Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation. Interim Report of the Blue Ribbon Task Force on Sustainable Digital Prese...

Technical Report · December 2008

7 authors, including:

Lorraine Richards
Drexel University
14 PUBLICATIONS **19** CITATIONS

Paul Courant
University of Michigan
66 PUBLICATIONS **1,791** CITATIONS

Daniel Rubinfeld
University of California, Berkeley
165 PUBLICATIONS **8,062** CITATIONS

Sustaining the Digital Investment:

Issues and Challenges of Economically Sustainable Digital Preservation

December 2008

Interim Report of the
Blue Ribbon Task Force on
Sustainable Digital Preservation and Access

Acknowledgements

This report represents the work of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access, a 17-member group of experts from economics, computer science, libraries, archives, museums, and related fields with funding and support from the U.S. National Science Foundation (NSF Award No. OCI 0737721), the Andrew W. Mellon Foundation, the U.S. Library of Congress, the U.K. Joint Information Systems Committee (JISC), the Electronic Records Archives Program of the National Archives and Records Administration, and the Council on Library and Information Resources. The report is authored by Brian Lavoie, Lorraine Eakin, Amy Friedlander, Francine Berman, Paul Courant, Clifford Lynch, and Daniel Rubinfeld with input, comments and consultation provided by all of the members of the Task Force. The Communications Group at the San Diego Supercomputer Center produced the document. Lorraine Eakin, an intern at the Council on Library and Information Resources, provided substantial research support for this report.

The views and opinions expressed herein represent the rough consensus among the members of the Task Force and should not be construed to represent those of the U.S. Government, the U.S. National Science Foundation, the Library of Congress, JISC, or any of the other sponsoring agencies and organizations. Much of the information from which these preliminary results, findings and lessons learned were drawn was provided by speakers who volunteered their knowledge and expertise to the Task Force. We are grateful to them for their contributions to the future that sustainable preservation and access of digital material will enable.

List of Members

Francine Berman, Director, San Diego Supercomputer Center, and High Performance Computing Endowed Chair, UC San Diego [co-Chair]

Brian Lavoie, Research Scientist, OCLC [co-Chair]

Paul Ayris, Director of UCL Library Services and UCL Copyright Officer, UCL (University College London), UK

G. Sayeed Choudhury, Associate Dean of Libraries, Johns Hopkins University, Baltimore, MD

Elizabeth Cohen, Academy of Motion Pictures Arts and Sciences, Stanford University, Stanford, CA

Paul Courant, University Librarian, University of Michigan, Ann Arbor MI

Lee Dirks, Director of Education & Scholarly Communications, Microsoft Corporation

Amy Friedlander, Director of Programs, Council on Library and Information Resources (CLIR), Washington, DC

Vijay Gurbaxani, Senior Associate Dean, Paul Merage School of Business, University of California at Irvine

Anita Jones, Professor of Engineering and Applied Science, University of Virginia, Charlottesville VA

Ann U. Kerr, Consultant, Vice Chair, International IEEE Mass Storage Systems and Technology Committee, AK Consulting, La Jolla CA

Clifford Lynch, Executive Director, Coalition for Networked Information (CNI), Washington, DC

Daniel Rubinfeld, Robert L. Bridges Professor of Law and Professor of Economics, University of California, Berkeley

Chris Rusbridge, Director, Digital Curation Centre, University of Edinburgh

Roger Schonfeld, Manager of Research, Ithaka

Abby Smith, Historian and Consulting Analyst to the Library of Congress, San Francisco, CA

Anne Van Camp, Director, Smithsonian Institution Archives, Washington, DC

Liaisons to Task Force

Martha Anderson
Library of Congress
Director of Program Management, National Digital Information Infrastructure and Preservation Program, Washington DC

Laura Campbell
Library of Congress
Associate Librarian for Strategic Initiatives
Washington DC

Robert Chadduck
Liaison for the National Archives and Records Administration, Computer Engineer, Principal Technologist Electronic Records Archives Program, The National Archives and Records Administration
College Park, MD

Chris Greer
Liaison for the National Coordination Office for Networking and Information Technology Research and Development
Arlington VA

Lucy T. Nowell
Office of Cyberinfrastructure
National Science Foundation
Arlington, Virginia

Donald J. Waters
Program Officer, Scholarly Communications
The Andrew W. Mellon Foundation
New York, NY

Table of Contents

List of Figures

List of Tables

List of Boxes

Executive Summary

There is general agreement that digital information is fundamental to the conduct of modern research, education, business, commerce, and government. Future accomplishments are accelerated through persistent access to data and digital materials, and their use, re-use, and re-purposing in ways both known and as yet unanticipated.

There is no general agreement, however, about *who is responsible* and *who should pay* for the access to, and preservation of, valuable present and future digital information. Creating sustainable economic models for digital access and preservation is a major challenge for all sectors, and the focus of investigation of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access.

The Task Force's particular focus is digital information ultimately in the public interest (including official and historical records, scientific research data, private data that may become part of the public record, etc.). Over 2008 and 2009, the Task Force's goal is

a) To sample and understand best and current practices for digital preservation and access, and to begin to synthesize major themes and identify systemic challenges. This has been our focus in 2008 and is the topic of this Interim Report.

b) To identify and develop useful economic models for digital preservation and access, and map these models to common institutional, enterprise, and community scenarios. This will be our focus in 2009 and the topic of the Task Force's Final Report. The goal is to provide actionable recommendations for decision makers seeking economic models for access and preservation that promote reliability, cost-effectiveness, trustworthiness, and compliance to relevant policy and regulation.

During 2008, as the Task Force heard testimony from a broad spectrum of institutions and enterprises with deep experience in digital access and preservation, two things became clear: First, **the problem is urgent.** *Access to data tomorrow requires decisions concerning preservation today.* Imagine future biological research without a long-term strategy to preserve the Protein Data Bank (PDB), a digital collection that drives new insights into human systems and drug therapies for disease, and represents an investment of 100 billion dollars in research funding over the last 37 years. Decisions about the future of the PDB and other digital reference collections -- how they will be migrated to future information technologies without interruption, what kind of infrastructure will protect their digital content against damage and loss of data, and how such efforts will be supported -- must be made *now* to drive future innovation.

Second, the difficulty in identifying appropriate economic models is not just a matter of finding funding or setting a price. In many institutions and enterprises, **systemic challenges create barriers for sustainable digital access and preservation.** These include:

- ***Inadequacy of funding models to address long-term access and preservation needs.*** Funding models for efforts that incorporate digital access and preservation are often not persistent – they may be "one time" (e.g. preservation supported by a research grant in the academic world), or assessed for support against other enterprise priorities and then abandoned as more critical short-term priorities emerge. *Reliable preservation can suffer no gaps* – data lost or damaged today often cannot be recovered tomorrow – pointing to the need for persistent digital access and preservation funding.

- ***Confusion and/or lack of alignment between stakeholders, roles, and responsibilities with respect to digital access and preservation.*** It is often the case that those who create and use digital information are distinct from those who serve as its stewards and support its preservation and access. Consequently, the costs are not necessarily shouldered by those who enjoy the benefits, and can lead to inadequate economic models for sustainability. For example, in the academic sector, many researchers expect free access to community reference collections such as the PDB, and assume that their support and maintenance will be handled by funding agencies, commercial enterprise, or someone else. Misalignment among roles -- who is accountable for digital data, who pays, who has rights, and questions about other key responsibilities – continues to be a major challenge in developing viable sustainable economic and stewardship models for digital data.

- ***Inadequate institutional, enterprise, and/or community incentives to support the collaboration needed to reinforce sustainable economic models.*** Digital preservation and access requires long-range planning and support, agreement on formats, standards and use models, interoperability of relevant hardware and software systems, and partnering among a diverse group of technologists, users, datacenter staff, compliance officers, financial managers, etc. In many environments, there are few incentives to develop the persistent collaborations and uniform approaches needed to support access and preservation efforts over the long-term.

- ***Complacency that current practices are good enough.*** The urgency of developing sustainable economic models for digital information is not uniformly appreciated. There is general agreement that leadership and competitiveness, if not institutional survival, in the information age depends on the persistent availability of digital information, making preservation of that information an urgent priority. Yet that urgency is often not translated or institutionalized into individual or group behaviors. Both "carrots," in the form of recognition that access to information is an investment in current and future success, and "sticks," in the form of the penalties for non-compliance, accounting of the explicit costs of lost opportunities or information, etc. are needed to make this clear.

- ***Fear that digital access and preservation is too big to take on.*** There is also general agreement that in its entirety, digital preservation is a *big* problem, incorporating technical, economic, regulatory, policy, social, and other aspects. But it is not insurmountable. Digital access and preservation may be as manageable as including a "data bill" as an explicit and fixed part of an institution's business model. Successes in sustainable digital access and preservation have involved both a willingness to make it a persistent line item on the part of stakeholders, and/or creative partnership solutions that spread costs effectively. Access and preservation of all valued digital information may be too big a problem for individual stakeholders to take on, but access and preservation of *your* valued information is not.

It is clear from the testimony and themes described in this interim report that **institutional, enterprise, and community decision makers must be part of the access and preservation solution**. Decision makers must make access and preservation a strategic and actionable priority, incorporating it into their planning, economic models, and interactions with constituent communities. Without their participation, it will be difficult to build on the critical foundation of digital information required for leadership and competitiveness in the information age.

Preface

The existence of a stable, agreed-upon **record** has enabled extraordinary advances in research and practice in all domains. Whether we turn to the publication of Copernicus' *De revolutionibus orbium coelestium* (1543) or Isaac Newton's *Principia Mathematica* (1687) or to the evolution of scholarly journals from a tradition of correspondence among scientists that Paul David delineated in his paper, "The Historical Origins of Open Science" (2008), we see how documentation of method, evidence, findings and implications and the communication of those ideas has advanced research and practice through debating, revising, verifying and ultimately building upon successive generations of work. The practical outcomes of this long historical process and their contributions to the public good are obvious in fields from epidemiology to entertainment.

The notion of the record was never confined to the explicit publication, as the correspondence among early scientists and the notebooks of both Leonardo and Galileo attest. In our own times, information technologies have revolutionized notions of the record, evidence and analysis; this is seen in a range of objects and activities from the creation of community datasets such as the National Virtual Observatory (aggregating digital sky surveys from the world's largest telescopes) or those managed for social scientists by the Inter-university Consortium for Political Science Research (ICPSR), to the data created by single-investigator or small team projects, to the more ephemeral online lab notebooks, wikis and notes now recorded in unstructured text files and spreadsheets that live on desktops and thumb drives.

These materials vary in their level of formality and anticipated audience, but the central tendency is obvious: *there is more and more heterogeneous digital information of importance to society and in the public good*. In the research community, the raw material of future scholarship has become especially diverse, and not all of it originates in labs and libraries. Imagine the work of a cultural anthropologist thirty years hence studying the United States 2008 presidential election without access to candidate websites and YouTube. Today's digital data cannot be narrowly defined but rather encompasses material created under many rubrics for which there may eventually be a public interest, either now or in the future.

It is the record of evidence, findings and analysis -- whether structured or unstructured, raw, semi-processed, formal or informal -- that concerns us. Preserving data for use tomorrow requires decisions today. Where traditional record keeping systems functioned reasonably well in the centuries of analog, they are inadequate for the digital age, characterized by rapidly changing technologies, explosion in the volume and heterogeneity of data and information, and global expansion in the conduct of science specifically, and scholarship more generally. The readerships for the *Principia* or even for James Clerk Maxwell's *A Treatise on Electricity and Magnetism* (1873) were modest and constrained by modern standards. Today's researchers consider digital technologies as fundamental as the wet lab and the print library. On the horizon, the potential contributions of citizen scientists and researchers enlarge the scale and scope of participation. The knowledge economy of the future can only expand the range of users of digital information in ways we can hardly imagine.

The sustainability of the long-term access and preservation of digital materials is a well-known challenge, and discussion frequently focuses on the difficult technical issues. Less clearly articulated are the organizational and economic issues implicit in the notion of sustainability which, at the risk of over-simplification, come down to two questions: ***How much does it cost?*** and ***Who should pay?***

Parsing these questions, and understanding how we might address them, led to the conceptualization and ultimate organization of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access. This Interim Report represents the work of this Task Force in its first year, and sets forth the conceptual framework that will guide its efforts in the second and final year. The synthesis of the Task Force's investigations will be the focus of its Final Report in 2009, which will provide an economic framework, and practical recommendations and strategies for sustainable digital preservation and access.

Chapter

1

Introduction

Today, you can scan one gigabyte of data or download it with a good computer system in a minute. But with current technologies, storing a petabyte would require about 1,500 hard disks, each holding 750 gigabytes. That means it would take almost three years to copy a petabyte database — and cost about $1 million.

Alex Szalay
Preserving digital data for the future of eScience
Science News: Magazine of the Society for Science and the Public
Web Edition, Monday August 18th, 2008
http://www.sciencenews.org/view/generic/id/35263/title/Preserving_digital_data_for_the_future_of_eScience.

I*nvestment* is about the future, and is almost always an uncertain proposition. The value of an investment will ebb and flow with the vagaries of shifting circumstances, priorities, and attention. Yet without investment, society cannot grow: in business, research, and culture, ongoing development and growth require that a portion of current resources be set aside for the purpose of maintaining, and ideally, *expanding*, our future productive and creative capacities. Despite the uncertainty, and despite the cost, we must invest now to secure the opportunity for a future that builds on and surpasses the achievements of the present.

Preservation of valuable assets is a form of investment; it is a way to ensure that the value-creating capacity of these assets remains available to us in the future. We are accustomed to making such investments on a personal level: in our automobiles, our homes, our health. We are also accustomed to making such investments at a societal level: in public libraries, historic buildings, archeological sites, roads and bridges, and the records of government. And we are now becoming accustomed to making such investments on an even greater scale: the "green movement" has sharpened awareness of the need to invest in the future of the environment and the planet. Yet most of us mix our support of such investments with a dose of practicality: nothing is free, there are costs associated with preservation, resources are limited, and consequently, we cannot save everything.

We live in an information-intensive age, so it is no surprise that many of the valuable assets we create today manifest themselves in the form of *information*. Books and articles, certainly, but also digitized and "born digital" scientific data sets, e-mail, images, music, software, records of transactions, Web sites, and blogs. Our capacity to capture, store, and access information about the economic, scientific, political, and cultural aspects of society has grown exponentially. More and more of our societal activities transpire in digital environments: business transactions, scientific research, education, recreation. These activities often consume digital information as input and produce digital information as output.

Many of the digital information resources we create and use today are assets with a value-creating capacity that persists far into the future. A scientific data set created today can serve as material for ground-breaking research years in the future (Figure 1.1). A campaign Web site from today's presidential election might be an important historical artifact and primary source for future historians. An electronic dialog on a scholar's blog might spark ideas that shift disciplinary paradigms. Sometimes the future value of an information resource is easy to predict; sometimes it is less so. Making investments often means "picking winners," and backing those predictions with money. Investments in managing information for the long-term are no exception, and the choices can be difficult. In some cases, the best we can do is to make an educated guess about future value; in other cases, our choices might be described as "hedges" against the future: an attempt to preserve the option of future use should an unanticipated need arise.[1]

---

[1] An interesting example of unanticipated use is found in a recent *London Times* article, which describes how centuries-old British Royal Navy log books, with their detailed descriptions of weather conditions around the world, are now being used as data sources for studies of climate change. See: http://www.timesonline.co.uk/tol/news/environment/article4449527.ece

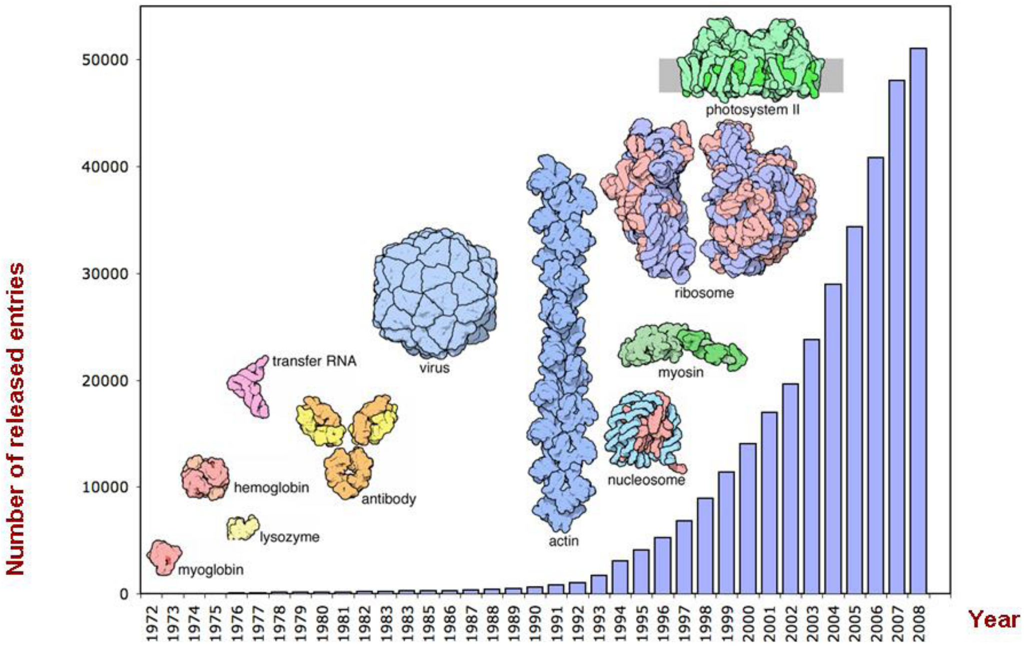

FIGURE 1.1: **Growth of the Protein Data Bank**
The Protein Data Bank (PDB) is an example of a key digital resource supporting scientific research and discovery. The PDB is a single international repository for all information about the structure of large biological molecules, together with an archival database of hundreds of thousands of users who depend on the data. This critical life sciences resource was begun in 1971 with 7 structures and is projected to contain 150,000 structures by 2014. As of January 1, 2008, the public archive contained 300,000 files and requires more that 70 GB of storage.
Source: H. Berman July 23, 2008. Used with permission.

Does the need to invest in the long-term preservation of digital assets represent a wholly new problem for society? Surely not: as mentioned earlier, we are accustomed to making preservation investments on both a personal and societal level. Moreover, institutions such as libraries, archives, and museums have traditionally embraced the mission of preserving the scholarly and cultural record. Businesses, who routinely manage assets for extended periods of time with a view to reuse them, also recognize the need to preserve aspects of their company history and transactions, and maintain archives for the purpose.

