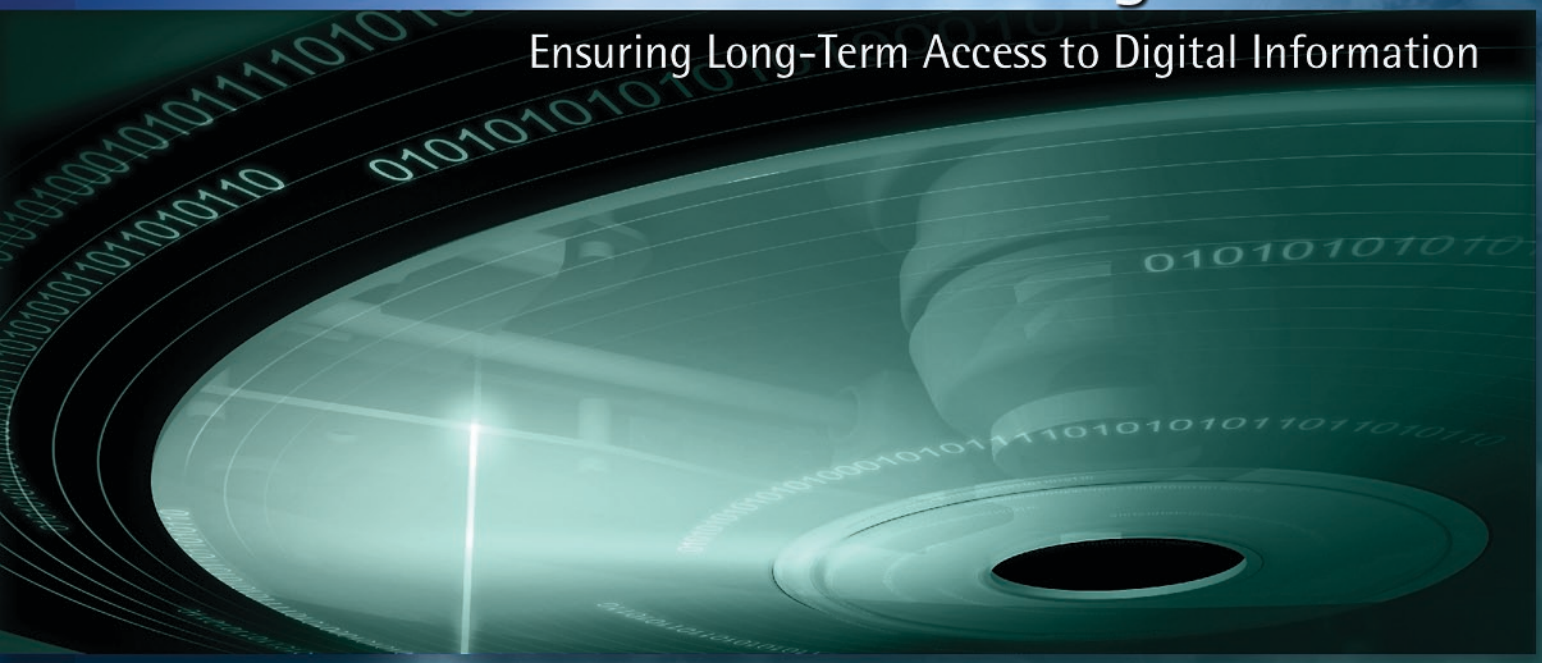
Sustainable Economics for a Digital Planet:

Ensuring Long-Term Access to Digital Information



February 2010

Final 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, 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, the Electronic Records Archives Program of the National Archives and Records Administration, and the Council on Library and Information Resources. The chairs would like to thank these organizations for their generous support of our work, as well as those organizations that granted time to Task Force members to participate in this work. Special thanks go to Abby Smith Rumsey, writer and editor of the report; our interns Lorraine Eakin and Elizabeth Bedford; Susan Rathbun, Jan Zverina, Ben Tolo, and Richard Moore of the San Diego Supercomputer Center; Kathlin Smith; and to the San Diego Supercomputer Center and OCLC for managing the Task Force funds. Finally, we are grateful for the written and oral testimony generously provided to the Task Force by economic and preservation experts over the course of our work.

The views and opinions expressed in this report represent the rough consensus among members of the Task Force and should not be construed to represent those of our sponsoring agencies and organizations.

List of Members

Francine Berman, Vice President of Research, Professor of Computer Science, Rensselaer Polytechnic Institute [co-Chair]

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

Paul Ayris, Director, UCL Library Services, University College London, UK

G. Sayeed Choudhury, Associate Dean, Director of the Hodson Digital Research & Curation Center, Johns Hopkins University, Baltimore, MD

Elizabeth Cohen, Acoustician, STEAM Education Evangelist (Science, Technology, Engineering, Arts and Math)

Paul Courant, University Librarian and Dean of Libraries, University of Michigan, Ann Arbor, MI

Lee Dirks, Director of Scholarly Communications, Microsoft Corporation

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

Vijay Gurbaxani, Taco Bell Endowed Professor, The Paul Merage School of Business, University of California, Irvine

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

Ann Kerr, Consultant to Library of Congress, 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 or the Digital Curation Centre, University of Edinburgh

Roger Schonfeld, Manager of Research, Ithaka

Abby Smith Rumsey, 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

Philip Bogden (2009-2010)
National Science Foundation
Program Officer, Office of Cyberinfrastructure
Arlington, VA

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

Robert Chadduck
National Archives and Records Administration
Computer Engineer, Principal Technologist, Electronic Records Archives Programs
College Park, MD

Chris Greer
National Science Foundation (2007)
Program Officer, Office of Cyberinfrastructure
Arlington, VA

Networking and Information Technology Research and Development (2008-2010)
Director of National Coordination Office
Arlington, VA

Lucy Nowell (2008-2009)
National Science Foundation
Program Officer, Office of Cyberinfrastructure
Arlington, VA

Sylvia Spengler (2009)
National Science Foundation
Program Officer, Division of Information and Intelligent Systems, CISE Directorate
Arlington, VA

Table of Contents

List of Figures

List of Tables

List of Boxes

Executive Summary

Digital information is a vital resource in our knowledge economy, valuable for research and education, science and the humanities, creative and cultural activities, and public policy. But digital information is inherently fragile and often at risk of loss. Access to valuable digital materials tomorrow depends upon preservation actions taken today; and, over time, access depends on ongoing and efficient allocation of resources to preservation.

Ensuring that valuable digital assets will be available for future use is not simply a matter of finding sufficient funds. It is about mobilizing resources—human, technical, and financial—across a spectrum of stakeholders diffuse over both space and time. But questions remain about what digital information we should preserve, who is responsible for preserving, and who will pay.

The Blue Ribbon Task Force on Sustainable Digital Preservation and Access investigated these questions from an economic perspective. In this report, we identify problems intrinsic to all preserved digital materials, and propose actions that stakeholders can take to meet these challenges to sustainability. We developed action agendas that are targeted to major stakeholder groups and to domain-specific preservation strategies.

The Task Force focused its inquiry on materials that are of long-term public interest, looking at four content domains with diverse preservation profiles:

> *Scholarly discourse*: the published output of scholarly inquiry
>
> *Research data*: the primary inputs into research, as well as the first-order results of that research
>
> *Commercially owned cultural content*: culturally significant digital content that is owned by a private entity and is under copyright protection; and
>
> *Collectively produced Web content*: Web content that is created interactively, the result of collaboration and contributions by consumers.

Economic analysis of digital preservation of these materials reveals structural challenges that affect all digital preservation strategies: (1) long time horizons, (2) diffused stakeholders, (3) misaligned or weak incentives, and (4) lack of clarity about roles and responsibilities among stakeholders. These risks, once identified, can be anticipated and provided for throughout the digital lifecycle. Major findings can be summarized as three imperatives for sustainable digital preservation.

Articulate a compelling value proposition.

> When making the case for preservation, make the case for use. Without well-articulated demand for preserved information, there will be no future supply.

Stakeholders for digital materials are often diffuse across different communities. The interests of future users are poorly represented in selecting materials to preserve. Trusted public institutions—libraries, archives, museums, professional organizations and others—can play important roles as proxy organizations to represent the demand of their stakeholders over generations.

