

D-Lib Magazine July/August 2004

Volume 10 Number 7/8

ISSN 1082-9873

Thirteen Ways of Looking at...Digital Preservation

[Brian Lavoie](#)

OCLC Research

<lavoie@oclc.org>

[Lorcan Dempsey](#)

OCLC Research

<dempseyl@oclc.org>

Introduction

Research and learning are increasingly supported by digital information environments. The as yet unfulfilled promise is a rich fabric of scholarly resources, learning materials, and cultural artifacts, seamlessly integrated and readily accessible, organized in ways that facilitate traditional uses and encourage new uses as yet undefined.

Fulfilling this promise requires the cultivation of stakeholder communities that, through their working and learning experiences, meaningfully engage with digital information environments. Meaningful engagement is, in turn, contingent on the following prerequisites [1]:

- *Predictability and comprehensiveness*: A critical mass of digital resources must be developed. Where coverage is intermittent and/or unpredictable, usefulness is diminished and stakeholder interest will not grow.
- *Interoperability*: Digital content must be easily shared between services or users; usable without specialist tools; surfaced in a variety of environments; and supported by consistent methods for discovery and interaction. Digital content should also be managed using well-understood practices, and supported by services that can be re-combined to meet new user needs.
- *Transactionability*: Mechanisms are needed to establish authoritatively the identity of content, services, and users interacting within the information environment, as well as to manage intellectual property rights and privacy, and to secure the integrity and authenticity of content and services.
- *Preservability*: The long-term future of digital resources must be assured, in order to protect investments in digital collections, and to ensure that the scholarly and cultural record is maintained in both its historical continuity and media diversity.

Of these four requirements, the last—preservation—has been the slowest to work its way

into digital information environments. That is not to say the issue has been ignored: in fact, there has been much concern and speculation regarding the prospects for long-term stewardship of digital materials. This has motivated an ambitious research agenda, shared by cultural heritage institutions, government agencies, and even private enterprise, aimed at identifying and resolving the challenges posed by digital preservation.

Much of this work approaches digital preservation as a self-contained problem, focusing on the technical obstacles that must be overcome in order to secure the long-term persistence of digital materials. Success, in this context, rests on the ability to prove that technical solutions, in one form or another, exist.

Even as this important and necessary work proceeds, our understanding of the totality of the challenges associated with maintaining digital materials over the long-term is coming more sharply into focus. New questions are emerging, having less to do with digital preservation as a technical issue *per se*, and more to do with how preserving digital materials fits into the broader theme of *digital stewardship*. These questions surface from the view that digital preservation is not an isolated process, but instead, one component of a broad aggregation of interconnected services, policies, and stakeholders which together constitute a digital information environment.

Digital preservation issues worked their way into the consciousness of cultural heritage institutions in the form of a sense of imminent crisis. Expressions such as "digital dark age" were put forward, with the implication that whole portions of the scholarly and cultural record were on the brink of disappearing. But accumulating experience in managing digital materials has tempered this view. While it is true that digital materials are inherently more fragile than analog materials, the degree of risk varies widely across classes of resources: there is appreciable risk, for example, that a Web site available today may be gone tomorrow, but there is little indication that the corpus of commercially published electronic journal content is under the same threat.

In this sense, the focus of digital preservation has shifted away from the need to take immediate action to "rescue" threatened materials, and toward the realization that perpetuating digital materials over the long-term involves the observance of careful digital asset management practices diffused throughout the information lifecycle. This in turn requires us to look at digital preservation not just as a mechanism for ensuring bit sequences created today are renderable tomorrow, but as a process operating in concert with the full range of services supporting digital information environments, as well as the overarching economic, legal, and social contexts. In short, we must look at digital preservation in many different ways. With apologies to Wallace Stevens [2], this article suggests thirteen ways of looking at digital preservation [3].

I. Digital preservation as...an ongoing activity

Preservation traditionally proceeds in fits and starts, with extended periods of inactivity punctuated by bursts of intensive effort—witness the Brittle Book campaigns of the 1980s, or recent efforts to save movies filmed on nitrate cellulose film stock. The pattern is one in which materials are left to approach a state of crisis, at which point the situation is remedied through large-scale intervention.