Investment in the preservation of *digital* assets does present new twists to the problem, especially in terms of *immediacy*, *scale*, and *uncertainty*. In the analog world, the rate of degradation or depreciation of an asset is usually not swift, and consequently, decisions about long-term preservation of these materials can often be postponed for a considerable period, especially if they are kept in settings with appropriate climate controls. The digital world affords no such luxury; digital assets can be extremely fragile and ephemeral, and the need to make preservation decisions can arise as early as the time of the asset's creation, particularly since studies to date indicate that the total cost of preserving materials can be reduced by steps taken early in the life of the asset (Figure 1.2). Our prolific capacity to create digital information amplifies the scale of the potential challenge: a recent report by the research firm IDC noted that 2007 marked the first year that the amount of digital information created, captured, or replicated exceeded available storage capacity (Gantz, 2008, Figure 1.3). The report also predicted that by 2011, nearly half of all digital information will not have a permanent home. Finally, investments in the long-term management of digital information also seem encumbered by exceptionally high levels of uncertainty, shrouding issues such as the level of required expenditure and the allocation of responsibility for making the investment.

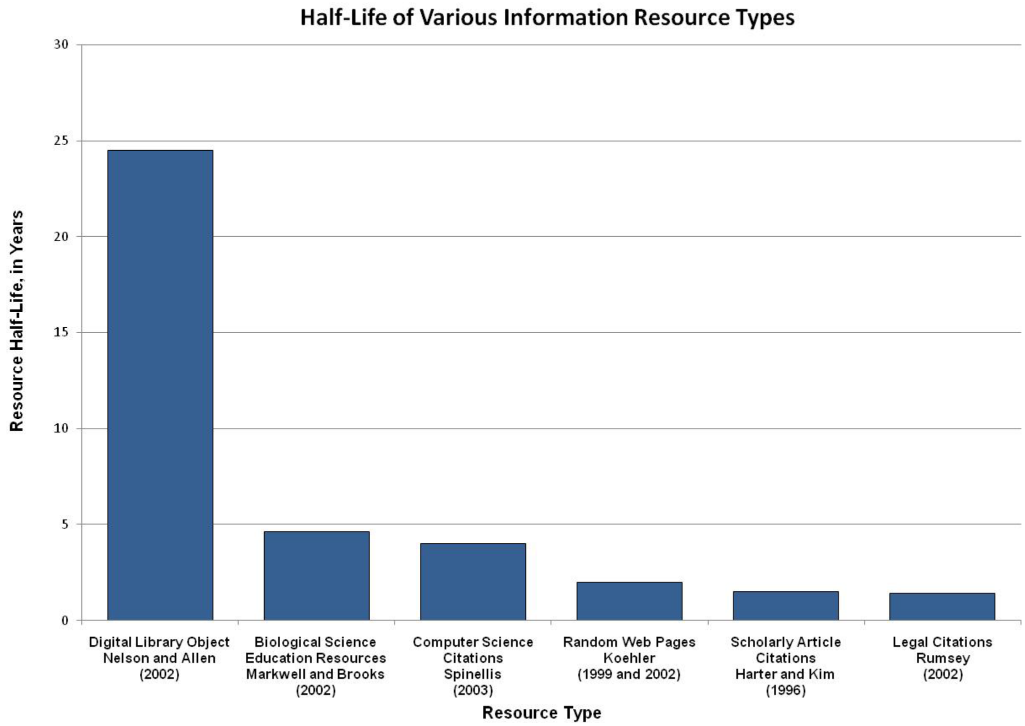

FIGURE 1.2: **Half-life of Various Digital Information Resource Types**
Different types of digital information objects exhibit differing "information decay" rates, but those managed in digital libraries survive the longest.
Source: L. Nowell March 2008. Used with permission.

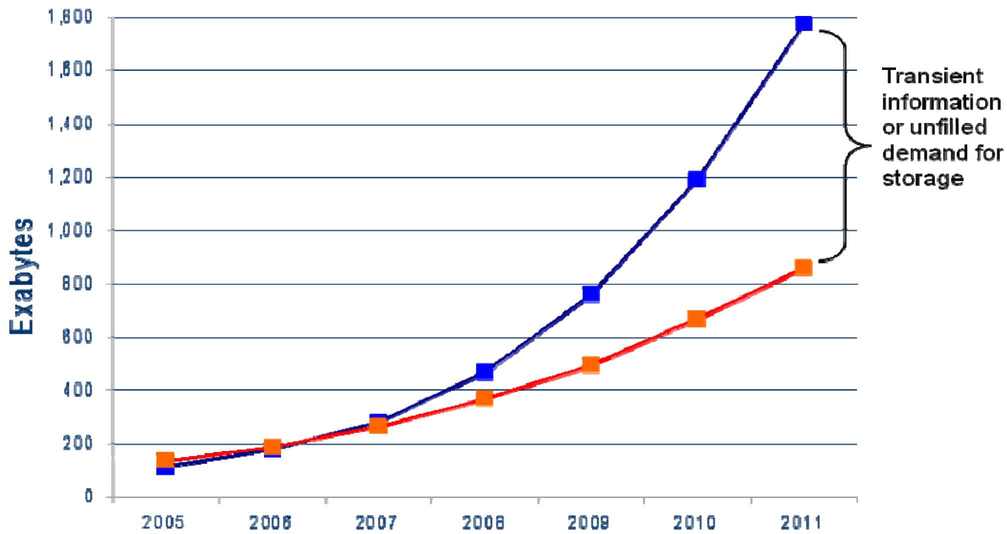

FIGURE 1.3: **Information and Storage**
Source: J. Gantz January 2008 (revised). Used with permission.

The long-term accessibility of digital resources cannot happen automatically and without cost; it will only happen as the result of deliberate decisions on the part of individuals and organizations to invest the resources necessary to shepherd digital information safely across time.

Although there is general agreement as to the positive impact of investing in the future, many digital environments do not have what their stakeholders would consider an adequate sustainable economic model for preservation. Over and above the issue of sufficient resources, there are often systemic institutional or community barriers that threaten digital preservation activities, and by extension, the long-term persistence of digital assets. These barriers are entrenched in many digital preservation contexts, and left unaddressed, can seriously impair efforts to organize economic mechanisms or sustain existing mechanisms to support long-term preservation. These barriers include:

- Misaligned incentives between those who are in a position to undertake preservation, and those who benefit from it;
- Lack of clear responsibility for digital preservation, and a prevailing assumption that it is "someone else's problem" (called the "free rider problem" in the literature);
- Challenges in valuing or monetizing the value of digital preservation, and translating this value into a willingness to pay;
- Ineffective coordination of preservation activity across diffused stakeholder communities;

- Lack of appropriate investment in developing critical preservation infrastructure and skill sets.

- Lack of continuous monitoring and treatment and/or reliance on last-minute rescues

- Persistent preservation activities funded using transient or "one-time" vehicles

This last barrier is particularly challenging. Organizations must secure sufficient resources to sustain their digital preservation activities beyond the next budget cycle or the end of a grant award. Resources might come from internal allocations, service revenues, donations, public funding, in-kind contributions, or other sources – all of which might be transient, uncertain, or both. *Lack of economic resources can make digital materials disappear just as surely as obsolete operating systems or aging magnetic tape* (Box 1.1).

The previous list is by no means exhaustive, however the urgency of the problem is clear: if we do not take steps to address these and other potential barriers to success, it will be difficult, if not impossible, to sustain the information required for future accomplishments. This in turn places our scientific, scholarly, and cultural legacy at economic risk. The purpose of the *Blue Ribbon Task Force on Sustainable Digital Preservation and Access* (hereafter, the **Task Force**) is to develop strategies to mitigate the economic risks attached to digital preservation.

BOX 1.1 **Computational neuroscience is dependent upon preservation funding**


EDITORIAL

Prying Open the Black Box

– Floyd Bloom

www.sciencemag.org SCIENCE VOL 314 6 OCTOBER 2006 17

In an October, 2006 article for Science, Floyd Bloom of the Scripps Institute discusses the rapidly evolving field of computational neuroscience. He notes that the emergence of large databases and the capacity for handling large sets of data create incredible opportunities to develop new neuroscientific models and to engage in scientific collaboration at a previously unheard-of scale. However, we will need to improve methods for maintaining that data and increase standardization to ensure easy transfer and re-use of it. He also notes that we will not be able to set up the necessary means for preserving and re-using this data without the necessary funding, and that such funding does not yet exist.

Source: F. Bloom October 2006.

Task Force Reports

In exploring the current state of knowledge in the area of economically sustainable digital preservation, the Task Force found little consensus or convergence of thinking on the general contours of the issue, and few practical recommendations on how to achieve economic sustainability in real-world digital preservation scenarios. This Interim Report is intended as a contribution toward the first gap; the second and Final Report is intended as a contribution toward the second. Taken together, the two reports are aimed at advancing our understanding of the economics of sustainable digital preservation and providing a solid foundation for further work in this area.

Note that we target both reports at senior decision-makers, acting as individuals or on behalf of organizations with an explicit or implicit responsibility for the long-term stewardship of digital materials, and/or an explicit or implicit stake in the long-term accessibility of digital assets. These individuals are fundamental to the creation and implementation of effective and sustainable strategies for digital preservation and access. We hope that others will find these reports a useful guide to the issues as well, and a starting point for more detailed analysis.

The remainder of Chapter 1 traces the general contours of the economics of digital preservation, and presents a detailed definition of what is meant by economically sustainable digital preservation activities. Chapter 2 discusses the ideas underpinning the notion of an "economic model," an analytical tool that can be used to represent, understand, and investigate issues surrounding sustainable digital preservation. Chapter 3 establishes some empirical context for the concepts discussed in Chapter 1, drawing on the existing literature and expert testimony to better understand the realities of building economically sustainable digital preservation activities. Chapter 4 offers some summary remarks, observations, and discusses additional issues related to economic sustainability.

1.1 The Economics of Digital Preservation

Consider the following description of the maintenance costs of the eighteenth-century British Royal Navy ship *HMS Astrea*:

> "The *Astrea* frigate was built at East Cowes in 1781 and cost £7,855, but she needed 'small repairs' three years later costing £3,414 and more totaling £4,271 two years after that. So the cost of the first five years' repairs, £7,685, was within £200 of her building price. She was then laid up for seven years but, with the war beginning, more repairs to the hull alone cost £5,347 and fitting her out another £5,677. Defects in 1794, 1795 and 1798 cost £539, £1,500 and £3,849. The copper sheathing was replaced for the third time in 1798 for £4,128. By then repairs had cost four times her original price …" (Pope, 1981)

Translated into present-day vernacular, this calculation becomes the familiar "total cost of ownership." The principle is the same now as it was in the eighteenth century: the true cost of an asset exceeds, often by a substantial margin, the expenditure required to create or acquire it. Left to the ravages of wind and weather, a wooden ship will soon fall into disrepair, its timbers rotting, its rigging frayed, its sails ragged. As the records of the *Astrea* indicate, ongoing expenditures were necessary to maintain the vessel in a condition that enabled it to fulfill its purpose – in other words, to produce the ongoing value that motivated the ship's creation in the first place. Significantly, the cost of maintaining the vessel over a period of less than two decades exceeded its original construction cost by 300 percent.

Without ongoing maintenance, digital assets will also fall into disrepair, succumbing to a host of "digital diseases" that impair or limit the ability to use them: bit rot (or degradation of the object so that it is no longer readable), technological obsolescence (which means that systems no longer exist that can read the encoding in which the data are represented and stored), or even outright loss. Consequently, preventive measures must be taken to insure that the media (tape, disk, and so on) are stable and the information encoded thereon can be read. Provisioning secure storage systems, refreshing aging media, fixity checks, replication in multiple systems and/or locations, format migration, and other techniques to keep information safe and accessible over time are costly and contribute to a digital "total cost of ownership" that is ongoing and potentially substantial. Funding this digital total cost of ownership requires a dedicated flow of sufficient resources built around a set of long-term goals regarding persistence, accessibility, and usability.

Ensuring that digital assets created today are available for use tomorrow is a multi-faceted problem. Technical issues comprise one aspect: building the systems, work flows, and operational strategies that secure the long-term persistence of digital information. Legal issues are another: in what ways do preservation strategies – especially those that involve replication or alteration of digital content – run afoul of intellectual property rights? There is a policy aspect as well, in the sense of articulating and enforcing requirements for preserving data, either within an organization or as a mandate encompassing a group of organizations. Financial institutions, pharmaceutical research laboratories, and manufacturers of aircraft are all examples of organizations that are required to maintain records which are increasingly created in digital form (Figure 1.4). And as the discussion of total cost of ownership implies, economics underlies effective solutions to the digital preservation problem.

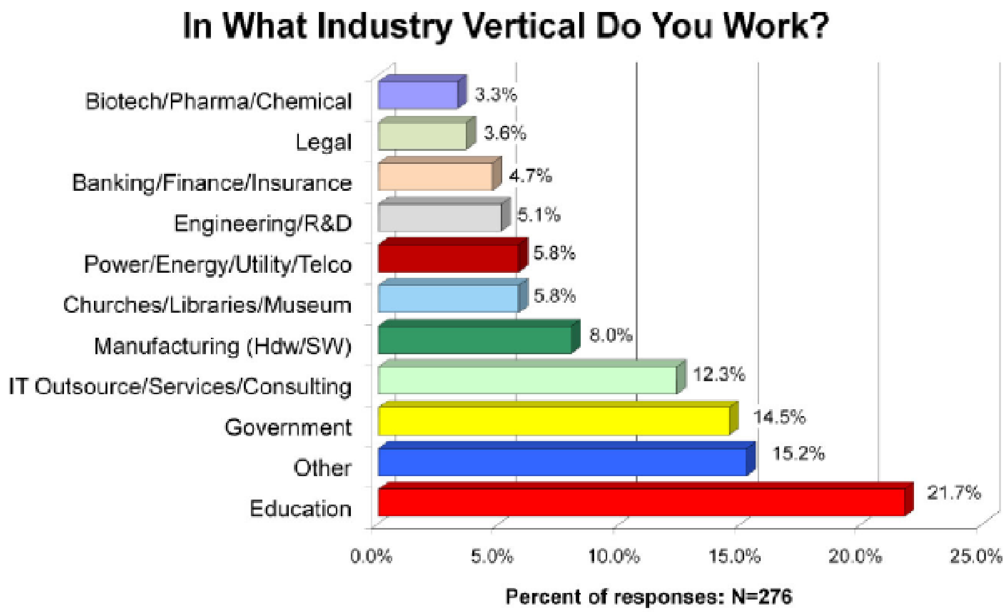

FIGURE 1.4: **Digital Preservation is Everyone's Problem**
The Storage Networking Industry Association (SNIA) recently conducted a survey to frame the business requirements surrounding long-term data retention. The chart above characterizes the broad set of industry affiliations of the survey respondents, indicating that digital preservation is an issue that impacts nearly every segment of business, government, and culture. In the survey, over 80% of the respondents reported a need to retain information over 50 years.
Source: P. Mojica October 2008. Used with permission.

1.2 Two perspectives on the economics of digital preservation

Assuring long-term access to digital assets imposes resource requirements on an organization. For most organizations, these resource requirements are new in the sense that allocating funds for digital preservation has only recently appeared – or for some organizations, has yet to appear – as a budgetary line item. In order to meet these resource requirements, an organization has two choices: it can find a new way to slice up its "economic pie" – that is, its budget – in order to release resources for digital preservation, or it can try to make the pie bigger. These choices represent distinct, yet complementary versions of the economic problem lying at the heart of digital preservation.

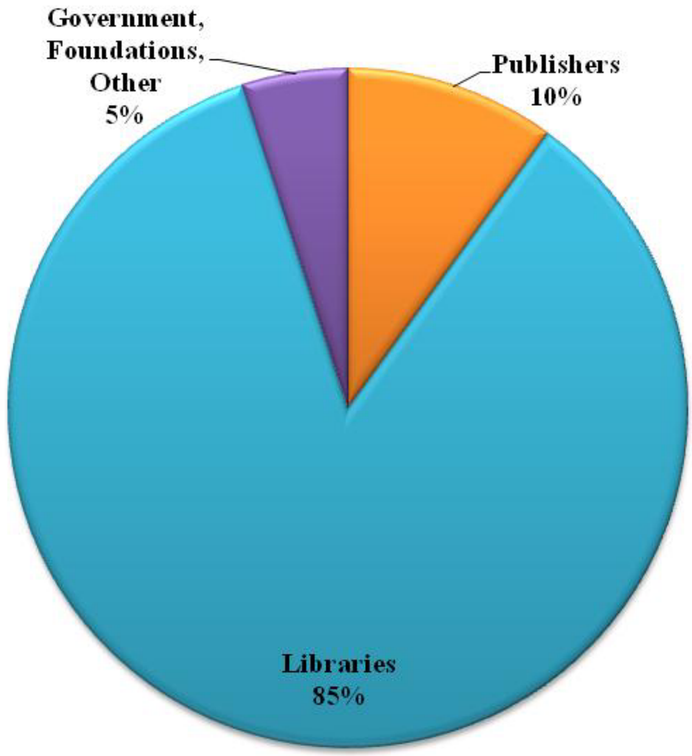

FIGURE 1.5 **Portico's Desired Funding Base**
Various organizations will choose to meet their resource requirements in different ways. E-publication preservation archive Portico is targeting to achieve a funding distribution close to the one shown above as it moves forward in time.
Source: Eileen Fenton November 2008. Used with permission.

Slicing Up the Pie

At one level, the economics of digital preservation is about *managing opportunity cost* within an economic pie that is largely fixed in size. To accommodate the new resource requirements imposed by the long-term stewardship of digital assets, it is likely that many organizations, at least in the short-run, will need to shift funds from one allocation to another within an effectively fixed budget. For example, a library might reduce its investment in services and infrastructure surrounding its print collection in order to release resources to support increased investments in the long-term stewardship of its digital collections. This re-allocation of resources might be accomplished through a variety of strategies, such as modifying service levels attached to the print collection, or seeking new efficiencies in collection management that permit service levels to be maintained at lower cost. A for-profit organization, on the other hand, may manage opportunity cost on its balance sheet: preserving important digital assets might take the form of a new and unavoidable cost of doing business that effectively shifts funds from profits to costs.

In general, if the size of the economic pie – i.e., the pool of resources from which funds for investment in long-term management of digital assets must be drawn – remains fixed, then the emergence of new requirements to manage digital assets over the long-term represents an opportunity cost – a trade-off – that must be managed by the organization (or organizations) paying for preservation. This perspective says, in effect: *If we are to meet our digital preservation goals, we must give up something else in order to pay for it.* This means that the economics of digital preservation is about confronting the opportunity cost associated with the need to invest in the long-term maintenance of digital assets and, therefore, understanding the scope of the trade-off that this opportunity cost presents.

From this perspective, the important economic questions involve the nature and scope of this opportunity cost: What are the key component activities involved in the long-term management of digital assets? What is the cost of these activities, in terms of the things we must give up in order to undertake them? What other activities must be sacrificed in order to preserve the opportunity of future use of a set of digital assets? What strategies can be invoked to ensure that this opportunity cost is as small as possible?