A decision to preserve now need not be thought of as a permanent or open-ended commitment of resources over time. In cases where future value is uncertain, choosing to preserve assets at low levels of curation can postpone ultimate decisions about long-term retention and quality of curation until such time as value and use become apparent.

Provide clear incentives to preserve in the public interest.

The lack of clear incentives to act will stymie timely preservation actions. Policy mechanisms can play an important role in strengthening weak motivations. Lowering barriers to efficient decentralized stewardship can be spurred by individual creators' use of nonexclusive licenses granting preservation rights to third parties.

Misalignment of incentives among stakeholders may occur between communities that benefit from preservation (and therefore have an incentive to preserve), and those that are in a position to preserve (because they own or control the resource) but lack incentives to do so. Policy mechanisms that can mitigate these problems include: financial incentives and other benefits to private owners who preserve digital materials for the benefit of the public; mandates to preserve when appropriate; and revision of copyright to enable preservation of privately owned materials by stewardship organizations acting in the long-term interest of the public.

Define roles and responsibilities among stakeholders to ensure an ongoing and efficient flow of resources to preservation throughout the digital lifecycle.

The strongest incentives to preserve will be ineffective without explicit agreement on the roles and responsibilities of all the actors—those who create the information, those who own it, those who preserve it, and those who make it available for use. Every organization that creates and uses data should implement policies and procedures for preservation, including: selection of materials with long-term value; preparation of data for archiving; and protocols to ensure a smooth and secure transfer of digital assets across organizational boundaries and between institutions.

There is a particular risk of "free riding" with digital materials because the cost of preservation may be borne by one organization but the benefits accrue to many. Effective governance mechanisms are needed to aggregate the collective interest into an effective preservation strategy that ensures that the effort and cost of preservation are appropriately apportioned.

Funding models should be tailored to the norms and expectations of anticipated users. They should leverage economies of scale and scope whenever possible. Digital assets do not need to be treated as a public good in all cases. Market channels are often the most efficient means of allocating resources for preserving many types of digital content.

Digital assets provided as market goods or otherwise privately held must have some provision for handoff to a trustworthy steward when the owner decides to stop preserving them, if those materials are of value to society. For materials that are not amenable to market provision and are at risk of loss—such as certain types of research data, Web-based materials, and digital orphans—public provision is necessary.

Finally, as the rate of digital information production continues to escalate, it is vitally important to reduce the cost of preservation for all types of digital assets. Reducing the cost of storing materials, developing sustainable sources of energy to power preservation systems, and engineering ways to lower the cost of preserving, curating, and providing access are all important.

There is a great diversity of preservation strategies among the content types that have long-term societal value. In the four domains evaluated, we were able to identify significant risks to sustainability and the near-term actions that stakeholders can take to remedy them.

Scholarly discourse: This is a fairly mature field, with well-developed preservation and funding strategies as a legacy of the print world. Disruptions are occurring to longstanding sustainability strategies as a result of digital preservation and distribution. There are particular needs to align preservation incentives among commercial and nonprofit providers; ensure handoffs between commercial publishers and stewardship organizations in the interest of long-term preservation of the scholarly record; and address the free-rider problem. Clarification of the long-term value of emerging genres of digital scholarship, such as academic blogs and grey literature, is a high priority. Research and education institutions, professional societies, publishers, libraries, and scholars all have leading roles to play in creating sustainable preservation strategies for the materials that are valuable to them.

Research data: There is a remarkable growth of data-intensive research in all knowledge domains. In most fields, there is high recognition of the benefits of preserving research data for various purposes and lengths of time. But there are few robust systems for making decisions about what to preserve; and there is often a lack of coordination of roles, responsibilities, and funding sources among those best positioned to preserve data (researchers) and the preservation infrastructure (curation and archiving services) that should support them. Research and education institutions, professional societies, archives, researchers, and the funding agencies that support data creation all have leading roles to play in creating sustainable preservation strategies.

Commercially owned cultural content: There are well-established preservation and access strategies undergoing fundamental changes as a result of new information technologies. This includes the creation, distribution, and consumption of cultural content, most evident in the emergence of interactive genres such as games and the creation of a long tail of use and reuse. As a result, there may be two forms of benefits—commercial and cultural, or private and public—that compete with one another. When that occurs, proxy organizations must step in to represent the public interest. Leading players in preserving this content include private creators, owners, and trade associations, stewardship organizations, regulatory authorities, and leading national and international institutions that can sponsor public-private partnerships to ensure the long-term access to our digital cultural heritage.

Collectively created Web content: The Web environment is marked by great dynamism, uncertainty about long-term value of digital content, and obscure ownership and rights issues for many collectively produced Web assets. The priority here is for stewardship organizations, content creators, hosting sites, platform providers, and users to model and test preservation strategies, and to provide clarification about long-term value and selection criteria.

The Task Force identified important next steps for each of these content areas; they are summarized in Table 5.2.

Sustainable preservation is a societal concern, however, and transcends the boundaries of any particular content domain. All parts of society—national and international agencies, funders and sponsors of data creation, stakeholder organizations, and individuals—have roles in achieving sustainability. Leadership is needed at all levels of society. Table 5.1 presents a summary of the action agendas for these major stakeholders.

Areas of priority for near-term action include the following:

Organizational Action

- developing public-private partnerships
- ensuring that organizations have access to skilled personnel, from domain experts to legal and business specialists
- creating and sustaining secure chains of stewardship between organizations over time
- achieving economies of scale and scope
- addressing the free-rider problem

Technical Action

- building capacity to support stewardship in all areas
- lowering the cost of preservation overall

- determining the optimal level of technical curation needed to operationalize an option strategy for all types of digital material

Public Policy Action

- modifying copyright laws to enable digital preservation
- creating incentives and requirements for private entities to preserve on behalf of the public (financial incentives, handoff requirements)
- sponsoring public-private partnerships
- clarifying rights issues associated with Web-based materials
- empowering stewardship organizations to protect digital orphans from unacceptable loss

Education and Public Outreach Action

- promoting education and training for 21st century digital preservation (domain-specific skills, curatorial best practices, core competencies in relevant science, technology, engineering, and mathematics knowledge)
- raising awareness of the urgency to take timely preservation actions

Sustainable preservation strategies are not built all at once, nor are they static. Sustainable preservation is a series of timely actions taken to anticipate the dynamic nature of digital information. Decision makers will always face uncertainties. Changes in technologies, policy environments, investment priorities, and societal concerns will unfold over the course of the digital lifecycle. But we can develop practices that resolve or anticipate uncertainties, that leverage resources among stakeholders, and above all, that leave options open for decision makers in the future. Sustainable preservation strategies will find ways to turn the uncertainties of time and resources into opportunities for flexibility, adjustments in response to changing priorities, and redirection of resources where they are most needed. **Commitments made today are not commitments for all time. But actions must be taken today to ensure flexibility in the future.**

Above all, sustainable digital preservation requires a compelling value proposition, incentives to act, and well-defined roles and responsibilities. Digital preservation is a challenge for all of society because we all benefit from reliable, authentic information now and into the future. Done well, all of society will reap the benefits of digital stewardship.

Preface

Digital preservation has received extensive attention both as a *technical* problem and, to a lesser degree, as a *policy* problem; yet to date, it has received comparatively little treatment as an *economic* problem. In its interim report, *Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation*, the Blue Ribbon Task Force on Sustainable Digital Preservation and Access traced the contours of digital preservation as an economic activity: that is, one where decision makers must make explicit and ongoing resource allocations in order to achieve long-term goals. The report offered a definition of economic sustainability in a digital preservation context, citing the conditions that must be met if digital preservation activities are to be economically sustainable over time. The report then examined a number of issues, challenges, and lessons learned, both institutional and systemic, attached to meeting these conditions, based on the testimony of speakers representing a broad swath of the digital preservation community, a review of published work, and of course, the expertise of the Task Force membership.

