

Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity

Shelby Sanett

The financial challenges of the processes involved in preserving electronic records into perpetuity are significant. Financial management tools will support the decision-making processes in which archives and libraries engage when preserving electronic records. Applying business concepts, in combination with archival precepts and collection management principles, to the challenge of preserving electronic records will assist institutions such as archives and libraries in making decisions that will support their mission statements and act in the best interests of their users. This article proposes that a cost model specific to preserving authentic electronic records be developed.



he preservation of digital data for the long term presents a variety of challenges from technical to social and organizational. The technical challenge is to ensure that the information generated today can survive long-term changes in storage media, devices, and data formats.

~Raymond A. Lorie¹

The financial challenges of the processes involved in preserving electronic records into perpetuity are significant. Literature in the field discusses cost-related issues that remain to be resolved in a meaningful, applicable manner. It is necessary to develop financial management tools that will support the decision-

making processes in which archives and special collections engage when preserving electronic records. This article argues that applying business concepts such as cost-benefit analysis, risk-benefit analysis, decision-making models, and cost models, in combination with archival precepts and collection management principles, to the challenge of preserving electronic records will assist large institutions such as archives and special collections in making decisions that will support their mission statements and act in the best interests of their users, both present and future.

Financial planning tools such as cost-benefit analysis, risk-benefit analysis, decision-making models, and cost models enable managers to make informed

Shelby Sanett is the Imaging and Preservation Services Manager at Amigos Library Services, Inc.; e-mail: sanett@amigos.org. The author gratefully acknowledges the encouragement and counsel of Robert M. Hayes, Professor Emeritus, Department of Information Studies, UCLA; Michèle V. Cloonan, Chair and Associate Professor, Department of Information Studies, UCLA; Gary S. Lawrence, Director, Library Planning and Policy Development, University of California, Office of the President; and Anne J. Gilliland-Swetland, Associate Professor, Department of Information Studies, UCLA.

decisions on a range of institutional and programmatic issues. Arguably, the most useful of such financial management tools for archivists is risk-benefit analysis undertaken as part of a risk assessment strategy. Gregory Lawrence et al. defined the term *risk assessment* as "a means of structuring the process of analyzing risk."² Archives are legally mandated to take in and

An archive does not have the choice of whether or not to preserve records, only the strategy it will use.

preserve records according to evidential requirements. Libraries, with the possible exception of special collections, have more choices as to what to acquire that support their collection development and how to replace or preserve items. Because of these differences in requirements, perspectives, and missions, preservation decisions in libraries can be cost driven and use cost-benefit analysis effectively. In contrast, preservation decisions in archives are often more complicated and involve more risk assessment. An archive does not have the choice of whether or not to preserve records, only the strategy it will use. If the record is electronic, the choice of preservation strategy is circumscribed by the requirement that the preservation process result in the archive's ability to produce an authentic copy of an electronic record (i.e., reproduced to a standard that enables the copy to be admissible into court as evidence).

Today's archivist has a suite of preservation strategies, such as reformatting (i.e., refreshing and migration) from which to choose but must determine the criteria by which particular preservation strategies are chosen. As part of that process, the archivist must weigh the trade-offs involved with choosing appropriate preservation strategies for particular records. When that choice has been made, *then* cost becomes an issue likely to influence the final decision.

Su Shing Chen stated the problem that provides the context for the development of these management tools:

Despite our information technology investments, there is a critical, cumulative weakness in our information infrastructure. Long-term preservation of digital information is plagued by short media life, obsolete hardware and software, slow read times of old media, and defunct Websites. Indeed, the majority of products and services on the market today did not exist five years ago. More importantly, we lack proven methods to ensure that the information will continue to exist, that we will be able to access this information using the available technology tools, or that any accessible information is authentic and reliable. This situation creates a fundamental paradox for digital preservation: On the one hand, we want to maintain digital information intact as it was created; on the other, we want to access this information dynamically and with the most advanced tools.³

Chen also asserted that decisions about preserving information should consider the costs of preservation. "We can use current technology to determine the costs of retaining information; however, both expenditures and technology will evolve. Whereas we can project the costs for basic elements of technology—such as digital media per unit volume of information and unit processing by computers—there are no proven techniques for estimating the costs of long-term digital information preservation."⁴

Archives, libraries with archival holdings, and special archival projects and programs in the United States and abroad are increasingly implementing the technology necessary to make accessible online both born-digital and reformatted electronic records.⁵ Moreover, the present and anticipated increase in the number of electronic records also has required concomitant attention be paid to the preservation of those materials that are retained and made available for legal, fiscal, administrative, cultural, or historical reasons.

In 1999, the author and Michèle V. Cloonan undertook a survey of thirteen preservation projects and programs to obtain information on digital preservation strategies in the United States and abroad with respect to electronic records.⁶ We carried out the survey on behalf of the InterPARES Project (International Research on Permanent Authentic Records in Electronic Systems).^{7,8} Only one project surveyed, located in the United Kingdom, has included the study of cost information and formulation of policy.

Because there are no extant cost models specifically addressing the unique issues that characterize the preservation of electronic records, this article proposes that a cost model specific to preserving authentic electronic records be developed.