Making the Pie Bigger

Rather than seeing the economic pie as fixed or static, a second perspective on the economics of digital preservation focuses on strategies to make the pie bigger. From the point of view of an organization undertaking digital preservation, this approach involves bringing in funds from external sources to support preservation activities, rather than finding these funds internally. For example, an organization might agree to preserve a set of digital assets on behalf of itself and several external stakeholders, with the cost of long-term management of the materials funded collectively by the partners. Or alternatively, an organization could establish a "third-party" archiving service, in which preservation is offered as a service to anyone willing to pay for it in the form of a market transaction. In either case, there is a transfer of resources across organizations: from those who benefit from, and are willing to pay for, digital preservation, to those who are willing and able to carry it out.

In these and similar scenarios, long-term maintenance of the digital assets is funded by a flow of resources *across organizational boundaries*: from those who benefit from or value preservation, to those who undertake activities required to maintain and preserve the digital materials. From the point of view of the organization performing the preservation tasks, the economic pie gets larger as a result of these flows. New sources of funds are tapped to meet the new resource requirements imposed by ongoing management of the digital information, and as a result, the opportunity cost of digital preservation for the preserving organization is at least partially mitigated by bringing in resources from external sources. If we assume that the benefits from preserving a particular set of digital materials redound not to a single organization, but to multiple organizations, and that the costs of preservation are shared in some way across these beneficiaries, this in effect makes the economic pie bigger for everyone. Since the cost of preservation is lessened for all by being shared by many, the opportunity cost for the beneficiaries of digital preservation is also reduced.

It is worth noting that "making the pie bigger" does *not* eliminate the problem of opportunity cost – resources always seem to be scarce, and choices must be made. But from a single organization's perspective, it may reduce the magnitude of what must be given up to achieve preservation objectives, and help sustain preservation activities in situations where it would otherwise be unsustainable.

These two perspectives – (1) opportunity cost (metaphorically speaking, slicing up the pie) and (2) transfer of resources across organizations (making the pie bigger) – capture core economic issues associated with long-term management and preservation of digital information. Cutting across both these perspectives is an additional set of issues – how to sustain the economic model over time. As discussed earlier, funding digital preservation activities is similar to making an investment, and like all investments, such funding decisions have a temporal aspect: it is a cost incurred now in anticipation of future benefits. More than this, *preservation is not a one-time cost*; instead, it is a commitment to an ongoing series of costs, in some cases stretching over an indefinite time horizon.

If we are to marshal sufficient resources to meet our long-term preservation goals, this ongoing series of costs must be matched by an ongoing series of resource allocations to preservation activities. To achieve this, we need to embed these resource allocations within an economic arrangement designed to maximize the likelihood that these allocations will continue to be forthcoming. The economic arrangements we choose must be robust enough to survive beyond the next budget cycle, the end of a grant award, or the transition to the next generation of decision-makers. In short, the economics of digital preservation must be *sustainable*.

1.3 Characterizing economic sustainability in a digital preservation context

Sustainability is a term that is bandied about frequently these days. We hear it attached to discussions of the environment, use of natural resources, and economic growth in developing countries, as well as in other contexts. The U.S. Environmental Protection Agency defines sustainability as "meeting the needs of the present without compromising the ability of future generations to meet their own needs."[2] Certainly this definition resonates with the general notion of investing current resources for the purpose of securing the long-term value-creating capacity of digital assets. But the focus of this report is specifically on *economic sustainability*, and in this regard, the simple "dictionary definition" of sustainability is perhaps more appropriate: "the ability to be

[2] See http://www.epa.gov/sustainability/.

maintained." Or more specifically, we are concerned with economic strategies to maintain digital preservation activities over long periods of time.

Even this definition is too general to impart much meaning to the concept of economic sustainability in a digital preservation context. Unfortunately, moving beyond this definition in search of a more rigorous explanation of what economically sustainable digital preservation entails is problematic. Despite some excellent recent work, there is no clear consensus on which to draw. At a fine level of detail, of course, economic sustainability resists precise definition and instead diffuses into manifold shadings of meaning linked to particular sets of circumstances. As a consequence, there are potentially many alternative pathways to achieving economically sustainable long-term preservation of digital assets. Given this ambiguity, the goal of this section is to supply a definition of economic sustainability detailed enough to trace the contours of the topic, yet sufficiently general to serve as a common framework under which to gather more nuanced interpretations. The definition also serves the dual purpose of (1) framing the topics that the Task Force is addressing in its work and (2) providing some common reference points to support a wider community discussion of economically sustainable digital preservation.

The definition articulated by the Task Force, stated in Box 1.2 below, underscores that economic sustainability in a digital preservation context is a multi-faceted enterprise. Underlying any sustainable digital preservation-related activity is recognition on the part of stakeholders of the benefits or value of preservation. These benefits must be translated into appropriate incentives that induce stakeholders to accept and act on responsibilities for long-term management and preservation, and to allocate sufficient resources to such activities. The discussion below Box 1.2 expands on each of the bullets in the definition.

BOX 1.2 **Definition: economic sustainability**

The set of business, social, technological, and policy mechanisms that encourage the gathering of important information assets into digital preservation systems, and support the indefinite persistence of digital preservation systems, enabling access to and use of the information assets into the long-term future.

Economically sustainable digital preservation requires:

- Recognition of the benefits of preservation on the part of key decision-makers;
- Incentives for decision-makers to act in the public interest;
- A process for selecting digital materials for long-term retention;
- Mechanisms to secure an ongoing, efficient allocation of resources to digital preservation activities;
- Appropriate organization and governance of digital preservation activities.

1.3.1 Recognition of the benefits of preservation on the part of key decision-makers

Allocation of resources to an economic activity proceeds from the recognition that the activity yields real benefit or value. Digital preservation is no different: to create and sustain a commitment of resources to preservation, the value from doing so should be clearly identified and articulated. Simply asserting that it is beneficial to invest in the long-term management and preservation of digital information is not enough, and carries with it no intrinsic acknowledgement or expression of the value being created by doing so.

In articulating the benefits of long term preservation of digital assets, it is important to remember that the term *digital preservation* merely refers to the processes and mechanisms by which we achieve certain outcomes. In general, these outcomes involve the ongoing *accessibility* and *usability* of digital assets over long periods of time. In particular circumstances, we can impart more specificity to these outcomes: for example, preservation of historical stock market datasets enables economists to compare current trends in stock price movements with historical patterns, and provides key input for market models that predict crashes and bubbles. Preservation of data on ozone levels is required to compute the size of the ozone hole, and contributes to a scientific characterization of global warming.

Regardless of the way preservation's benefits are determined, expressed, or employed, decision-makers need to be aware of the value-creating opportunities preservation affords, if they are to be persuaded to undertake a sustained allocation of resources to digital preservation activities.

1.3.2 Incentives for decision-makers to act in the public interest

Digital preservation activities often involve action on the part of a variety of stakeholders, each of whom may bring different sets of responsibilities, motivations, and incentives to the table. In some cases, key decision-makers associated with the preservation activity may not represent those who will benefit most from ongoing access to the materials. For example, those who hold the right to preserve, and are in the best position to carry out long-term preservation (e.g., by virtue of holding custody of the materials) may be distinct from those who perceive benefits from – and are willing to pay for – preservation. In these circumstances, the decision to undertake preservation may hinge on a willingness to act in the public interest, rather than on an organization's perceived "self-interest".

Even when all stakeholders perceive a value in preserving a set of digital materials, the magnitude and duration of the incentive to preserve may differ markedly from stakeholder to stakeholder. For example, a media company's incentive to preserve may persist only over the perceived and presumably limited economic life of a digital asset such as a movie or sound recording; the incentive to preserve embodied in a library or archive, on the other hand, may extend much further in keeping with a mission to preserve the scholarly or cultural record indefinitely.

Recognition of the benefits from preservation leads to an incentive to preserve, but different stakeholders may have different appraisals of value, leading to different incentives to preserve, and ultimately, different expressions of willingness to allocate resources to preservation activities. Sustainable economic models for digital preservation need to ensure that the over-arching public interest in long-term preservation is supported by the complex patterns of stakeholder relationships and incentives attached to a particular set of digital materials.

1.3.3 A process for selecting digital materials for long-term retention

Scarcity of resources imposes constraints on our capacity to carry out any economic activity, and digital preservation is no exception. The mantra "preserve everything for all time" is unlikely to be compatible with a sustainable digital preservation strategy. The mechanism for aligning preservation objectives with preservation resources is *selection* – determining which materials are "valuable enough" to warrant long-term preservation.

Stakeholders employ a variety of methods for selecting materials for preservation, and emphasize different factors in appraising value. In general, selection should incorporate a consideration of risk, where the benefits of preservation weight future value by the likelihood of loss. Thus, materials perceived to be at high probability of loss but of moderate future value may offer greater benefits from preservation than another set of materials considered to be of high future value, but with low probability of loss.

Selection might also take into account the notion of future costs avoided: preservation actions taken now prevent the consequences of irreversible damage or loss to neglected materials. Some decisions about selection or assessments of value are static, once-and-for-all decisions, while in other cases perceptions of value will change over time, necessitating the revisiting of earlier selection decisions.[3]

1.3.4 Mechanisms to secure an ongoing, efficient allocation of resources to digital preservation activities

Orchestrating the pattern of incentives across a range of stakeholders must be done in such a way as to secure an ongoing allocation of resources to support long-term preservation. Too often, digital preservation activities are funded through one-off, limited-term sources on a project-by-project basis. Sustainability requires an uninterrupted flow of resources to appropriate stakeholders, sufficient to meet preservation-related objectives.

---

[3] The perceived value of digital content is often the single most important element in choosing a preservation technique, as such techniques can range from the inexpensive to the very expensive. In libraries and museums, the decisions about value and technique can be seen in the amount of effort that is put into a rare book or unique manuscript, as opposed to the more generic treatment, such as reinforcing a slip cover or re-binding, that is given to mass circulation books and journals.

Economists have defined many market-based and non-market-based mechanisms for this purpose, aimed at the provision of a variety of categories of goods: private goods, public goods, and goods that are private in some aspects and public in others. The issue, therefore, becomes one of choosing the appropriate mechanism to sustain a preservation-related activity in a particular context. In the case of market mechanisms, this issue begs several questions: What pricing strategies can be devised to support a sustainable business model for the provision of preservation services? In the case of non-market solutions, what policies can be enacted to spread the cost burden of digital preservation equitably across stakeholders? Does the responsibility for ensuring that preservation takes place reside with the appropriate stakeholder or stakeholders, given a particular choice of mechanism for obtaining resources to support a digital preservation activity? Sometimes the best method of coordination will involve a transfer of preservation responsibility from one entity to another.

In addition, it is important that resources allocated to preservation be used as efficiently as possible. The economic concept of *efficiency* is not about reducing quality or cutting corners; it means obtaining the maximum output from a given set of resources, or alternatively, using the minimum amount of resources to produce a specified output. Economists' notion of efficiency touches on issues such as reducing duplicative effort and reinvention of wheels, and exploiting economies of scale and scope (Box 1.3). Key efficiency questions to consider for digital preservation include

- Are activities to support digital preservation more efficiently organized as distributed capacity replicated across many institutions, or as a centralized service leveraging economies of scale?

- Can economies of scope be realized by co-locating and integrating preservation, access, and distribution services?

- Are there economic advantages to implementing digital preservation activities within an organization as a single, "monolithic" process, or can costs be reduced by "unbundling" the process into discrete activities performed by different parties?

- If unbundling is called for, are there ways to devise an efficient division of labor across suppliers of digital preservation services? For example, can the digital preservation process be segmented into capital-intensive, "infrastructure" activities that are best performed at scale, and labor-intensive services in which specialized attention rather than economies of scale is key?

All of these issues must be considered to identify mechanisms that support an ongoing allocation of resources to digital preservation activities, as well as efficient usage of these resources.

BOX 1.3 **Definitions: efficiency, economies of scale, and economies of scope**

Efficiency

The term "efficiency" refers to a situation in which one is producing a good or service at the lowest cost possible, everything else being equal. The "everything else being equal" clause is quite important. If, for instance, the price of one of the resources used to produce the good goes down, the resulting cost decrease does not indicate an increase in efficiency. Likewise, if one is able to reduce the cost of production by reducing the quality of the good, this is not an increase in efficiency. If, however, one can find a new technique that allows one to produce the same, identical good at a lower cost, (with no changes in the price of inputs in the market having taken place) an increase in efficiency will have occurred. Efficiency is not the same as "cheap." In many cases, the most efficient way to produce is still very expensive.

Economies of Scale

The term "economies of scale" refers to a situation in which the average cost of producing a good (or service) declines as the scale of production increases. This could happen, for instance, if a firm can buy in bulk, taking advantage of lower unit costs on its inputs, or by allowing more specialization of its workforce, allowing each worker to become more efficient. Economies of scale occur because the organization can spread its fixed costs over a larger and larger level of output as it expands in scale. If a particular industry experiences economies of scale, this suggests that one very large firm can produce the product at a lower average cost than a number of smaller firms could.

Economies of Scope

The term "economies of scope" refers to a situation in which the average cost of production is lower when an organization produces a wider range of products, rather than just one. This occurs because inputs can be spread over several different products rather than allocated to just one product. For example, building a range of different collections may lead to reduced costs per document, because activities such as metadata creation, web development, and storage can be shared across the collections.

1.3.5 Appropriate organization and governance of digital preservation activities

Digital preservation stakeholders can be organized in a variety of ways. For example, digital preservation activities could be carried out by:

- an organization that has no private interest in the long-term persistence of the content of the materials in question (e.g., a "third party service" like Iron Mountain); or

- an organization with a private interest in the preservation of the materials, that provides preservation services on behalf of itself and other organizations with a similar private interest (e.g., a research library or public television station); or

- an organization operating under a mandate to preserve, conferred by public policy and aimed at fulfilling a stated public interest (e.g., a state archive).

It is important to evaluate the advantages and disadvantages of managing digital preservation-related activities through private, public, and quasi-public organizations. In some cases, it may be that one organizational form will be demonstrably advantageous in comparison to others; in other cases, preservation-related objectives can be achieved equally well through alternative approaches. Each of these may have different implications in terms of which stakeholders take on the task of preservation, the value propositions these stakeholders recognize, and the incentive structures public policy may need to reinforce in order to motivate a sufficient and ongoing allocation of resources.

Regardless of the organizational form chosen, governance issues must be considered – broadly speaking, issues of responsibility, authority, accountability, and trust. There are a number of questions that must be addressed:

- Who will articulate preservation goals, and what mechanisms will be implemented to determine whether or not these goals are adequately met?

- What are the procedures for adapting or even changing the organizational form in response to evolving environmental conditions?

- If certain stakeholders choose to abdicate or are unable to discharge their preservation responsibilities, what safety nets are in place to ensure preservation activities continue?

- Do the organizational and governance forms chosen for a particular activity generate an appropriate level of trust among stakeholders?

These and other questions are common to all types of digital preservation activities, but their answers will depend on whether digital preservation is undertaken as a private good transacted among private parties, a public good provided through public agencies, or through some other form of economic organization.

In summary, all of the components of the definition in Box 1.2 must be addressed, in one form or another, if economic sustainability is to be achieved. Neglecting any one of them would seriously jeopardize the long-term prospects of a digital preservation activity.

1.4 Scope

The preceding discussion of the economics of digital preservation and the more specific issue of economic sustainability, while intended to help demarcate the topics of interest to the Task Force, nevertheless still circumscribes a dauntingly wide range of issues. While there are certainly advantages to approaching these issues from a general perspective – with no presupposition or emphasis on a particular preservation context, class of stakeholders, motivation for preservation, and so on – it is also apparent that an analysis that attempts to be everything to everyone will quickly become unfocused and too shallow to be of practical use. In light of this, this section briefly outlines some additional criteria the Task Force invoked in order to draw a tighter circle around the scope of its discussions by describing the categories of materials and domains of interest.

1.4.1 Digital materials in scope

The discussion so far has referred to "digital data," "digital materials," and "digital assets." The use of these generic terms is by design; the problem of economic sustainability extends to all forms and categories of digital objects: text, audio, video, websites, scientific data sets, e-journals, computer programs, and so on, and the preservation process, rather than the specific format and content is the materials is our focus here. What is relevant to our discussion is who will benefit by securing the long term accessibility and usability of the data.

During Task Force discussion, there was reasonable consensus that persistent access to digital materials ultimately in the public interest constitutes one of the most pressing problems for preservation, and one of the most challenging problems for economic sustainability. For this reason, the Blue Ribbon Task Force investigations have focused on *sustainable economic models for digital materials for which there is a public interest* in their long-term persistence. "Public" in this sense can mean anything from a very small set of stakeholders – for example, a group of researchers in a specialized scientific sub-discipline – to society as a whole.

The distinction between "public interest" and "private interest" was chosen as a general rule to help circumscribe the preservation contexts the Task Force would include within the scope of its discussions. However, it is important to acknowledge at the outset that this distinction is not a clean one, and there are many "gray cases" along the boundary between these two spheres. One particularly important issue in this regard is the dynamic nature of perceived value. Some digital materials can be identified at the time of their creation as being valuable and worthy of long-term preservation; the "public value" of other materials becomes apparent only over time. In the latter case, economic sustainability requires not only a flow of resources to support preservation, but also a mechanism for the transfer of heretofore "private assets," and preservation responsibilities, to organizations willing to provide long-term accessibility in the public interest.

What kinds of materials does "public interest" exclude? Examples include purely internal (and non-regulated) business records and data; personal digital photos and other "family" memorabilia (excepting those with acknowledged historical significance); private assets exchanged in market transactions (e.g., the catalog of digital songs in the iTunes inventory)[4]; non-regulated personal e-mail messages, personal

[4] Although the iTunes catalog is out of scope for the purposes of the Task Force, it should be acknowledged that it is easy to imagine situations where personal collections of iPod songs might be of

records (health, taxes, and so on), and other materials where the benefits from preservation – as well as the decision whether or not to pursue preservation – resides with a single private party. For these materials, the decision whether or not to invest in long-term preservation is a purely private one, where the benefits and costs of the decision are fully internalized within a particular organization or individual, who will choose to preserve (or not) based on their private assessment of the merits of doing so.

For other kinds of materials, the decision to preserve – and by extension, the decision to allocate resources to preservation – is more complicated, blurred by factors such as the existence of multiple beneficiaries, and uncertainty over allocations of preservation responsibilities and costs. These complications in turn impact the prospects for economic sustainability, which will ultimately depend on *who values the digital materials and for how long, who can pay, and who will benefit by their preservation.*

1.4.2 Relevant domains

Digital data in the public interest is found in a variety of domains -- from research and education to cultural and creative communities, from the public sector to private enterprise. We describe digital data in the public interest in some of these domains below:

- *Research and education:* Valuable digital data is being collected and used in the earth sciences, life sciences, engineering, and other disciplines. Stakeholders for such data include professional associations that support science constituencies, international collaborative laboratories that are creating comparative data that can be shared across space and time, and individual scientists collecting and using data. Libraries, archives, laboratories and research centers often house this data on behalf of communities of researchers, educators and practitioners; a significant portion of this data, however, is at present not curated at all.[5]

- *Cultural heritage*: The mission of libraries, archives, and museums includes the preservation of the scholarly and cultural record. An ever-increasing portion of this record is manifested in digital form, including e-books, e-journals, databases, digital art work, and other types of materials. In addition to "born-digital" materials, there is also an increasing corpus of digitized versions of print materials. Some of this material is in the custody of cultural heritage institutions; however, a significant portion is not.[6]

---

interest to researchers – ethnographers, for example, who are interested in what these collections contain, how they are organized, and what is their connection to over-arching social issues.