In its final report, the Task Force builds on its earlier findings by developing a general economic framework for analyzing digital preservation as an economic problem, and employing this framework as a tool for analyzing the economic conditions and implications intrinsic to four key digital preservation contexts: scholarly discourse, research data, collectively produced Web content, and commercially owned cultural content. This analysis serves as a foundation for a body of findings and recommendations advanced at the end of the report. The recommendations address issues at a variety of levels—local and system-wide—and are relevant to a variety of different types of digital information.

The analysis in this report views digital preservation through an economic lens. While this perspective is quite revealing, it also has limitations. Real-life decision makers are not the abstract economic agents of textbook theories and models. Sometimes they will make choices that differ from what economic theory says they should do, and they may do so for good reasons. Readers should keep this caveat in mind. Also, our analysis is not a quantitative accounting of the costs of digital preservation, although the Task Force recognizes that understanding preservation costs, represented in terms of current and future monetary outlays, is a matter of considerable importance. Readers looking for an estimate of the cost of preserving a terabyte of data over a specified time period—a task over-laden with the details of a particular implementation and preservation context—will not find it here. Instead, we consider the broader economic setting in which preservation decisions are made, in particular the perceived *value* of a set of digital materials, the *incentives* to act to preserve valuable digital materials, and the *roles and responsibilities* of stakeholders sharing an interest in long-term preservation. As our report makes clear, these are the fundamental economic elements that ultimately shape the prospects for achieving long-term sustainability in a digital preservation context.

Solving the economic challenges of long-term digital preservation is not an easy task, nor is it an insuperable one. As we hope this report makes clear, economically sustainable digital preservation is a tractable problem that can be organized into a set of well-defined

elements, summarized by the Task Force's definition of sustainability. The Task Force's final report offers findings and actions agendas we hope will be useful to decision makers as they address these sustainability elements. If there is one finding that perhaps merits special attention, it is that **sustainable economics for digital preservation is not just about finding more funds. It is about building an economic activity firmly rooted in a compelling value proposition, clear incentives to act, and well-defined preservation roles and responsibilities**. Lacking these ingredients, digital preservation efforts—and the materials in their care—have little prospect of persisting over time; with them, our digital heritage will have a sound economic foundation for the future.

Brian Lavoie
Fran Berman
Task Force Co-Chairs

Abby Smith Rumsey
Writer and Editor

Chapter 1

Purpose and Background of the Report

Without preservation, there is no access

Economically sustainable preservation—ensuring the ongoing and efficient allocation of resources to digital preservation—is an urgent societal problem. It is urgent because digital information is inherently fragile, prone to information loss and degradation. Preservation insures against multiple risks to information assets over time. Such assets must be actively managed for sustained periods of time, using best practices for data stewardship across the full lifecycle of creation, description and curation, deposit in secure storage, use, and reuse. Some digital materials require relatively intensive levels of preservation to ensure usability, and others much less. But in all cases, access to information tomorrow depends on preservation actions taken today. A fundamental fact of digital sustainability is that without preservation, there is no access.

Digital preservation is a societal challenge because information is a vital resource in our knowledge economy; it is critical to science, research and education, public policy, the creative industries, and the cultural heritage sectors that are the focus of our report. In the past decade, leadership organizations have engaged a broad spectrum of stakeholder communities to address digital preservation challenges from organizational, policy, and technical perspectives.[1] And the storage industry has tracked data trends to anticipate archiving requirements in a burgeoning digital universe, only to find that the scale of digital creation is far outpacing the capacity to store the data. Figure 1.1 presents one widely credited projection of current trends.[2]

[1] See, for example, *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age* by the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine of the National Academies. Available at http://www.nap.edu/catalog.php?record_id=12615.
See also *Preserving Our Digital Heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report. A Collaborative Initiative of the Library of Congress,* forthcoming from the Library of Congress. More information on the Library's digital preservation efforts can be found at http://www.digitalpreservation.gov/.
See also *The Digital Dilemma*, by The Science and Technology Council of the Academy of Motion Picture Arts and Science. Available at http://www.oscars.org/science-technology/council/projects/digitaldilemma/index.html.

[2] See J. F. Gantz, *The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth through 2001*: International Data Corporation (IDC), 2008.

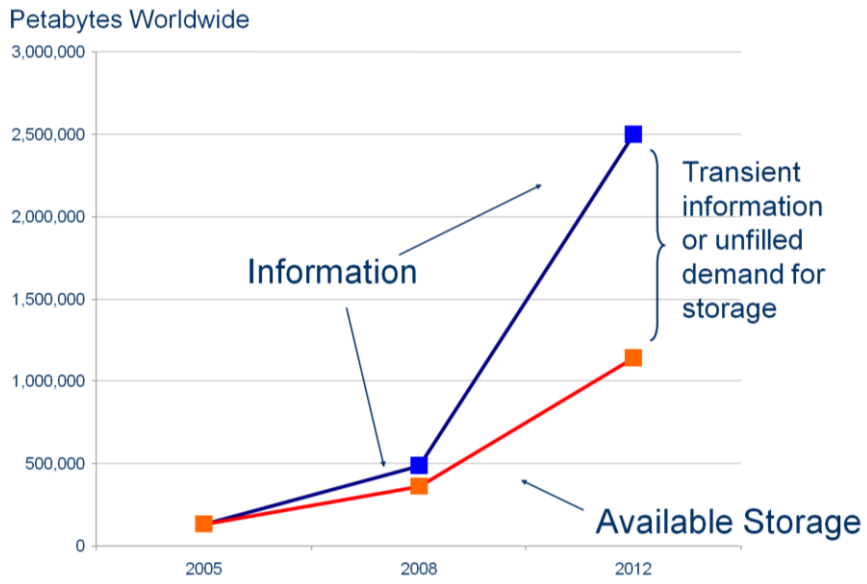


FIGURE 1.1 **Growth of Information and Storage Trends**
Projected growth of global information creation outpaces growth of available storage.
Source: IDC Digital Universe White Paper, sponsored by EMC, May 2009. Used with permission.

1.1 Purpose of the Report

Critical questions remain:

- What digital information should we preserve?
- Who will preserve it?
- Who will pay for it?

They are unanswered because we do not fully grasp the opportunities, constraints, and realities of sustainability from an economic point of view.[3] These are the questions we address in this report. In the first phase of work, the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (the Task Force) gathered information about current preservation activities and best practices. Our findings are published in an interim report, *Sustaining the Digital Investment: Issues and Challenges of Economically Sustainable Digital Preservation.*[4]

In our second and concluding phase, we developed an economic framework based on our findings to get at the questions facing all preservation decision makers, such questions as: What is an appropriate investment in preservation and what is the return

[3] This was the challenge put to the Task Force by Dan Atkins, founding dean of the University of Michigan School of Information and founding director of the NSF Office of Cyberinfrastructure, at a Task Force session held July 29-30, 2009.

[4] See http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf.

on such investment? What are the tradeoffs between investments to create and use information today versus investments that enable future uses? How will a stakeholder community pay for preservation? As a result, we have been able identify a set of principles that should inform all preservation decision making and propose an agenda for further action.

1.1.1 Scope of the Task Force Work

This report addresses concerns that affect a wide spectrum of the digital universe. Digital information has become a fundamental and valued resource in fields that range from the natural sciences to the creative sector. **Throughout this report, we use *value* in the broadest sense—including historical, ethical, aesthetic, scholarly, public policy, and institutional, which may or may not be financially quantifiable**. Stakeholders in all information communities face a set of tough decisions about long-term management of the digital materials they create, use, add value to, and distribute for use. Each sector needs to assess the long-term value of its digital assets and understand the tradeoffs between benefits and costs in preserving its data, handing the data to some other stakeholder for long-term stewardship, or abandoning the resource.

Decisions about longevity are made throughout the digital lifecycle. Bench scientists face a choice at the end of every research project about what to do with their data. Filmmakers wrapping up production must decide which production elements promise future productivity as long-term assets and how to pay for what they keep. Contributors to open-source code create a collective good that is everyone's responsibility—and hence no one's in particular. The curatorial team at a museum that is bringing down an exhibition for which many original images and design materials were created faces similar decisions about what to retain and whose responsibility it is to provide and fund long-term retention. Preservation decision makers run the gamut from university provosts, foundations, and philanthropic funders to anyone who has created a website that has potential value for reuse.