But digital materials generally do not afford the luxury of procrastination. The fragility of digital storage media, combined with a high degree of technology dependence, considerably shortens the "grace period" during which preservation decisions can be

deferred. Issues of long-term persistence can arise as soon as the time digital materials are created: for example, in choosing between a widely-used, stable digital format, and one that is obscure or on the verge of obsolescence. This sense of urgency is driven largely by the fact that it is problematic to apply digital preservation techniques *ex post*—i.e., after deterioration has set in. While a print book with a broken spine can be easily re-bound, a digital object that has become corrupted or obsolete is often impossible (or prohibitively expensive) to restore. Digital preservation techniques are most effective when they are pre-emptive.

This suggests that as more and more digital materials come under the stewardship of collecting institutions, preservation will become less like an *event* occurring at discrete intervals, and more like a *process*, proceeding relatively continuously over time. As a consequence, it will become more difficult to distinguish preservation activities from the routine, day-to-day management of digital materials.

It is important that the sudden ubiquity of preservation processes in digital collection management does not interfere unduly with other components of the digital information environment. Implementation of preservation measures should be as transparent as possible to users of digital materials, and should not represent obstacles to access and use. In the print world, preservation of rare book collections is achieved in part by restricting usage: materials are accessed under the supervision of a librarian and off-premises circulation is prohibited. While these measures undoubtedly prolong the life of these valuable materials, they do little to promote their use. In the case of digital materials, mechanisms to ensure long-term persistence should operate harmoniously with mechanisms supporting dissemination and use.

II. Digital preservation as...a set of agreed outcomes

It is one thing to recognize that actions must be taken to secure the long-term persistence of digital materials; it is another to articulate precisely what the outcome of preservation should be.

This issue is not confined to digital materials. [Nicholson Baker](#) (2001), for example, has decried reformatting efforts that result in the loss of the original item; to Baker, preservation of the original is the measure of successful preservation. To others, however, destructive microfilming meets their preservation needs, in that content is transferred to a medium with a life expectancy of half a millennium.

Similar questions are attached to the preservation of digital materials, but the issues involved are amplified. Digital content often embodies a degree of structural complexity not found in physical materials. It can subsume multiple formats, being at once text, images, animations, sound, and video; it can be interactive, providing tools for the user to create alternate views of the content, or link to new content; it is mutable, in that it can be updated or enhanced over time; it can be broken apart, with the pieces distributed and used individually, or re-combined to create new resources. In short, digital content can incorporate features with no equivalent in the analog world. How many of these features can or should be preserved?

Unfortunately, there is no single answer to this question. For some purposes, a preserved digital object must be a perfect surrogate for the original, replicating the full range of functionality, as well as the original "look and feel". But for other purposes, intensive preservation of this kind is unnecessary: perpetuating the object's intellectual content

alone, or even a diminished approximation of the original object, is enough. The period of archival retention is also a point of debate. For some, nothing less than retention in perpetuity constitutes successful preservation; for others, a finite period is sufficient.

These considerations suggest that the choice of preservation strategy will need to reflect a consensus of all stakeholders associated with the archived digital materials. Achieving such a consensus is difficult, and in some circumstances, impossible. A second-best solution is for the digital repository to articulate clearly what outcomes can be expected from the preservation process. These outcomes should in turn be understood and validated by stakeholders. Communication between the repository and stakeholders, either to promote consensus on preservation outcomes, or for the repository to disclose and explain its preservation policies, mitigates the risk that the repository's commitments are misaligned with stakeholder expectations.

III. Digital preservation as...an understood responsibility

The likelihood that digital preservation activities will proceed continuously throughout the information lifecycle suggests that preservation responsibilities will extend beyond traditional stewards of the scholarly and cultural record. If, for example, preservation considerations must be taken into account at the time of a digital object's creation, it is authors and publishers, rather than libraries and archives, who must take the first steps toward securing the long-term persistence of digital materials.

The need for entities beyond collecting institutions to play a role in preservation is not new: the publishing industry, in response to the brittle books crisis, recognized and acted on the necessity to produce printed materials on acid-free paper. In the digital realm, entities who do not regard preservation as part of their organizational mission will find the scope for their involvement in the preservation process greatly expanded. Consequently, the responsibility for undertaking preservation will become much more diffused.

The rapid take-up of networked digital resources, obtained through license or subscription, has led to portions of the scholarly and cultural record—e.g., electronic journals, e-books, and Web sites—lying outside the custody of collecting institutions. This has prompted anxiety about the long-term stewardship of these materials, in particular when economic value has diminished while cultural importance has not. Since the value of certain digital materials can persist indefinitely, those who have custody of these materials during the various stages of the information lifecycle must recognize and act upon the need to manage them in ways compatible with long-term preservation.

