

## **Addressing Methods of Cost Reduction in Digital Preservation**

Ryan Downie

Morgridge College of Education, University of Denver

LIS 4000: Foundations of Library, Archival, and Information Science

Professor Christine Dyar & Professor Mary Stansbury

July 13, 2020

### **Abstract**

Digital preservation of born digital data and digitized analog data is an essential aspect of modern libraries, archives, cultural heritage organizations, as well as all other organizations dealing with archived information and materials. It involves the digitization of books, manuscripts, images, photographs, and other analog documents, as well as managing the maintenance needs of digitized and born digital data (data that from its first inception has been digital in nature). These tasks are time and resource intensive, with a majority of small- and mid-sized institutions, as well as many larger ones, finding financial limitations the greatest obstacle in their preservation efforts, with poor infrastructure, policies, and insufficient training being other major difficulties. Despite many institutions' lack of resource allocations for their preservation systems, active and proactive actions are necessary for the vitality of digital data preservation: media formats and access technologies become obsolete, digital data can suffer from bit rot and corruption with time, and hesitation on behalf of institutions to establish policies and infrastructure.

The necessity of digital preservation grows in tandem with the use of digital resources for users' information needs, both remotely and on-site, which have skyrocketed in the past decade. In order to overcome the challenges of preservation, libraries, archives, and other institutions involved in digital preservation should implement methods to minimize their resource expenditure, increase their workflow and ingest, supplement staff training, and increase their storage capacity. The best methodologies for these outcomes are the simplest: collaborative partnership systems of libraries and institutions can range from two partners to more than 200,

vastly increasing overall capabilities with lower costs; cheaper and more efficient technology can be used in place of commercial options; and better workflow policies can be established.

Annotated Bibliography

Altenhöner, R. (2012). Tailoring digital preservation services for practice: Workflow

development and cost modeling. *DESIDOC Journal of Library & Information Technology; Delhi*, 32(4), 302-305.

<https://search-proquest-com.du.idm.oclc.org/docview/1413409178/abstract/1F879F95A3974B35PQ/1?accountid=14608>

- This report from the German National Library, or the Deutsche Nationalbibliothek (DNB), was written in response to difficulties encountered in the process of implementing digital preservation services into pre-existing workflows. Their reported challenges were caused by: several independent workflows capable of handling far fewer objects than were demanded of them, services requiring high levels of IT proficiency, and unadjustable systems that proved insufficient when presented with large numbers of materials.
- These presented issues, as well as the study's proposed solutions to them, are of note not only for my paper but for any organization or system pursuing an expansion of their digital preservation workflow. It also demonstrates the effects of institutional policy changes and their potential impacts on libraries, as is seen by the DNB's 2006 law change which forced the library to enact strong changes in a comparatively brief period of time.

Keskitalo, E. (2011). Costs and benefits of a shared digital long-term preservation system. *LIBER Quarterly*, 21(1), 69-85. <http://doi.org/10.18352/lq.8008>

- A cost-benefit analysis looking at digital long-term preservation (LTP) in the Finnish National Digital Library Project (NDL) through a shared system of as many as 209 institutions estimated “remarkable benefits,” such as a cost-benefit advantage of about €100 million over a 12-year period of time. The proposed system would include 209 institutions sharing information technology, human resources, organizational structures, policies, and modes of funding to “ensure the usability of information.” At the time of the projects surveys in 2009, all of the respondents felt “very unsure” regarding the quantity of digital data that they had and by how much that would increase annually, with a total estimated amount of data for 2011 being approximately 700 terabytes with an annual increase of fifteen percent. To account for this quantity of data the annual expenditures for data space such as storage tape and hard disks, as well as electricity, would remain static following the assumption that as costs of these materials decreased and efficiency increased, the system’s storage capacity growth would match that of the data.
- The relevance of this analysis to digital preservation cost reduction is clear from the estimated cost advantage alone, but by examining the methods taken to reach that estimate and establish the hypothetical system, we see much more relevance from this analysis. Smaller partners of the system gain more resources in terms of assistance, training, and especially in terms of data storage access, which is more secure (with data being stored in multiple copies in multiple locations) and cheaper. Furthermore, this system study shows that a system as large as the NDL Project at 209 partners of various sizes is feasible, and provide all-around benefits to small-, mid-, and large-sized institutions.

Kilbride, W., & Norris, S. (2014). Collaborating to clarify the cost of curation. *New Review of Information Networking*, 19(1), 44-48.