While digital preservation is a critical issue for every sector that creates and uses information, we focus our inquiry on digital materials that are of greatest importance for science and the humanities, research and education, public policy making, cultural heritage, and the creative industries. Within that range, we look at information in which there is a public or collective interest in long-term access and use, whether created in the private or public sectors. We address four specific information types:

1. *Scholarly discourse*—the published output of scholarly inquiry, including the ideas, theories, analyses of data, assessments of previous scholarship, and conclusions that collectively form the scholarly record

2. *Research data*—the primary inputs into scientific and other research, as well as the first-order results of that research

3. *Commercially owned cultural content*—culturally significant digital content that is owned by a private entity and is under copyright protection

4. *Collectively created Web content*—Web content that is created interactively, the result of collaboration and contributions on the part of consumers.[5]

Our investigation and findings focus on preservation in the United States and the United Kingdom. The complexity and variety of political economies and legal systems worldwide preclude any comprehensive or significant analysis of sustainable digital preservation in other countries. However, by focusing on foundational economic matters, we believe that the Task Force's framework for analysis and recommendations are applicable across the globe in most domains of digital content. Digital preservation is an imperative that spans national boundaries; so, too, do its challenges.

1.2 Background of the Report

1.2.1 Phase 1 Findings

The Task Force brought the expertise of preservation specialists, knowledge domain experts, and economists to bear on assessing the choices that preservation decision makers face. The first phase of our work was empirical. Through a literature review, testimony from experts, and additional research, we developed case studies of real-world preservation activities in the domains of scholarly discourse, research data, commercially owned cultural content, and collectively produced Web content. These case studies, numbering close to two dozen, form the basis of our analysis in this report; they can be found in a separate report on the Task Force website.

We identified five conditions that must be met for preservation to be sustainable.

BOX 1.1 **Conditions for Sustainable Digital Preservation**

Five conditions required for economic sustainability are:

- recognition of the benefits of preservation by decision makers;
- a process for selecting digital materials with long-term value;
- incentives for decision makers to preserve in the public interest;
- appropriate organization and governance of digital preservation activities; and
- mechanisms to secure an ongoing, efficient allocation of resources to digital preservation activities.

[5] Public and corporate records are vital to preserve, but such records fall outside the scope of our work because there are well-articulated mandates for preservation and well-defined organizations with clear roles and responsibilities. For more about records, see Appendix 1.

We found a number of practitioners who have well-developed strategies for analog preservation and are grappling with the challenges of sustainability as they play out in the digital realm. But even in domains with long histories of success in analog preservation and access, such as academic publishing and many of the creative industries, practitioners reported systemic problems in reconfiguring their preservation strategies for the realities of the digital marketplace because of deep structural disruptions in the nature of production, dissemination, and consumption of information. Many new forms of digital content are at special risk: Web-based genres such as blogs, collectively created sources such as Wikipedia and Flickr, and, in the sciences, a whole range of new data sets that result from digitally enabled forms of observation, analysis, and modeling. Digital preservation strategies face the following challenges:

- uncertainty about selection criteria for assessing long-term value, especially with large-scale data sets, small "hand-crafted" digital collections, and the emerging genres of collective authorship on the Web;
- misalignment of incentives between those who are in a position to preserve and those who benefit from preservation and access;
- lack of clear responsibility for digital preservation, coupled with a prevailing assumption that it is someone else's problem;
- little coordination of preservation activities across diffused stakeholder communities;
- difficulty in separating preservation costs from other costs, that is, in distinguishing between the processes of making things available now and making things available in the future; and
- difficulty in valuing or monetizing the costs and benefits of digital preservation, which are necessary to secure funding and investment.

In this report we develop an economic framework for sustainable digital preservation based on the conditions necessary for sustainability (Box 1.1). The framework emerges from analysis of preservation services and preserved assets as economic goods and services, with special attention to the dynamics of supply and demand.

Three essential components stand out as vital throughout our discussion of sustainable preservation:

1. **value and benefits derived from preservation**
2. **incentives to preserve**
3. **roles and responsibilities among preservation stakeholders**

These three components are pivotal in developing preservation strategies that will be flexible enough to adapt to changes in uses, technologies, and stakeholder communities over the course of the digital lifecycle. Each component prompts decision makers to take actions to resolve uncertainties, leave open options for future stakeholders, and significantly lower barriers to sustainability.

BOX 1.2 **Value, Incentives, Roles and Responsibilities**

Actions necessary for sustainability include:

- **assessing the value** of preserved information, selecting materials for long-term preservation based on that assessment of value, and articulating a compelling value proposition to community stakeholders;
- **providing incentives** for stakeholders to preserve digital assets directly or provide preservation services for others, and tailoring those incentives to the prevailing community norms and to information policy regulations and privacy considerations; and
- **defining the roles and responsibilities** of individuals and institutions in preservation, specifying how actors and stakeholders are organized and how resources flow among them to ensure preservation.

These components map to the five conditions necessary for sustainability. For the sake of simplicity, we focus on this trio of components for our analysis. We shall return to the five conditions that define sustainability again in our conclusion.

1.2.2 Audiences for This Report

Our report is intended to support and inform preservation decision makers including, but not limited to:

- *digital information creators*, from scientific research teams to independent filmmakers, digital artists, and bloggers;
- *funders of digital information creation*, such as federal or corporate research agencies, the creative industries, and private foundations; and
- *managers of digital information*, located in organizations that range from universities and research labs to music production companies and social networking platforms that are responsible for data management in their day-to-day business.

All who play a role in a preservation strategy are *actors*. All who benefit from access to and use of preserved information, or who support or fund those who do, are *stakeholders*. A distinctive feature of digital information is that to some degree, all actors in preservation are *decision makers*, whether they are aware of it or not. This is important to keep in mind when thinking about the organization of preservation activities

because even actors who may be unaware of their implicit roles can be brought into the implementation of responsible digital stewardship activities. Digital preservation is a challenge for society because we all benefit from reliable, authentic information now and in the future. Done well, all of society will reap the benefits of digital stewardship.

1.2.3 How to Navigate the Report

Digital preservation is a challenge for society because we all benefit from reliable, authentic information now and in the future

Report Structure

In Chapters 2 and 3, we develop a framework for and analyze digital preservation from an economic perspective. In Chapter 2, we focus on the nature of supply and demand for preservation and identify attributes of preserved information as an economic good that influence the choices available to preservation decision makers. In Chapter 3, we look at the risks posed by these attributes and the remedies that can address them. In Chapter 4, we look broadly at four preservation scenarios through an economic lens to identify some of the leading choices and tradeoffs facing decision makers in each context. We propose specific actions and recommendations that the leading stakeholders in each scenario can take to achieve desired preservation outcomes. In Chapter 5 we synthesize our findings into general principles of sustainable digital preservation, propose specific recommendations for concrete next steps, and identify the areas of further work and research that promise to advance both the knowledge and practice of sustainable digital preservation.

Recommendations for Action

Recommendations are proposed within the body of the report when appropriate: in Chapter 3 we propose remedies to preservation risks; in Chapter 4 we set forth domain-specific recommendations; and in Chapter 5 we set forth global recommendations, a set of recommendations aimed at organizational and individual actors, and a summary of recommendations (Tables 5.1, 5.2).

Technical Concepts

We draw on expertise from two knowledge domains with highly developed methods and specialized vocabularies—economics and preservation. When we use a specialized term, we explain its meaning in context. We make use of several economic concepts that are explained at greater length in a series of appendices, referred to at relevant junctures in the report. For the sake of readability, we have not included lengthy citations within the report. We recommend that readers consult the Task Force's bibliography, available online.[6] This includes a link to the interim report, which contains a literature review for digital preservation.