The division of labor for preserving print materials is well-established. The division of labor in regard to *digital* preservation has yet to be determined—for example, clarification of legal deposit requirements for digital materials will be a key factor in determining how much of the digital preservation burden will be allocated to national libraries or archiving agencies. But the distribution of digital preservation responsibilities is almost certain to include decision-makers outside the cultural heritage community. It is important that these decision-makers understand the necessity of taking steps to secure the long-term persistence of the digital materials under their control.

IV. Digital preservation as...a selection process

Preservation of print materials is both a benign by-product of production and distribution

modes, and a process of active decision-making and intervention. Preservation of digital materials will reflect a similar mix, although the dividing line between benign by-product and active decision-making remains to be drawn. But as the volume of information in digital form continues to expand rapidly, an issue emerges that will surely require active decision-making and intervention: what should be preserved?

It is safe to assume that preserving everything is not an option. Digital preservation is expensive, and it is therefore impractical to make every bit of information in digital form the subject of active preservation measures throughout its entire lifecycle. Given this, two options remain. One is to collect as many digital materials as possible and deposit them into mass storage systems. The stored materials could then be sifted over time, with selections for more intensive preservation periodically made as need and/or interest arises.

The "save now, preserve later" strategy is feasible only through the unique characteristics of digital information, where the steady decline in storage cost makes it conceivable to save everything. The chief criticism of this approach is summarized by the adage "saving is not preserving"; there is considerable uncertainty concerning the extent to which preservation techniques can be applied retrospectively to digital materials that have resided untouched in storage for long periods of time.

The second strategy is selection: that is, determining from the outset which digital materials should be preserved and taking steps to curate them throughout their lifecycle. The choice of which materials to preserve is a difficult one, and will depend on a number of factors, including institutional mission, cultural preferences, economic practicality, and risk management policies. The question will also hinge on the digital medium's impact on the scholarly and cultural record...is an e-mail discussion list, for example, part of the scholarly record, and if so, should it be preserved with as much care as the contents of a peer-reviewed journal?

Selection is not just a "preserve or not preserve" issue. It also involves the level of desirable intervention for a particular set of digital materials. Is it necessary to go to the trouble and expense of preserving a digital object in its original form? Or is preservation of the intellectual content enough? This issue presents difficult choices, but in a world of scarce preservation resources, these choices must be confronted.

V. Digital preservation as...an economically sustainable activity

Two key economic challenges plague efforts to preserve digital materials. First, allocation of funds to digital preservation has been insufficient; [Neil Beagrie](#) (2003) observes that in the context of funding decisions, the need to take immediate and frequent actions to preserve digital collections usually is overshadowed by the desire to create and disseminate new forms of digital content. Second, funds that are made available are usually provided on a temporary basis, often as grants to support one-off undertakings or special projects. Few institutions have allocated ongoing, budgeted resources for the long-term care of digital materials.

The impulse to fund digital preservation activities is dampened by the expectation that the costs will be formidable. It is difficult to forecast the precise magnitude of these costs, which will depend on factors such as system architecture, length of archival retention, scale, and preservation strategy. But regardless of their form, digital preservation activities will require a substantial resource commitment to sustain them over time.

Economic sustainability is the ability to marshal sufficient resources, on an ongoing basis, to meet preservation objectives. There are many avenues by which sustainability can be achieved. An institutional commitment to budget a continuous supply of funds to support digital preservation is one; these funds might be used to extend a pilot project originally funded through seed money from a grant-giving organization. Digital preservation activities might also be self-sustaining, generating revenues as a by-product of day-to-day operations. In these circumstances, economic sustainability might be defined in terms of cost recovery, or a minimum level of profitability.

Strategies for attaining economic sustainability must be built on a sound empirical footing; consequently, much more data on the costs of digital preservation is needed. Digital preservation is still in its infancy, and much of the available data is heavily skewed toward upfront costs: reformatting, setting up the digital repository, ingestion of materials, etc. As projects mature, empirical descriptions of digital preservation's complete cost trajectory will emerge. This data must be consolidated and synthesized to produce reasonable benchmark estimates of the cost requirements associated with various forms of digital preservation.