<https://doi-org.du.idm.oclc.org/10.1080/13614576.2014.898543>

- Through a survey of “gap analysis and stakeholder needs” the EU-funded project Collaboration to Clarify the Costs of Curation (4C) was designed to “ help public and private organizations invest more effectively in digital curation and preservation.” 4C found that cost models, though touted as generic and viable for many institutions, are, in fact, highly specialized to their establishing institution, not transparent with all material and investment fees, and suffer from poor usability. A tangible outcome of 4C is the Curation Costs Exchange (CCEX), which is a virtual resource designed for interested parties to view “other project deliverables, as well as other information on the costs of curation.”
- One of the key takeaways from this article is the conclusion that a library, archive, or other institution involved with preservation should not rely on a cost model system from another institute or organization without first adjusting it to their own needs, potentially combining several models to create one that is pertinent. There is also the establishment of the CCEX and its assistance for European organizations in creating cost-affordable plans; similar programs may be established by other governments of varying levels, or by library collaborations themselves for the benefit of individual branches.

Koehn, S. L., & Hawamdeh, S. (2010). The acquisition and management of electronic resources:

Can use justify cost? *The Library Quarterly*, 80(2), 161-174.

<https://doi.org/10.1086/651006>

- This article looks at a case-study in the Tulsa City-County Library system to review the “cost-effectiveness of acquiring digital resources,” as well as whether usage of digital serials and subscriptions warrant their rising costs. Their findings support the consensus that library users are increasingly using digital information at a rate higher than the costs related to the material subscriptions, resulting in the average cost-per-search to plummet. However, the study also concludes that vendors are reluctant to, or inefficient in, sharing data and lack standard definitions; therefore databases should not be compared against one another, but against themselves through year-to-year data, and libraries should be more proactive in approaching vendors for what they need.
- Although the statistics and costs referenced in this article are now slightly dated in respect to modern trends, they are still relevant and indicative of users’ needs, actions, and preferences. Additionally, the inferences made by the study’s authors regarding library practices in working with vendors and in pursuing an increase in wider data collection are of importance.

Maceli, M., & Cocciolo, A. (2017). Monitoring environmental conditions with low-cost single

board computers. *Preservation, Digital Technology & Culture; Berlin*, 46(4), 124-131.

<http://dx.doi.org.du.idm.oclc.org/10.1515/pdtc-2017-0008>

- Environmental conditions for analog material is a critical, but easily overlooked part of every archive that is commonly controlled with HVAC systems and data-loggers. What Maceli and Cocciolo investigate in this article is whether a DIY data-logger built from a Raspberry Pi 3 Model B can be as effective at recording temperature and humidity records as a commercially available PEM2 for less than 25% the price. Their results showed that the Raspberry Pi did perform just as well as the commercial alternative, however it required more IT expertise to install additional software and hardware. Additionally the Raspberry Pi, being a single-board computer, is easily adaptable to performing any other function, such as monitoring air quality for gas or particles.
- Providing options for libraries and archives with less funding to maintain safe environments for their materials is paramount, and this experiment is useful in gauging ways for smaller libraries and various institutions to find cost-cutting ways in order to divert more resources. And, though the impacts discussed within this article focus on the benefits to analog data safety, the temperature and humidity levels of libraries are important to data stored digitally as well, primarily data stored on obsolete media systems.

May, C. A. (2017). InDiPres: A statewide collaborative approach to digital preservation. *Digital Library Perspectives; Bingley*, 33(3), 221-230.

<http://dx.doi.org.du.idm.oclc.org/10.1108/DLP-08-2016-0035>

- The Library Services Technology Act's grant funding was demonstrated in this article by May to have been used for the establishment of the Indiana Digital Preservation



(InDiPres), a partnership of libraries established to “manage and sustain a low-cost, secure, and geographically distributed archive.” InDiPres is mainly orientated towards assisting its small- and mid-sized partners in the collaboration, which, during a 2015 Digital Preservation Readiness Survey conducted prior to InDiPres’s establishment, found that just over half of the survey respondents had a “lack of digital preservation workflow,” with nearly a quarter of respondents unsure of how large their own digital collections were, and 78% being interested in a statewide collaboration.

- Surveys conducted prior to the establishment of InDiPres found many of the state’s libraries lacking in digital preservation skills, workflows, and knowledge, with 73% of them citing financial limitations as being the greatest challenge to their digital preservation. InDiPres, however, has allowed smaller and mid-sized institutions to gain resources previously attainable only by large public and academic libraries, while maintaining lower costs; proving to be another example of the overall benefits for libraries of all sizes to join collaborations.

Rosenthal, D. (2020). *Archival cloud storage pricing*. DSHR’s blog.