[5] See, for example, the National Science Board's Long-Lived Digital Data Collections: Enabling Research and Education in the Twenty-First Century. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf.

[6] See, for example, the Library of Congress' *Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program: a Collaborative Initiative of the Library of Congress.* http://www.digitalpreservation.gov/library/resources/pubs/docs/ndiipp_plan.pdf.

- *Government agencies:* Governing and regulatory institutions, from local, to national, to international, must maintain digital data that reflects a faithful record of their activities and ensures their long-term accountability. Citizens and organizations rely on this data as an important bulwark of a free and open society. Caretakers of data in this realm include public agencies, archives, and libraries.[7]

- *Private enterprise*: Private (often commercial) organizations manage a variety of digital assets in the course of their ongoing operations, some of which may have a lasting value to a wider community and perhaps even society as a whole. Examples include aircraft design data, the long-term preservation of which has implications for public safety; and the output of film and television studios, which forms an integral part of society's cultural record. These digital data are often managed by the private organization itself, which may or may not be bound by a public mandate specifying minimum preservation requirements.[8]

Clearly these domains are not mutually exclusive, but they represent broad segments of society in which digital preservation, and by extension, *economically sustainable* digital preservation, is a critical issue. Different domains may employ different mechanisms to generate and coordinate the flow of resources, but the imperative to create and sustain a viable economic model is the same.

[7] See, for example, *Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for Initial Development* by the Committee on Digital Archiving and the National Archives and Records Administration. http://www.nap.edu/catalog.php?record_id=10707.

[8] See, for example, *The Digital* Dilemma, by The Science and Technology Council of the Academy of Motion Picture Arts and Science. http://www.oscars.org/council/digital_dilemma/index.html.

Chapter

2

Economic Models and Their Properties

To explain in what has consisted the revenue of the great body of the people, or what has been the nature of those funds, which, in different ages and nations, have supplied their annual consumption, is the object of these Four first Books. The Fifth and last Book treats of the revenue of the sovereign, or commonwealth.

Adam Smith
Introduction and Plan of the Work
An Inquiry into the Nature and Causes of the Wealth of Nations, Vol. I
(London: Methuen & Co., Ltd., 1904, 5th ed., first published 1776)
http://www.econlib.org/library/Smith/smWN1.html#B.I,%20Introduction%20and%20Plan%20of%20the%20Work.

Adam Smith (1723-1790) is widely considered the father of modern economics, and his much cited *Wealth of Nations* the first modern economics text. In the short excerpt above, Smith provides a plan for his five-volume work, where he proposes to lay out a model for the relationships among income, consumption and governance at the national level that is generalized above the specific data he considered. Thus, the notion of a *model* – a level of abstraction above the description of a given set of circumstances that sets forth the underlying relationships – is intrinsic to the modern study of economics.

> BOX 2.1 **Economic models**
>
> Economic models are stylized representations of how economic processes work. They are a means to abstract an economic process down to the essential details that are important for 1) understanding how the process works, and 2) identifying the aspects of the process that can be influenced by outside intervention, such as public policy.

The *economic model* (see box 2.1) is the primary tool the Task Force will use to examine the issue of economically sustainable digital preservation. In exploring alternative economic models for digital preservation, we are interested in understanding the key features of different mechanisms by which digital preservation activities can be sustained over the long-term (i.e., how they work), and how different policy instruments might be used to shape the outcomes of these processes in socially beneficial ways (i.e., how they can be influenced). This chapter discusses in more detail what an economic model is, distinguishes *economic models* from *business models* and *cost models*, and lists several properties of economic models important for analyzing sustainable digital preservation.

2.1 Building an economic model for sustainable digital preservation and access

Although in common usage, the term "economics" is applied to pretty much anything that involves money, economists see themselves as studying human behavior in which people (including aggregations of individuals such as businesses or governments) make choices about the use of scarce resources. The fundamental idea is that if we use something of value (say, the electricity, materials and workers' time and skill needed to store a terabyte of information for a year) for one purpose, we are not using it to do something else of equal value.

In one of the simplest and most widely used textbook examples of an economic model, a society is depicted as choosing between guns and butter – metaphorically, between security and the standard of living. One of the important insights of this simple model is that one can turn guns into butter, and vice versa, by redeploying land, labor and capital from one set of activities into another. Similarly, by taking the money that we would use in storing a terabyte of data for a year at some specified level of accessibility, society would be able, as a technical matter, to produce more guns, more butter, or more of anything else.

An economic model provides a compact representation of some economic process that allows us to make predictions and judgments regarding how the process works, and, most important, how its performance can be affected by changes in the environment, including changes in policy. Crucially, when the process under consideration is digital preservation and access, economic models illuminate questions of who is willing to pay how much for what level of service over time (the *demand* side of the model), and what resources are required to deliver given levels of service over time (the *supply* side).

Both the demand and supply of digital preservation will depend on institutional arrangements, including copyright law, and upon existing and future technologies. Moreover, because preservation is an activity that takes place over time, an economic model of digital preservation will recognize that both the value and the cost of preserving something in the future will depend on how (and whether) it was preserved previously. It follows that how things are preserved (or are not preserved) in the present will affect both the cost and value of their preservation in the future.

Decisions regarding preservation are "path dependent", a term some economists use to mean a course of action in which subsequent outcomes depend upon specific actions and are in some sense not reversible. In effect, it means that time matters uniquely. Thus, preservation actions taken today provide the *option* of continued preservation tomorrow (or not). But when something is lost, or an action is not taken, the item is lost forever. In one well-known example, the tapes for the U.S. federal census were almost lost because appropriate preservation actions were not implemented. Once a preservation path is chosen, there is no going back. This phenomenon – also termed "irreversibility" in economics – must be incorporated into a well-specified model of digital preservation. And as observed in Chapter 1, this notion is bound up with notions of investment and hence in expectations about the future.

As we know, there are multiple approaches to digital preservation, ranging from simple bit storage to complex (and costly) techniques such as emulation. One decision that might be built into economic models is the cost/benefit aspects of preserving the option to invest down the road. For example, we can preserve (nearly) everything to a degree in the sense that we can simply transfer digital materials with modest or minimal treatment to large-scale storage systems at a comparatively modest cost. Doing so preserves the option to invoke more costly methods of processing and preservation at a future date, should the need or incentive arise. The trade-off for the information manager or decision-maker concerns whether the cost of preserving the option to invest (i.e., the cost of paying for simple bit storage over time) is worth the possible future benefits (i.e., future opportunity for more intensive preservation should the need arise). As this example illustrates, by acting in the present, even in a limited way, the opportunity to act more flexibly in the future is preserved. This allows an information manager to provide for changes in demand from his or her users as well as to respond to advances in technology that may alter the costs of preservation.

2.2 Economic, business and cost models

Economic models can be distinguished from *cost models* and *business models*, each of which is useful and may be essential for understanding an economic process, but neither of which can be used reliably except in the context of a broader economic model. In Chapter 1, we discussed ways that inflows, or revenues, might be balanced against outflows, or costs, and ways in which perception of the balance between revenues and costs might be reconciled with growth and future investments. In this balance between inflows and outflows, the frequently-asked question, "what does digital preservation (and access) cost?" implies a *cost model* (which associates enterprise components with itemized costs), rather than the more general economic model in which the cost model may be embedded, or the business model that might be adopted by a given organization or institution.

There is a great deal of literature[9] that does a careful and excellent job of determining the cost of preservation of specific digital assets using specific technologies, at a specified level of reliability, and in a specific institution, usually a research library. When we have this kind of information, we have a good cost model for the narrow problem in question. But such models have severe limitations. For one thing, a cost model that applies to a particular technology will often fail to tell us what we could get

[9] A bibliography of publications relevant to digital preservation and access can be found on the Blue Ribbon Task Force website: brtf.sdsc.edu.

by spending somewhat more or somewhat less. For another, as technology changes the methods used for preservation will change as well and, almost certainly for the foreseeable future, a given quality of preservation and access will become less expensive over time. At the same time, migration of data from one platform to another is itself costly, and also poses risks of loss.

For all of these reasons, a good cost model for a specific technology and desired outcome at a point in time may not be very useful as part of an economic model designed to help make policy over a longer period of time. But the more flexible economic model will, perforce, be less accurate, as it will have to incorporate assumptions about an uncertain future. Thus the right answer to the question, "what does preservation cost?" is, "it depends," representing decisions concerning a spectrum of choices, trade-offs, and predictions.

A *business model* is generally less formal than either an economic model or a cost model. It is a description of how sufficient resources flow to the activity at hand to keep things running. Like cost models, business models generally specify a particular technology or set of technologies. By way of example, the business model for newspapers involves selling advertising space in widely distributed print copies of newspapers.

Digital preservation involves many business models, because the entities that might undertake such preservation are widely varied, including financial institutions, the video and audio recording industries, commercial and public producers of scientific data, publishers, academic and public libraries, the Library of Congress, and many others. Each of these has different ways of bringing resources to bear on the activities and products that they produce, and each will have different business models.

In examining sustainable digital preservation and access, then, useful economic models will need to account for the multiplicity of business models and business types. Note that the business model for digital preservation may often involve moving digital resources from one type of entity (e.g., a scientific laboratory, or a video production company) to another (e.g., a library). Developing mechanisms such that valuable materials are not lost when the entities that produce them no longer find them worth keeping is one of the difficult challenges of sustainable digital preservation, in part because business models for such handoffs are not well developed (Box 2.2).

BOX 2.2 **Description: economic model versus business model**

Business Model

A business model is a description of the ways in which an organization does its business to achieve its mission. A formal business model includes items such as the market being served, its product and service offerings, the perceived value delivered to the market, sources of revenue, financial and cost models that will support inflows and outflows of funds, and its strategic alignment, policies, and procedures. In general, however, business models are not formally expressed except at start-up, when funding bodies require proof that a coherent and sustainable plan does exist.

Frequently, the term "business model" is used in everyday parlance as a way to express a general intuition of the ways in which an organization raises its money to sustain itself.

Economic Model

An economic model, on the other hand, abstracts from the overall environment and explains how scarce resources are allocated, given the constraints on the various "actors" in the economy. For example, the wide variety of business practices found within archiving may be described as abstract types such as a pay-per-use model, an endowment model, a voluntary fee model, etc., and actors may be defined in terms of their goals, such as consumers, patrons, archives, and regulatory agencies. The economic model then analyzes the interactions that occur between these actors to predict outcomes based upon the constraints they face (such as their budgets) and their incentives to engage in certain types of action.

The development of a good business model relies upon a well conceived economic model. The economic model "describes" how economic reality works and the business models provide "templates" for acting within that reality. With a good economic model, organizations can assess which types of business models will work best for them, given their missions, the environments in which they operate, and the types of products, services, and customers they face.

2.3 The Minimum set of properties

Useful economic models that address the sustainability of preserving and providing long-term access to digital data will have the following minimum set of properties:

- They will account for the resources used to produce sustainability and access.
- They will pay special attention to the role of time, in both the simple sense of the elapsed time that leads to bit rot, and in the more complicated sense that over time ownership of the data and available technologies may change.
- They will enable us to examine the effects of different organizational and technical strategies on the quality of preservation and access.
- They will enable us to assess the technical and the economic risks of losing data.
- They will allow us to evaluate alternative policies, including changes in intellectual property law.
- They will allow us to evaluate the implications of the five components of our sustainability definition, both individually and collectively.

In short, economic sustainability requires a balancing of inflows and outflows, supply and demand, and perception of value and willingness to pay. Business models focus on the inflows and the demand side of an economic activity like digital preservation;

cost models focus on the outflows and the supply side of the activity. Economic models subsume both inflows and outflows, supply side and demand side, and the important interactions that bind the two aspects of an economic activity together (Box 2.2 above). They also highlight the parameters intrinsic to the activity that could be influenced by policy. Economic models provide a framework for a well-informed society to evaluate the costs and consequences of investing resources in the long-term preservation of digital data.

In Chapter 1, we noted that the long-term persistence of digital assets is exposed to systemic barriers – creating risks that manifest themselves in the form of a variety of potential points of failure, which individually or collectively could impair the ability of a digital preservation activity to sustain itself over time. The goal of the Task Force is to develop recommendations to help decision-makers mitigate the risks involved in digital preservation. Economic models are the tools with which these recommendations can be articulated, explained, and analyzed.

Models and frameworks are used to understand and discuss other aspects of digital preservation. For example, the widely-accepted Open Archival Information System (OAIS) reference model provides a high-level, stylized representation of the functional components of a digital archiving system (Figure 2.1). OAIS has proved exceedingly useful in the digital preservation community both as a means to galvanize conversation about digital archive architectures and functionalities, and as a high-level guide to aid in system implementation. The proliferation of models like OAIS have greatly assisted in identifying and resolving key technical issues associated with implementing digital archiving systems. We need similar models to help us identify and resolve the economic issues associated with building sustainable digital preservation activities.

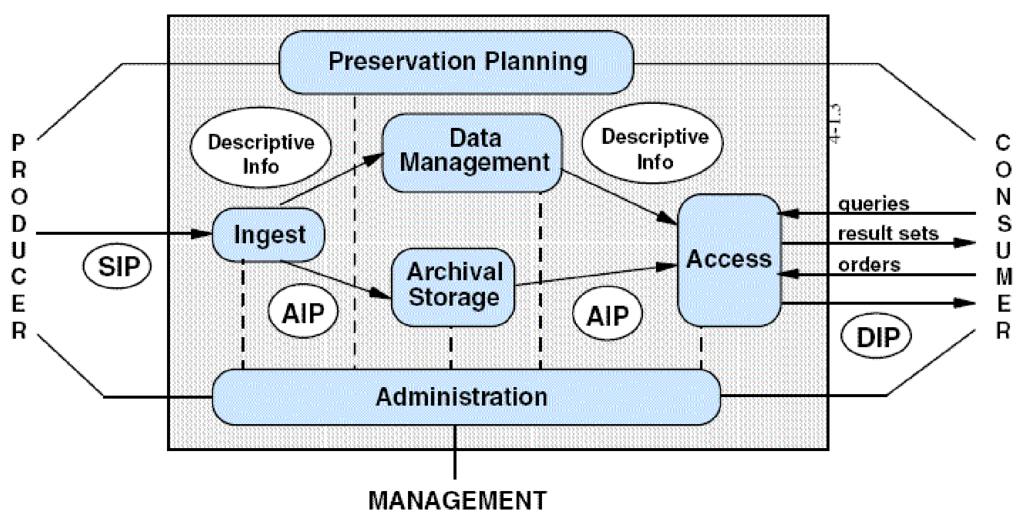

FIGURE 2.1: **The OAIS Reference Model**
http://public.ccsds.org/publications/archive/650x0b1.pdf, Page 4-1.
Source: Consultative Committee for Space Data Systems January 2002.

Chapter 3

Prior Experience and Preliminary Lessons Learned

Nature, to be commanded, must be obeyed.

Francis Bacon
Organum, bk. 1, aph. 129 (1620)
http://www.bartleby.com/66/14/5114.html.

"Like eggs" observed the economist Paul Samuelson, "there are only two kinds of theories: good ones and bad ones. And the test of a theory's goodness is its usefulness in illuminating observational reality. Its logical elegance and fine-spun beauty are irrelevant. Consequently, when a student says, 'That's all right in theory but not in practice,' he really means 'That's not all right in theory,' or else he is talking nonsense" (Samuelson, 1948, p. 8). Or to paraphrase Francis Bacon, Nature -- in this case, shorthand for experience – always wins.

Samuelson's remarks are a good reminder that in any field of analysis, models and other abstractions are useful only to the extent that they accurately capture the salient features of the "real-world" processes, relationships, or behaviors they represent and in doing so, provide a sound basis on which decision-makers and practitioners can consider important implications and practical strategies. Economic models for long-term management of digital assets are of course no exception: any analysis of the economic mechanisms appropriate for sustaining digital preservation activities over the long-term should be preceded by an examination of the *economic context* in which these models are expected to operate.

Like all economic activities, digital preservation does not take place in a vacuum. Instead, it is embedded within a complex pattern of motivations and incentives, organizational goals, and economic constraints. Moreover, the nature of the economic context surrounding digital preservation often varies significantly from domain to domain: a business seeking to manage the long-term preservation of digital assets, access to which is transacted for in the open marketplace, faces a different economic problem than a library aiming to fulfill its mission of preserving the scholarly and cultural record. An economic model designed to achieve long-term sustainability in the first instance may be entirely unsuitable for the second. Consequently, it is useful to identify some of the major contexts in which long-term access to digital information is an important concern, and draw out some of the key aspects of these contexts that might help in guiding the choice of a suitable economic model to support sustainability.

To develop an understanding of current and best practices in digital preservation (represented in this Interim Report), the Task Force used two strategies. First, the Task Force undertook a selective review of the existing literature on the economics of digital preservation. This review was largely concentrated in the literature associated with cultural heritage institutions (libraries, archives, museums); these institutions provide a rich accumulation of materials of addressing the costs and related issues associated with digital preservation, and also an opportunity to trace the evolution of work related to the economics of digital preservation in a particular domain. The literature review provided a base-line understanding of key economic issues associated with digital preservation as they have evolved over the last decade, which can then be compared to experiences in other domains. The Task Force also solicited the views of a number of industry and domain leaders experienced in digital preservation and access. These leaders presented testimony during three meetings over the course of 2008, and represented a variety of sectors, areas of expertise, and digital preservation contexts and content.

A synthesis of the information from the selective literature review and the Task Force's interaction with the speakers are reported in this chapter. Together, they provide an initial "body of evidence" that will inform the next phase of the Task Force's work.

3.1 The Economics of Digital Preservation: A View from the Literature

The goal of the literature review is to provide a baseline understanding of the current state of digital preservation sustainability practice and research, particularly as it relates to fundamental economic aspects of preservation activity: e.g., who the stakeholders are, the value proposition for preservation of particular types of materials, how "successful" preservation can be defined, and what organizational structures and funding sources best serve to achieve sustainable digital preservation. To keep the length of this report manageable, we focus on a small number of key projects from this review in this section. An extended literature review and bibliography of related work are posted the Task Force's website at brtf.sdsc.edu to supplement the information herein. In addition for convenience, a summary of findings from the literature review is presented in Table 3.1.

