

Getting to the Bottom Line: 20 Digital Preservation Cost Questions

Matt Schultz
Grand Valley State University
1 Campus Drive
Allendale, MI 49401
1-616-331-5072
schultzm@gvsu.edu

Aaron Trehub
Auburn University Libraries
231 Mell Street
Auburn, Alabama 36849
1-334-750-1695
trehuaj@auburn.edu

Katherine Skinner
Educupia Institute
1230 Peachtree St, Suite 1900
Atlanta, GA 30309
1-404-783-2534
katherine@educopia.org

ABSTRACT

Getting to the Bottom Line: 20 Cost Questions for Digital Preservation is a cost-gathering resource created by the Outreach Committee of the MetaArchive Cooperative in Spring 2015. Launched during an Association of Southeastern Research Libraries (ASERL) webinar (<https://vimeo.com/121926212>) on March 11, 2015, this resource has been shared broadly with libraries, archives, and other institutions that have an interest in procuring digital preservation services. The easy-to-use resource is designed to equip institutions with questions that they can use to identify the full range of costs that might be associated with any particular digital preservation service—proprietary, community-sourced, or otherwise. For a variety of reasons, services of all types do not always make their costs as transparent as institutions might prefer. Using the *Getting to the Bottom Line* question-set will help ensure that institutions do not leave any stones unturned when evaluating their options and that they gather the information that they need to make informed choices that lead to sustainable solutions. Institutions are encouraged to make free use of the questions, adapt them as needed, and provide feedback on their usefulness. Going forward, the resource will serve as a foundation for building additional and more sophisticated cost transparency resources targeted toward the digital preservation community.

General Terms

Institutional opportunities and challenges; Frameworks for digital preservation; Preservation strategies and workflows; Training and education.

Keywords

Economics. Digital Preservation Costs.

1. INTRODUCTION

The question of cost—and the need for good cost models—has received extensive attention in the literature on digital preservation, including the 1996 Garrett-Waters report on “Preserving Digital Information”, the 2010 final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (“BRTF-SDPA”), and the published volume *Aligning National Approaches to Digital Preservation* (Educupia, 2012). There has been no shortage of cost models for digital

preservation, including (in roughly chronological order) products of the LIFE 1, 2, and 3 projects (UK: 2006-2012), Charles Beagrie’s Keeping Research Data Safe (KRDS) projects (UK: 2008-2011), the Cost Model for Digital Preservation (CMDP) project (Denmark: 2009-2012), the APARSEN/APA project (EU: 2010-present), the California Digital Library-Total Cost of Preservation (CDL-TCP) project (USA: 2012-present), and the 4C project (EU: 2013-2015), not to mention the work of David Rosenthal (LOCKSS), Serge Goldstein and Mark Ratliff (the Pay-Once-Store-Forever formula), Adrian Brown (“estimated value of digital assets”), and others.

This work has proven influential at-and-beyond the field level, and some of it has produced tools that people are using today—for example, the 4C Project’s Web-based Curation Costs Exchange (CCEX) calculator. There is, however, a lack of comprehensive, comparable, and reliable cost and pricing information on the digital preservation solutions—commercial, quasi-commercial, and community-based—that have emerged in the past decade and that are vying for the attention (and the money) of decision-makers at cultural memory institutions, media companies, and government departments. These decision-makers are interested in more than digital preservation theory. They want to know what a given solution is going to cost, and how that cost compares with the costs, both up-front and “hidden”, of other solutions in the marketplace.

Getting to the Bottom Line: 20 Cost Questions for Digital Preservation builds upon an “Action Session” on this topic at the Aligning National Approaches to Digital Preservation II (ANADP II) conference in Barcelona, Spain in November 2013 and, more recently, a webinar given in March 2015 as part of an ongoing series of webinars on issues in digital preservation sponsored and hosted by the Association of Southeastern Research Libraries (ASERL). As we will discuss, this list of twenty cost questions for digital preservation solutions compiled by the MetaArchive Cooperative Outreach Committee in late 2014-early 2015 provides a much-needed basis for practical cross-comparison between digital preservation solutions. Herein, we begin by providing a brief overview of transparency problems in digital preservation. We then describe the list of questions, as well as the initial results of applying them against extant solutions and the implications thereof.