Key Terms

The terminology for digital materials and preservation processes varies among stakeholder communities. As a rule members of the scientific community refer to digital materials as *data*; further, they refer to activities that enable use and long-term accessibility as *curation and archiving*, which taken together, are called *stewardship*. In cultural domains and the humanities, digital materials are more often referred to as

[6] See http://brtf.sdsc.edu/bibliography.html.

content, and the activities that ensure their long-term availability are called *preservation and access*. Finding a common vocabulary among the many domains we address is not possible; unless speaking in a domain-specific context, we use the terms *digital assets*, *materials*, and *information* interchangeably. And unless indicated otherwise, references to *preservation* always mean *digital preservation*. Frequently used terms and acronyms are explained in the Glossary.

Chapter

2

The Economic Perspective on Digital Preservation

Without well-articulated demand for access to preserved digital assets, there will be no supply of preservation services

We turn now to preservation from an economic perspective. If the problems in sustainable digital preservation point to systemic failures, what can economics tell us about the systemic and structural problems? To start, if preservation experts tell us that without preservation there is no access, economists tell us that without the demand for access, there will be no preservation. Without well-articulated demand for access to preserved digital assets, there will be no supply of preservation services.

Two themes emerge from this chapter. The first is the specific nature of preserved digital assets as economic goods, and how that nature affects supply and demand. The second theme is the role that time plays in the supply of and demand for those assets. The long time horizon of preservation raises the specter of uncertainty. A sustainable preservation strategy must be flexible enough to span generations of data formats, access platforms, owners, and users. The next chapter addresses the topic of time in greater detail. In this chapter, we describe preservation as an economic good and preface our economic analysis with a few clarifications about supply and demand. The dynamics of value, incentives, and roles and responsibilities are crucial to understanding how preservation supply and demand operates.

2.1 Supply and Demand

We know far more about the problems of supply of preservation services than we do about the nature of demand for preserved digital goods. This poses a problem, since without a well-articulated demand there can be no sustainable supply. This should be neither surprising nor disturbing, given how recently digital information has gained its foothold and how rapidly its patterns of creation and use have changed in just a matter of decades. On the supply side, there is a robust and well-developed preservation infrastructure to manage analog artifacts; and to the extent possible, it has been modified, updated, and reconfigured to manage digital assets. But analog materials—books and journals, maps and manuscripts, film and audio—differ from digital assets in significant ways with respect to patterns of consumption, relative scarcity of supply,

and copyright and privacy regulations. Appendix 2 describes in greater detail significant differences between analog and digital preservation.

Even as the preservation community works to understand the implications of the differences between analog and digital preservation, it has moved quickly to address some of the fundamental challenges—technical, organizational, and policy related—to building out the digital preservation infrastructure. The community has done this, by and large, in the absence of well-articulated demand for long-term reuse, but with evidence of fairly heavy short-term reuse. But its understanding remains provisional in matters concerning new genres, such as social networking sites, and in managing scientific data produced on a scale that is unprecedented. As it turns out, the demand for reuse of data over even short periods of time—a year or less—requires some measure of data management that looks very much like preservation, and this has taught preservation managers valuable lessons. One of them is that preservation decisions can often be made "on the margin," as an incremental cost, and are often indistinguishable from decisions made to meet current demand.

Preservation services can be supplied by one institution, or distributed across many. Actors and stakeholders can work individually or in small groups to effect long-term access, or they can make use of trusted proxy organizations to achieve their ends, using specialized skills and exploiting economies of scale. Preservation actors may be guided by the invisible hand of the market, or by centralized command and control. They can be open and permeable, or create their own walled gardens. How coordination among actors and stakeholders is best achieved depends upon the conditions prevailing in a particular context. However configured, though, no preservation system will be sustainable without a strategy that specifies clear lines of responsibility among the many actors, supported by a reliable flow of resources. **Whatever preservation strategies are used, demand for preserved information must be articulated well enough to ensure there is sufficient supply**.

We describe here what we mean by value, incentives, and roles and responsibilities in general terms designed to illustrate how these components affect decision making. These descriptions are by no means comprehensive, and readers can learn more about how these components operate in specific preservation contexts in Chapter 4, "Sustainable Preservation in Context."

2.1.1 Value and Benefits

When speaking about value, economists like to ask "Who benefits?" or "Who cares?" because well-articulated demand starts with a clear and compelling value proposition about the benefits to be gained by having, in our case, access to information at some point in the future. The value of information is not to be confused with its monetary or financial value per se, although it can often be denominated in currency. The value of digital assets is best understood as what digital materials are good for, and that is usually understood as the ways that the materials are used—to advance knowledge, entertain or bring pleasure, help solve problems, or inform public policy.

Selection is an expression of value

Each user community will identify its own set of values and benefits in the digital materials they demand. For example, in scholarly discourse there is a clear community consensus about the value of e-journals over time. There is much less clarity about the long-term value of emerging forms of scholarly communication such as blogs, products of collaborative workspaces, digital lab books, and grey literature (at least in those fields that do not use preprints). Demand may be hypothesized—*social networking sites should be preserved for future generations*—but that does not tell us what to do or why. Demand is diffuse, and the criteria for selection have not been formulated. An entity willing to supply such materials would not know what to collect, for whom, or for what purpose. **The first challenge to preservation arises when demand is diffuse or weakly articulated**. Addressing the matter of demand is always the first step in developing sustainable preservation strategies.

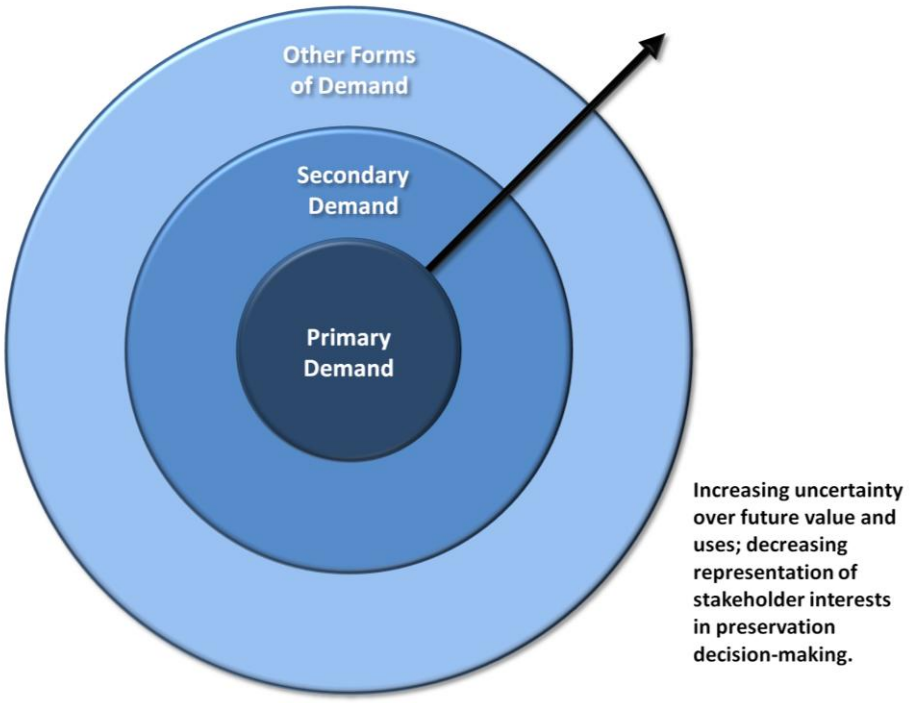


FIG. 2.1 **Layered Demand**
Figure 2.1 plots the layers of demand that reflect perceptions of value and influence selection processes. Stakeholders in the inner circle, who are closest to the assets because they create, own, or control them, will have the greatest say over what is preserved. Representing the interests of those whose demand is diffuse or can be anticipated only in the future becomes a central challenge for sustainable preservation strategies.

Value propositions always include judgment about priorities for preservation. Selection is an expression of value. Deciding which digital materials to preserve over time means discriminating among many competing collections that demand limited resources. In some user communities, selection criteria favor small, discrete, highly curated sets of digital materials, for example, a reference collection such as the Worldwide Protein

Data Bank.[7] In other user communities, the scale of the materials under consideration is vast, and samples must be chosen. This is the case in selecting for longitudinal studies that rely on observational data, and it could also apply in the case of social network sites.