The reports discussed in this review highlight the difficulty in taking full advantage of individual project findings: the concrete models that have been developed remain largely organization-specific and are difficult to reconcile with one another, i.e. they largely do not support "apples to apples" comparisons. For the most part, when seeking to develop detailed cost assessments, organizations have only had their own data to fall back on, and this is reflected in the literature by cost models and assessments that are largely atomistic. Studies also structure themselves differently; they define costs differently and assign different units of measurement; different formats are captured; and decisions regarding which costs and cost adjustments to include and exclude vary from project to project. When publishing project updates authors typically do not create economic "crosswalks" between their and others' frameworks. Even for those projects that explicitly build on earlier work, it is clear that within any given project the costs captured are generally focused upon only a small subset of activities within the digital preservation lifecycle (for example, storage costs). In short, the structure for previous studies rarely supports direct comparisons.

Nonetheless, over time, the discussion has become more sophisticated, recognizing that costs are embedded in a larger framework that considers stakeholder interests, organizational structure, and cultural milieu. In particular, two recent projects highlight the increasing economic sophistication we are now seeing: (1) the cost model developed by the LIFE (Life Cycle Information for E-Literature) project, and (2) the recently published model developed by Beagrie, Chruszcz, and Lavoie (2008).

TABLE 3.1 **Economics of digital preservation: summary findings**

DESCRIPTION	COST
2001	
Roquade Project *Dekker et al.*	**Basis for assessment:** Experiential estimates, published literature reports **Unit of measurement:** Cost per information item • Personnel costs of assigning metadata: approximately 10 euros • Processing SIP's: approximately 10 euros per information item • 5,000 items per year added: 6 PC's with a network card and AV facilities: 1500 euros each + professional serer: $5000 euros • Total hardware costs: approximately 32,000 euros, depreciated over 4 years • Software and licensing fees: 15,000 euros per year using proprietary software • Maintenance support costs: 2,000 euros per year • Technical support: 0.2 FEs = 9,000 euros per year • Data refresh every 5 years @ 1 euro per MB; if DIPs are kept for 20 years and DIP is about 500 kB, cost - about 2 euros per information item, that is, 10,000 euros per year for all information items • Total per information item costs: 29 euros per item

2003	
Harvard Depository Chapman *Format:* Microfilm, Book	**Basis for assessment:** Billing model *Excludes ingest costs, excludes access costs* **Unit of measurement:** Billable square feet • $0.08 per 332-page (microfilm) volume per year in the standard vault • $0.19 per 332-page (microfilm) volume per year in the film vault • $0.31 per 332-page (book) volume in the standard vault
OCLC, Inc. Chapman *Format:* ASCII text, 600-dpi 1-bit page images	**Basis for assessment:** Billing model *Excludes ingest costs, includes access costs* **Unit of measurement:** Total GB of data deposited • $0.01-0.06 per 332-page ASCII text • $0.47/$1.01/$1.89 per 332-page 600-dpi 1-bit page image (variable rate, based upon total amount of data deposited per account)
2005	
Digital Preservation Testbed, Nationaal Testbed Digitale Bewaring Archief of the Netherlands *Format:* Email, Text, Spreadsheet, Database	**Basis for assessment:** Literature review, testbed experience, and external project cost information *Includes an estimate of 20% overhead* **Unit of measurement:** Annual costs (total), Email batch costs • Creation of a batch of 1000 records (assuming 50kb per email, 100 kb per text document, 250 kb per spreadsheet, and 2 Mb per database): 333 euros • "Repair" of a batch of 1000 records (assuming 50kb per email, 100 kb per text document, 250 kb per spreadsheet, and 2 Mb per database): 10,000 euros • Acquisition and input of metadata for "normal" email: 1.41 euros • Acquisition and input of metadata for XML email: 0.06 euros
2006	
Riksarkivet National Archives of Sweden *Palm* *Format:* 1-bit 600-dpi files in A4 format, 8-bit grey-scale at 297 dpi, Audiovisual	**Basis for assessment:** Audiovisual digitization costs *Includes storage system, staff operations, staff data input, service/support, and premises - averaged over 5 years* **Intertemporal adjustment:** 3% interest on hardware included in intertemporal calculations **Unit of measurement:** Cost per year per 1 Gb stored; Total costs per year • 1 Hierarchical Storage Management System (i.e., HSM) (2003 price + 3% interest per year): 449,694 euros over five years Storage medium for additional 40 Tb/year: 43648 euros over five years • Staff operations costs: 132240 euros over five years (0.6 FTE) • Staff ongoing data input: 88160 euros over five years (0.4 FTE) • Total annual input cost: 131808 euros over five years (staff & storage medium included)

	• Facilities ("Premises") (100 square meters): 66228 euros over five years • Service/support: 138300 euros over five years • Digitization of paper materials (1-bit 600 dpi files in A4 format): 0.10 euro per file, with 5 million images scanned annually • Scanning of large-format drawings and maps (8-bit grey-scale at 297 dpi, in manually fed scanners): 0.61 euro per file, with 1,321,000 image files created annually • Production costs for 1 Gb 1-bit digitized information: approximately 17 euros per Gb • Production costs for 1 Gb 8-bit digitized information: approximately 30 euros per Gb • Production costs for Audiovisual information: approximately 11 euros per Gb
LIFE *Ayris et al.*	*Includes full life cycle costs* **Intertemporal adjustment:** 7% inflation factor for materials costs, 3.5% cost of living increase for staff costs • First year of e-monograph's life: lifecycle cost = 19 pounds (English) • Tenth year of hand-held e-monograph's life, lifecycle cost = 48 pounds (English) *predicted* • First year of hand-held serial's life, lifecycle cost = 19 nth year of hand-held serials life, lifecycle cost = 14 pounds (Engligh) pounds (English) *predicted* • First year of non hand-held e-monograph's life, lifecycle cost = 15 pounds (English) • Tenth year of a non hand-held e-monograph's life, lifecycle cost = 30 pounds (English) *predicted* • First year of non hand-held e-serial's life, lifecycle cost = 22 pounds (English) • Tenth year of a non hand-held e-serials life, lifecycle cost = 18 pounds (English) *predicted* • First year cost for new website = 21 pounds (English) Tenth year cost for new website = 6,800 pounds (English) *predicted* • First year of e-journal's life, lifecycle cost = 206 pounds (English) Tenth year of an e-journal's life, lifecycle cost = 3,000 pounds (English) *predicted*

2007	
Academy of Motion Picture Arts and Sciences AMPAS Science and Technology Council *Format:* "all film" production; film-captured, digitally finished production at 4K; digitally captured, digitally finished production using HDCAM SR at 1920 x1080; digitally captured, digitally finished production using uncompressed system at 2K; digitally captured, digitally finished production using uncompressed system at 4K	**Basis for assessment:** Data storage costs as reported by the San Diego Supercomputer Center; Annual (total) storage costs *Excludes initial inspection and access costs for both film storage and digital storage* **Intertemporal adjustment:** Amortized cost of YCM separation master manufacture ($800/year) **Unit of measurement:** Total Annual Storage Costs • "All film" production generating no digital assets, annual storage costs for archival master: $1059 • A film-captured, digital finished production at 4K, annual storage costs for archival master: $12,514 • Digitally captured, digitally finished production using HDCAM SR videotape as the capture medium at 1920 x 1080, annual storage costs for archival master: $1,830 • Digital captured, digitally finished production using an uncompressed digital data capture system at 2K, annual storage costs for archival master: $1,955 • Digitally captured, digital finished production using an uncompressed digital data capture system at 4K, annual storage costs for archival master: $12,514
Cambridge University *in Beagrie, Chruszcz, and Lavoie*	**Basis for assessment:** Literature review, survey questionnaire-based interviews, three in-depth case studies; Archaeology Data Service Charging Policy **Unit of measurement:** Staff time, Total cost in pounds or pence • Project initiation costs: 2-5 days per project for 2 FTEs at Grade 8 • Creation costs: 1-5 days at Grade 8 and above • Metadata creation: 5% of Grade 6 post (recurrent) • Acquisition: 5 days for one-off standard terms plus additional recurring negotiations • Outreach support: 20% of ongoing effort for repository manager (Grade 8) and Support and liaison officer (Grade 6) • Ingest: 10% of ongoing effort (recurrent) • Metadata upload and integration: 20% of two Grade 8 posts for 3 months • Integration with other campus systems: 300,000 English pounds over 3 years • Preservation planning: 0.5 FTE at Grade 8 (recurring)\ • First Mover Innovation: up to 5% of two Grade 8 personnel (recurring)

	• Sun Fire X4500 x64 Server: about 69,200 English pounds • DELL/EMC CX3-20c FC4 SPE DAE4P-OS for CX3-20: about 107,100 English pounds • Initial staff costs for 2 FTE graduate students Grade 6: 16,656 English pounds • Production staff costs for 1 10% FTE Graduate student Grade 6: 3331 English pounds • Production staff costs for 1 25% FTE Computer Officer Grade 8: 12,339 English pounds
King's College London *in Beagrie, Chruszcz, and Lavoie*	**Basis for assessment:** Literature review, survey questionnaire-based interviews, three in-depth case studies; Archaeology Data Service Charging Policy **Unit of measurement:** Staff salaries • Archive manager: 45,000 English pounds • Half-time system administrator: 24,000 English pounds • Collections officer: 35,000 English pounds
Archaeology Data Service (ADS) *in Beagrie, Chruszcz, and Lavoie*	**Basis for assessment:** Literature review, survey questionnaire-based interviews, three in-depth case studies; Archaeology Data Service Charging Policy **Unit of measurement:** Staff days, Total cost in pounds or pence, Per megabyte (Mb) cost in pence • Text and image file deposits: 1-10 files = minimum of 1 day • Text and image file deposits: 11-100 files = minimum of 2 days • Text and image file deposits: 100+ files = minimum of 4 days • Mixed files including GIS, CAD, Geophysics, Databases, etc.: 1-10 files = minimum of 2 days • Mixed files including GIS, CAD, Geophysics, Databases, etc.: 11-100 files = minimum of 3 days • Mixed files including GIS, CAD, Geophysics, Databases, etc.: 100+ files = minimum of 6 days • Queriable database: about 1000-5000 English pounds • Fully-functional GIS interface: as much as 10,000 English pounds • Cost of a gigabyte of disc storage: as low as 7 English pence; in 5 years, as low as 1 pence; approaching zero cost • Refreshment costs: for five year retention period: about 13 pence per Mb • Refreshment costs: for ten year retention period: about 22 pence per Mb • Refreshment costs: for fifteen year retention period: about 227 pence per Mb • Refreshment costs: for twenty year retention period: about 28 pence per Mb • Ongoing refreshment costs: about 30 pence per Mb

University of Southampton *in Beagrie, Chruszcz, and Lavoie*	**Basis for assessment:** Literature review, survey questionnaire-based interviews, three in-depth case studies; Archaeology Data Service Charging Policy • Staffing: 4 RA's = 332,000 English pounds • Staffing: Department Service experimental officer = 90,000 English pounds • Staffing: Department Self Service RA = 83,000 English pounds • Staffing: 3 PhD Research students = 90,000 English pounds • Lab instrumentation capital cost 45,000 English pounds (@ 10%) • Maintenance: 2,000 English pounds • Repair (averaged over 10 years): 10,000 English pounds • Raw data storage: 1,200 English pounds • Consumables: 4,000 English pounds • Assuming 2,000 datasets collected per annum, cost per crystal structure = 328.60 English pounds
2008	
SHERPA-DP IR *Ayris et al. (LIFE^2)*	**Basis for assessment:** Case study / workflow analysis *Includes full life cycle costs; Excludes interest rate, depreciation* **Unit of measurement:** Costs measured at the unit for which metadata is created (e.g., per object cost for analogue, per page cost for digital) • Year 1: 18.40 English pounds per year total cost • Year 5: 9.70 English pounds per year total cost • Year 10: 8.10 English pounds per year total cost
SHERPA-LEAP IR *Ayris et al. (LIFE^2)* ~Goldsmiths ~Royal Holloway ~UCL	**Basis for assessment:** Case study / workflow analysis *Includes full life cycle costs; Excludes interest rate, depreciation* **Unit of measurement:** Costs measured at the unit for which metadata is created (e.g., per object cost for analogue, per page cost for digital) Goldsmiths: • Year 1: 31.50 English pounds per year total cost • Year 5: 32.00 English pounds per year total cost • Year 10: 32.20 English pounds per year total cost Royal Holloway: • Year 1: 23.10 English pounds per year total cost • Year 5: 23.60 English pounds per year total cost • Year 10: 23.90 English pounds per year total cost UCL: • Year 1: 15.00 English pounds per year total cost • Year 5: 16.50 English pounds per year total cost • Year 10: 16.70 English pounds per year total cost

British Library Newspapers Digitization Project *Ayris et al. (LIFE²)*	**Basis for assessment:** Case study / workflow analysis *Includes full life cycle costs; Excludes interest rate, depreciation* **Unit of measurement:** Costs measured at the unit for which metadata is created (e.g., per object cost for analogue, per page cost for digital) • Digital: 1,045,587 English pounds total project cost • Analogue: 1,820,702 English pounds total project cost

The LIFE project has developed a comprehensive model oriented toward library operations and provides a list of elements for costing a digital library project or initiative. Its goal is to develop a framework that will allow one to determine the long-term cost of digital preservation, the long-term costs of library partnerships within the higher education community, a comparison of paper versus digital preservation for any given publication, and an understanding of the relative risks of digital versus paper archiving. The project developers also want to be able to use their framework to determine who will engage in preservation and when it is appropriate to switch from paper to digital formats for publications (Ayris, Mcleod, & Wheatley, p. 6). One key finding of this project has been that the upfront (i.e., "one-time") costs of a project are often distinct in structure from the recurring maintenance aspects of the same project. The second phase of the project, LIFE², which was concluded in the summer of 2008, refined preliminary results presented in the first phase of work. In the second phase, the project added data from three additional case studies; commissioned an independent economic review, which validated the approach; and presented a refined model (Box 3.1).

BOX 3.1 **Stages of the LIFE² Model**

$$L_T = C + Aq_T + I_T + BP_T + CP_T + Ac_T$$

L_T: Life cycle (i.e., the total cost of maintaining the objects over their "life")

C: Creation or purchase (i.e., the cost of purchasing or creating the objects)

Aq_T: Acquisition (i.e., selection, submission agreements, IPR & licensing, ordering, invoicing, obtaining, and checking in)

I_T: Ingest (i.e., quality assurance, metadata, deposit, holding update, and reference linking)

BP_T: Bitstream Preservation (i.e., repository administration, storage provision, refreshment, backup, and inspection)

CP_T: Content Preservation (i.e., preservation watch, preservation planning, preservation action, re-ingest, and disposal)

Ac_T: Access (i.e., access provision, access control, and user support)

Source: P Ayris et at. August 22, 2008 p. 16.

The model developed by Beagrie, Chruszcz, and Lavoie (2008), henceforth referred to as the BCL model, is oriented toward enabling institutions of higher education to develop a digital preservation cost model. This model builds upon the work of LIFE and other cost models, and generalizes the findings from a variety of case studies to come up with a comprehensive, concrete model that accounts for the variety of functional phases within the preservation environment, enables context-specific adjustments to occur (such as accounting for inflation or depreciation of capital assets), and maps to the OAIS and TRAC[10] models. Like the LIFE model, the BCL model focuses upon long-term preservation costs within higher education institutions. It also follows LIFE in recognizing that fixed and variable cost structures may differ, and supports tracking each of these cost types separately. Additionally, the BCL model includes "First-Mover Innovation" costs (p. 6): the often substantial costs involved in being the originator or developer of a new tool, technology, or method that moves the state of practice forward. Finally, a key finding of the BCL case studies is that the costs of preservation increase, but at a decreasing rate, as the retention period increases. That is, the *pace* of the increase tends to fall over time after an initial period of growth.

The two-phased LIFE project and the BCL model arise from the higher education sector in the U.K. and its accounting framework. While this framework greatly assists analysis of indirect costs, it is not obvious that the cultural assumptions that underpin the U.K. higher education sector will map easily to different nations with different academic traditions, or to other sectors. In particular, institutional culture and the implications of that culture on the mission of the institution matters. In addition, review of the cost literature to date also points out what may now seem obvious, but was not always recognized: format matters, and scale matters. In one particularly well-documented example, the Academy of Motion Picture Arts and Sciences called attention to the potentially variable nature of costs across different sectors by highlighting the preservation environment in a number of different industries (The Science and Technology Council of the Academy of Motion Picture Arts and Sciences, 2007). The authors examined strategic issues in archiving digital motion picture materials, and concluded that the annual cost of preserving digital materials per title is approximately eleven times greater than the cost of preserving film material per title (pp. 1-2).

The increasing sophistication of economic models over time has led to a series of discoveries about the costs of digital preservation. For example, by focusing upon the lifecycle of digital objects, various projects have determined that costs may be unevenly distributed over that lifecycle. Beagrie and colleagues noted that the costs of ingest carry especially heavy weight, leading to a recognition that if done right, the incremental costs of preservation can reduce dramatically over the long run. Of course, if the preservation activities were not done right, then the costs of repair are exceedingly expensive, as pointed out by the National Archives of the Netherlands (Testbed Digitale Bewaring, 2005).

---

[10] The acronym "TRAC" here refers to the "Transparent Approach to Costing," a standard approach to costing in UK higher education. See http://www.jcpsg.ac.uk/guidance/about.htm.

Palm's 2006 paper "The Digital Black Hole," together with studies on the growth of the information universe (Gantz, 2008) and the storage industry itself (Peterson, Zasman, Mojica, & Porter, 2007), suggest that storage is a more textured component than early studies may have indicated. In addition, power costs are continuing to rise and to rise more quickly than the costs of new servers (Gantz, 2008, p. 4).[11] Greater capacity will not necessarily solve the problem for custodial institutions since the creation of new information is accelerating faster than capacity to store it (Gantz, 2008, p. 2), and the engineering of cooling large, complex servers are complex and hence power requirements to operate and maintain such systems are considerable and the costs not only of upgrading but also of operating these systems require consideration. Both for-profit and not-for-profit institutions are adopting strategies such as data center consolidation, use of more efficient equipment, or virtualization. At the Spring 2008 Preservation and Archiving Special Interest Group (PASIG) meeting, participants discussed the need to reconsider storage architecture in a fundamental manner. At the least, it is essential to consider the costs of power and cooling requirements explicitly, especially given that the cost of energy is likely to continue to rise in the future.

The key studies described herein, the literature review at brtf.sdsc.edu, and related work are beginning to form the basis of a literature on the economics of digital preservation. Such literature is critical to illuminate the issues and inform discussion on practical solutions and necessary investigations; it is an important part of a landscape in which viable economic models for preservation will be developed and implemented.

3.2 Speaker Testimony

In addition to reviewing existing studies on the economics of digital preservation, the Task Force invited sixteen speakers (Box 3.2), representing a variety of domains and areas of expertise, to present and discuss emerging trends, perceived challenges, and current solutions.