VI. Digital preservation as...a cooperative effort

The fact that digital preservation is expensive, funding is scarce, and preservation responsibilities are diffused suggest that digital preservation activities would benefit from cooperation. Cooperation can enhance the productive capacity of a limited supply of digital preservation funds, by building shared resources, eliminating redundancies, and exploiting economies of scale.

In order to persuade institutions to invest in bringing digital collections online, and to make these collections a meaningful part of research and learning experiences, there must be assurance that the collections will persist. But long-term stewardship may be beyond the means of an individual institution. Aggregating collections into "union archives", maintained and funded as a shared community resource, would serve the dual function of promoting shared access and distributing the costs of long-term maintenance over a larger stakeholder community. The fact that both the benefits of access and the costs of long-term maintenance are shared by a large number of institutions would furnish a strong incentive to contribute materials to these shared digital collections.

Cooperation would also minimize redundancy. The characteristics of digital information are such that relatively few archived copies of a digital resource will likely be required to meet preservation objectives. The rationale for this assertion can be framed as follows. Sharing analog materials is generally more expensive than sharing digital materials: to access an archived copy of a print book, users must either travel to the book's location, or request that the book be shipped via interlibrary loan. To reduce access costs, it is desirable to preserve many copies of the same print book in geographically dispersed locations. In contrast, the ease with which digital information can be replicated and shared over networks suggests greater scope for preserving a particular digital resource in a single location, rather than preserving copies in multiple locations [4]. This can introduce significant cost savings by minimizing the incidence of redundant, fragmented efforts, multiple learning curves and reinvention of wheels.

Finally, cooperation opens possibilities for realizing greater efficiencies through economies of scale. Maintaining digital materials over the long-term will require an elaborate and costly technical infrastructure, as well as specialized human expertise. It is

economically impractical for every collecting institution to develop local digital preservation capabilities. A coordinated approach promises to be more cost effective, by spreading fixed costs over a greater number of institutions. It also might make certain kinds of highly specialized, or "niche", digital preservation activities economically feasible, by expanding them to a sufficiently large scale to bring costs in line with benefits. These activities might be otherwise impractical if done piecemeal on a small scale.

VII. Digital preservation as...an innocuous activity

In some circumstances, digital preservation is perceived as a threat to intellectual property rights. Much of this resistance can be attributed to the current ambiguity surrounding copyright law as it pertains to digital materials; the principles of fair use and legal deposit are in particular need of clarification.

Digital materials purchased through license or subscription, such as electronic journals or e-books, illustrate the collision between the need to intervene to preserve digital materials and the need to protect intellectual property rights. These materials are typically accessed over the Web through a central server controlled by the content provider, rather than through locally maintained copies. In these circumstances, the entities who perceive the need to preserve—i.e., collecting institutions—are often distinct from the entities that hold the right to preserve, as well as custody of the materials. Publishers are reluctant to distribute digital copies of their revenue-generating assets, even for preservation purposes, to individual licensees or subscribers; few institutions would have the resources to preserve the materials even if they did.

This presents two options: the content provider must be persuaded or enjoined to preserve the materials in their custody; or alternatively, the content provider must cede the right to preserve to another entity who is willing and able to assume responsibility for preservation. Currently, the latter approach seems to be in ascendance, evidenced by the emergence of "escrow repositories" or "archives of last resort". For example, the publisher Elsevier has agreed to transfer a copy of the content available through its ScienceDirect service to the National Library of the Netherlands [5], with the understanding that the Library will maintain this material in perpetuity and assume the responsibility for making it available should circumstances prevent Elsevier from doing so through its own systems.

Other issues remain to be resolved. In order to meet preservation objectives, the archiving agency may have to alter the archived content in some way—for example, by migrating it to another format in order to keep pace with changing technologies, or by disaggregating complex objects into more granular resources, such as breaking up an issue of a journal into its constituent articles. In these circumstances, appropriate permissions must be obtained from the rights holders in order to give the repository sufficient control over the archived materials to carry out its preservation responsibilities.

Striking a balance between the interests of content providers and collecting institutions may best be achieved through appropriately designed contracts. In the United States, copyright law is generally superseded by contract law; therefore, regardless of current interpretations of fair use or legal deposit, all stakeholders in a set of digital materials may address preservation requirements through provisions included in licensing or subscription agreements. An example of this is found in the UK's Model License governing digital materials licensed to UK higher education institutions. The Model License includes archiving clauses which identify the need for libraries to have continued access to purchased materials following the license's expiration, and commits the publisher to

address this need as part of the licensing agreement [6].

