

PRIORITIZING STORAGE MEDIA FOR DIGITAL ARCHIVING AND PRESERVATION

Leo Konstantelos

*University of Glasgow
United Kingdom
Leo.konstantelos@glasgow.ac.uk*

Emma Yan

*University of Glasgow
United Kingdom
Emma.Yan@glasgow.ac.uk*

Abstract – This paper summarizes our efforts to-date at Archives & Special Collections, University of Glasgow, to develop a methodology and tool for prioritizing archival processing of digital collections stored in physical storage media. We present the sources and process we used to develop the methodology, and outline the functionality of a prototypical tool to generate prioritization scores.

Keywords – digital archiving, legacy storage media, prioritization, selection and appraisal

Conference Topics – From Theory to Practice, Immersive Information.

I. INTRODUCTION

This paper summarizes our efforts to-date at Archives & Special Collections, University of Glasgow¹, to develop a methodology and tool for prioritizing archival processing of digital collections stored in physical storage media.

The motivation and requirement for this endeavor arose from our ongoing work on digital archiving, archival forensics, and preservation of large at-risk born-digital collections [12-15]. Since 2019, the University of Glasgow Archives have been maintaining a register of digital assets deposited and maintained in the collections as physical storage media. As the asset register continued to grow, it became evident that the micro-appraisal approach we had thus far followed to select storage media assets for processing on a case-by-case basis, which was based on predominantly empirical criteria for selection and prioritization, were insufficient to deal with growing volumes. A recurring concern was the further degradation of storage media, the majority of which being already legacy and obsolete, in

potentially aggravating conditions, to the point that they became inaccessible.

These problems are by no means new. The fragility of storage media, alongside the need to establish methodologies for addressing the preservation of the digital objects they contain, have been signaled repeatedly [e.g. 1, 4, 5, 6]. Extant studies from collecting institutions internationally, and community/web resources, have provided insight into the longevity, average lifespan and susceptibility to damage of different kinds of storage media [2, 6-11]. The 'Bit List' of Digitally Endangered Species [3] incorporates, extends, and contextualizes many of the concerns expressed in these studies, in a resource that is maintained and reviewed by the global digital preservation community.

The goal of the work presented here has been to encapsulate the knowledge/guidance deriving from these resources into a methodology that aligns with our workflows [12-13]; and use this methodology as the foundation for a simple tool to generate a priority score for processing physical storage media, that takes into account both community practice and other evidence-based criteria.

II. METHODOLOGY

In defining the criteria for prioritizing digital archiving and preservation of physical storage media, we drew from the methodological approaches suggested in [2] and [5]; and our work was further informed by the issues identified in [6]. To align the methodology with our workflows, and to keep the criteria as succinct and flexible as possible, we agreed on the following assumptions:

¹ www.gla.ac.uk/myglasgow/archivespecialcollections/

- The majority of digital assets contained in the storage media we hold in our collections are unpublished and have been transferred to our care without detailed file manifests or other descriptive documentation as to their exact contents.
- The digital assets contained in the storage media are as valuable, important, or otherwise intrinsic to the rest of the collection that they belong to, until they have been appraised.
- The digital assets contained in the storage media are unique, there is no other copy of the contents other than that on the storage media we hold, unless we have specific information to the contrary in our records.
- The initial focus will be on storage media that we can currently hold in our collections and can process via our archival forensics capability; and will exclude such legacy media as punch cards.
- Environmental conditions in which storage media were stored prior to being deposited to the University Archives, will be considered as 'aggravating' unless we hold information to the contrary in our records.

Three storage media-specific criteria were deemed as the most important in prioritizing digital archival processing and preservation:

- Average lifespan of the medium, as indicated in the examined literature [6-11].
- Year of production of the medium, as a measure of longevity and obsolescence [2].
- Environmental conditions in which the medium has been stored *after* being deposited to the University Archives. These draw from information recorded in [3] and [4].

We used the classification and – to the most part – the terminology adopted by the 'Bit List' of Digitally Endangered Species [3] as a fourth criterion, so as to inform the methodology with community practice; and avoid duplication of existing effort.

We assigned a score from 1-5 for each criterion, as shown in Table 1.