BOX 3.2 **Speakers who presented to the Blue Ribbon Task Force in 2008**

Helen M. Berman, Board of Governors Professor of Chemistry and Chemical Biology at Rutgers University, Co-Founder of the Protein Data Bank (PDB)

Eileen Fenton, Executive Director, Portico

John Gantz, Chief Research Officer, IDC

Myron P. Gutman, Professor of History and Information and Director of the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan.

Melissa Levine, Exhibits and Outreach Librarian, University of Michigan

Rick Luce, Vice-Provost and Director of Libraries at Emory University

[11] At the May, 2008 meeting of the Preservation and Archiving Special Interest Group (PASIG), one of Sun's chief engineers offered a slide with back of the envelope calculations that show that providing power and cooling for a petabyte of data could cost $1 million a year based on California utility rates.

Andy Maltz, Director, Science and Technology Council, Academy of Motion Picture Arts and Sciences

Edmond Mesrobian, Chief Technology Officer, RealNetworks, Inc.

Peter Mojica, Vice President, Product Product/Strategy and Business Development, AXS-One Inc., SNIA-DMF Storage Networking Industry Association Data Management Forum

Kris Carpenter Negulescu, Director, Web Group, Internet Archive

Paul Ratnaraj, Director of Advanced Initiatives for WRDS at the Wharton School of the University of Pennsylvania

Nan Rubin, Project Director, Preserving Digital Public Television, Special Projects/ Technology Planning, Thirteen/WNET

Rick Zuray, PLM Technical Principal of The Boeing Company

Stijn Hoorens, Senior analyst, RAND Europe

Stuart McKee, National Technology Officer, U.S. Public Sector, Microsoft Corporation Strategic Advisory Council Members

Victoria Reich, Director, LOCKSS Program

The testimony from these speakers addressed five general questions, touching on a variety of aspects of digital preservation and economic sustainability (Box 3.3). A summary of the testimony for each question is given below the text box.

BOX 3.3 **Speaker questions**

1. **What is the nature of the materials being preserved?**
   E.g., source(s), content, volume, format, copyright restrictions, frequency of use, etc.
2. **Who are the stakeholders for these materials?**
   Describe the users/communities that benefit from the preservation of the preserved materials.
3. **What is the "value proposition" for this preservation effort?**
   Why are stakeholders interested in the long-term preservation of the materials? What are the anticipated future uses of the materials? Is the "value proposition" perpetual, or does it expire within a finite time frame?
4. **What are the key features of long-term preservation for these materials?**
   In other words, what are the elements of "successful preservation" for these materials? E.g., bit preservation only? Format migration? Rich description/documentation with metadata? What we are looking for here are the key features of a preservation service that would meet the preservation requirements of the stakeholder community.
5. **What are the "economic aspects" of digital preservation?**
   Given the preservation requirements for the materials, what are the key cost categories associated with a preservation strategy aimed at meeting these requirements?

3.2.1 The Nature of the Materials

The quantity of digital information is growing faster than our ability to store it (see Figure 1.2, Chapter 1). It is estimated that 281 exabytes (281 quintillion bytes) of digital information were produced in 2007. For context, that would be equivalent to the amount of information stored on over 1.1 billion PCs, each equipped with a standard 250 GB hard drive. Of course not all of that information is worth preserving; a lot of it exists only temporarily. In 2007, IDC estimated that approximately 20 percent of that data would be "preservation intensive" that is, is anticipated to be valuable enough to store for 10 years or more. By 2011, IDC estimates that percentage will rise to about a third. The distinction drawn between temporary and preservable information calls attention to the importance of processes of appraisal and selection, which are tied to the mission of the collecting organization and the framework in which it functions.

Speakers represented a variety of organizations and the materials covered were thus also diverse, ranging from highly detailed scientific gene structures to aircraft product definition data, electronic scholarly literature, and the gamut of materials presented on the Web. A number of organizations accepted material in a variety of formats, which in turn tended to increase costs at the time of ingest. In addition, data sources varying from public to semi-private to private were represented as well. For some organizations, issues of privacy and confidentiality, as well as ownership and copyright, were cited as obstacles to undertaking preservation activities. With respect to scientific data, copyright was less of an issue, but the desire to ensure open access to research materials represented a perceived constraint on the types of economic solutions available to that community; for instance, charging for access was often considered an undesirable option.

The size of datasets discussed by the speakers ranged from about sixty gigabytes to around four petabytes. Commercial entities often manage collections of data at significant scale, in comparison to scientific or cultural organizations. Organizations managing relatively large data collections have become increasingly conscious of the impact of power requirements on the costs of preservation; this dovetails with similar findings from the literature review. Availability of appropriately skilled human resources was also mentioned as a concern, especially in regard to rapidly growing data collections.

3.2.2 Who are the Stakeholders

All of the speakers explicitly noted that ultimately, society as a whole (both present and future) were stakeholders in the preservation of, and long-term access to, their organizations' digital information. In addition, more narrowly-defined stakeholder communities included academia, consumers of government research, federal data producers, public and private research sponsors, and individual university and research communities, as well as students in both pre-collegiate and university settings, teachers, television producers, journalists, federal, state and local government agencies, librarians, archivists, and documentarians.

3.2.3 What is the Value Proposition

Value propositions varied depending on the specific stakeholders that organizations represented. For example, speakers representing commercial entities cited ease of use, time savings, and reduced costs to consumers as primary benefits. It was also noted that in many cases, digital preservation activities undertaken by corporations are driven by risk management considerations – having to do with regulatory compliance, legal protection, and so on – rather than the prospect of future access and use (Figure 3.1).

A scientific repository representative noted that they consider their primary value in terms of opportunity cost: what would it cost scientists in terms of time and effort to re-construct this data on their own rather than to have it available to them freely? In this case, many scientists can save upwards of *two years* of data collection and analysis by having the contents of the repository available. One representative of an organization mandated to preserve their materials noted that their primary value proposition was *compliance*, although they recognized that time savings for users was also key.

In most cases, speakers pointed to a single crucial insight: *their value propositions are always painted in terms of access and use* (rather than solely in terms of preservation). As one speaker stated, "They [data users] never ask us how much they're preserving; they never ask what's being preserved. They only ask what they are downloading and what it would cost to download by the piece."

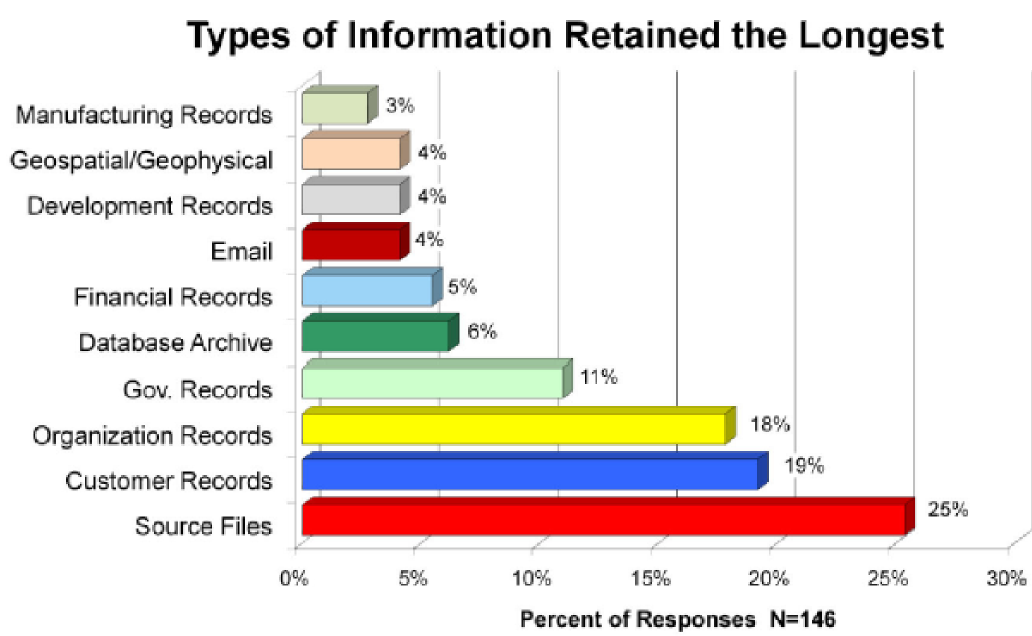

FIGURE 3.1: **Enduring Value**

The Storage Networking Industry Association (SNIA) survey asked respondents to identify the kinds of information retained the longest within their organizations. "Source files" refer to original files or information objects (e.g., documents, spreadsheets, etc.). Retention policies depend on a variety of organization-specific legal, compliance, and business considerations.

Source: P. Mojica October 2008. Used with permission.

3.2.4 The Nature of "Successful" Preservation

Most speakers focused on the process-oriented steps that needed to be instituted or refined by their organizations in order to ensure successful preservation (Figure 3.2). They discussed what specific activities would need to be in place or what features would need to exist, such as increased search capabilities, guarantees of perpetual access to ex-license holders, open access after a trigger event, redundancy of data, continuous testing of plausible risk scenarios, and widespread institution of consumer Service Level Agreements. In contrast, there was much uncertainty in regard to the *economic* steps needed to ensure sustainability of their preservation activities, other than the need to tie their value proposition to access. It was also clear that institutional cultures play a significant role in shaping preservation decision-making, especially in regard to preservation motivations and objectives.

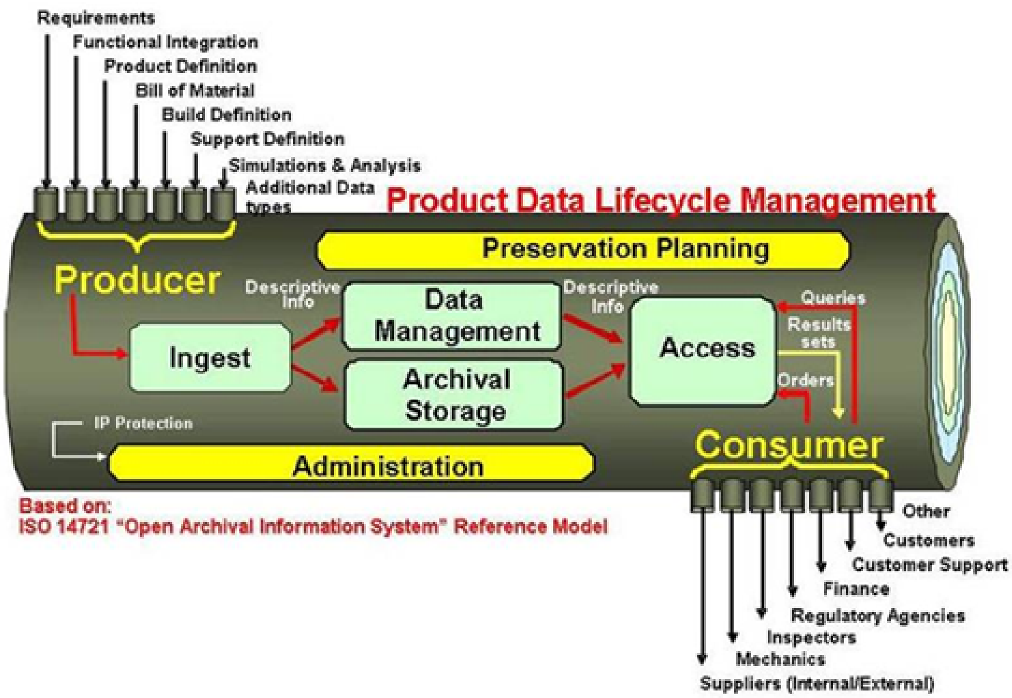

FIGURE 3.2: **An Example of an OAIS-Based Preservation Process**
Long Term Archival & Retrieval (LOTAR) of digital product & technical data represents a single set of harmonized standards that addresses the storage, retention and retrieval of 3D Product Definition Data across the entire Aerospace Industry. The Team was chartered in December 2002 and was Co-chaired by Rick Zuray, from Boeing, who presented testimony to the task force. and Jean-Yves Delaunay, from Airbus. The International team meets five times a year and has developed several parts to the base Standard which will be released under the name EN9300-Part-001 for Europe and NAS 9300-Part-001 for Americas. As this diagram shows, the team adopted the familiar OAIS framework (Reference Model ISO 14721) and contextualized in the airspace industry to meet both the types of assets (requirements, production definitions, and so on) and the needs of its user groups or consumers (customers, regulatory agencies, and so on).
Source: Rick Zuray July 2008. Used with permission.

3.2.5 Economic Aspects: Organization and Funding

Critical to success and survival of the institutions speakers represented were organization and funding. For example, ICPSR, a large-scale social science data archive, reported a diversified funding pool that includes membership charges from a stable base of educational institutions (Figure 3.3).

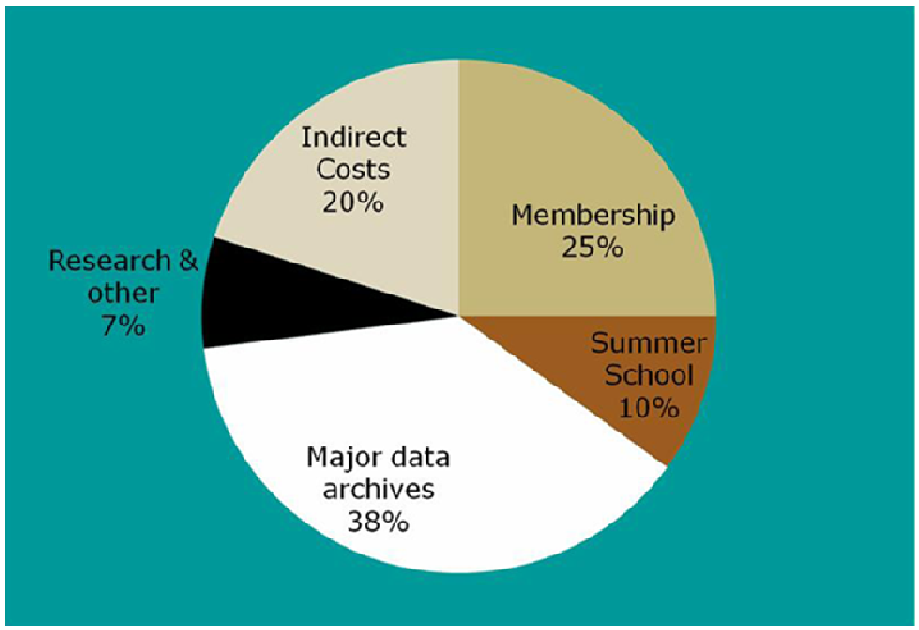

FIGURE 3.3: **ICPSR: Who Pays the Bills?**
The Inter-University Consortium for Political Science Research (ICPSR) was founded in 1962 by political scientists to share data and teach methods. It currently has more than 650 members, primarily colleges, universities and research centers and an annual budget of about $11.5 million. It holds about 3 terabytes of data in multiple formats in its distribution archive and about 6 terabytes of data in its archival storage. *Source: M. Gutmann July 2008. Used with permission.*

A large web archiving organization reported that 65%-80% of their budget is covered by fees for services, with the remaining amount a combination of grants and individual contributions. This organization charges flat rates for some services, and a flat annual subscription fee for a certain volume of data archived, with a volume cap. At the same time, however, on philosophical grounds they do not charge institutions for *access*. They do not wish to be considered a service provider but rather, a community member engaging in reciprocal, *open* sharing.

Several speakers indicated that they did not have data available regarding the costs of digital preservation *per se*, largely because their organizations are in the very early stages of understanding exactly what digital preservation entails for them. Often, the preservation activities are tied to other production or access-related activities and determining costs specific to preservation would require a largely manual endeavor of parsing the activities from budgets that are not structured to highlight preservation as a separate activity. Likewise, because voluntary labor is used by several of the organizations, assessing the true costs in absence of fee schedules would require an intensive analysis that would be difficult and time-consuming to undertake.

Several speakers (e.g., from Portico and LOCKSS) represented activities in which publishers and preservation entities are working together to address the preservation of certain classes of material, such as the e-journal literature. In these cases, there is a tension between the long-term preservation goals of the activity, and publishers' economic interest in current access. Often, this tension is reconciled through a preservation model in which preserved content is released to the public only upon occurrence of a pre-defined "trigger event" – for example, if the publisher should suddenly become unable to provide access to the materials through its own services.

A theme that ran through all of the speaker presentations, and which was openly voiced several times, was the hope that the Task Force itself would provide some answers to questions about economic sustainability. Whereas in many cases technical and even business-process oriented solutions had been developed, there was much uncertainty on the part of the speakers about how to ensure that the required funding would remain available over the long-term. There was a strong sentiment that economic sustainability will require an emphasis on cost minimization, and a strong articulation of value propositions for funding decision-makers. Moreover, the value proposition should emphasize access and use; in more than one case, experience indicated that expressing value solely in terms of preservation was relatively ineffectual in attracting attention and resources. Speakers agreed that how to formulate persuasive arguments about the need to invest in long-term preservation, access, and use is critical for them and remains a challenge.

3.3 Observations and Preliminary Lessons Learned

The literature review and speaker testimony, as well as internal discussion within the Task Force, yielded a number of important insights relating to the economics of digital preservation. Taken together, they form a useful backdrop to any discussion of sustainable economic models for digital preservation – a "reality check," in a sense, where theoretical models must observe the practical boundaries imposed by real-world implementation. It is easy to talk about the digital preservation imperative in the abstract, and the need to secure adequate funding to meet preservation goals. Translating theory into practice, however, requires certain economic elements in the surrounding environment to be taken into account. We offer the following observations about experience and findings so far as a starting point for discussion.

(1) It is easier to "sell" outcomes than processes.

The definition of economic sustainability discussed earlier emphasizes the importance of articulating the value or benefits associated with long-term preservation. It is important to keep in mind, however, that in most circumstances the true value of digital preservation is not preservation *per se*. Digital preservation is a process – a means by which outcomes are achieved. We can recognize the need to invest in reliable repository architectures, detailed descriptive and administrative metadata, and robust, distributed data stores, but ultimately, these are not the ends we are trying to achieve. *Rather, it is the future activities that preservation permits: for example, ensuring persistent access to, and re-use of, digital assets. It is the prospect of avoiding future cost: for example, by eliminating the need to re-create important data sets by repeating expensive experiments, or attempting to salvage corrupted or technologically obsolete digital objects. And it is the possibility of expanding usage of digital assets by securing the option to re-purpose them in ways as yet unanticipated.*

As discussed earlier, allocating funds for preservation is an investment in preserving, and perhaps enhancing, the value-creating capacity of digital assets. It is the specific nature of that value-creating capacity – in other words, the prospective outcomes – that should be emphasized when articulating the value or benefits of digital preservation.