Background

Articles regarding accounting, costs, and cost models within a library setting from a variety of perspectives have been written about the usefulness of managerial accounting in libraries, modeling specific to the information science field, and cost studies and methods related to online bibliographic searching and evaluating information retrieval systems.⁹⁻¹² With respect to cost models, two particular articles should be mentioned: a cost model developed to estimate the costs of options associated with acquiring periodicals within libraries, and a cost model developed to assess the feasibility of storage and weeding programs within academic libraries and to provide information for planning and budgeting of such programs when they are shown to be feasible.^{13,14}

Digital library researchers Arturo Crespo and Hector Garcia-Molina have written about costs as they relate to the design of an archival repository.^{15,16} The authors defined an archival digital repository as "a repository that guarantees long-term data survivability."¹⁷ The concepts the authors discussed seem applicable to a digital library as well: "We believe our design framework and ArchSim/C can help librarians and computer scientists

make rational and economical decisions about preservation, and help achieve better archival repositories."¹⁸

Much less work, however, addresses the costs of preservation of electronic records in archives. So far, cost modeling for both archives and digitization initiatives has received scant attention. The present focus appears to be on budgeting for digital conversions rather than on preserving authentic electronic records for the long term (and thus in developing a model that will factor in the ongoing costs of the preservation process). Two exceptions from the library perspective are studies by Tony Hendley and by Kelly Russell and Ellis Weinberger.

In "Comparison of Methods & Costs of Digital Preservation," Hendley provided a Table of Digital Preservation Cost Elements that was compiled by Neil Beagrie, Daniel Greenstein, and the Arts and Humanities Data Service.¹⁹ In it, the cost elements involved in developing and preserving digital collections are keyed to the three life cycle stages of a digital resource. For example, the first life cycle stage is that of data design and data creation. Cost elements for the first stage include publications, training events, and consultancy (table 1).

Hendley also discussed seven key cost factors that must be taken into account when deciding on the preferred preservation strategy for each category of digital resource. The digital resources produce specific types of data that must be taken into consideration when making preservation decisions. The key cost factors include costs associated with data creation, data selection and evaluation, data management, resource disclosure, data use, data preservation, and rights management.²⁰ Generally, the breakdown of cost factors corresponds to the life cycle stages of the data.

Hendley discussed the development of a cost model for digital preservation that can be used to assess the ten categories of digital resources and compare the costs of the preferred methods of preservation for each category, as shown in table 2.²¹

TABLE 1
Digital Preservation Cost Elements

Life Cycle States	Cost Elements
1. Data design, data creation	1. Publications 2. Training events 3. Consultancy
2. Data accessioned into collections	1. Acquisition 2. Accessioning 3. Catalog records and documentation 4. Data processing 5. Data storage 6. Preservation 7. Monitoring reports 8. Interface design 9. Administering commercial use 10. Withdrawal fees
3. Data use and administration	1. Distributing data and documentation 2. Access and administration 3. User support 4. Royalties 5. Training 6. Publications

Source: Table of Digital Preservation Cost Elements, compiled by Neil Beagrie, Daniel Greenstein, and the Arts and Humanities Data Service in "Comparison of Methods & Costs of Digital Preservation" (1997). For a complete discussion, please refer to the study.

The cost model references the framework that divides the digital collection management into seven modules, one of which is preservation. The costs are reviewed, and those that are directly or indirectly related to preservation are totaled up to enable the (definition of) preservation costs as a percentage of the total costs of digital collection management.²²

In their article, "Cost Elements of Digital Preservation," Russell and Weinberger stated that the ongoing costs of digital preservation span a more extended time frame than those for traditional preservation and thus will require resource commitments of a different nature.²³ Different preservation strategies may necessitate different costing time frames and schedules. Russell and Weinberger stated that current cost models have yet to reflect this more complex environment where, "The creation of a digital object is the true starting point for

digital preservation." They also discussed cost issues: "These costs need to be weighed against the desirability/necessity of preserving the object. For digital materials, the preservation of complex functionality may prove considerably more costly than preservation of the basic intellectual content. In general, the more complex the digital object, the more involved (and resource intensive) the digital preservation. The question that must be asked is whether the object's perceived long-term value is worth the expense of preserving the 'bells and whistles.'" As well, the authors discussed access in terms of various levels, perhaps in response to the various types of prospective users and the information needs they bring. Thus, "access can be at a variety of levels for digital materials ranging from access to the full range of functionality and content to simply access the 'bare bones' intellectual content."

Russell and Weinberger wrote that it might not be easy or possible to separate the costs of preservation from the costs of access. They noted that the institution's investment in technical infrastructure for providing access to digital materials also supports a preservation function. Thus, the costs of both preservation and access are shared. The authors made a rare mention of archives among the work that has been done in this field in connection with costs for resource discovery and delivery of materials (i.e., costs will vary depending on the extent to which the archive is integrated into existing collection management functions where access arrangements are shared across a range of collections). Selection decisions may be based on existing policy documents or taken on an object-by-object or collection-by-collection basis.

Several cost elements presented by Russell and Weinberger incorporate a library perspective (items 1, 2, and 3 in table 3), and the preserved digital materials are stored in an archive (items 5 and 8 in table 3).

To estimate a budget for image acquisition, Anne R. Kenney and Oya Y. Rieger referred to the RLG Worksheet for Esti-

imating Digital Reformatting Costs in their book, *Moving Theory into Practice: Digital Imaging for Libraries and Archives*.²⁴ This worksheet, in combination with an assessment of costs derived by Cornell University's Department of Preservation,

In translating records into the digital environment, we have neglected to carry with them many of the features of records and record keeping that previously guaranteed their reliability and authenticity.

identifies costs for image acquisition in six cost categories: personnel, equipment, cataloging, supplies, contingency, and overhead/indirects.

