

RLG DigiNews

August 15, 2003, Volume 7, Number 4

ISSN 1093-5371

The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks

[Shelby Sanett](#)

Amigos Library Services, Inc

“Within the U.S. and elsewhere, funding agencies are advancing digital preservation as a serious research area. Digital preservation projects and cooperative international efforts have increased significantly over the past decade. Examples include the US National Science Foundation (NSF), collaborative international programs with the UK Joint Information Systems Committee (JISC), with the Deutsche Forschungsgemeinschaft (DFG), and with the European Union (EU); and the international InterPARES Project, which has received funding from a number of countries. These have spurred the development of an interdisciplinary domain that has as its primary goal ensuring long-term access to materials in digital format for legal, economic, and cultural purposes. This domain unites the interests of librarians, archivists, museum specialists, and other preservation professionals with digital object creators, computer scientists, lawyers, publishers and others. The issues cut across government, non-profit, commercial, and academic sectors.”

Editor's Note, [RLG DigiNews](#), October 15, 2002

The economics of digital preservation underlies many projects and programs exploring how to identify and resolve various practical and theoretical problems of preserving digital objects. There is a need to provide scalable, workable solutions quickly. Large numbers of born-digital and born-again electronic records and materials require immediate attention,^[1] and more will be produced over time. Along with technological advances goes a responsibility on the part of the creators and the preservers to develop both an economic framework and a context within which these processes can assure continuing access to information preserved in electronic form.

This paper explores issues of cost modeling and proposes a possible methodology to evaluate costing frameworks and models to preserve authentic electronic records. The methodology could be adapted by institutions interested in the costs of the preservation strategy under consideration. For the purposes of this paper, the term electronic materials will refer to authentic electronic records in born-digital or born-again (reformatted) digital form.

Currently several research projects and institutional initiatives are investigating a broad spectrum of issues in preserving electronic materials. The emphasis in research so far has been on the development of software and hardware to support the implementation of long-term preservation strategies. Significant funding has been provided to various projects to assess whether and how authentic electronic records can be preserved and to address other questions that have arisen from previous research. Assuming there are workable strategies for maintaining digital information, I believe we must now consider how to evaluate costing strategies, develop policies to ensure continued preservation and access, and formulate other long-term mechanisms for digital preservation.

Rationale for a Proposed Methodology to Evaluate Cost Models

Over the past several years a number of cost models and costing frameworks for the preservation of electronic materials have been advanced that consider a variety of ways to determine the full extent of the costs, including possible hidden ones.

Cost models facilitate an informed decision-making process. Over the past several years a number of cost models and costing frameworks for the preservation of electronic materials have been advanced that consider a variety of ways to determine the full extent of the costs, including possible hidden ones. Some relate costs to the life cycle of the records (Hendley),[2] the OAIS model, or a particular project[3]; identify elements of the digital preservation process; or otherwise attempt to determine categories of costs.

It is expected, however, that the full costs of preserving electronic records will be high and will extend over a long period. Therefore it is particularly important for decision makers to use a methodology to evaluate the various frameworks and models, because they must have information that is as specific as possible. This information will support the choice of a preservation strategy (indicative of the full range of

costs) or suite of strategies appropriate to a particular institution, its mission, and anticipated use of the materials.

An evaluative process is needed that can be applied when the decision-making process has begun. In the end this process should facilitate making an appropriate choice from among the cost models. So far, such a methodology to evaluate across models has not been addressed in the literature.

The requirements for an evaluative strategy of this type are complex. The proposed methodology must be flexible enough to be applied to a broad spectrum of extant models, yet credible so that the results have merit. It must be applicable to costing frameworks and models not yet developed. As well, the methodology should be user friendly and easy to apply. A daunting prospect.

A Proposed Methodology to Evaluate Cost Models

For an evaluative methodology in this area to be effective, it must be straightforward. Costing models and frameworks can be evaluated in terms of (1) acquisition and preservation-related activities, and (2) access-related activities.