Time is another variable of decision making. It may be a high priority to curate and preserve digital materials for relatively short periods of time, as is the case of the modeling data generated by simulations of early stages after the Big Bang.[8] But those data may be abandoned after five or ten years in favor of newer or more precise data. In other cases, the priority is to maintain materials for the indefinite future. This is typically the case with culturally significant collections of music, still and moving images, literary texts, endangered language corpora, archeological evidence, seismic event records, and many historical collections. Metrics of value derive directly from the benefits of using the assets; they reflect the values of specific user communities and these values may well change over time.

In some cases, the benefits of preservation may be most compellingly expressed in terms of *negative benefits*—the costs incurred if data are not preserved. These costs may reflect the time and effort needed to recreate the information or, if it cannot be recreated, the kinds of uses that would then not be possible. For classes of data that carry ethical issues—human subject and animal research, archaeological sites, or extinct species and languages—the benefits of preservation are often better framed as mitigating the risks of unacceptable loss—unacceptable because the loss violates shared ethical standards.

The value proposition is not a one-time declaration. Benefits can decline or be eclipsed by other priorities, and the value proposition must be revisited and re-articulated over the course of the digital asset lifecycle. But in all cases, the ultimate threat to persistent access to digital assets occurs when those responsible for preserving the materials decide that the cost of preservation exceeds the perceived benefits to them of long-term access.

2.1.2 Incentives

Incentives to preserve are strong when there are clear self-interests involved. When the creator of the digital asset is the primary user as well as the owner, incentives to preserve are well aligned: the owner has both the incentive to preserve and the right. (This does not mean that they have the wherewithal to do so, however; incentives can motivate people to secure funding for preservation, but they do not in and of themselves provide funding.) In the case of e-journals, scholars as a community know that a complete record of scholarship is important for the advancement of knowledge and to validate scholarship built on past knowledge. In addition, individual scholars have incentives to preserve journals or monographs they have authored, as this product is a signal of their accomplishments and a metric used in the advancement of

[7] See http://www.wwpdb.org.

[8] See http://lca.ucsd.edu/portal/software/enzo.

their careers. Similarly, an artist or musician who produces and distributes under the auspices of a firm shares with that firm a strong incentive to preserve, as they both receive benefits. It is important to underscore that benefits are not by definition strictly monetary. There are many rewards that come from having access to preserved information. In the case of the artist and scholar, for example, enhancement of reputation and prestige, and the promise that an audience will always have access to their work can be powerful motivators.

Two problems with incentives are commonly found in preservation contexts:

- **lack of a clear incentive to preserve; and**
- **misalignment of incentives between communities that benefit from preservation (and therefore have an incentive to preserve), and those that are in a position to preserve (because they own or control it) but lack incentives to do so.**

Public policy can play an important role in providing appropriate incentives to preserve valuable materials in the public interest

Incentives can weaken or become misaligned when assets no longer serve their original purpose. For example, data sets created for research may be eclipsed by new, presumably better, data sets. The incentive to preserve the former data set may disappear, and the materials may be abandoned to redirect scarce resources to assets yielding greater benefits to researchers. In another example, a digital film or a collection of songs may be created that provide great entertainment value for a period of time. The revenues they generate support their creation and preservation. Once the owner finds that the asset is no longer generating enough to cover the cost of its preservation, the asset is at risk of loss.[9] In both cases, a library, representing the future demand of historians or lovers of film or music, may have an incentive to preserve that content in the public interest, but it does not have the right to. This is where a handoff from those who have no incentive to preserve becomes necessary.

Such misalignment is not unusual between the incentives of private entities to preserve content, and the greater societal interest in preserving the same content for the public benefit. As this last example shows, incentives to supply preservation services depend on clarity of ownership and rights, areas that are sometimes obscure in the digital realm and sometimes constrained by intellectual property laws and privacy issues. Copyright laws have worked somewhat efficiently with analog materials but are perpetuating unintended negative consequences for digital materials. No responsible institution will preserve materials to which it has no legal right. Libraries, representing future users, are able to preserve analog films and music because copyright law gives them rights to do so when they own physical copies.[10] This is not the case when libraries license materials, the norm with digital content. We will return to the risks inherent in incentive gaps and misalignments, but note here that public policy can play an

[9] This period has lengthened considerably into a so-called *long tail of demand* because of digital reissues of backlist titles.

[10] In U.S. copyright law, these are sections 107 and 108.

important role in providing appropriate incentives to preserve valuable materials in the public interest.

2.1.3 Roles and Responsibilities

Even the strongest incentives to preserve will be ineffective without explicit agreement on the roles and responsibilities of all the actors—those who create the information, those who own it, those who preserve it, and those who make it available for use. There are many different ways of allocating those responsibilities among stakeholders, and often the allocations occur naturally. This is the case in much of higher education, for example, where universities support scholars to produce scholarship, and they support libraries to preserve that scholarship and make it accessible over generations.

Preservation goals sometimes need to be supported by mandates to preserve. For some who receive such mandates—from a funder, for example—preservation is a relatively new responsibility. In such cases, there may not be a natural allocation of roles and responsibilities among those who create data, those who are in a position to preserve the data, and those who are responsible for making it available. Grant recipients will be able to abide by the mandate only if they can secure sufficient funds to prepare the data for deposit and identify an appropriate repository to receive the data. This structural problem in the allocation of those responsibilities is also seen in the arena of public and corporate records, where mandates to preserve can be stymied by a lack of capacity to do so.

BOX 2.1 **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.

With respect to the ongoing and efficient (cost-reducing) allocation of resources for preservation, **even when incentives are in place, actors and stakeholders often do not have the means—organizationally or financially—to preserve valuable materials. Stimulating growth of capacity and funding to meet the demand is crucial**. In the case of independent artists and musicians, who work outside of organizations with deep capacity for data management and preservation, the choice is between taking time and money to prepare their materials for preservation and finding someone who will take them—in theory, highly desirable—versus finding funds for the next production. Preservation priorities seldom prevail in such tradeoffs. There may be funding models that could provide resources to support preservation, but in some cases community norms will argue against using some funding models. A well-known example is open access. In principle, everyone would benefit from unfettered access to scholarly discourse, especially scholarship produced with public funds. The potential downside for such an access policy is that if there is no provision for sustaining the data over time, preservation becomes an unfunded mandate. Open access is like any other form of access: without preservation, there will be no access, open or otherwise.

2.2 Digital Preservation as an Economic Good

We turn now to the economic nature of the goods and services under discussion. There are four *core attributes* that are common to all preserved digital assets. There is also a wide variety of *context-specific attributes* that further characterize preservation goods and services within certain knowledge domains and communities of users. Together, the core and context-specific attributes shape the choices available to decision makers and inform the nature of the tradeoffs facing each decision maker crafting a sustainable preservation policy.

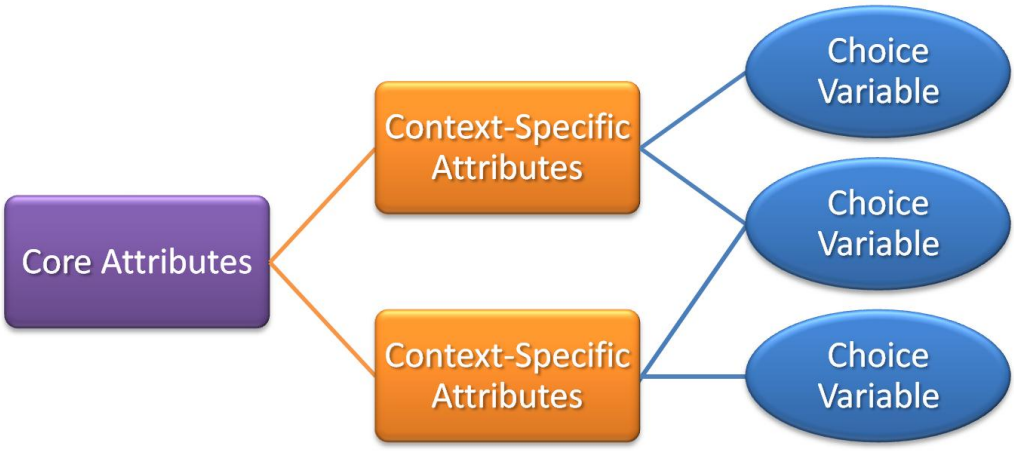


FIGURE 2.2 **Core Attributes, Context-Specific Attributes, and Choice Variables**
Choices available to decision makers are conditioned by core attributes common to all preserved digital assets and those that apply only within specific contexts.