In 1996, the Commission on Preservation and Access and the Research Libraries Group (RLG) published the report of the Task Force on Archiving Digital Information (available online at: <http://www.rlg.org/ArchTF/index.html>). The task force focused on materials already in digital form and "envision[s] the development of a national system of digital archives which it defines as repositories of

TABLE 2
Digital Resources and Data Types

Digital Resources	Data Types
Data sets	Alphanumeric data
Structured texts	Alphanumeric data; markup data; tags to graphics
Office documents	Alphanumeric data; raster & vector graphics; moving graphics
Design data	Alphanumeric data; raster & vector graphics
Presentation graphics	Alphanumeric data; raster & vector graphics; moving graphics
Visual images	Raster graphics
Speech & sound recordings	Audio data
Video recordings	Video data
Geographic/mapping data	Alphanumeric data; raster & vector graphics
Interactive multimedia publications	Audio/video data; moving graphics; alphanumeric data; raster & vector graphics

Source: Table of Digital Resources and Data Types, compiled by Daniel Greenstein in "Comparison of Methods & Costs of Digital Preservation" (1997). For a complete discussion, please refer to the study.

TABLE 3
Cost Elements for Digital Preservation

1. Selecting a particular digital object for preservation
2. Negotiating the right to preserve the object
3. Negotiating the right to provide access to the preserved object
4. Determining the appropriate technical strategy for preservation and continuing access
5. Validating the completeness of the object on delivery to the archive
6. Producing metadata
7. Storing files
8. Administering the archive

Source: Table of Cost Elements for Digital Preservation, from Russell and Weinberger, *Cost Elements of Digital Preservation* (2000). For a complete discussion, please refer to the study.

digital information that are collectively responsible for the long-term accessibility of the nation's social, economic, cultural and intellectual heritage instantiated in digital form." The report explored cost modeling and concluded that "little systematic understanding has yet emerged of the actual costs of digital archiving." In connection with the Yale Cost Model, the report explored whether the costs of storage and access of digital texts would be less expensive than the traditional paper environment. The conclusion was that it would not be. Cost factors include equipment, maintenance, overhead, printing, and delivery.

A relatively recently initiated project, The CAMiLEON Project (Creative Archiving at Michigan and Leeds: Emulating the Old on the New) is a three-year project jointly funded by JISC (Joint Information Systems Committee) in the United Kingdom and the NSF (National Science Foundation) in the United States. It began in October 1999, and its central aim is to evaluate emulation as a digital preservation strategy. Among several issues the project is investigating are the costs to develop an emulator and a cost-benefit of emulation versus migration.²⁵

In sum, to date, very few resources are available on this topic. However, the literature recognizes that determining cost factors in connection with digital preservation as being important, and research that incorporates exploring costs is being funded.

The InterPARES Preservation Process Model and Identification of the Need for Cost Models

Whether a document is authentic depends on its transmission and how it is preserved. Electronic records give no reassurance of authenticity because they are so easily altered at numerous points of transmission. In translating records into the digital environment, we have neglected to carry with them many of the features of records and record keeping that previously guaranteed their reliability and authenticity. The issue of maintaining the reliability and authenticity of electronic records is the challenge that has been taken up in this research project being spearheaded by members of the archival profession and, on the basis of this knowledge, to formulate model (s), policies, strategies, and standards to ensure their preservation in and over time.

To meet this goal, the InterPARES project's research plan has been divided into four interrelated domains of investigation addressed by task forces: appraisal, authenticity, preservation, and strategies. The work undertaken by the Preservation Task Force addresses cost implications most directly.

The InterPARES Preservation Task Force selected two approaches to use in order to meet its goals. First, it used a modeling method called Integration Definition for Function Modeling (IDEF) to map the preservation activity for authen-

tic electronic records. Second, Michèle V. Cloonan and Shelby Sanett undertook a survey of thirteen preservation projects and programs to obtain information on the range of preservation strategies in use in the U. S. and abroad. Cost categories for preserving authentic electronic records were mapped upon the preservation process model developed by the Preservation Task Force using IDEF notation and methodology. The Preservation Task Force designated this preservation process model, IDEF-Ø. This approach facilitates the development of a cost model that parallels the preservation process model that will ultimately be tested.

Function Modeling Method

A function model represents the functions, activities, or processes incorporated within the modeled system in a structured manner. The purpose of using modeling in InterPARES was to articulate the functions, information, and resources required to preserve permanent, authentic electronic records.²⁶

The resulting preservation process model has been constructed within the framework established by the Reference Model of an Open Archival Information System (OAIS), which is an ISO International Standard and includes those activities specifically required for the preservation and delivery of electronic records. OAIS was accepted as an ISO International Standard in April 2002.

The primary strength of IDEF is that the notation and methodology has proven effective in detailing the system activities for function modeling. Activities can be described by their inputs, outputs, controls, and mechanisms (ICOMs). Additionally, the description of the activities of a system can be easily refined into greater and greater detail until the model is as descriptive as necessary for the decision-making task at hand.

One of the observed problems with IDEF models is that they often are so concise that they are understandable only if the reader is a domain expert or has participated in the model development. One

other limitation is the tendency of IDEF models to be interpreted as representing a sequence of activities. In reality, the preservation activities would overlap and that is not readily apparent from the model.