VIII. Digital preservation as...an aggregated or disaggregated service?

For the most part, digital preservation systems have been designed "holistically", combining raw storage capacity, ingest functions, metadata collection and management, preservation strategies, and dissemination of archived content into a physically integrated, centrally administered system. But other organizational structures are also possible: for example, digital preservation activities might adopt a "disaggregated" approach, where the various components of the preservation process are broken apart into separate services distributed over multiple organizations, each specializing in a focused segment of the overall process.

A digital preservation system can be deconstructed into several functional layers. The bottom layer includes hardware, software, and network infrastructure supporting the storage and distribution of digital content. The next layer includes more specialized services to manage the archived content residing in the system, including metadata creation and management, and validation of materials' authenticity or integrity. Preservation measures are implemented in the next layer of services, including monitoring the repository's environment for changes that could impact the ability to access and use archived content, as well as initiating processes such as migration or emulation to counteract these changes. The top-most layer includes services that support browsing or searching, access requests, validating access permissions, and arranging for delivery.

This range of functions can be offered as separate yet interoperable services that can be combined in various ways to support different forms of repository activities. For example, some digital materials might require only "bit preservation"—i.e., an assurance that the bit streams constituting the digital objects remain intact and recoverable over the long-term. Other materials, however, may require more sophisticated preservation services: i.e., migration to new formats, or the creation of emulators to reproduce the content's original look, feel, and functionality. Some preservation efforts will require "active archives", characterized by a relatively continuous process of ingest and access; other efforts might submit materials for preservation at irregular and widely-spaced intervals, with little or no user access.

These preservation activities utilize various combinations of some or all of the services described above. A fully integrated system may find that one or more services end up under-utilized and therefore of insufficient scale to realize technical or cost efficiencies. On the other hand, entities that specialize in only a few of these services may be able to spread them over a larger collection of digital materials, and in doing so, attain the necessary scale to realize economies within the limited sphere of their chosen service layer. This reflects Adam Smith's classic argument for specialization in production, or a *division of labor*. Determining the extent to which digital preservation can benefit from a division of labor, in the sense of finding 1) a sensible deconstruction of the digital preservation process into a set of more granular services, and 2) the optimal degree of specialization across preserving institutions, is a key issue in the design of digital repository architectures.

IX. Digital preservation as...a complement to other library services

Although much work remains to be done to resolve the challenges specific to preserving

digital materials, it is not too early to begin thinking about how digital preservation mechanisms will be integrated with, and operate alongside of, the wide range of other services which, taken together, constitute a digital library.

The notion of "dark archives", supporting little or no access to archived materials, has met with scant enthusiasm in the library community. This suggests that digital repositories will function not just as guarantors of the long-term viability of materials in their custody, but also as access gateways. Fulfilling this dual mission requires that preservation processes operate seamlessly alongside access services. Preservation should not impede access or reduce the scope for sharing information. Careful records of the outcome of preservation processes must be kept: for example, in cases where material is migrated to new formats, users must understand which versions of a particular digital resource are available for access, and what alterations, if any, have been made to these versions as a consequence of preservation.

As preservation assumes a more prominent role in the day-to-day management of digital collections, preservation activities will co-exist, and at times, operate in concert with, other routine collection management functions, such as acquisition, description, and ILL fulfillment. When a new digital resource is acquired, it is simultaneously ingested by the digital repository's archival system. At the same time that the resource is being prepared for circulation, it must also be prepared for long-term retention. Not only must the resource be surfaced in the library's access environments (e.g., through a new record in the OPAC), it must also be surfaced in the library's "preservation system". Digital content management systems must find ways to integrate preservation tools and services into their environments.

It is essential that preservation actions be as transparent as possible to users of archived digital materials. It would be unfortunate if the preservation process were such that the scope for sharing digital materials across systems, institutions, and users was reduced. In the print world, preservation often exacts a heavy toll on users' ability to access material, by removing books from the shelves while they are re-bound, filmed, or scanned; by placing rigorous restrictions on circulation; or even removing the materials from circulation entirely. The characteristics of digital information are such that access and use of archived materials can be supported without comprising preservation objectives, but achieving this in practice requires explicit recognition of the impact of preservation on access (and *vice versa*) in the design and implementation of digital library systems.

