422-2264
luntb@byu.edu

ABSTRACT

The most fundamental component of digital preservation is managing the digital objects in archival repositories. Preservation Repositories must archive digital objects and associated metadata on an affordable and reliable type of digital storage. There are many storage options available; each institution should evaluate the available storage options in order to determine which options are best for their particular needs.

This poster examines three criteria in order to help preservationists determine the best storage option for their institution: Cost, Longevity, and the Migration Time frame. While Richard Wright maintains that "storage is becoming the lowest cost in a digital repository," Cost is probably the single most important factor when considering long term storage. Cost may be a limiting factor in the number objects that are preserved. Chapman asserts that repository storage costs "must be affordable and manageable or content owners will withhold materials from deposit." Storage costs, even if they are declining, may influence decision makers to select a low-cost storage option, at the expense of essential preservation needs. DeRidder, in her presentation "Considerations for Storage and Protection of Content", lists Cost as the first factor in choosing a storage media option.

Figure 1, included at the end of page 2, shows the costs of institutional storage, cloud storage, and alternative types of digital storage that we looked at when considering storage possibilities. (The author gathered this information from internet sources or from the storage providers directly).

Another very important criterion regarding digital preservation is the average lifespan of digital media. Selecting long-lived media for archiving digital content affects not only the end costs, but the long-term safety of the objects as well. Typical digital storage media have an expected lifespan of 3 – 10 years, though failure could occur at any time. Short-lived media, when combined with ineffective backup procedures, can result in the permanent loss of digital content.

Figure 2, Average Lifespan of Digital Media – Years, shows the often-quoted potential lifespan in years for different types of media. This figure also shows the realistic lifespan of the same media from our experience. Evidence compiled so far through observation and institutional experience show that the actual lifespan is often far below the advertised lifespan.

The only exception to this discrepancy is the M-Disc. This disc, developed at Brigham Young University and available in DVD and Blu-ray formats, is the only method that makes an irreversible physical change to a digital medium. The M-Disc is highly resistant to any of the normal factors that degrade digital objects, such as light, heat, humidity, temperature change, magnetism, bit rot and bit flips.

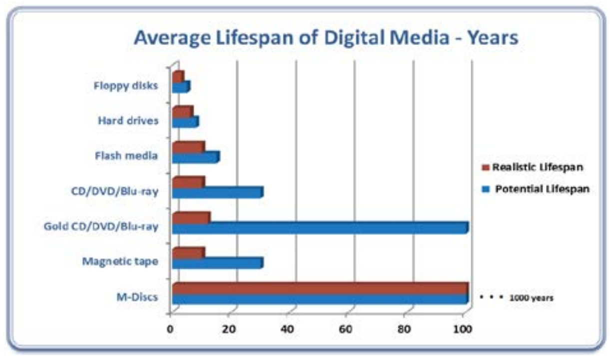

Figure 2: Average Lifespan of Digital Media – Years

The third criterion that involves both the lifespan of the media and the resulting costs of digital preservation is the Migration Time Frame. Every digital medium and every digital system has a limited lifespan. The media will eventually fail or the system will become obsolete. In order to preserve digital objects beyond the expected lifespan of the media, most digital media need to be refreshed or migrated regularly.

Figure 3 shows the expected migration time frame for hard drives, computer tapes, and the long-lived M-Disc optical discs. Each round of migration has an additional risk of data loss, and repeated migrations increase the probability of loss.

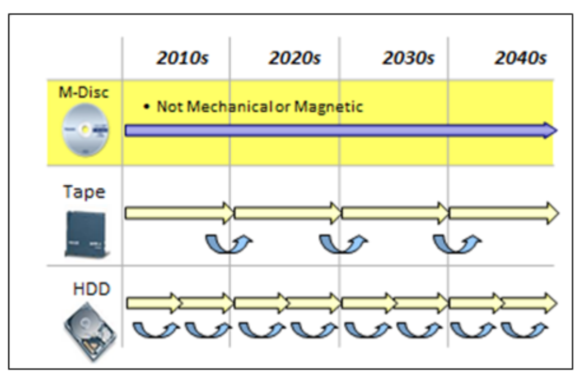

Figure 3: Migration Time Frame

The cost of migrating from one generation of media to another, or from one type of media to another type can be significant. Tape devices, such as with LT0 tapes, can only write to the current and one previous tape generation, and read the two previous generations. Thus, in three generations of LTO tapes, which currently is approximately ten years, the tapes and drives could become obsolete. Migration would include the costs of the new media and the new systems, but it must also include the personnel costs to manage the replacement and verification processes so that there is no data loss or degradation. Hard drives have a limited lifespan; they must be replaced regularly.

The Library of Congress archiving website recommends creating new media copies every five years or when necessary to avoid data loss. Since it is not possible to predict accurately when a drive or media will fail, it is important to refresh or migrate your digital media every few years. However, long-lived optical discs, such as the M-Disc DVD or Blu-ray, do not require refreshing or migration, thus adding to the cost savings.