The Preservation Task Force developed a functional model of the process of preserving authentic electronic records following the Integrated Definition (IDEF) method prescribed by the InterPARES International Team. Specifically, it used IDEF to describe processes or functions involved in preserving electronic records. Because InterPARES was concerned with the preservation of electronic records that had been selected for preservation after they were no longer needed for the practical purposes for which they had been created, the process described in the model began with the transfer of the records from their creator, or from an agent acting for the creator, to a person whose primary responsibility is that of preserving authentic electronic records. When developing the InterPARES preservation model, the Preservation Task Force began with raw data (generally interview results with domain experts, such as those gathered from the survey). By grouping together activities that are closely related or functionally similar, a hierarchy began to emerge. Modeling software allowed the researchers to see immediately, when making changes on one level of the model, the effect of these changes both below and above that level.

Benchmark and baseline requirements to prove the authenticity of electronic records were developed by the Authenticity Task Force. For much of the duration of the InterPARES project, the Authenticity Task Force was developing the requirements. As the preservation process model was being developed, requirements to prove the authenticity of electronic records were being developed as well. Yet, the Preservation Task Force could not delay work to wait for outcomes from the Authenticity Task Force. Thus, the process model developed by the Preservation Task Force is essentially neutral, that is, a framework in which require-

ments for authenticity and authentication can be introduced and integrated within the process model via the use of “placeholders.”

The requirements for assessing and maintaining the authenticity of electronic records fall into two groups: those that support the presumption of the authenticity of electronic records before the records are transferred to the custody of the preserver, and those that support the production of authentic copies of electronic records that have been transferred to the custody of the preserver.^{27,28} All of the requirements included in the baseline requirements must be met before the preserver can attest to the authenticity of the electronic copies in its custody. This is why the requirements for the production of authentic electronic copies are termed “baseline” requirements. Satisfaction of

these baseline requirements will enable the preserver to certify that copies of electronic records are authentic. With electronic records, the difficulties related to preservation make it prudent for the preserver to produce and maintain documentation relating to the manner in which it has maintained the records over time as well as the manner in which it has reproduced them to support its attestation of authenticity.

The three baseline requirements developed by the Authenticity Task Force are satisfied by the framework of the preservation process model as presented in table 4.

The preservation process model developed by the Preservation Task Force took an approach contrary to that discovered by the survey in many projects. Rather than giving priority to the technological challenges of preserving electronic records,

TABLE 4
Baseline Requirements Developed by the Authenticity Task Force

Requirement	Satisfied by
1. Controls over records transfer, maintenance, and reproduction	Terms & conditions of transfer; preservation action plans; preservation methods; record reproduction methods
1.a. Unbroken custody of the records is maintained.	Selection of creator's records for preservation; terms and conditions for transfer; transfer and preservation of records
1.b. Security and control procedures are implemented and monitored.	Access control and access privileges of a database management system
1.c. The content of the record remains unchanged after reproduction.	Preservation methods that preserve the content of electronic records
2. Documentation of reproduction process and its effects	Preservation action plans store a record of updates to digital components. The effect of a preservation action plan on the reproduction of form and content is assessed.
3. The archival description of the fonds containing the electronic records includes information about changes the electronic records of the creator have undergone since they were first created.	Information supporting the presumption of authenticity of transferred records; preservation action plans store records of refreshment of digital media and updates to digital components.

TABLE 5
Costs for Preserving Electronic Records

Part 1. Capital Costs
Software development
Hardware (for preservation processing)
Research and development
Facilities
Interface design for processing electronic records
Part 2. Direct Operating Costs
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)
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)

the model enables the preserver to select appropriate preservation strategies over time, thus including those technologies that have not yet been developed. It should be noted that what is being preserved is not the electronic record but, rather, the ability to reproduce the record. Any preservation strategy or technology chosen must be feasible, and feasibility includes affordability. Developing a cost model was not one of the original goals of the Preservation Task Force; however, the preservation process model provides for the application of a cost model in developing preservation strategies, evaluating preservation action plans, and evaluating their execution.

The IDEF-Ø preservation process model incorporates four activities:

1. Manage the preservation function.
2. Bring in electronic records.
3. Maintain electronic records.
4. Reproduce electronic records.²⁹

As has been noted by the authors previously discussed, cost categories related to the preservation activities must be identified. Three broad cost categories were established:

- costs for preserving electronic records (table 5), which include capital costs, direct operating costs, and indirect operating costs (overhead);
- costs for use (table 6), which are costs associated with the institutional use of the preserved records;
- user populations (table 7), which provide information relating to access and the user's use of the preserved records.

TABLE 6
Costs for Use of Preserved Electronic Records

Part 1. Capital Costs for Use

Equipment, software, user training, facilities, interface design, etc.

Part 2. Direct Operating Costs for Use

Storage, royalties, communications, record access mechanisms

Staff for monitoring, user query response and services, records access management

Part 3. Indirect Operating Costs for Use

Indirect staff, facilities, amortization of capital costs, general and administrative

Costing categories were then established in the first two categories in combination with the IDEF-Ø preservation process model and generally accepted accounting principles. Types of information that could be generated and used to provide access and user services are identified in table 7.

Cost Elements Connected to the IDEF-Ø Preservation Process Model Developed by the InterPARES Preservation Task Force

A. The capital costs for preserving electronic records (table 5, part 1) are costs incurred at the beginning. They must be amortized over a time period, such as five years, which then can be used as the period for present value calculations.

It is impossible to predict how archives and libraries will be organized, and thus managed, a hundred years from now.

B. Indirect and direct operating costs for preserving electronic records (table 5, parts 2 and 3) are costs incurred on a yearly basis. They should be brought to present value (the value now 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 replacing at five years, rather than earlier.

C. The sum of A and B gives the total costs of 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 6) are incurred on a yearly basis. These costs should be brought to present value.