In a sense, this point is more about "message" than economics. But often the message is a key impetus to economic decision-making. Consider, for example, that it is possible to have preservation without access – the so-called dark archive – but it is not possible to have access without preservation. Potential funders may be reluctant to pay for "preservation" as an end in itself, but they might see greater incentives to pay for long-term access. Our ultimate aim is not to build "secure digital repositories," but to ensure ongoing access to important digital materials. Framing the benefits from preservation in ways that emphasize outcome rather than process helps place the cost/benefit analysis underpinning digital preservation investment in its proper perspective.

(2) Avoid excessive discounting of the benefits from digital preservation.

Digital preservation is a dynamic process, unfolding over an extended period of time. The very nature of digital preservation as an economic activity suggests an inherent "future aspect" to the benefits derived from preservation investments. Like most investments, preservation can be characterized as incurring a cost now (e.g., to build a secure repository) to realize a benefit in the future (e.g., to achieve long-term access). While this characterization is generally true, we must be careful not to focus too narrowly on the expected benefits redounding to future stakeholders, at the expense of giving short shrift to the benefits realized by current stakeholders. Preservation is often cast as an investment on behalf of future generations, somehow detached from the interests or concerns of current stakeholders. In the context of digital preservation, this perception creates two problems. First, it is not accurate. The preservation activities undertaken to secure ongoing access and use of digital assets operate on a much shorter time horizon – often in the range of 5 to 10 years – than most analog or physical items. Actions taken to ensure a digital asset is available for use ten, or even five years from now secures a benefit that is realized by today's stakeholders, rather than some vague future constituency.

The second problem is that today's decision-makers are primarily and justifiably concerned with the interests of today's stakeholders – i.e., today's researchers, students, customers, business partners, and so on. Allocations of funds reflect this emphasis on current rather than future need. More generally, when we make investments, we typically do so on the assumption that we will be around to reap the return on the investment; put another way, economic decision-makers are primarily interested in investments that will yield value within their own management horizon. It is important to make clear that digital preservation, in most circumstances, falls squarely in that category of investment. Too often, digital preservation is perceived as an activity that is separable from the interests of today's stakeholders, aimed instead at the needs of future generations. But in practice, digital preservation is very much part of the day-to-day process of managing digital assets in responsible ways; it is much more about ensuring that valuable digital assets can be handed off in good condition to the next succession of managers or stewards five, ten, or fifteen years down the road than it is about taking actions to benefit generations of users a hundred years hence.

In light of this, we must avoid discounting the future benefits of digital preservation too heavily. In accounting, dollars are worth less – i.e., are discounted more – the further into the future they are received. A dollar received ten years now is worth less than a dollar received five years from now, which in turn is worth less than a dollar received today. By over-emphasizing the long-term aspects of digital preservation, the benefits from preservation are implicitly discounted in ways that distort their true distribution over time. It is important to make clear that realization of the benefits from digital preservation is a matter of *current* as well as future interest.

(3) Separating preservation costs from other costs is difficult.

The previous set of observations, summed up as a "lesson" – that it is inaccurate to cast the outcomes of preservation as a discrete set of future benefits separable from the interests of current stakeholders – finds a corollary on the "cost side" of the ledger. The costs of preservation are generally not entirely separable from the costs of "day-to-day" management of digital assets. *There is no clear dividing line between investments made to support current use, and those made to support future use.* If we invest in a secure data store that supports current access, are we not also making an investment in the ongoing availability of the digital assets residing in the store? If we perform fixity checks on digital objects to confirm that their bit streams remain uncorrupted, we are ensuring that the objects persist in an authentic form over time, but are we not also ensuring that current access and use of the objects is uncompromised? In short, while it seems intuitive to talk in the abstract of "making things available now" and "making things available in the future" as two distinct processes, in practice, the distinction is much harder to articulate.

If digital preservation is not an entirely separable process from other aspects of digital asset management, it follows then that the decision to allocate resources to digital preservation is not entirely distinct from the decision to allocate resources to the overall digital asset management process (represented, for example, by the LIFE^2 model). This suggests that in many circumstances, it may be difficult to treat digital preservation as a separate economic activity with a clearly circumscribed set of costs and benefits, which in turn can be weighed in the context of a distinct economic decision.

Instead, digital preservation might be better cast as an "incremental decision" built on top of the larger question of digital asset management. So for example, the question might not be "should I allocate resources to digital preservation?" but instead, "given what I am doing now to secure the ongoing availability and use of digital materials, what are the additional costs and benefits of either a) extending the time horizon over which I can be reasonably confident the materials will persist in their current condition, or b) reduce the likelihood that the materials will cease to be available or usable within the current time horizon?" In either case, *the decision on whether or not to invest in digital preservation is not a completely separate decision, but more akin to a parameter within a broader decision.*

(4) Diversity of funding streams is important for sustainable digital preservation.

It is a common tenet of systems development that "single points of failure" should be avoided – that is, elements of the system, which if they cease to function, bring the entire system to a halt. Such is the case with economically sustainable digital preservation: if a digital preservation activity relies on a single source for most or all of its funding, elimination of that source could end the activity forthwith. ICPSR, which has provided archiving and preservation services to the social and political science research communities (including to government agencies) since the 1960s, maintains multiple revenue streams (Figure 3.3). A counterexample is the experience of the U.K.'s Arts and Humanities Data Service (AHDS), which included digital preservation activities as part of its service portfolio. In March 2007, AHDS's primary funder, the Arts and Humanities Research Council, announced it would cease funding AHDS after March 2008. The AHDS ended most of its activities in 2008 after a dozen years of operation.[12]

Another example of the perils of relying on a single funding source is the common practice of funding activities on a project-by-project basis through "soft money" or grants. In these circumstances, the horizon of sustainability extends only as far as the duration of the grant award, and all too often, no arrangements are made to secure sufficient resources to carry the activities forward beyond the grant's expiry. These and other examples reinforce the conclusion that, where possible, diversified streams of funding should be sought to enhance the economic sustainability of digital preservation activities.

Diversification of funding can be manifested in several ways. First, an activity can implement several different funding models to cover its resource requirements. For example, an activity might seek philanthropic donations or awards as one source of funding, while simultaneously offering a portfolio of fee-based data services that yield a sufficient margin to subsidize a substantial fraction of the costs of long-term digital preservation. In this case, elimination of either the philanthropic or market-based component of the activity's overall funding stream would not necessarily spell the end

[12] An example of an exception proving the rule: one part of the AHDS, the Archaeology Data Service, did have access to multiple streams of funding and survived.

of the activity; it could proceed, albeit perhaps on a more limited scale, on the basis of the surviving funding model. Another way that funding streams can be diversified is by diffusing them over multiple sources within a single funding model. In this way, the activity could spread its funding streams over multiple funding agencies, so that the elimination of funding from any single agency would not substantially impact the capacity of the activity to continue its operations. In either case, a strategy of diversifying funding streams is no different than the familiar practice of diversifying a portfolio of investments: rather than investing everything in an "all or nothing" gamble on a single funding source, the goal is to create a stable flow of resources over time from multiple sources which collectively smooth out the peaks and valleys of individual funding opportunities.

(5) Non-monetary incentives are important.

In an address to the World Economic Forum in January 2008, Microsoft's Bill Gates outlined a vision of something he termed *creative capitalism*: "an approach where governments, businesses, and non-profits work together to stretch the reach of market forces so that more people can make a profit, or gain recognition, doing work that eases the world's inequities." The phrase "gain recognition" is key to Gates' vision: "… [P]rofits are not always possible when business tries to serve the very poor. In such cases, there needs to be another market-based incentive – and that incentive is recognition. Recognition enhances a company's reputation and appeals to customers; above all, it attracts good people to the organization. As such, recognition triggers a market-based award for good behavior. In markets where profits are not possible, recognition is a proxy; where profits are possible, recognition is an added incentive" (Gates, 2008). Gates' remarks underscore an idea that is important for sustainable digital preservation activities: non-monetary incentives, such as recognition or reputation-enhancement, are potentially important motivators with which to engage the interest and support of prospective funders.

For many organizations, such as libraries, archives, and museums, non-monetary incentives are a familiar refrain; these and other "mission-driven" institutions are accustomed to non-monetary expressions of value or benefit as incentives to action. But it is also useful to examine the scope for employing similar incentives to spur action on the part of profit-driven organizations; as Gates suggests, incentives based on recognition and reputation enhancement may be of particular importance in this context. Preservation bestows general societal benefits to research, learning, and culture; is there a way to engage private enterprise in the production of these benefits, by articulating a range of non-monetary incentives for them to do so?

Of course, mission-driven organizations are accustomed to asking for sponsorships and donations from private enterprise; the novel element here might be asking for such support on the basis of the digital preservation imperative specifically – for example, is it possible to engage the interest of a corporation to "sponsor" the long-term preservation of a particular set of digital materials? Gates notes that "if you give people a chance to associate themselves with a cause they care about – they will pay more, and that premium can make an impact." It remains for advocates of responsible digital preservation to make the case that securing the long-term persistence of digital materials is a cause that people should care about, and to devise innovative ways for mutually beneficial forms of corporate participation to occur. In an age where terms like "corporate philanthropy" and "socially responsible corporations" are entering common usage, there is an opening to solicit private enterprise to directly support investment in digital preservation activities as an important contribution toward the public good.

(6) Consider the full range of options when selecting an economic model to support digital preservation.

It is important to bear in mind that when selecting an economic model to sustain an economic activity, it is the characteristics of the activity itself that should chiefly guide the choice, not the attributes of the overarching organization within which the activity is embedded. The reason for this is two-fold. First, different digital preservation activities within the same organization might have dramatically different features, and therefore, lend themselves to different mechanisms for channeling economic resources in a sustained way over time. Second, just because an organization is of a particular type – say, a cultural heritage institution – does not mean that it is necessarily limited to choosing from a particular class of economic models normally associated with organizations of that kind. There are a variety of mechanisms for achieving economic sustainability, many of them well-tested in other domains. Organizations tasked with preserving digital assets should consider the full range of options when selecting an economic model to support digital preservation.

This last point may seem obvious, but it can be difficult to achieve in practice. For example, some organizations or activities are accustomed to offering services to users free of charge, in keeping with a public-spirited mission normally funded through donations, awards, in-kind grants, and so on. In such a culture, shifting some services to a for-fee model may be difficult to countenance, and indeed, may be perceived by some as an infringement on the stated organizational mission. Such concerns should not be dismissed lightly: care must be taken that a particular funding model does not collide with the perceived mission of the organization, and in doing so, jeopardize other important sources of funding.

At the same, however, opportunities to generate revenue from digital preservation services, either on a cost recovery basis, or even to produce a reasonable margin, should not be disregarded either. If a compelling social value can be expressed in monetary terms, and if a funding mechanism can be devised with which those who share in this value can be identified and charged a reasonable fee to support its ongoing provision, then a viable model for achieving economic sustainability may exist, and should be explored. Just as profit-seeking organizations should consider opportunities to respond to non-monetary incentives, mission-driven organizations should consider judicious use of fee-based models to support digital preservation.

There is still much to be learned about the economic aspects of digital preservation. The Task Force's findings to date, summarized in this report, are but an initial step toward filling in the empty space that currently surrounds the issue of economically sustainable digital preservation. Yet even at this early stage, it is possible to step back and offer some perspective on what we do know about the economics of digital preservation, and to connect it to a larger context. We turn to this in Chapter 4.

BOX 3.4 **Preliminary lessons learned (summary)**

1. It is easier to "sell" outcomes than processes.
2. Avoid excessive discounting of the benefits from digital preservation
3. Separating preservation costs from other costs is difficult
4. Diversity of funding streams is important for sustainable digital preservation
5. Non-monetary incentives are important.
6. Consider the full range of options when selecting an economic model to support digital preservation.

Chapter

4

Understanding Economic Sustainability: Observations, Gaps and Opportunities

One can describe a landscape in many different words and sentences, but one would not normally cut up a picture of a landscape and rearrange it in different patterns in order to describe it in different ways. Because a photograph is not composed of discrete units strung out in a linear row of meaningful pieces, we do not understand it by looking at one element after another in a set sequence. The photograph is understood in one act of seeing; it is perceived in a gestalt.

Joshua Meyrowitz
"The Blurring of Public and Private Behaviors," No Sense of Place: The Impact of Electronic Media on Social Behavior, Oxford University Press (1985).
http://www.bartleby.com/66/69/39469.html.

This chapter synthesizes the discussion of the Blue Ribbon Task Force in 2008 and offers some observations and perspectives about what is known, what is currently unknown, and where opportunities for research lie in regard to economically sustainable digital preservation.

4.1 The Current landscape

Probing the notion of economic sustainability in the context of long-term management and preservation of digital material demonstrated that the concept has several dimensions, each with its own economic parameters. Technology, social behaviors and expectations, organizational context, and public policies interleave in interesting and complex ways. As a result, an early and necessary step was to understand and disentangle the various elements so that the economic thread could be seen clearly and become subject to analysis. This process and its results have been laid out in Chapter 1.

Preservation is about present and future use of materials created in the past where the past may be as recent as yesterday and as distant as centuries. From an economic perspective, preservation is about actions today that will have value in the future, which invokes the notion of investment. As explained in the first chapter, the way that the economic considerations are defined is embedded in a larger context that is divided, for purposes of explanation, into a *static* model, where the size of the economic pie is fixed and costs of preservation are for the most part reallocated from other activities, and a *dynamic* model, in which the pie is growing and the costs of preservation may be at least partially accommodated by growth. Both cases require forecasts about the future based on information from the past, where prior actions and their consequences may be helpful but are not necessarily predictive, particularly given the probability of change and the associated uncertainties. The tension between prior experience and its ability to predict future outcomes is inherent in any type of forecasting. Chapter 2 amplifies the conceptual framework outlined in Chapter 1 by explaining the notion of an "economic model" in more detail.

Perhaps not surprisingly, it proved easier to examine the tension between investing in the future and assessing past performance by starting with costs, which can be derived from prior experience. Indeed, a number of studies have grappled with modeling costs over the past 15 years. These are briefly described in Chapter 3 and more fully discussed in a companion paper that is posted to the Task Force website brtf.sdsc.edu.

Of these, perhaps the LIFE project best represents the current state of the art. Spread over two phases, which took place between 2005 and 2008, this joint project by the British Library and University College, London with funding from JISC and LIBER, has made a substantial contribution to understanding the costs of long-term management and preservation of resources. It models the costs through a life cycle approach (see Box 3.1 in Chapter 3) and offers institutional users a downloadable template that allows them to do the calculations based on local experience and anticipated needs. Validated by an independent economic review, the model has been adopted by the Royal Danish Library, State Archives and the State and University Library, Denmark as well as by the partners in the project itself. The research leading up to articulation of the model, as well as the process of review and evaluation, constitute a body of work in their own right and has been made fully available to the public.

Overall, the various studies (see Chapter 3, Table 3.1) illustrate a progression in thinking and sophistication that reveals increasing awareness of the complexities of managing digital libraries and similar teaching, collecting and preserving institutions as well as the challenges specific to preserving digital materials. A simple example serves to illustrate some of the ambiguities: How are total labor costs determined in institutions that rely heavily on volunteers and where mobilizing those volunteers performs an important societal mission? Are the costs really low? Or does an apparently low cost reflect uncounted hours? If a proxy is to be developed to price those hours, what are the sources for determining that proxy and is that metric legitimate across all institutions? That is, if we were to compare total labor costs, including volunteers, across libraries, are we really looking at the same value? Clearly, these issues can be resolved, but they do need to be resolved and resolved in a way that is satisfactory to the concerned groups.

While the LIFE[2] model stands out for its sophistication, depth and testing, consideration of the work that preceded it highlights the issue of time. Historians, in particular, are acutely aware of the implications of the passage of time. It is inherent in the notion of "path dependence," which, as noted in Chapter 2, is fundamental to modeling digital preservation. Simply looking at the sequence of studies shows not only greater awareness of more dimensions to the challenge of undertaking digital preservation activities but also suggests the extent to which the specific time – the roughly three decades from 1980 to the present – may have influenced our perceptions and thinking. In this period, personal computing, the Internet, and the World Wide Web emerged, along with a much broader adoption of information technology in personal and organizational settings. The ways in which content was created and distributed changed radically and repeatedly.

This has several consequences. The emergence of modern information technologies as a tool with which to manage and retain digital materials has created rapid changes in within libraries, museums, repositories and cultural heritage institutions. As information technologies themselves have matured, and institutions have integrated their capabilities into professional practice, the sense of upheaval has largely subsided. Mature technologies have other benefits: costs become more predictable and often decline, including the costs of acquiring, installing and using relevant technological systems and tools, as well as the costs of learning associated with new technologies. In essence, organizations and even individuals are learning to manage technology evolution. Introducing new equipment and software upgrades into the mix then becomes a routine part of an institution's cycle of maintenance.

Consider two examples. In 1980, discussions about bit stream preservation focused substantively on the physical properties of various kinds of media and their shelf-life; now we take continual evolution of both media and the technology that reads and writes this media for granted, and migrate bits from one generation of media to the next in a routine fashion. In the 1980s and 1990s, facing a very dynamic, distributed and rapidly evolving industry, vast numbers of word processing and document formats were put forward, many of which were frequently orphaned when software developers abandoned specific programs. During this time, the preservation concern began to shift from the ability to carry bits into the future to the question of being able to interpret these bits. We now have an enormous base of digital documents, a software industry in this area with a small number of dominant players who are very concerned with migration paths, backwards compatibility and similar questions, and increasing use of open standards and similar developments.

Indeed, where digital preservation once seemed alien, there is already evidence of its incorporation into commercial products, evidenced, for example, by the standards work surrounding the PDF/Archives (ISO 19005-2) and the Portable Document Format (ISO 32000-1), which is described as a standard that, among other attributes, encourages "the propagation and dissemination of a common technology that cuts across systems and is designed for long-term survival" (Quarterly, 2008). Of course, managers of digital preservation activities will constantly face new formats and technologies connected with new genres of digital material that emerge – think of virtual worlds, massively multiplayer games, simulations, CAD models and the like as highly immature and rapidly evolving examples. But preservation managers are likely to be more comfortable in such situations, and the systems with which they work will likely be more sophisticated and responsive to their needs.

In addition to the distortions in perception that inevitably arise from experiences derived from a period of rapid change is the emergence of issues that take on greater resonance or changes in public mindset and affect the ways in which materials are appraised, selected and made available. The current "green" movement is a case in point. After a period of enthusiasm in the 1970s, environmental issues receded from public view only to experience a resurgence of interest in the last year or so. As a result, there is greater attention to energy costs, including energy required to operate IT systems, at a time when those costs are rising generally and the implications of the heating and cooling of large scale storage systems are becoming apparent. Earlier studies, dating from the 1990s, not surprisingly, did not extract the costs of heating and cooling, and storage architectures were generally thought to be sufficient. This view is changing at the societal level as well as at the specific levels associated with IT facilities. It is likely that the technology will evolve to become more energy efficient as well as more capable.[13] This example suffices both to demonstrate a gap in prior work that has become evident with the passage of time, and changes in attitude. It also points to an aspect of the economic environment for digital preservation in which future costs are unknown and probably cannot be easily predicted based on current information.