Until recently, our institution stored full resolution digital collections in three ways: on optical discs (gold CDs and DVDs); on a variety of tape formats; and on external hard drives. Since each of these types of media experienced failures, multiple copies of every archived collection were required. Here are some problems we encountered with these media:

- Name brand gold archival CDs had an advertised life expectancy of 100 or 300 years, depending on the manufacturer. Our yearly check of collections burned to disc since 1995 found that between 2% and 5% of the discs failed annually.
- Long-term tape storage, such as Advanced Intelligent Tape (AIT) or Linear Tape-Open (LTO), usually read only two prior generations of tape, so the tape drives are often upgraded every 10 years, making the tapes obsolete sooner than the anticipated life expectancy. We have AIT2 tapes and the tape drives, but the drives are difficult to connect and to use.
- External hard drives and raid arrays have also failed and caused data loss; in one case, 8TB of master images were lost. With some important collections, one-third of external drives failed during the first year.

Typically there is no warning that digital storage is about to fail, so knowing when to refresh or to migrate media becomes a guessing game. Change too early and money is wasted. Wait too long and there is the potential of data loss. These experiences show why our institution now uses the M-Disc as one of our archival copies of long-term collections.

Each institution may have different storage policies and environments. Not every situation will be the same. By considering the criteria above (the storage costs, the average lifespan of the media and the migration time frame), institutions can make a more informed choice about their archival digital storage environment.

General Terms

Infrastructure opportunities and challenges; Frameworks for digital preservation

Keywords

Digital Preservation; Digital Storage; Storage Costs.

REFERENCES

[1] Chapman, Stephen. Counting the Costs of Digital Preservation: Is Repository Storage Affordable? *Journal of Digital Information*. 4, 2 (February. 2004), n. pag. DOI= https://journals.tdl.org/jodi/index.php/jodi/article/view/100.

[2] DeRidder, Jody L. Considerations for Storage and Protection of Content. *ASERL Webinar: Library Publishing.* (February. 2012), 14. DOI=http://www.aserl.org/wp-content/uploads/2012/02/Store_protect.pdf.

[3] Wright, Richard, Ant Miller, and Matthew Addis. The significance of storage in the "cost of risk" of digital preservation. *International Journal of Digital Curation*. 4, 3 (2009), 104-122. DOI=http://www.ijdc.net/index.php/ijdc/article/view/138.

	1 TB			100 TBs				200 TBs			
Digital Storage Costs. Simple projection only	First Year	Yearly Charge	20 Year Projected	First Year	Yearly Charge	10 Year Projected	20 Year Projected	First Year	Yearly Charge	10 Year Projected	20 Year Projected
Campus Storage											
Campus Data Center (5 yr replace)	$1,250	$0	$5,000	$125,000	$0	$250,000	$500,000	$250,000	$0	$500,000	$1,000,000
Liibrary T (40 TB for $7,500). (5 yr replace)	na	na	na	$18,750	$0	$37,500	$75,000	$37,500	$0	$75,000	$150,000
Cloud Storage											
Amazon S3 – Regular	$360	$360	$7,200	$35,406	$35,406	$354,060	$708,120	$70,812	$70,812	$708,120	$1,416,240
Amazon S3 Reduced & copy in Glacier	$288	$288	$5,760	$28,325	$28,325	$283,248	$566,496	$56,650	$56,650	$566,496	$1,132,992
Glacier storage only; no retrieval	$120	$120	$2,400	$12,000	$12,000	$120,000	$240,000	$24,000	$24,000	$240,000	$480,000
DuraSpace – Preservation	$1,875	$1,875	$37,500	$71,175	$71,175	$711,750	$1,423,500	$142,350	$142,350	$1,423,500	$2,847,000
DuraSpace –Plus. (S3+Glacier)	$2,000	$2,000	$40,000	$140,600	$140,600	$1,406,000	$2,812,000	$281,200	$281,200	$2,812,000	$5,624,000
DuraSpace – Enterprise Plus	$5,875	$5,875	$117,500	$124,675	$124,675	$1,246,750	$2,493,500	$249,350	$249,350	$2,493,500	$4,987,000
M-Discs											
DVD (@4.7 GB = 250 Discs / TB)	$638	$0	$638	$63,830	$0	$63,830	$63,830	$106,383	$0	$106,383	$106,383
BD (@25GB = 40 Discs / TB)	$200	$0	$200	$20,000	$0	$20,000	$20,000	$38,400	$0	$38,400	$38,400
BDXL (@100GB = 10 Discs / TB)	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0	$0
HLDS Storage											
Quoted Purchase Price	na	na	na	$39,500	$0	$39,500	$79,000	$72,500	$0	$72,500	$72,500
(Including server, switch, and 1.2 uplift)	na	na	na	$56,400	$0	$65,400	$83,400	$96,000	$0	$105,000	$123,000
Digital Preservation Network (DPN)											
Total DPN storage including free	$5,500	$0	$5,500	$467,500	$0	$0	$467,500	$935,000	$0	$0	$935,000

Figure 1: Estimated Costs of Digital Preservation Storage