2.2.1 Core Attributes

Preserved digital assets share four essential attributes as economic goods.

1. **The demand for digital preservation is a *derived demand*.**
2. **Digital materials are *depreciable durable assets*.**
3. **Digital assets are *nonrival in consumption* and create a free-rider potential.**
4. **The digital preservation process is *temporally dynamic and path-dependent*.**

Three of these attributes are common to all forms of preserved assets, including physical objects such as sculpture and paintings, books and maps, and historic buildings and structures. There is one attribute common to all digital information that is not shared by physical objects: its consumption is nonrival. This becomes a critical difference between preservation strategies that are effective for analog materials and those that are effective for digital assets.

Preservation is a Derived Demand

The value that preservation services produce is a *derived demand*—an activity that people undertake in the service of something else that they value. In the case of digital information, people care about the possibility of future access and use, and preservation creates that potential. To have access, they must also have preservation. For example, astronomers might place substantial value on preserving data captured by deep-space observation systems because having access to such data will help them advance their understanding of the universe in the future. The preserved data reflect past historical states, so loss of the data means that possibilities to reuse those data,

now and in the future, are lost forever. Because supply of preservation services follows from a well-articulated demand for preserved materials, it is important to frame the demand for preservation servicess as a demand for the product of those services—the digital materials—for future use.

Derived demand is a well-studied subject in economics and flows from an analysis of the production of goods and services. To illustrate, the demand for labor is a derived demand, for we are really interested in the fruits of the labor. Just as labor is an element of the cost of widgets or washing machines or a four-course meal at a three-star restaurant, so too is preservation an element in the cost of being able to use scholarly literature in research or watch *Casablanca* streamed into your home. The demand for physical capital (or infrastructure) is similarly a derived demand.

It is important to frame the demand for preservation services as a demand for the product of those services—the digital materials—for future use

The mechanisms by which derived demands are articulated in the marketplace are highly variable. Economic principles tell us that markets sometimes do a good job in providing for derived demands, but in other cases markets fail. When they do, nonprofits and public entities have an important role to play. This is particularly true for digital preservation. One reason that markets have limited success with preservation is that the demand is not articulated until well into the future, or comes from stakeholders who are not well represented in the decision-making process. Unless actions are taken in the present, before demand is articulated, there will be no supply of preserved digital materials. There are few market incentives that reward investing in the unknowable value of nonrivalrous goods (we talk about nonrivalrous consumption later). Another reason that markets have limited success with preservation is that demands for digital preservation are context-specific: they vary across disciplines, genres, and types of data. Each context taken separately may not make enough of a market to spur supply. This, combined with the natural human tendency to focus on short-term gains at the expense of the longer term, means that the market today is unlikely to adequately represent the perspectives and values of future generations. Appendix 3 provides general economic thinking about why markets can fail in this situation. When markets fail to meet the demand for societal benefit, the public sector typically steps in.

Because preservation is a derived demand, the decision to preserve will ultimately be based on the perceived value associated with the digital materials over time.

Preserved Digital Materials are Depreciable Durable Assets

A *depreciable durable asset* is something long-lasting that produces a flow of value over time, with the quality and quantity of the flow declining over time if actions are not taken to maintain the viability or productivity of the asset. Thus, any digital object for which we contemplate future use and which will require some degree of preservation to ensure usability in the future can be thought of as a depreciable durable asset.

Depreciation of digital assets occurs in two ways that have economic consequences. First, there is simple physical depreciation, which typically takes the form of substrate deterioration, file degradation, or both. When that occurs, an object may no longer be

identical to the object as it was first created, analogous to a wooden chair drying out and cracking, causing the arms to loosen and the legs to wobble over time.[11] The second type of depreciation is intellectual or logical, which affects the ability to make sense of a digital file that is physically intact, but is in an old or incompatible file format. Dependencies on software and hardware, both subject to relentless change, make technical obsolescence a constant threat. File format incompatibility between different versions of the same software can render a file illegible. As a rule, the more complex the file format, the more effort must go into preserving it. Having well-documented file formats is essential for the future intelligibility of data, so metadata and format documentation are integral elements in preservation. This may add upfront costs for the creation of sustainable digital assets and their deposit into a repository, and it imposes ongoing costs on preservation.

Deciding which form of preservation is most suitable for which digital asset is intimately tied to value. It is important to understand the essential feature of the information that makes it most useful and thus which technical system will keep the digital asset most usable. For some data, migration from one system to another over time will suffice. For others, in which the "look and feel" of the original software and hardware are crucial to its value, emulation is better suited. Experiential and aesthetic criteria usually suggest emulation, as is the case of art works or games. In the case of records, a court will demand that the digital evidence be submitted in the form in which it existed at the time the action under litigation occurred.

Mitigating both the threat of physical degradation and loss of functionality and intelligibility can increase the cost of preservation, but without such safeguards, the ability of users to access and use the information is at risk. We recommend that research be conducted on how to reduce the costs of preservation by engineering automated metadata extraction, reducing power consumption by servers, and instituting other engineering optimizations.

Because digital materials are depreciable durable assets, we must make ongoing investments in their maintenance if we are to sustain their value-creating capacity over time.

Preserved Digital Materials are Nonrival in Consumption

Thus far, our discussion of preservation as an economic good could apply to both analog and digital objects. The third key economic characteristic of digital assets is crucially different: digital assets are *nonrival in consumption* as a general rule. This means that one person's use of a digital asset does not impede or detract from another person's use of the same. Once on a server, a digital asset can be consumed by many users simultaneously; the cost of adding one or dozens more users of the asset is close to zero.[12] Such is not the case for analog objects. If I am reading a copy of a book, you

[11] In preservation, the perfect identity of a current object with its original state is referred to as *authenticity*, that is, when an object is what it purports or appears to be, and is usually highly prized as such.

[12] Of course, there can be degraded performance and congestion if many people hit the same server at once. And there are usually some marginal costs, associated with accounting, management, and so forth, of adding

cannot read the same copy. If a film studio does a wide film release in many theaters, it must make many copies to play in those many theaters. In contrast, once created, digital materials can be shared at essentially no incremental cost, creating significant economies of scale. The economic consequences of this technical characteristic are enormous. Generally, goods that are nonrival in consumption will not be efficiently produced and distributed via free markets.[13] Moreover, from the perspective of producers, the fact that such goods are very easy to copy means that any market model of production and preservation must include provisions to prevent unauthorized (and unremunerated) reproduction.

The fact that digital assets are nonrival in consumption can be of enormous benefit from the viewpoint of users. Data useful to one scientist will be useful to many, and discoveries made from research of those data may be better or more efficient if shared. Astronomical and physical data generated by large surveys and experiments can be posted in ways that make it easily available to scientists around the world. But even this happy scenario poses a *free-rider* problem. For an object of value that is nonrival in consumption, it is in the interest of all actual and potential consumers, now and in the future, that someone pay to make the object available now, and that someone preserve it for future use. But the best outcome for any individual user or potential user is that *someone else* undertake the requisite expense; hence the free-rider problem.

Even where there is a sustainable model that involves charging for access to preserved digital assets—to obviate the free-rider problem—the result can be economically inefficient. Some people who would benefit from access, and would be willing to pay something to get it (though not the fee demanded), will be denied access even though the cost of providing it to them is zero. Consider a museum that charges for entry, excluding people who cannot afford the entry price even though the cost of an additional visitor is often relatively small.[14] The museum policy is designed to cover the long-term and fixed costs of preservation and access. As a technical matter, when goods of any kind are nonrival in consumption, exclusion of individuals is inefficient. This is because the benefits or value the person would realize by being allowed to enter the museum exceed the costs of allowing them to do so. Data services supported by subscription fees are similar. All social scientists could conceivably benefit from unrestricted access to data held in social science research archives, such as the Inter-university Consortium for Political and Social Research (ICPSR).[15] But ICPSR has chosen to impose a fee on users of its archives to support the service it provides. Hence, nonrivalrous consumption creates a tradeoff: often a sustainable economic

more users to a system. There may be even greater costs as system capacities are approached or exceeded; for example, new hardware might need to be procured and software/algorithms may need to be revised to scale past current capacity limitations. But for the most part, one person's reading of a journal article or viewing of a YouTube video or downloading of a book or a movie has no effect on the ability of many others to do exactly the same thing at the same time.