X. Digital preservation as...a well-understood process

There is as yet little consensus on best practice for carrying out the long-term preservation of digital materials. Prospects for cultivating a shared view on this issue hinge on three factors: identification and development of standards to support digital preservation; suitable benchmarks and evaluative procedures for assessing the outcomes of digital preservation processes; and mechanisms for certifying adherence to a minimum set of practices on the part of digital repositories.

The emergence of standards would benefit many aspects of the preservation process. Some progress can already be reported. The [Open Archival Information System reference model](#) (2002), which details a conceptual framework for an archival repository, as well as the environment in which it operates and the information objects it manages, has been well-received and extensively applied in the digital preservation community. But many other areas remain to be addressed, ranging from preservation-quality digital formats to optimal

preservation strategies for various classes of digital materials.

Digital preservation would also benefit from the articulation of benchmarks or metrics for evaluating the efficacy of preservation processes as they unfold. Preservation activities necessarily require institutions to incur costs well in advance of realizing benefits. How can decision-makers be assured that investments to preserve digital collections are producing tangible results? It would be useful to devise a widely accepted set of evaluative procedures, similar to a quality assurance audit and based on measurable aspects of the preservation process, that would serve as a reliable indicator of how well preservation activities are progressing toward meeting preservation objectives.

Finally, well-understood processes for preserving digital materials must be paired with mechanisms for assessing whether a particular digital repository commands the expertise and resources to carry them out. Preservation requires institutions to transfer valuable (and often, rare and priceless) materials into the custody of the repository and its staff. These transfers must be accompanied by a high degree of confidence that the materials will be preserved according to well-known, established procedures. Such conditions exist in preservation microfilming, where fragile printed materials such as old newspapers and books are entrusted to service providers with the understanding that the materials will be returned unharmed. A similar element of trust must be cultivated in the digital preservation community. One way to contribute to this is through the establishment of certification procedures for digital repositories. Certification would indicate that a repository has met certain minimum requirements in its curatorial policies and procedures, including conformance to what is regarded as current best practice in digital preservation.

Development and take-up of standards and evaluative metrics, along with certification of digital repositories, will help dispel fears that scarce resources devoted to preservation will be wasted as digital materials are managed using non-standard or outmoded practices, and as a consequence, fail to release their value in use.

XI. Digital preservation as...an arm's length transaction

The responsibility for ensuring the permanence of the scholarly and cultural record is deeply rooted in the library, museum, and archival communities. But the characteristics of digital materials—their fragility, technology-dependence, and networked access—has unsettled preservation's traditional division of labor.

While it is certain that collecting institutions will continue to serve as the primary stewards of society's memory, it is unlikely that every collecting institution responsible for the curation of digital materials will have the resources and expertise to implement the entire digital preservation process locally. Part of the preservation responsibility may be taken up by third-party services specializing in the preservation of digital materials. In this event, digital preservation activities would be conducted as an arm's length transaction between separate parties. This raises several questions concerning how such a transaction would take place.

An obvious issue is pricing. The costs of digital preservation are subject to the vagaries of numerous factors, chief among which is the constantly evolving technological environment to which digital materials are so closely intertwined. The more rapid the pace of technological change, the costlier it will be to ensure that archived digital objects remain usable. Given the uncertainty over the pace and direction of technological change, it is difficult to estimate future preservation costs, and therefore, suitable pricing scales.

Wide-spread use of relatively stable digital formats and technology would mitigate this problem, but not eliminate it.

Sustainable pricing models must also be developed. Several possibilities exist: the repository could charge a one-time, upfront capitalized archiving fee; alternatively, it could distribute the fees over time, perhaps as an annual fee. Pricing models must strike a balance between customers' preferences (e.g., inability to pay a large upfront fee, or desire to avoid budgeting ongoing funds) and those of the repository (e.g., difficulty in collapsing future preservation costs into a one-time fee, or need to invest large sums upfront to meet future preservation commitments).

A related question concerns what is supplied in exchange for payment. What preservation *guarantees* can the digital repository offer? To what compensation is the depositor entitled if promised outcomes are not achieved? Should the repository guarantee a specific *outcome* associated with its preservation process ("these digital objects will be renderable, using contemporary technology, in fifty years"), or should only the *process* itself be guaranteed ("these digital objects will be recorded on up-to-date digital storage media, refreshed at regular intervals, and maintained under environmentally controlled conditions")? Resolution of these issues must emerge from a convergence of customer expectations and repository commitments.