Finally, studies to date tend to consider or model costs in terms of the arrival of material at the institution which has taken on or been charged with the mandate for preservation in perpetuity. Thus, the LIFE^2 model begins with "acquisition," which may be acquisition from the creator, by gift, as a legal deposit, or by purchase. This term could be easily expanded to encompass acquiring material from another custodial institution -- for example, the acquisition of the "morgue" of a newspaper that had preserved its electronic editions or the deposit of records by an institution subject to public mandates requiring preservation for a specified period of time after the mandatory period had elapsed. But the costs may be substantially affected by existing preservation practices.

Work by Beagrie, Chruszcz and Lavoie (2008) and funded by JISC suggests that accessioning and ingest costs are high relative to ongoing costs over the long-term.

[13] See, for example, Sun Microsystems' work on Project Blackbox, http://www.sun.com/emrkt/blackbox/story.jsp, and the more recent Sun Modular Datacenter S20, http://searchdatacenter.techtarget.com/news/article/0,289142,sid80_gci1296642,00.html; this issue is also discussed in slightly more detail in Chapter 3.

This preliminary finding (the authors note that their work is early and suggestive rather than exhaustive and definitive.) points to the importance of effective management strategies early in the life cycle of information, confirming archivists' long held belief, based on their experience, that "preservation begins at creation." Not all material acquired may have been created with preservation in mind, however; the observation is most relevant for organizational settings where there is a requirement and a commitment to maintaining a record for the long-term. Further, acquisition and ingest tends to have a high "setup" cost, and particularly in the early days of digital preservation, every acquisition seems unique, requiring specialized lengthy analysis and processing strategies. Acquisition of at least partially processed material will become more routine and, most likely, more standardized over time. Under a scenario in which digital material may be held by a sequence of custodial institutions, the total cost of ownership (a concept explained in Chapter 1) over the entire life of the material may perhaps be seen at the system level but borne differentially by the separate entities that manage the data.

Adding to the complexity are different institutional missions and user communities. An institution that maintains electronic records primarily for purposes of regulatory compliance may not be much concerned with format migrations or other curatorial activities that make the materials easier to access, manipulate and reuse. In such situations, managers may be deeply concerned with the efficient storage of the information as a hedge against the risk of a lawsuit; the motivation is minimizing risk, not the utility of the content, and the costs of a lawsuit are such that corporations will go to substantial lengths to preserve their data whether or not they will actually be required to provide access to the information.

Practices are likely to differ when preservation is undertaken with the intention of providing future, reliable access to the content. Where up-to-date and dynamic data (e.g. sensor readings to reflect present state) is central to the ongoing conduct of research, commerce or other activities, users ask the question, *"Is it current?"* and reliability has one set of tests. In the case of ICPSR, where the contents of collections are individually static and where the boundaries of those collections are important for establishing the integrity of the materials, the users might ask, *"Is it authentic?"* In other words, is this collection that represents the 1960s U.S. federal census the correct one and does not include material from, say, the 1970 U.S. federal census? Both types of users require reliable and accurate data, but the measures by which they understand these terms and the boundary conditions that obtain are substantially different, as are the characteristics and economics of sustaining the different types of resources.

As a practical matter, the condition of materials as they are transferred from one custodial agency to another may profoundly affect the costs over the lifetime of the material as well as for each entity in this food chain of collecting and preserving institutions. Some indication of the magnitude of such transfer costs is suggested by Portico, which maintains a detailed set of steps for acquiring and managing the e-journal collections received from various publishers. The standards process may be of some help in some of the aspects of this set of challenges by changing the circumstances under which materials are managed, imposing a threshold of uniform treatment and good practice, and aligning short, medium and long-term preservation goals. Such measures will enable institutions to anticipate the approximate condition of at least some classes of materials, thus reducing ambiguity and providing a measure of stability. Nevertheless, the larger question of hand-offs and their costs remains a topic on which relatively little information has been identified.

4.2 Some of the unknowns

As the preceding remarks suggest, much is still unknown and in some cases has gone largely unmentioned. To some degree, and as this section will discuss further, these ambiguities are inherent in any system that seeks to manage resources for the indefinite future, as the discussion of the notion of investment in Chapter 1 explains. Still, it is possible to suggest some of the places where we can anticipate changes that affect the way that libraries, archives and museums must plan for the management and economic sustainability of their digital collections and that has not been captured in the work thus far.

Survivability

First, and most concretely, it is surprising that more attention has not been paid to the economic aspects of threat models to survivability, especially in the wake of both 9/11 and Katrina, and the associated risks or exposures to those threats. Indeed, the 2003 report *Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for Initial Development*, by the Computer Science and Telecommunications Board of the National Research Council specifically recommended, "The risks of the various possible causes of data loss—such as malicious acts, natural disasters, software bugs, human error, and hardware failures—should be assessed and used to make informed engineering cost-benefit trade-offs" (Committee on Digital Archiving and the National Archives and Records Administration et al., 2005a, p. 83). A subsequent letter report in 2003 encouraged the US National Archives and Records Administration (NARA) to "specify an explicit threat model be developed early in the ERA's life cycle," noting that a draft specification for follow-on work to the 2003 study "makes occasional mention of measures that might help in avert threats … but it includes no overall requirement that the system be capable of surviving an attack or incident" (Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies, 2003, p. 88). In a second report, NARA briefly addressed the importance of threat modeling and threat countering in the context of a general discussion of record integrity and authenticity (Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies, 2005b).

FIGURE 4.1: **Threats to Long-Term Persistence**
A disaster management plan needs to consider the possible threats to all types of information, including that which is digitally rendered.
Source: State of Iowa. 2008. Cedar Rapids Public Library.
http://www.statelibraryofiowa.org/archive/2008/aug08/flood/crpl3/

On the one hand, it is well understood that storage repositories should be backed up routinely, replicated in geographically distinct locations, and synchronized regularly, and these costs have, in some cases, been accounted for. For example, the $LIFE^2$ Model does allow for back-up, replication and synchronization, and a partnership involving NARA, the University of Maryland and the San Diego Supercomputer Center proposed a model for a persistent archive that addressed risk management and disaster recovery as well as technology evolution (Moore, JaJa, & Chadduck, n.d.). However, there has been no analysis of the economic issues addressing, for example, what the optimum number of replication facilities would be when balanced against the probable occurrence of various kinds of natural or man-made disasters. Basic geographic dispersion of data may well not protect against events such as electromagnetic pulse.

Much data loss is due to human error; a very large number of attacks are carried out by insiders. And archives and libraries have often been targets in overt or covert wars. Consequently, there is every reason to expect that this will be the case with digital archives of key cultural materials. So the threats are real at the level of the trans-institutional system but highly unpredictable for any given element in that system. Any model for sustainability and for the costs associated with it must take such unpredictable considerations into account, if only to allow for contingency budgeting.

Recovery

Related to the question of failure is the question of recovery. Again, substantial work could be done through case studies of massive failures that might provide some parameters to be used for the suggested contingency budgeting. In the case of Katrina, for example, a close analysis might be undertaken to parse the steps and associated costs of reinstantiating the massive systems that were wiped out in hospitals, banks, and so on. While such efforts may not have been technically considered preservation, they are instances of recovery in the wake of disaster and might contribute to putting dimension around the vague problem of recovery and to acknowledging the importance of contingency planning as part of managing digital assets over the long-term. In the near term and as a purely practical matter, *organizations should have contingency budgets and provision for recovery*. Also, we need to abandon the belief that recovery is a routine process that leads to perfect reconstitution; in real world cases, there will often be extensive damage assessment, attempts to reconstruct or re-verify data, and sometimes recovered data will be of questionable quality, but may be used anyway because it is all that is left.

Recovery clearly involves more than simply buying equipment, re-installing programs, and copying data onto the correct place, all items that may show up in the accounting systems as investments. In some instances, having to re-build occasions revisiting existing workflows and a different organization emerges as a result of the process. This is one reason why an environmental disaster like an oil spill can look like economic growth. New resources and investments may be brought to bear in a local economy where none previously may have existed so the size of the local economy looks like it has grown. Implementation of new technology is more than simply exchanging one system for another. Word processing applications, spreadsheets, and e-mail systems are well known examples of the ways that new end user technologies altered workflows, and the learning curve that may be specifically associated with a new technology can have ramifications well beyond the specific purpose for which it may have been intended, thus affecting the overall health and efficiency of the organization. Such costs and benefits have been traditionally difficult to capture and have not been reflected in the material identified to date or in the testimony before the Task Force by experts who are already managing digital collections although, clearly, all of them routinely migrate their collections to new hardware and software environments.

End Users and Institutions

It is likely that similar changes should be expected as tools to enable preservation are developed and become part of end users' workflows, much as has been envisioned in the development of the PDF standard previously cited. Indeed, *Nature* has already called for such awareness in its 4 September 2008 editorial in a special issue on data, noting, "Researchers need to be obligated to document and manage their data with as much professionalism as they devote to their experience." The editorial calls for greater support for such endeavors, noting that the number of publicly funded databases with preservation responsibilities is relatively small: "Universities and funding agencies need to provide and support curation facilities, tools and training" (Nature, 4

September 2008, p. 455). Again, modeling the system from workbench to archival repository and its economic implications, including the value of the data that is part of the flow, have not been addressed. Reducing complexity by widespread adoption of standards might be one way in which costs might be rationalized and even reduced. But the first step requires a change in behavior and then understanding and modeling that behavior so that the economic dimensions can be understood.

Institutions are embedded in society and culture and the way that policies are formulated, understood and implemented reflect the tenor of the milieu. These, too, have implicit risks that affect the ability of institutions to manage their collections. For example, the attitude toward energy and conservation and the implications for IT systems, including those that support long-term management and preservation of digital information, has already been mentioned. Societal trends that affect appraisal, selection, and access can be anticipated but not quantified.

Privacy

One obvious shift concerns privacy and its tension with access. Changes in the understanding of privacy are immense as recently illustrated in an issue of *Scientific American* (Vol. 299, No. 3, September 2008) devoted to the topic and clearly will have ramifications for the way video and social networking sites are collected, archived, preserved, and ultimately made available. There exists a fundamental tension between protection of personal privacy and personally identifiable information and certain kinds of epidemiological research, particularly as search technology advances and it becomes possible to identify individuals from pools of data in which identifying information has supposedly been removed – and indeed was removed given the state of the art at the time the data were processed. There exists a cluster of competing, legitimate concerns, namely: research that requires contextualizing highly granular information in social groups, the desire to protect individuals' confidential information, advances in technology, and the laws and regulations attempting to govern those relationships. Clear guidance is lacking and perception and societal values will change, resulting in an inherently unstable equilibrium that inevitably leaves the management of the collecting agency vulnerable, as the museums that have custody for anthropological collections are already discovering.

The degree to which a custodial institution may be affected by these societal shifts is likely to vary depending on mission, regulatory context and the nature of the material. The larger point is that digital collections are targets for a wide range of legal and public relations attacks and are likely to become victims of physical attacks, whether from natural disasters, random electronic surges, or outright malice. The frequency of these challenges and the costs of dealing with them can be unpredictable but very large. The outcomes of litigation are also unpredictable and potentially life-threatening for otherwise sustainable preservation strategies. Several of the speakers from whom the Task Force heard acknowledged that their revenue streams may be precarious. The risk is that the unpredictable elements – threats from natural disaster, changes in perceptions of value, accidents and malice – will tip an institution from viable to failing.

Organization

Much of the cost modeling that has been done for preservation has focused on trying to quantify the relatively predictable cost factors. Substantial progress has been made and our ability to parse the challenge has become more refined. Serious problems remain, particularly when we think about preservation across long periods of time, notably, the unpredictable but inevitable: the "black swans," to use Taleb's term (2007) – the very low probability events, the high-cost legal challenges, the threats that were not considered in the threat model but came to pass anyway. Some of this can be handled by contingency planning and the development of contingency budgets and strategic reserves; other parts may lend themselves to insurance approaches. Yet another question is the choice of appropriate scale of preservation activities: they need to be big enough to have flexibility to respond to challenges but not so large that their failure is catastrophic.

But it seems clear that there is no substitute for a flexible, committed organization dedicated to preserving a corpus of material. This organization must be able to make choices and devise strategies to deal with unexpected problems of all types. If necessary, it can conduct triage and make compromises. Modeling predictable lifecycle costs and arranging funding streams to support these costs is necessary but is clearly not going to be sufficient. As implied in the definition of economic sustainability that guides this study (Box 1.1), the design of appropriate organizations, the economic implications of the organization as more than the sum of a series of flows, and the organization's placement in legal, public policy and cultural settings are clearly going to be key to achieving long-term sustainability of digital collections.

FIGURE 4.2: **Raising the Barn**
Ensuring economic sustainability for digital preservation activities requires organization and commitment to shared goals.
Source: T. Abate 2007. http://minimediaguy.org/2007/02/21/the-amish-approach-to-new-media/

4.3 Looking Ahead

This Interim Report is the first of two that the Task Force will publish. Its purpose is to frame the general contours of economically sustainable digital preservation as a topic of both practical importance and intellectual interest. To this end, we have explored and synthesized past studies and analyses pertaining to the economics of digital preservation, the perspectives of domain leaders and subject experts in the field, and discussions within the Task Force. The findings of this report will serve as a basis for the Task Force's work over the coming year; it should also contribute to the broader discussion of economic issues regarding digital preservation and access.

The Task Force's Final Report, to be published at the end of 2009, will identify and analyze a range of economic models suitable for achieving economically sustainable digital preservation activities. We plan to anchor this work in real-world practicality by defining a set of stylized digital preservation scenarios representing common conditions under which the long-term preservation of digital assets takes place. For each scenario, we will articulate economic models particularly suited for supporting long-term preservation under scenario conditions. For each economic model, we will discuss its strengths, weaknesses, policy context and, where they exist, provide examples of real-world implementations. In addition, for each digital preservation scenario and its associated economic models, the Task Force will provide analysis and recommendations on implementation issues, and suggestions for future work in the economics of digital preservation and access.

Taken together, the Task Force's two reports should contribute to the discussion on the economics of sustainable digital preservation and access, and provide practical information for decision makers to address this critical problem.

References Cited

Ayris, P., Mcleod, R., & Wheatley, P. (2006). *Lifecycle Information for E-literature: Full Report from the LIFE Project*. London, U.K.: University College London and the British Library.

Bartleby.com. (1996a). The Columbia World of Quotations, Number 5114. Retrieved November 19, 2008, from http://www.bartleby.com/66/14/5114.html

Bartleby.com. (1996b). The Columbia World of Quotations, Number 39469. Retrieved November 19, 2008, from http://www.bartleby.com/66/69/39469.html

Bartleby.com. (1996c). The Columbia World of Quotations, Number 54757. Retrieved November 19, 2008, from http://www.bartleby.com/66/57/54757.html

Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). *Keeping Research Data Safe: A Cost Model and Guidance for UK Universities*. London: JISC.

Bloom, F. (2006). Prying Open the Black Box. *Science, 314*, 17.

Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies. (2003). *Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for a Long-Term Strategy* Washington, D.C.: National Academies Press.

Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies. (2005a). *Building an Electronic Records Archive at the National Archives and Records Administration : Recommendations for a Long-Term Strategy* Washington, D.C.: National Academies Press.

Committee on Digital Archiving and the National Archives and Records Administration, Computer Science and Telecommunications Board, & National Research Council of the National Academies. (2005b). *Building an Electronic Records Archive at the National Archives and Records Administration: Recommendations for a Long-Term Strategy* Washington, D.C.: National Academies Press.

Consultative Committee for Space Data Systems. (2002). *Reference Model for an Open Archival Information System (OAIS): Recommendation for Space Data System Standards: Blue Book*. Washington, D.C.: CCSDS Secretariat, Program Integration Division, National Aeronautics and Space Administration.

Copernicus, N. (1543). *De revolutionibus orbium coelestium*. Nuremberg: Johannes Petrius.

David, P. A. (2008). The Historical Origins of 'Open Science': An Essay on Patronage, Reputation and Common Agency Contracting in the Scientific Revolution. *Capitalism and Society, 3*(2).

Gantz, J. F. (2008). *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011*: International Data Corporation (IDC).

Gates, B. (2008). A New Approach to Capitalism in the 21st Century. *Creative Capitalism: A Conversation* Retrieved November 19, 2008, from http://creativecapitalism.typepad.com/creative_capitalism/2008/06/bill-gates-crea.html

Gutmann, M. (2008). Digital Preservation at ICPSF: The Background. Unpublished Presentation. University of Michigan.

Leake, J. (2008). Captains' Logs Yield Climate Clues: Records Kept by Nelson and Cook are Shedding Light on Climate Change. *TimesOnline*.

Maxwell, J. C. (1873). *A Treatise on Electricity and Magnetism* (First ed.). London: Macmillan and Co.

Mojica, P. (2008). Blue Ribbon Task Force on Sustainable Digital Preservation and Access. Unpublished Presentation. SNIA-DMF Storage Networking Industry Association.

Moore, R., JaJa, J., & Chadduck, R. (n.d.). A Pilot Persistent Archive: Partnership between SDSC, UMD, and NARA. Retrieved October 3, 2008, from http://www.archives.gov/era/pdf/it-conference-jaja-workshop.pdf

National Science Board. (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. Arlington, VA: National Science Foundation.

Nature. (4 September 2008). Editorial. *Nature, 455*.

Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica*. London: Joseph Streater.

Nowell, L. (2008). Data Preservation and Access: A Global Challenge. Unpublished Presentation. National Science Foundation.

Palm, J. (2006). *The Digital Black Hole*. Stockholm, Sweden: Riksarkivet/National Archives.

Peterson, M., Zasman, G., Mojica, P., & Porter, J. (2007). *100 Year Archive Requirements Survey*: Storage Networking Industry Association (SNIA).

Pope, D. (1981). *Life in Nelson's Navy*. Annapolis, MD: Naval Institute Press.

Quarterly, I. S. (2008). *Information Standards Quarterly, 20*(3).

Samuelson, P. (1948). *Economics: An Introductory Analysis*. New York: McGraw-Hill.

Science and Technology Council. (2007). *The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials*: Academy of Motion Picture Arts and Sciences (A.M.P.A.S.).

Smith, A. (1904). *An Inquiry into the Nature and Causes of the Wealth of Nations* (5th Edition ed.). London, UK: Methuen & Co., Ltd.

Szalay, A. (2008). Preserving Digital Data for the Future of eScience. *Science News: Magazine of the Society for Science and the Public*(Web Edition).

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Testbed Digitale Bewaring. (2005). *Costs of Digital Preservation*. The Hague, Netherlands: Nationaal Archief of the Netherlands.

The Science and Technology Council of the Academy of Motion Picture Arts and Sciences. (2007). *The Digital Dilemma: Strategic Issues in Archiving and Accessing Digital Motion Picture Materials*. Beverly Hills, CA: The Science and Technology Council of the Academy of Motion Picture Arts and Sciences.