Table 1. Criteria and scores for storage media prioritization

Bit List' of Digitally Endangered Species
--

Classification	Score
Lower risk	1
Vulnerable	2
Endangered	3
Critically Endangered	4
Practically extinct	5
Average lifespan	
Lifespan	Score
1-3 years	5
3-5 years	4
5-10 years	3
10-20 years	2
More than 20 years	1
Conditions	
Conditions	Score
Optimal conditions	1
Good conservation practice	2
Minimal conservation practice	3
Some aggravating conditions	4
Mostly aggravating conditions	5
Year of production	
Produced	Score
Within the last 5 years	1
More than 5 years ago	5

We collated information from the studied sources to generate a list of storage media types that are currently help by the University Archives and assigned prioritization criteria to each medium (Table 2). The list summarizes our current knowledge on average life span and contemporaneity of storage media; and reflects the status identified in the 'Bit list' as of the time of writing. The conditions of storage were purposefully left out of this summary list, as they are bound to differ per medium – for instance, a current portable HDD may have been stored in either optimal or aggravating conditions and this can only be gauged on a case-by-case basis.

Table 2. Collated types of storage media, with prioritization criteria assigned.

Medium	Produced	Bit list status	Average lifespan (years)
Current internal HDD	Within the last 5 years	Vulnerable	3-5 years
Current internal SSD	Within the last 5 years	Vulnerable	3-5 years
Non-current internal HDD	More than 5 years ago	Critically Endangered	3-5 years
Non-current internal SSD	More than 5 years ago	Critically Endangered	3-5 years
Current portable HDD	Within the last 5 years	Endangered	3-5 years
Current portable SSD	Within the last 5 years	Endangered	3-5 years
Current optical media (CD, DVD, BlueRay)	Within the last 5 years	Endangered	5-10 years

Medium	Produced	Bit list status	Average lifespan (years)
Current magnetic tape	Within the last 5 years	Endangered	10-20 years
Current Flash storage (USB stick, SD card)	Within the last 5 years	Vulnerable	3-5 years
Floppy disk	More than 5 years ago	Critically Endangered	1-3 years
Non-current magnetic tape	More than 5 years ago	Critically Endangered	10-20 years
Cassette tape	More than 5 years ago	Critically Endangered	10-20 years
lomega zip disk	More than 5 years ago	Critically Endangered	10-20 years
Non-current optical media (CD, DVD, HDVD, Laser disc)	More than 5 years ago	Critically Endangered	5-10 years
Non-current portable SSD	More than 5 years ago	Critically Endangered	3-5 years
Non-current Flash storage	More than 5 years ago	Critically Endangered	3-5 years
Locally managed network storage	Within the last 5 years	Vulnerable	10-20 years
Cloud storage (third-party)	Within the last 5 years	Vulnerable	5-10 years
Current locally hosted web resources (websites, online databases)	Within the last 5 years	Vulnerable	5-10 years
Current externally hosted websites (websites, online databases)	Within the last 5 years	Endangered	5-10 years
Non-current locally hosted web resources (websites, online databases)	More than 5 years ago	Critically Endangered	3-5 years
Non-current externally hosted websites (websites, online databases)	More than 5 years ago	Critically Endangered	3-5 years

Lastly, we developed a five-point priority score that is meant to indicate the time period within which digital archiving and preservation action should be taken (Table 3).

Priority score	
Score	Priority level
1	Low priority - action within 3 years
2	Low priority - action within 1 year
3	Medium priority - action within 6 months
4	High priority - action within 3 months
5	Extreme priority - immediate action

III. PRIORITIZATION TOOL

The prioritization tool is a simple proof-of-concept, which calculates a priority score (Table 3) for each type of storage medium, based on the scores identified for the individual criteria plus the conditions that a medium has been held in (Figure 1).

GENERATE PRIORITY RATING	
Select storage medium:	Non-current portable SSD
What conditions has the storage medium been	Minimal conservation practice
Priority rating: 4	
Priority action: High priority - action within 3 months	

Figure 1. Screenshot of the storage media prioritization tool.

In its current version, the tool calculates a priority score using equal weights (25%) for each of the four prioritization criteria.

IV. FURTHER WORK

Being based to community-generated resources, guidance and practice, the prioritization methodology and tool are equally open to community feedback and discussion. Our aim with this piece of work is not to epitomize practice in this area, but rather invite dialogue and create a space for both further insights on handling computer storage media as archival records; and for reusing community-maintained resources, such as the 'Bit list'.

Recent discussions within our teams and with the wider digital preservation community on this topic, have highlighted issues with the score weighting. Specifically, it has been suggested that storage conditions should be given a higher weight, as it can adversely impact all other criteria.

The list of storage media that we have collated is neither complete nor comprehensive – and it is bound itself by obsolescence. Changes in community guidance, and the findings of future studies on the longevity and susceptibility of storage media, will require respective changes to the current scores. In this sense, it is an ongoing piece of work that provides the means to inform decision-making, rather than an end in itself.