XII. Digital preservation as...one of many options

An implicit assumption attached to most discussions of digital preservation is that materials currently in digital form must be preserved in digital form. For some materials—i.e., born-digital materials with no obvious print equivalent—there may be no choice but to preserve them as digital objects. But a large class of materials, including digital surrogates of analog items, as well as born-digital objects for which analog equivalents can be easily produced, present other options in addition to digital preservation. Indeed, analog manifestations of digital materials may already be the subject of preservation efforts, even as their digital equivalents are perceived to be at risk. Efforts to preserve digital materials must take into account potential overlap with analog preservation activities, as well as circumstances where preservation in analog form may be preferable to digital preservation.

A document in digital form comprised solely of text and static images can be easily reproduced as a paper document with little or no loss of information. In making this document part of the permanent scholarly or cultural record, which form should take precedence? For example, most researchers in the digital preservation community are familiar with the Council on Library and Information Resources (CLIR) reports in the maroon covers. These reports are available in print form, and may also be downloaded from the Web in digital form. Which copy should be the focus of preservation activity? In this case, the print and digital versions are, for all intents and purposes, perfect substitutes.

In cases where digital and analog versions differ, preservation issues become more complex. Even minor differences, such as pagination, may elicit questions as to which version should be considered the authoritative version for scholarly citation. For example, print magazine articles are easily cited, by volume, issue, and page. However, online versions of these same magazines often omit pagination, presenting each article as one HTML file of unbroken text. More significant differences between digital and analog versions, impacting appearance, functionality, or content, amplify the problem. If one institution collects the analog version, while another collects the digital version, which institution holds the official "copy of record"? Should both versions be preserved, or just

one? Who decides?

Preservation decision-making in regard to materials existing simultaneously in digital and analog form often must be informed by a longer view. Are multiple versions of the same item expected to co-exist indefinitely, or is this merely a transitional state, with analog versions gradually supplanted by digital equivalents? In the latter case, preservation of only the digital version may be appropriate; in the former case, preservation of both versions might be necessary, or an authoritative version must be selected for preservation.

The decision to preserve in digital or analog form may turn on a simple cost comparison of the two approaches, but ideally, it should also take into account the preferences of users. Librarians discovered some time ago that users were resistant to replacing paper publications such as newspapers and magazines with microfilm copies, despite the advantages the latter format offered in terms of prolonging the longevity of the materials and reducing storage space requirements. In the same way, users may prefer that certain information resources be preserved as analog objects, and others as digital objects. User preferences, such as concerns about ease of access, may override purely economic factors.

XIII. Digital preservation as...a public good

Few would disagree that preserving an information resource benefits its owner, whether a library, museum, archive, publisher, or private collector. But preserving a resource, and in so doing, making it part of the permanent scholarly or cultural record, also confers benefits on society at large, by securing the resource's continued availability for use by current and future generations of researchers and students. An institution that preserves the last copy of a resource has performed a service of potentially incalculable value to the public. In these circumstances, the benefits from preservation are widely distributed; unfortunately, the costs of preservation are not.

A preserving institution can generate societal benefits extending well beyond its immediate stakeholders. The costs of producing these extra benefits often remain uncompensated. In the analog world, inequities in the distribution of preservation costs have little impact on collecting institutions' incentives to preserve. This partly reflects the mission of these institutions, which includes the responsibility to act as stewards of society's memory. But other factors also play a role. Institutions directly own, and have physical custody of, one or more copies of the analog materials in their collections. The institution is therefore uniquely placed to undertake the preservation of these materials, and this enhances the incentives to preserve.

Another factor that strengthens preservation incentives for analog materials is that the distribution of the benefits from preservation are, in a sense, self-limiting. Analog items, such as print books, can be difficult and/or expensive to access by individuals outside the collecting institution's direct user community. For example, inter-library loan can cost as much as \$30 - \$50 per item. Extremely rare or valuable materials may not be circulated at all, further reducing the scope for access by outside users.

The factors that enhance incentives to preserve analog materials—physical custody and limited opportunities for sharing—break down in the digital world. Rather than being purchased outright and transferred into the custody of each collecting institution, digital resources are often obtained through license or subscription, and then accessed by users from all institutions via a central Web server operated by the publisher. Institutions, while considering the licensed digital materials part of their collections, nevertheless do not have

physical custody, and therefore little or no opportunity to undertake their preservation.