E. The sum of C and D gives the total present value for preservation and use of electronic records. The cost per use is $(C + D) / (\text{total use of electronic records over five years [or the period used for present value calculations]})$.

These cost categories are specific to the IDEF-Ø preservation process model developed by the InterPARES Preservation Task Force. However, there are elements in common with the work of Hendley et al., Russell and Weinberger, and Kenney and Rieger. Moreover, there are implications for further research to determine cost elements related to multimedia materials/records and music in electronic form.

Gap in Existing Cost Models to Preserve Electronic Records

So far, there is no published research related to modeling the costs of preserving

TABLE 7
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 Web

Part 4. User statistics

authentic electronic records. Some cost elements in the models discussed above may map onto the preservation process. For example, in general, cost models that are designed for nonarchival institutions may not transfer satisfactorily because of fundamental differences in professional mission and perspective, including adherence to evidentiary requirements and the needs of secondary users of the preserved electronic records. For that reason, existing cost models that are not specific to long-term preservation of electronic records must undergo extensive adaptation in order to meet the costing needs of long-term preservation of electronic records. For example, the cost categories described by Russell and Weinberger and by Kenney and Rieger might be adapted and augmented through the addition of categories, such as costs involved in providing access to the materials and costs related to long-term maintenance of digital materials, production of metadata, administration, and research and development.

Requirements that are applicable to archival environments must operate as standards in the prospective preservation cost model; for example, staffing, education, and equipment must be considered as well. These additional costs are different from those that libraries might incur due to the archives' need to meet evidential requirements. For example, costs that are specific to meeting these standards might include extra metadata creation and constant integrity checking, both of which would impact the labor/staffing cost categories. Increased or additional costs could occur as a result of the limited choices of preservation strategies available to archives.

The most challenging area for adapting existing cost models is to successfully incorporate the predictive component. Although most library-oriented cost models incorporate a five- to one-hundred-year span as the forecast period, the unique qualities of electronic records require forecasting costs into perpetuity. Calculating the costs of staff, equipment,

and storage could remain strictly a mathematical exercise because influences on these expenses over the long term derive from a variety of sources. The value of money also will change over time because of inflation and other factors.

Although there are tables to calculate future values, the issue remains as to how far into the future these calculations will be reliable, at least to the extent that an archive can make a long-term decision that would require it to commit institutional resources. Social structures change, too. It is impossible to predict how archives and libraries will be organized, and thus managed, a hundred years from now. How useful will the cost model be beyond fifty years? Should it focus on one type of preservation mechanism or on a suite of preservation tools? Traditionally, preservation decisions have been made on the basis of use, cost, and available remedies. Following in this tradition, a useful cost model should be flexible and incorporate options that enable the preservationist to choose from a suite of preservation remedies available now and from those available in the future, but of which we have no knowledge at this time. By bringing the costs described in tables 5 and 6 to present value figures, it may be possible to address this issue.

The effect of Moore's Law should be considered in developing a cost model, particularly in this arena which is dependent on current technologies and looks toward future technological advances as sources of preservation tools and remedies.³⁰ Moore was the first to observe the relationship between cost and the amount of information that could be stored on siliconintegrated circuits. He theorized that the amount of information that could be stored in this way would double every year until 1975. At that point, the doubling period slowed to eighteen months and that rate has held through the 1990s. Although it is not within the scope of the article to debate the difference between a law, a model, and an algorithm, Moore's Law does demonstrate a general rule or principle that is thought to be true or held

to be binding.³¹ Over the years, the question of whether there is a finite limitation to Moore's Law has arisen several times. As chip production processes become smaller and smaller, even down to the atomic (or molecular) level, physical limitations of Moore's Law may be reached by 2017.³² The question arises as to whether the industry is pursuing or following Moore's Law.

Computer prices have fallen even as computer capabilities have risen—and consumer spending on computers has risen for the last twenty years at an average annual rate of 24 percent. In 1995, Moore reviewed microchip progress and saw “increasingly difficult” technical roadblocks to staying on the path predicted by his law. However, he was most worried about the rising cost of manufacturing increasingly complex chips. He noted that, “Capital costs are rising far faster than revenue ... the rate of technological progress is going to be controlled [by] financial realities.” The implication is that some technical innovations may not be economically feasible.³³

There are implications for developing a reliable cost model as well. If, in fact, Moore's Law is in effect through 2017, for example, can the cost of equipment accurately be calculated after that year? Can the rate increase of costs to produce chips be determined and factored into a preservation cost model? As of this writing, during the past twelve months, orders for durable goods, which include computers, electronics and autos, plummeted 9.1 percent in comparison to the previous year and the U.S. economy is currently in a recession. What effect might these events have on equipment prices in the coming year? Three years? Five years? More? The time frame of “perpetuity” provides a certain amount of cushion because it can be argued that these “bumps” average out over time, similar to the bumps experienced by the stock market over time. Being able to predict equipment costs accurately is crucial to developing a reliable preservation cost model because these expenses constitute substantial initial and

ongoing expenses, certain to tax institutional budgets and stun appropriations committees.

The other side of this discussion is the potential role of depreciation for equipment and its effect as a component of the preservation cost model. Depreciation is the estimated decrease in value of a fixed asset over its useful life. It enables the allocation of the cost of the equipment, which will be of use to the organization for more than one year, into accounting periods into the future. Equipment wears out or becomes technologically obsolete, and depreciation provides a means of writing off the costs of these assets over their remaining useful life. From a financial view, depreciation is a source of cash because it is an expense that reduces taxes.³⁴ If, according to Moore's Law, the equipment will become outdated every eighteen months, equipment costs allocated into future accounting periods due to depreciation will become an important issue because the life of the equipment is short and replacing the equipment will require a substantial capital investment.