<https://blog.dshr.org/2020/03/archival-cloud-storage-pricing.html>

- Rosenthal addresses “interruptions in the money supply” due to the current COVID-19 pandemic in this blog post as the greatest vulnerability currently facing long-term archival data storage. As a corollary to this, Rosenthal discusses the risks and misleading pricing of cloud storage systems as options for archival storage. His main focus in this blog article is that month-by-month pricing is subversive for long-term intentions (his

example of Wasabi's monthly payment program compared to its Reserved Capacity Storage option over a five-year timeframe had a cost difference of over \$10,000 to the benefit of RCS), and bears, albeit at incredibly small odds, risks of data becoming inaccessible or deleted.

- Cloud storage systems are often overlooked for their capacity to store archived data by larger organizations and institutions, and they are not ideal primary storage systems for gross volumes of data, best used as secondary or tertiary storage systems. They are, however, viable options for smaller foundations lacking the financial or infrastructure capacities to house their own data on-site or elsewhere, as well as for organizations in phases of transition who could best benefit from temporary month-by-month payment plans.

Velte, A., & Wikle, O. (2020). Scalable born digital ingest workflows for limited resources: A case study for first steps in digital preservation. *Preservation, Digital Technology & Culture; Berlin*, 49(1), 2-13. <http://dx.doi.org.du.idm.oclc.org/10.1515/pdtc-2020-0004>

- With a focus on born digital collections (i.e., "files that were created, used, and maintained digitally throughout their lifecycle without ever being physically manifested" as per the authors' definition), this study looks into issues hindering digital preservation and uses the University of Idaho's Library Special Collections and Archives (SPEC) as a model for establishing a budgeted workflow with the efficacy of a well-funded organization. Funded only by a single seed grant, SPEC used one-time purchases of hardware tools in conjunction with free, user-friendly softwares, with the remaining

funding used for the hiring of an undergraduate student assistant. Though not as efficient as well-funded institutions' workflows, the budgeted methods put forward by SPEC and this study are viable options for under-funded organizations facing risks of losing data in obsolete formats.

- This provision is an excellent method relevant to my discussion topic for the preservation of born digital data at high risk of bit rot due to technical and financial neglect.

Additionally, it contains an in-depth discussion of the unique dangers facing born digital and digitized data, such as bit rot, technical and media obsolescence, and an absence of infrastructure policies.

Zarnitz, M., Bähr, T., & Arning, U., (2019). Ten years of strategic collaboration of libraries in digital preservation. *LIBER Quarterly*, 29(1), 1-22. <http://doi.org/10.18352/lq.10278>

- The German National Specialist Libraries collaborate together to enhance their digital preservation capacities, with one partner operating as host for the preservation and other partners creating their own workflows and ingesting their digital materials. Each library benefits from “synergy effects” such as reduced costs, similar (but independent) workflows and ingests, and sharing networking and training. This paper’s goal is to portray the German National Specialist Libraries system’s digital preservation cooperation, governance, and networking “as a case study of good practice,” which it does with great self-applied praise. However the system is very efficient and effective, and does, in fact, serve “as a case study of good practice,” and the article addresses the rising importance of digital preservation in concert with the rise of professionals and

students relying on digital resources, as well as how the collaborative system managed to simplify traditional difficulties facing digital preservation.

- This article provides an example of how successful cooperation can be beneficial for libraries in terms of digital preservation, cost reduction, resource reduction, and expedite workflow and ingest. The presented model is best suited for large institutions seeking to collaborate; however, it can easily incorporate a large, well-funded partner serving as “host” to smaller, less-funded partners.

Zhou, Y. (2016). Fulfill your digital preservation goals with a budget studio. *Information Technology and Libraries*, 35(1), 26-50. <https://doi.org/10.6017/ital.v35i1.5704>

- In an undercutting move to the scanner industry, Zhou compares the efficacy of an overhead image scanner to a commercial digital single-lens reflex (DSLR) camera in the imaging of archived analog documents. The resulting comparisons show the DSLR camera, purchased at \$4,500 in 2014, to be vastly superior to the scanner, bought for \$55,000 in 2007 with an additional \$8,000 annual maintenance fee, in nearly every test.
- Focusing specifically on comparing a scanner to a commercial camera is appropriate for this discussion, as it exemplifies a larger issue of libraries, archives, and other organizations using already strained resources for the purchasing, training of use, and maintenance of technology that has cheaper, more user-friendly, and more efficient alternatives. Already well-funded libraries can benefit from spending less money on hardware that is inferior, and likewise poorer-funded libraries can gain greater digitization access, improving their workflow and ingest.