In addition to diminishing the notion of physical custody, digital materials are also more easily shared than analog materials. Resources can be made available online and accessed from all over the world, making an institution's user community potentially limitless. In these circumstances, there may be some resistance to underwriting expensive preservation activities that benefit a large pool of users, most of whom make no contribution to the preserving institution's resource pool (via tuition, taxes, etc.). Incentives to preserve are further reduced if the materials in question are not unique, but instead held by multiple institutions. Which institution should go to the trouble and expense of preservation, when the benefits, in terms of making the materials part of the permanent scholarly or cultural record, will accrue to all?

As [Donald Waters](#) (2002) points out, digital preservation exhibits characteristics of a *public good*, chief among which is the difficulty in excluding those who do not contribute toward the provision of the good from enjoying its benefits. Once a digital resource has been preserved by one institution, it has, in a sense, been preserved for all. In an era of rising costs and shrinking budgets, activities that confer uncompensated benefits outside the institution's immediate stakeholder community may diminish in priority. Also, as preservation responsibilities diffuse beyond collecting institutions, preservation incentives will become even less assured: in the absence of a formal preservation mandate, incentives to preserve digital materials without compensation for the benefit of society as a whole may be weak indeed [7].

Conclusion

Preserving our digital heritage is more than just a technical process of perpetuating digital signals over long periods of time. It is also a social and cultural process, in the sense of selecting what materials should be preserved, and in what form; it is an economic process, in the sense of matching limited means with ambitious objectives; it is a legal process, in the sense of defining what rights and privileges are needed to support maintenance of a permanent scholarly and cultural record. It is a question of responsibilities and incentives, and of articulating and organizing new forms of curatorial practice. And perhaps most importantly, it is an ongoing, long-term commitment, often shared, and cooperatively met, by many stakeholders.

As experience in managing the long-term stewardship of digital materials accumulates, there will likely be even more ways we will need to look at digital preservation in the course of building digital information environments that endure over time. But this should come as no surprise: after all, Wallace Stevens found at least thirteen ways of looking at a blackbird.

Notes

[1] Remarks in this section are adapted from Lorcan Dempsey and Dan Greenstein, 1999. "The Fabric of Culture and Learning: A Draft Briefing Paper" (Unpublished)

[2] "Thirteen Ways of Looking at a Blackbird". Available at:
<<http://www.poets.org/poems/poems.cfm?45442B7C000C07020B77>>.

[3] This is not to say that there are only thirteen ways of looking at digital preservation!

[4] Of course, some redundancy may be desirable as a disaster-recovery measure.

[5] See <<http://www.kb.nl/kb/pr/pers/pers2002/elsevier-en.html>> for a description of this agreement.

[6] To learn more about the Model License, visit <<http://www.nesli2.ac.uk/>>.

[7] For more detailed discussion of the incentives to preserve digital materials, see Brian Lavoie (2003) *The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making*. Available at: <<http://www.oclc.org/research/projects/digipres/incentives-dp.pdf>>.

References

Nicholson Baker, 2001. *Double Fold: Libraries and the Assault on Paper*. New York, NY: Random House.

Neil Beagrie, 2003. *National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity*. Washington, D.C.: Council on Library and Information Resources. Available at: <<http://www.clir.org/pubs/abstract/pub116abst.html>>.

Consultative Committee for Space Data Systems, 2002. *Reference Model for an Open Archival Information System (OAIS) (CCSDS 650.0-B-1 - Blue Book)*. Washington, D.C.: CCSDS Secretariat. Available at: <<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>>.

Donald Waters, 2002 "Good Archives Make Good Scholars: Reflections on Recent Steps Toward the Archiving of Digital Information", in *The State of Digital Preservation: An International Perspective*. Washington, D.C.: Council on Library and Information Resources, pp. 78-95. Available at: <<http://www.clir.org/pubs/abstract/pub107abst.html>>.

Copyright © 2004 OCLC Online Computer Library Center

[Top](#) | [Contents](#)
[Search](#) | [Author Index](#) | [Title Index](#) | [Back Issues](#)
[Editorial](#) | [First article](#)
[Home](#) | [E-mail the Editor](#)

[D-Lib Magazine Access Terms and Conditions](#)

doi:10.1045/july2004-lavoie