According to Moore's assessment in 1995, the cost of equipment may not actually decrease even though capacity is doubled. How practical will it actually be to replace equipment this often? What sort of capital investment will need to be made and budgeted for? Moore's Law must certainly be taken into consideration in developing a preservation cost model, but for how long? What will be the effect of rising costs to produce microchips on cost projections for preservation equipment? What will be the effect of depreciation of preservation equipment on developing a predictive cost model? Ideally, the preservation cost model would address depreciation into perpetuity.

Traditionally, in preservation arenas, funding for preservation work has been obtained from grants. Issues include the variable amounts, research constraints, and projects built around funding. Due to the anticipated ongoing and high expense of preserving authentic electronic records, perhaps now is the time to look

at developing cooperative preservation and access partnerships with other institutions, across the boundaries of libraries, archives, museums, historical societies, and other cultural heritage institutions. Additionally, perhaps now is the time to pursue a legislative mandate at state and federal levels to fund preservation efforts by these institutions on an ongoing basis.

Profitability is an issue to be explored with respect to archives and libraries. As preservation-related costs impact institutional budgets, it is foreseeable that funders will look ahead to exploring opportunities to generate revenue in order to offset costs and develop a rate of return on the institutional investment. Possibilities might include instituting access and copy fees. Most national archival institutions are not permitted to charge for the actual document, but charges can be assessed for copies. However, when the copies are requested online, some thought will have to be given to managing the process.

Finally, there must be a middle ground in developing a cost model to preserve electronic records, one that is neither so broad that it is unrealistic and lacks credibility nor so narrow in scope that it is perceived to be of limited use.

A Difference in Perspectives

According to Ken Thibodeau of the National Archives and Records Administration, "One of the key elements that distinguishes archives from other institutions which preserve information is that archives' essential responsibility is to preserve and deliver authentic records to subsequent generations of users."³⁵ Ensuring the authenticity of the preserved record is a requirement that is unique to the archival field because the preserved electronic records must meet evidential requirements for admissibility into court.

It is the responsibility of archivists to ensure the authenticity, reliability, and long-term accessibility of permanent electronic records.³⁶ Traditionally, archivists have done so by gathering documents,

establishing provenance, and maintaining and demonstrating an unbroken chain of custody in an evidence-based approach to managing records.³⁷ Is it possible to ensure authenticity and reliability of records regardless of their formats? If so, which technologies and techniques are archivists using to preserve electronic records at the time of their creation as well as throughout their life cycle?

These principles apply, even though over the years the media being preserved has expanded from paper to digital material, audiovisual and sound materials, and electronic records. Much of the information about properties necessary to be preserved, uses and users, and the unique needs of various types of institutions has been obtained as a result of various projects and programs that have been funded.

Are there differences in how librarians and archivists determine costs to preserve electronic records? Quite likely. Librarians are usually responsible for materials that, if they become missing or damaged, can generally be replaced. Archivists have custody of unique materials that are not replaceable. When a record is destroyed, it is really gone, whether the destruction is due to a disposal schedule or an accident. Because certain cost elements, such as operating costs, equipment, and administrative costs, are fairly consistent between and among the institutions, it is realistic to envision a cost model that will be relevant to the practices and needs of both of these types of institutions.

A trend that is becoming apparent in large academic libraries is that the libraries are becoming the legacy holders of their university's administrative records, which are usually in a combination of paper and electronic formats. Organizationally, these new duties often fall within the purview of the Department of Special Collections in libraries. University archives become the repositories of these materials and are often positioned as a department within the Department of Special Collections or are attached to the library organization laterally. A fascinating aspect of this trend is

that librarians must begin to think like archivists; that is, they will not have a choice of the records to take in and will become responsible for ensuring that the records are stored, maintained, preserved, and available for access to an evidential standard for as long as is needed. The issues of ensuring authenticity of the records, which have arisen for archivists, will arise for librarians as well. Unfortunately, unlike archivists, librarians are not equipped by training to meet these evolving requirements. It is quite possible that, eventually, the impact of electronic records and other digital materials will enable a blending of the two professions on some level—or at least provide the impetus to spur a review of job descriptions and training programs.

The amount of electronic materials that are being created and require preservation is increasing. Moreover, the rate of increase will continue to accelerate.

The Role of Financial Analysis in Preserving Electronic Records

As we have seen, technical processes cannot be separated from economic issues. The library and archival professions have not fully explored the implications of economic influences on preservation decisions. Approaches for evaluating the full economic impact of digital preservation should be identified.

The preservation of authentic electronic records does not operate within an information vacuum. Rather, it is part of a larger infrastructure composed of policy and decision-making tools, including cost models. The process of preserving electronic records and providing access to them will require institutions to make many decisions. These decisions will involve a certain amount of risk that cannot be avoided but might be minimized by attention to the financial implications of each decision. Ideally, cost models assist management in making these decisions by grouping factors and variables that are related to a particular organizational function into an easy-to-use, flex-

ible formula that results in predictive cost information related to that function.