Earlier I proposed a cost framework that includes three categories: (1) Costs for Preserving Electronic Records (table 1), which include capital costs, direct operating costs, and indirect operating costs; (2) Costs for Use (table 2), which are costs associated with the continued institutional use of the preserved records; and (3) User Populations (table 3), which provides information relating to access and the users' use of the records. This activity includes gathering various types of information that could then be used to provide access and user services.

Costing categories were then established in the first two categories in combination with the preservation process model developed by the Preservation Task Force of the InterPARES 1 Project.[4] The components of the activity categories may shift as necessary in the future, but the cost categories themselves are consistent with generally accepted accounting principles.

Table 1

Costs for Preserving Electronic Records

Part 1.	Capital Costs <ul style="list-style-type: none"> • Software development • Hardware (for preservation processing) • Research and development • Facilities • Interface design for processing electronic records
Part 2.	Direct Operating Costs <ul style="list-style-type: none"> • Identify potential records • Evaluate/Examine (negotiate intellectual property issues and rights) • Acquire records (staff and purchase or royalty payment) • Establish inventory record • Process (prepare for preservation, confirm authenticity/integrity of record) • Produce metadata • Preserve (select and implement appropriate strategy) • Storage (container/other) • Maintenance (refresh/migrate) • Monitor • Evaluate
Part 3.	Indirect Operating Costs (Overhead) <ul style="list-style-type: none"> • Indirect staff (supervision, clerical support, benefit times, training times, unallocated times) • Facilities (rent, utilities, off-site storage of records) • Amortization of capital costs • General and administrative (human resources, accounting, funding development and grant writing, staff training and professional development, partnerships with other institutions, policy development)

Table 2

Costs for Use of Preserved Electronic Records	
Part 1.	Capital Costs <ul style="list-style-type: none"> • Equipment, software, user training, facilities, interface design, etc.
Part 2.	Direct Operating Costs <ul style="list-style-type: none"> • Storage, royalties, communications, record access mechanisms • Staff for monitoring, user query response and services, records access management

Part 3.	Indirect Operating Costs (Overhead) <ul style="list-style-type: none"> • Indirect staff, facilities, amortization of capital costs, general and administrative
----------------	--

Table 3

User Populations	
Part 1.	Mission statement, legal mandate
Part 2.	Target user population
Part 3.	Unintended audience, i.e., as a result of exposure to records on the Web
Part 4.	User statistics

- A. The capital costs for preserving electronic records (table 1, part 1) are costs incurred at the beginning. They must be amortized over a time period, such as five years, that can then be used as the period for present value calculations.
- B. Indirect and direct operating costs for preserving electronic records (table 1, parts 2 and 3) are costs incurred on a yearly basis. They should be brought to present value (the current value of a sum of money expected to be received in the future). The period of five years is suggested because the magnitude of the investment in hardware and software is great enough to justify replacement at five years, rather than earlier.
- C. The sum of A) and B) is the total cost for preserving electronic records brought to present value. The cost per item preserved is $(A+B)/(\text{total number of items preserved})$.
- D. Operating costs for the use of preserved electronic records (table 2) are incurred on a yearly basis. These costs should be brought to present value.
- E. The sum of C) and D) is the total present value for preservation and use of electronic records. The cost per use is $(C+D)/(\text{total use of electronic records over five years [or the period used for present value calculations]})$.

To apply the proposed evaluative methodology, acquisition and preservation-related activities would include the following:

Table 4

Costs for Acquiring and Preserving Electronic Records	
Part 1.	Capital Costs <ul style="list-style-type: none"> • Software development • Hardware (for preservation processing) • Research and development • Facilities • Interface design for processing electronic records
Part 2.	Direct Operating Costs <ul style="list-style-type: none"> • Identify potential records

	<ul style="list-style-type: none"> • Evaluate/Examine (negotiate intellectual property issues and rights) • Acquire records (staff and purchase or royalty payment) • Establish inventory record • Process (prepare for preservation, confirm authenticity/integrity of record) • Produce metadata • Preserve (select and implement appropriate strategy) • Storage (container/other) • Maintenance (refresh/migrate) • Monitor • Evaluate • Delete
Part 3.	<p>Indirect Operating Costs (Overhead)</p> <ul style="list-style-type: none"> • Indirect staff (supervision, clerical support, benefit times, training times, unallocated times) • Facilities (rent, utilities, off-site storage of records) • Amortization of capital costs • General and administrative (human resources, accounting, funding development and grant writing, staff training and professional development, partnerships with other institutions, policy development)