2. The Problem of Cost Transparency

Digital preservation and digital archiving services and solutions are becoming widely available. Services and solutions traverse a spectrum from the strictly commercial (e.g. Amazon Glacier, Preservica, Google Cloud Storage Near Line, etc.) to the community-based (e.g., LOCKSS, MetaArchive Cooperative,

iPres 2015 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a [copy of this licence](#).

Alabama Digital Preservation Network, etc.). There are also services that fall somewhere between the two ends of this spectrum, representing a mixture of community-developed technologies that are then hosted and offered from within commercial environments such as Amazon or Rackspace (e.g. DuraCloud, ArchivesDirect, etc.).

Each service/solution offers its own unique technical approach to preserving digital information and does so via very different business models. Institutions have, for the first time, a range of maturing options that can help them to address the challenge of preserving their unique digital assets. However, understanding the significant differences between these services is challenging, even for seasoned professionals (see e.g., the POWRR Project). Establishing a clear understanding of what features each service offers, how those services might fit together to inform a preservation workflow, and what the costs associated with each service (or combination of services) will be is daunting at best, and nearly impossible to project at worst.

Understanding at a distance the range of costs that might be associated with any given service/solution is critical in the early stages of evaluating options. For a variety of reasons, this information is not always easy to obtain. Service providers may not have a fixed pricing schedule and instead prefer to negotiate pricing individually with customers. Still others withhold pricing information when it does not place them at a competitive advantage for a particular service offering (e.g., storage or subscription/licensing/membership fees). And finally, many services withhold pricing information in order to ensure that prospective customers will speak directly to them about their service offerings rather than relying on a cost sheet alone. These are all very common and familiar barriers to cost transparency, and they are increasingly encountered more generally in the library services marketplace. There is a noted culture of silence has begun to permeate many service offerings--database subscriptions and scholarly journals pricing to note a couple of prominent examples, where confidentiality clauses have helped to occlude differences in costs assessed against various institutions for access to the same content (see e.g. Bergstrom, McAfee, and Courant's work in this area).

As academic research libraries and data centers, public libraries, historical societies, museums, and other scholarly/cultural institutes seek digital preservation services/solutions, they must consciously demand cost information, and they must use that information to critically evaluate both the services available and the costs of those services. The *20 Cost Questions* are intended to empower institutions; helping them to gain the information they need regarding a range of digital preservation costs. Our hope is that their use within the community will help the entire field to avoid the longstanding transparency problems we have faced in other areas of service procurement.

Transparency in costs for digital preservation--ensuring that institutions can make sustainable choices and avoid hidden costs that might undermine their preservation missions--is vitally important as this field continues to mature and more and more services become available. Admittedly, cost transparency is often viewed as a risk factor for service/solution providers. However, when the full relationship between costs and service offerings is made more widely available it provides each service/solution provider with important information about how their offerings compare to those of others, and data that can be used for improvements and standardization of both services and business models. Furthermore, transparency around costs does not have to

equate to a race to the bottom when it comes to pricing, rather it is an opportunity for a service/solution provider to clearly argue for the excellence and return on investment of their unique approach to solving the challenges of digital preservation.

3. Tackling the Transparency Problem

Getting to the Bottom Line: 20 Cost Questions for Digital Preservation includes the diverse perspectives of academic and public library representatives who met over the course of several months to actively discuss the barriers to transparent cost gathering that they have experienced. The libraries involved in its development included Auburn University, Greene County Public Library, Indiana State University, Purdue University, and the University of Tennessee, Knoxville. Each of these libraries has been an early adopter of digital preservation services and solutions. In some cases, these libraries have chosen to experiment with and/or use multiple service offerings for the sake of comparison and benchmarking. All have ample anecdotal and evidential information from several years worth of their own efforts to advance their digital preservation agendas. In addition to identifying common barriers, they also clearly delineated the full range of digital preservation activities to which services/solutions tend to assign fees. As these institutions worked together to craft questions that other institutions could use to navigate the waters of cost transparency, they made intentional efforts to incorporate the concerns of smaller, under-resourced institutions. Through structured dialogues and interviews with smaller institutions, concerns around sustainability, requirements for local expertise, and availability of support services, among other concerns, were emphasized and given proper recognition via the questions.