After costs are determined, two financial planning tools support using that information to make the best possible decisions for the institution and the electronic records, in conjunction with the organization's mission statement. The first is decision tree analysis, which combines decision points with probabilities and costs to produce better information. It is particularly effective in assisting decision making in relation to buildings or equipment. Using this method of analysis, it is possible to look at a variety of alternatives. With this method, each decision point has more than one choice. When the decision points are charted, they look like branches—hence the name “decision tree.”³⁸

Cost-benefit analysis is the second financial planning tool, which will supply information to support decision making. A cost-benefit analysis measures the relationship between anticipated returns and losses and the anticipated return on investment. The benefit applies to assets that will benefit the organization in the long term (i.e., more than one year), such as equipment, buildings, or research. A cost-benefit analysis also can be used to compare two different proposals or projects, especially if they are for tangible items.³⁹

These tools can be customized for the institution and its needs, and software is available that supports the analysis. After costs are determined, both of these tools can be used to maximize the use of information so that a more informed decision can be made. Decision tree analysis helps to determine payoff for two or more alternatives under various economic conditions through the use of cost, revenue, and probable demand. Cost-benefit analysis can make use of probabilities and utilizes comparisons.⁴⁰

On analysis of investment decisions, Erich A. Helfert wrote:

Business investment decisions are made continuously within the larger context of business strategy.

This context evolves over time.... Risk is inherent in all estimates of future conditions because of the uncertainty about most variables.... Inflation and specific price changes in revenues and costs can complicate both the estimating process and the level of the investment standards, and they must be handled consistently in both. Analytical techniques can provide ranges of results and quantitative insights of considerable sophistication, but they can't supplant qualitative business judgments that reflect the broader context of strategy and risk assessment.⁴¹

Avenues of cooperation between and among institutions should be explored. Shared/cooperative preservation facilities will enable a rosier financial picture to be achieved because allocating preservation costs among institutions will reduce costs for individual institutions. The cooperative model is especially advantageous when applied to small- to medium-sized institutions and to low circulating records, which cost the same to process, maintain, and preserve as high circulating records do. Low-use records do not realize a high rate of return via access or copy fees. Cross-institutional cooperation could offset these costs to some extent. The role of publishers in this cooperative schema should be explored as well. This concept of cross-institutional cooperation, although not new, is also worth considering because it must be assumed that at some point, the funders of the expensive and ongoing preservation process will expect to explore areas where they can realize a return on their investment. It is prudent to plan for this eventuality at the outset.

A predictive, yet credible, preservation cost model and a combination of financial planning tools, qualitative business judgment, and knowledge of the unique requirements of preserving electronic materials will together support decisions made regarding the beneficial use of available funding.

Conclusion

The rush to develop the technological processes necessary to preserve authentic electronic records appears to have come at the expense of first addressing cost and policy. Electronic records are unique, as are the preservation requirements. "Traditional" preservation of paper-based and other physical media incorporates passive preservation strategies that are intended to last, perhaps, a hundred years. Digital preservation is action oriented and extends into perpetuity. In the United States, we do not yet have an infrastructure that adequately supports digital preservation. The data indicate that we have been focusing on techniques rather than developing planning strategies and policies, including cost models.⁴²

Funding for preserving electronic records will remain an issue to be addressed and resolved. The amount of electronic materials that are being created and require preservation is increasing. Moreover, the rate of increase will continue to accelerate. Even without an extant cost model that is specific to preserving electronic records, it is clear that preserving authentic electronic records into perpetuity is far too expensive a process to rely on being subsidized by grants. As well, the mandate of archives to preserve records to an evidential standard truly falls within a legislative purview. Soft funding should be used to finance special or priority projects. The softfunding scenarios of the past are no longer enough to assist institutions that must take the long view of preserving and providing access to electronic materials. These institutions cannot be expected to bear the burden of financing the preservation of societal memory by themselves on a grant-by-grant basis.

"There are numerous challenges before us, but also enormous opportunities to contribute to the development of a national infrastructure that positively supports the long-term preservation of digital information. Such an infrastructure is a desirable outcome that will benefit us only if we conceive and structure it to benefit those served by our successors'

successors."⁴³ Preservation of electronic records needs to operate within a context of strategic planning, infrastructure, and policy, which has yet to be developed for digital media. Components to be developed would include preservation policy, preservation funding planning and development, preservation decision-making models, and preservation cost models

that integrate the unique properties which electronic materials and the institutions that house them share. An agenda for the not-too-distant future includes the need to develop and then incorporate preservation of electronic materials and the related components of that process into the big policy picture on global, national, and organizational levels.