Costs associated with access-related activities, including the institution's own use would include:

Table 5

Costs for Institutional Use/Outside Access of Preserved Electronic Records	
Part 1.	<p>Capital Costs for Use</p> <ul style="list-style-type: none"> • Equipment, software, user training, facilities, interface design, etc.
Part 2.	<p>Direct Operating Costs for Use</p> <ul style="list-style-type: none"> • Storage, royalties, communications, record access mechanisms • Staff for monitoring, re-appraising records with each new migration, deleting records, user query response and services, records access management
Part 3.	<p>Indirect Operating Costs for Use</p> <ul style="list-style-type: none"> • Indirect staff, facilities, amortization of capital costs, general and administrative
Part 4.	<p>Mission statement, legal mandate</p>

Part 5.	Target user population
Part 6.	Unintended audience, i.e., as a result of exposure to records on the Web
Part 7.	User statistics

Thus the categories referenced in tables 1, 2, and 3 have been reduced to two tables (4 and 5) when an activity-driven evaluative methodology of those categories is applied. The discussion of the process to arrive at costs to acquire, preserve, and access the records is revised as follows:

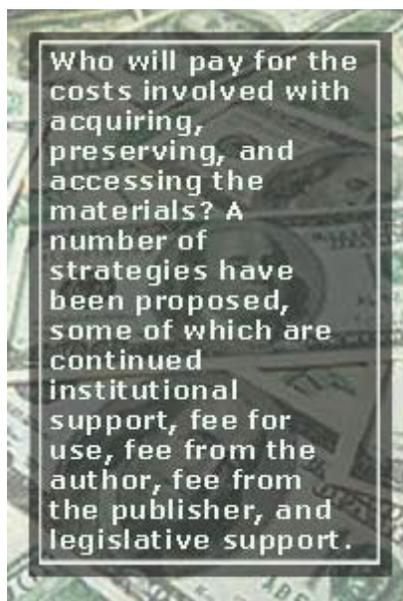
- A = The capital costs for preserving electronic records (table 4, part 1) are costs incurred at the beginning. They must be amortized over a time period, such as five years, which can then be used as the period for present value calculations. This section remains the same.
- B = Indirect and direct operating costs for preserving electronic records (table 4, parts 2 and 3) are costs which are incurred on a yearly basis. They should be brought to present value (the current value of a sum of money expected to be received in the future). To remain consistent with the original framework, the period of five years is suggested. This section also remains the same.
- C = The sum of A and B is the total cost for preserving electronic records brought to present value. The cost per item preserved is $(A+B)/(\text{total number of items preserved})$. This section also remains unchanged.
- D = Costs for institutional use/outside access of preserved electronic records (table 5, parts 1, 2, and 3) are incurred on a yearly basis. These costs should be brought to present value.
- E = The sum of C and D is the total present value for acquisition, preservation, and access to electronic records. The cost per use is $(C+D)/(\text{total use of electronic records over five years [or the period used for present value calculations]})$.

This evaluative strategy can be applied to extant and future cost models to determine the costs to be incurred to preserve electronic materials. Using this activity-driven methodology, one can compare similar categories of costs across the cost models and frameworks and against the context of a particular preservation strategy or suite of preservation strategies being examined, e.g., cost-related decisions can be determined within a context of other models as well as according to the requirements of a particular preservation strategy.

The methodology can also be customized to a particular institution by adding or deleting appropriate components of the categories. For example, an institution may determine that it must delete records or re-appraise records with each new migration and consider each of these actions to be an institutional use activity. The activity can be added to the direct operating costs in table 5 and would include the cost (planned or actual) allocated for staff to accomplish the task. If the institution determined that costs for the activity were associated with the acquisition and preservation of electronic records, the cost category would be added to table 4, part 2 (direct operating costs).