Below are some examples of the 20 questions we encourage organizations to present to prospective solution providers.

1. What are the solution provider's licensing, subscription or membership fees?
 - Have these fees increased or decreased over the past three years, and why?
 - How often is the fee structure reviewed? And how are fees set?
 - How are customers/subscribers/members consulted during any such reviews?
2. Is there a minimum licensing/subscription/membership term?
3. On average, how long does it take to begin using the solution once a contract or service license agreement (SLA) has been signed?
 - What steps are involved?
4. In terms of sustainability, does the solution provider have a strategic plan, succession plan, or disaster recovery plan?
 - If so, how up-to-date are such plans?
 - Has the solution provider engaged in any audits or risk assessments?
 - Are any of the plans or audit/assessment results publicly available?

The full set of cost questions is available here: <http://www.metaarchive.org/cost-questions>.

4. MOVING FORWARD

The *Getting to the Bottom Line* question set was published in early 2015 and has already gained interest and currency. The MetaArchive and its extended community of like-missioned institutions look forward to gathering further feedback on the questions and taking this timely work on cost transparency to the next stage: namely, a Web-based matrix for collecting and comparing costs for various digital preservation solutions, using an agreed-upon set of cost elements derived from the question set and community feedback. We propose that the matrix be hosted and maintained at a community-driven and oriented organization, of which there are several respected candidates. For example, the Open Preservation Foundation (<http://openpreservation.org/>), the Digital Preservation Coalition (DPC), or the Community Owned Digital Preservation Tool Registry (COPTR: http://coptr.digipres.org/Main_Page) to name just a few. The matrix itself and the right organization can help to ensure that the resource is actively used and maintained by both the digital preservation community as well as solution/service providers from whom cost information would need to be solicited.

5. ACKNOWLEDGMENTS

Our thanks to the MetaArchive Outreach Committee for their efforts to improve transparency in the digital preservation environment.

6. REFERENCES

- [1] 4C project (EU: 2013-2015). <http://4cproject.net>
- [2] APARSEN/APA project (EU: 2010-present). <http://www.alliancepermanentaccess.org/index.php/knowledge-base/digital-preservation-business-models/costbenefit-data-collection-and-modelling>
- [3] Beagrie, Neil, Brian Lavoie, and Matthew Woollard (2010), *Keeping Research Data Safe 2* <http://www.jisc.ac.uk/media/documents/publications/reports/2010/keepingresearchdatasafe2.pdf>
- [4] Bergstrom, Ted; McAfee, R. Preston; and Courant, Paul. Big Deal Contract Project. <http://www.econ.ucsb.edu/~tedb/Journals/BundleContracts.html>
- [5] Blue Ribbon Task Force on Sustainable Digital Preservation and Access. *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*. February 2010. http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
- [6] Brown, Adrian. "Cost Modeling: The TNA Experience." The National Archives (UK). PowerPoint slides presented at the DCC/DPC joint Workshop on Cost Models, held July 26, 2005. <http://www.dpconline.org/docs/events/050726brown.pdf>
- [7] California Digital Library-Total Cost of Preservation (CDL-TCP) project (USA: 2012-present). <https://wiki.ucop.edu/display/Curation/Cost+Modeling>
- [8] Cost Model for Digital Preservation (CMDP) project (Denmark: 2009-2012). <http://www.costmodelfordigitalpreservation.dk>
- [9] Goldstein, Serge J., and Mark Ratliff (2010), *DataSpace: A Funding and Operational Model for Long - Term Preservation and Sharing of Research Data*, <http://arks.princeton.edu/ark:/88435/dsp01w6634361k>
- [10] LIFE 1, 2, and 3 projects (UK: 2006-2012). <http://www.life.ac.uk/>
- [11] Rosenthal, David (2011), "Modeling the economics of long-term storage," *dshr's blog*, September 27, 2011 <http://blog.dshr.org/2011/09/modeling-economics-of-long-termstorage.html>
- [12] Wheatley, Paul, P. Bright, and Rory McLeod (2008), *LIFE Generic Preservation Model v.1.4* <http://discovery.ucl.ac.uk/14128/1/14128.xls>