REFERENCES

- [1] Digital Preservation Coalition, "Digital Preservation Handbook: Legacy media", 2nd Edition, 2015.
<https://www.dpconline.org/handbook/organisational-activities/legacy-media>
- [2] A. Brown, "Selecting Storage Media for Long-Term Preservation", The National Archives, August 2008.
<https://cdn.nationalarchives.gov.uk/documents/selecting-storage-media.pdf>
- [3] Digital Preservation Coalition, "The 'Bit List' of Digitally Endangered Species", 3rd Edition, November 2023.
<https://www.dpconline.org/digipres/champion-digital-preservation/bit-list>
- [4] A. Brown, "Care, Handling and Storage of Removable media", The National Archives, August 2008.
<https://cdn.nationalarchives.gov.uk/documents/information-management/removable-media-care.pdf>
- [5] R. Erway, "You've Got to Walk Before You Can Run: First Steps for Managing Born-Digital Content Received on Physical Media", OCLC Research, 2012.
<https://www.oclc.org/content/dam/research/publications/library/2012/2012-06.pdf>
- [6] D. Elford, N. Del Pozo, S. Mihajlovic, D. Pearson, G. Clifton, C. Webb, "Media Matters: developing processes for preserving digital objects on physical carriers at the National Library of Australia", 74th IFLA General Conference and Council, August 10-14, 2008.
<https://archive.ifla.org/IV/ifla74/papers/084-Webb-en.pdf>
- [7] C.L. Erickson and B.M. Lunt, "Alternatives for Long-Term Storage of Digital Information", Proceedings of the 12th International Conference on Digital Preservation (iPres 2015), November 2-6, 2015.
https://digitalpreservation.lib.byu.edu/wp-content/uploads/2015/11/iPres_poster_2015_resubmit2-Erickson_Lunt.pdf
- [8] R. Pinciroli, L. Yang, J. Alter and E. Smirni, "Lifespan and Failures of SSDs and HDDs: Similarities, Differences, and Prediction Models," in IEEE Transactions on Dependable and Secure Computing, vol. 20, no. 1, pp. 256-272, 1 Jan.-Feb. 2023, doi: 10.1109/TDSC.2021.3131571.
- [9] B.M. Lunt, "How Long Is Long-Term Data Storage?", IS&T Archiving Conference 2011 (Archiving 2011), May 16-19 2011.
https://www.imaging.org/site/PDFS/Reporter/Articles/2011_26/REP26_3_4_ARCH2011_Lunt.pdf
- [10] National Park Service, "Digital Storage Media", *Conserve O Gram*, vol. 22, no. 5, October 2010.
<https://www.nps.gov/museum/publications/conserveogram/22-05.pdf>
- [11] Arcserve, "Data storage lifespans: How long will media really last?", Arcserve Blog, July 2022.
[https://www.arcserve.com/blog/data-storage-lifespans-how-long-will-media-really-last#:~:text=Most%20hard%20disk%20drives%20\(HDD,desktop%2C%20or%20an%20external%20HDD.](https://www.arcserve.com/blog/data-storage-lifespans-how-long-will-media-really-last#:~:text=Most%20hard%20disk%20drives%20(HDD,desktop%2C%20or%20an%20external%20HDD.)
- [12] Archives & Special Collections, "Digital archiving workflow (high-level)", v0.6, University of Glasgow, November 2022.
[https://coptr.digipres.org/index.php/Workflow:Digital_archiving_workflow_\(high-level\)](https://coptr.digipres.org/index.php/Workflow:Digital_archiving_workflow_(high-level))
- [13] Archives & Special Collections, "Archival Forensics workflow (storage media deposit)", v0.2, University of Glasgow, November 2022.
[https://coptr.digipres.org/index.php/Workflow:Archival_Forensics_workflow_\(storage_media_deposit\)](https://coptr.digipres.org/index.php/Workflow:Archival_Forensics_workflow_(storage_media_deposit))
- [14] L. Konstantelos, E. Yan and C. Paterson, "Integrating archival forensics with digital archiving workflows", BitCurator Forum 2023, March 2023 (forthcoming).
- [15] L. Konstantelos, E. Yan and C. Paterson, "Evaluating Digital Preservation Capability With Large At-Risk Collections. Lessons Learnt From Preserving The Nva Archive", Proceedings of the 18th International Conference on Digital Preservation (iPres 2022), September 12-16, 2022, pp. 282-86.