Notes

1. Raymond A. Lorie, "Long-term Preservation of Digital Information," in *Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries, June 24–28, 2001, Roanoke, Virginia, U.S.A.* (New York: ACM, 2001), 346.
2. Gregory Lawrence et al., *Risk Management of Digital Information: A File Format Investigation* (Washington, D.C.: Council on Library and Information Resources, 2000), 5.
3. Su-Shing Chen, "The Paradox of Digital Preservation," *Computer* 34, no. 3 (Mar. 2001): 1. Available online from <http://is.gseis.ucla.edu/us-interpares/R3ChenS.pdf>.
4. *Ibid.*, 6.
5. "Born-digital electronic records" refers to those records that originated in electronic form; "reformatted electronic records" refers to those records that originated in an analog form and were subsequently transformed into digital form.
6. Michèle Valerie Cloonan and Shelby Sanett, "Comparing Preservation Strategies and Practices for Electronic Records," *New Review of Academic Librarianship* 6 (2000): 205–16.
7. Web sites of the U.S. team (<http://is.gseis.ucla.edu/us-interpares/>) and the international team (<http://www.interpares.org/>). The InterPARES 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 related to the permanent preservation of authentic electronic records. The project is investigating and developing theoretical frameworks, methodologies, and prototype systems required for the permanent preservation of authentic electronic records. It focuses on the permanent preservation of inactive electronic records; that is, records that are no longer needed for day-to-day business activity but need to be preserved for administrative, legal, or historical reasons. Examples of such records might 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-mails with attachments, materials incorporated into a database management system, and PDF viewer materials.
8. This study was prepared, in part, with support from the National Historical Publications and Records Commission (NHPRC).
9. Robert M. Hayes, "Simulation and Modeling in the Information Sciences," a talk given at the 2nd Annual Conference on Urban Planning Information Systems and Programs at the University of Pittsburgh, Sept. 24–26, 1964: MR-16, 1964.
10. ———, "Managerial Accounting in Library and Information Science Education," *Library Quarterly* 53, no. 3 (1983): 340–58.
11. Michael D. Cooper, "Evaluation of Information Retrieval Systems: A Simulation and Cost Approach" (PhD. diss., School of Librarianship, University of California at Berkeley, May 1971), 47–78, 177–84.
12. Michael D. Cooper and Nancy A. DeWath, "Technical Report: The Cost of On-line Bibliographic Searching," *Applied Communication Research* (Dec. 1975): 1–46.
13. Vernon E. Palmour, Marcia C. Bellassai, and Robert R.V. Wiederkehr, *Costs of Owning, Borrowing, and Disposing of Periodical Publications* (Arlington, Va.: Center for Naval Analyses, 1977), 1–65.
14. Gary S. Lawrence, "A Cost Model for Storage and Weeding Programs," *College & Research Libraries* 42 (Mar. 1981): 139–47.
15. Arturo Crespo and Hector Garcia-Molina, "Modeling Archival Repositories for Digital Libraries," extended version, in *Proceedings of the Fourth European Conference on Digital Libraries (ECDL)* (Sept. 2000). Available online from <http://www-db.stanford.edu/~crespo/publications/index.html/>.

16. ———, "Cost-driven Design for Archival Repositories," in *Proceedings of JCDL '01, June 24–28, 2001, Roanoke, Virginia, U.S.A.* (New York: ACM, 2001), 363–72.
17. *Ibid.*, 364.
18. *Ibid.*, 372.
19. Tony Hendley, "Comparison of Methods & Costs of Digital Preservation" (1998): 96. Available online from <http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html>.
20. *Ibid.*, 45–55.
21. *Ibid.*, 86–87.
22. *Ibid.*, 78–79.
23. Kelly Russell and Ellis Weinberger, "Cost Elements of Digital Preservation." Available online from <http://www.leeds.ac.uk/cedars/documents/CIW01r.html>. [CEDARS Project]
24. Anne R. Kenney and Oya Y. Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, Calif.: Research Libraries Group, 2000), 166.
25. Stewart Granger, "Digital Preservation & the CAMiLEON Project." Available online from <http://dspace.dial.pipex.com/stewartg/cam-london.htm>. [CAMiLEON Project]
26. The IDEF notation and methodology is a U. S. Federal Information Processing Standard. In December 1993, the Computer Systems Laboratory of the National Institute of Standards and Technology (NIST) released IDEF as a standard for Function Modeling in FIPS Publication 183.
27. Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records.
28. Baseline Requirements for the Production of Authentic Copies of Electronic Records.
29. Much of the material in this section was obtained from the drafts of the Authenticity Task Force Final Report and the Preservation Task Force Final Report. For a complete discussion, see the reports at <http://www.interpares.org/>.
30. Gordon E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics* 38 (Apr. 19, 1965). Available online from <http://www.intel.com/research/silicon/moorespaper.pdf>.
31. In 1965, Moore addressed the implications regarding cost writing that "Reduced costs is one of the big attractions of integrated electronics, and the cost advantage continues to increase as the technology evolves toward the production of larger and larger circuit functions on a single semiconductor substrate."
32. Michael Kanellos, "Moore Says Moore's Law to Hit Wall," *Business Week Online* (Sep. 30, 1997). Available online from <http://news.cnet.com/news/0-1003-202-322592.html>.
33. Charles C. Mann, "The End of Moore's Law?" *Technology Review* (May/June 2000). Available from http://www.technologyreview.com/articles/mag_toc_may00.asp.
34. James O. Gill and Moira Chatton, *Financial Analysis: The Next Step*, revised edition (Menlo Park, Calif.: Crisp Publications, Inc., 2001), 50.
35. Kenneth Thibodeau, "Building the Archives of the Future: Advances in Preserving Electronic Records at the National Archives and Records Administration," *D-Lib Magazine* 7, no. 2 (Feb. 2001): 1. Available online from <http://www.dlib.org/dlib/february01/thibodeau/02thibodeau.html>.
36. For various definitions of authenticity and authenticity of digital information, see Nancy Brodie, "Authenticity, Preservation and Access in Digital Collections," *New Review of Academic Librarianship* 6 (2000): 225–38.
37. Anne Gilliland-Swetland, *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment* (Washington, D.C.: Council on Library and Information Resources, 2000), 11–12.
38. Gill and Chatton, *Financial Analysis*, 81, 92.
39. *Ibid.*, 97.
40. *Ibid.*, 100.
41. Erich A. Helfert, *Techniques of Financial Analysis: A Modern Approach*, 9th edition (New York: Irwin/McGraw Hill, 1997), 187–237.
42. Cloonan and Sanett, "Comparing Preservation Strategies and Practices for Electronic Records," 214.
43. "Preserving Digital Information," Report of the Task Force on Archiving of Digital Information commissioned by the Commission on Preservation and Access and the Research Libraries Group, May 1, 1996, 44. Available online from <http://www.rlg.org/ArchTF>.