When the methodology is applied, it may in fact turn out to be less expensive for an institution to continue to maintain the records if its mission permits. In walking through this exercise, not all the costs are allocated evenly. Individual institutions knowing their own priorities and situation best would choose where to allocate a number of the capital and indirect costs. However, this approach is a step toward identifying commonalities among the extant models. Having that information should result in more-informed choices on the part of the decision makers.

How Does This Fit into the Larger Picture?



Now is the time to develop frameworks against which cost, policy, and other issues may be examined and answered. It is clear that the soft-funding scenario of the past and present is not sufficient to fund present and projected activities to preserve electronic materials. The issue of institutional sustainability in preservation must be discussed and resolved. Who will pay for the costs involved with acquiring, preserving, and accessing the materials? A number of strategies have been proposed, some of which are continued institutional support, fee for use, fee from the author, fee from the publisher, and legislative support. Infrastructure funding should be explored as well, to determine whether strategies may be successfully applied by other institutions, e.g., in universities, to determine how computer networking and other costs are paid or funded. Not all of these are possible solutions for all institutions. If institutions had a realistic idea of costs, they could plan accordingly. A cost model makes intelligent planning possible.

We must develop a strategic plan for the future to fund the long-term preservation of the world's digital and born-again digital materials. This plan should include preservation process models; costing frameworks; preservation policies; a financial, organizational, and economic infrastructure to support ongoing preservation efforts; a pedagogical platform to train future preservation administrators; a centralized funded agency to coordinate these activities; and a blueprint to develop a model of coordinated cross-institutional cooperation and regional repositories.

Footnotes

[1]Born-digital electronic records refer to those that originated in electronic form; born-again digital electronic records refer to those that originated in an analog form and were subsequently transformed, e.g., reformatted, into digital form.([back](#))

[2]In appendix 3 of his paper, Hendley provides a Table of Digital Preservation Cost Elements compiled by Neil Beagrie, Daniel Greenstein, and the Arts and Humanities Data Service.([back](#))

[3]Sanett, Shelby. "Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity," *College & Research Libraries* 63(5) (September 2002): 388-404.([back](#))

[4]Both the [US team](#) and the [International team](#) have Web sites. The InterPARES 1 Project is an international research initiative that involves national archives, university archives, and various government agencies working together with industry representatives and a team of academic researchers in archival science, preservation, and computer science to address important issues of permanent preservation of authentic electronic records. The mandate of the InterPARES 1 Project was to investigate and develop theoretical frameworks, methodologies, and prototype systems. The InterPARES 1 Project focused on the permanent preservation of inactive electronic records, that is, records that are no longer needed for day-to-day business activity, but needed to be preserved for administrative, legal, or historical reasons. Examples of such records include organizational records, legal records, and research data. Among the electronic forms these records might take are ASCII text files, graphics, video and audio material, moving graphics, e-mail with attachments, materials incorporated into a database

management system, and PDF viewer materials. The InterPARES 2 Project is currently under way. ([back](#))

Publishing Information

RLG DigiNews (ISSN 1093-5371) is a Web-based newsletter conceived by the RLG preservation community and developed to serve a broad readership around the world. It is produced by staff in the Department of Research, Cornell University Library, in consultation with RLG and is published six times a year at www.rlg.org.

Materials in RLG DigiNews are subject to copyright and other proprietary rights. Permission is hereby given to use material found here for research purposes or private study. When citing RLG DigiNews, include the article title and author referenced plus "RLG DigiNews, ." Any uses other than for research or private study require written permission from RLG and/or the author of the article. To receive this, and prior to using RLG DigiNews contents in any presentations or materials you share with others, please contact Jennifer Hartzell (jlh@notes.rlg.org), RLG Corporate Communications.

Please send comments and questions about this or other issues to the [RLG DigiNews editors](#).

Co-Editors: Anne R. Kenney and Nancy Y. McGovern; *Associate Editor:* Robin Dale (RLG); *Technical Researcher:* Richard Entlich; *Contributor:* Erica Olsen; *Copy Editor:* Martha Crowe; *Production Coordinator:* Carla DeMello; *Assistant:* Valerie Jacoski.

All links in this issue were confirmed accurate as of August 15, 2003.

[RLG DigiNews](#)

[BROWSE
ISSUES](#)

[SEARCH](#)

[RLG](#)