[13] See Appendix 3 on market failures in these circumstances.

[14] In practice, museums often use multipart pricing models distinguished by fee reductions for certain times of the week, or for students or senior citizens. Tiered pricing is frequently used in such circumstances to mitigate the societally undesirable exclusion of people with limited means.

[15] See http://www.icpsr.umich.edu.

model can be devised that excludes use by those unwilling to pay a given fee; however, exclusion in this way results in some loss of economic benefit to those excluded, at no saving in cost, and hence there is a loss of economic efficiency.

Preserved digital assets are nonrival in consumption because once one party preserves the assets, they are for all intents and purposes preserved for all. In these circumstances, the incentive for any single party to incur the cost of preservation is weakened, since the other parties can *free ride* on the benefits.

Our ability to use preserved materials in the future, and the cost and quality associated with that use, are affected by what we do today

Preservation is Temporally Dynamic and Path-Dependent

There is one more attribute essential to the economic analysis of digital preservation, one that complicates all other attributes: preservation is *temporally dynamic*—it takes place over time. This simple and obvious fact means that our ability to use preserved materials in the future, and the cost and quality associated with that use, are affected by what we do today. Moreover, the relationship between current behavior and future possibilities is always at stake: it is modified by what has been done in the past and is always modifying what can and will be done in the future. In formal economic terms, preservation decisions are *path-dependent*.

When devising preservation strategies, time is best understood not by the calendar, but by a conceptual framework, such as the information lifecycle model. Here we see time compressed into imprecisely defined spans, such as now, soon, and later; or stages of the process, such creation, use and reuse, and storage. By making a set of reasonable simplifying modeling choices, we can reduce the infinite range of possibilities into a finite and tractable set of alternatives. The moments of greatest vulnerability occur with regularity at certain junctures in the lifecycle—at the point of creation, when decisions are first made about the long-term value of the data and preliminary decisions are made about whether to preserve or abandon an asset; at the point when the current owner or custodian of the asset determines that the cost of preservation outweighs the benefits of access to the asset; and at any moment when there is a transfer in physical custody or ownership of the assets.

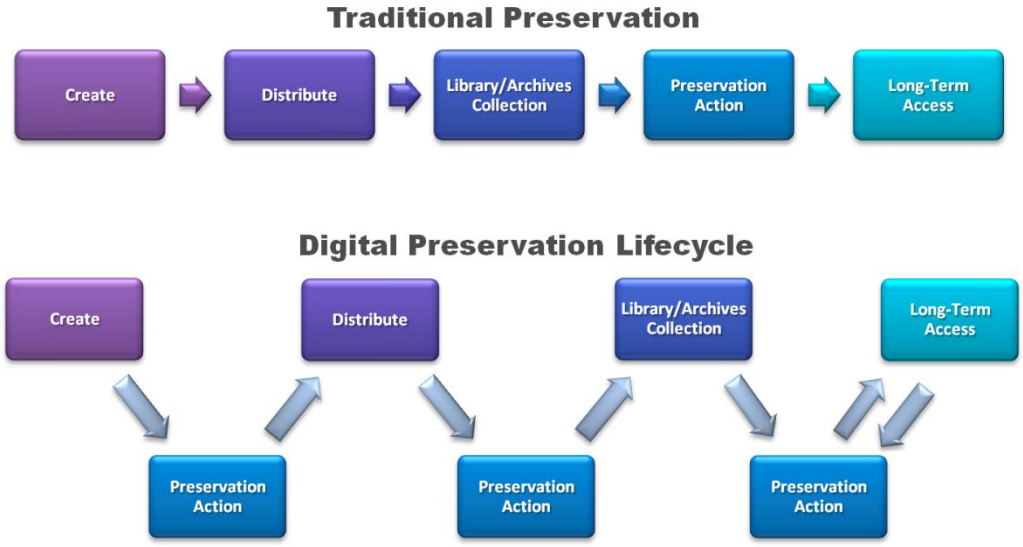


FIGURE 2.3 **Traditional and Digital Information Lifecycles**
In contrast to traditional preservation, digital preservation is a dynamic process with multiple actions taken over the course of the digital lifecycle.
Source: Preserving Our Digital Heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report. A Collaborative Initiative of the Library of Congress, forthcoming. Used with permission.

Figure 2.3 shows that analog preservation is largely linear with a focus on fixing information to physical objects; the conservation of these physical objects becomes the mode of preserving the information. Microfilming a deteriorating book or copying an audiotape to preserve information becomes important from time to time. But such copying or reformatting of physical artifacts is not a common occurrence, unlike the copying and backup that is necessary in digital archives. Digital preservation requires more active management of materials over time and demands different resources at different moments. Above all, the legal provisions that protect ownership and access in digital assets create technical barriers for preservation where there are none for analog preservation. (For more detailed discussion of the salient differences between these two types of preservation, see Appendix 2.)

However one chooses to reduce this infinite range of possibilities into more finite increments, the temporal nature of digital preservation poses unavoidable problems. They include *irreversibility* and the *coincidence and divergence of interests*.

Irreversibility is a well-known if still troubling problem. If data (and enough information to render them) are lost, they are lost forever. Further, if the metadata and other documentation that renders them intelligible are lost, even well-preserved bits may not be usable. But time can wreak havoc not only on data, but on stakeholder interests as well. Over time the interests of stakeholders sometimes align and sometimes diverge. In each preservation scenario we can distinguish three pivotal roles—*users* or beneficiaries who demand preserved digital objects, *owners* of digital assets, and *archives* which together supply preserved digital objects. Preservation works

best when the interests and actions of these users, owners, and archives can be aligned in an economic strategy and operationalized in a business model. When the interests (perceptions of value) and actions (incentives) are in alignment, sorting out the roles and responsibilities of the three entities can be straightforward, even if there is jockeying among the three about who pays "the most," that is, more than they want to. By the same token, long-term preservation is more difficult when the roles are diffuse or fragmented. Preservation is most difficult when these three roles are diffuse and change over time.

The critical nature of temporal dynamics and path-dependency is particularly evident at junctures when handoffs of digital assets occur, handoffs that usually result from changing perceptions of value and incentives. One set of handoffs is technical in nature, when the systems for physical or logical preservation must be updated. At such points, there will likely be some level of data loss, and skilled preservation personnel think deeply about "what we can afford to lose" when they design their systems. Another type of handoff occurs when there is a change in organizational custody. If a commercial company were to let go of its assets because they are no longer profitable, or because the company is going under, it becomes imperative that assets determined to be of long-term public value move to an organization that can and will assume their stewardship. Domain experts also report vulnerabilities when a stewardship system that requires personnel with highly specialized knowledge and skills undergoes staff turnover and is unable to replace needed expertise. In short, there are technical risks, inter-institutional risks, and intra-institutional risks that will cascade over time.

The path-dependent nature of preservation decision making means that decisions made at any time shape future conditions and determine the range of future choices.

Context-Specific Attributes

In addition to core attributes, there are features specific to a data type or user community that constrain choices among preservation strategies. *Context-specific attributes* vary in significant ways among scenarios, because they map to the particular type of digital content being preserved, such as observational data of natural phenomena, individual works of creative expression, or interactive games with multiple players.

Engaging context experts to identify these context-specific attributes is critical for a designing a sound preservation strategy. Understanding what it takes to sustain data generated by satellites in Earth's orbit or by the Large Hadron Collider requires knowledge not only of the data and its technical requirements, but of the needs of knowledge-domain experts, be they climate scientists or particle physicists. Developing strategies for the preservation of cultural materials requires similar expertise, because they, too, come with an array of considerations that determine the choice variables for preservation strategies. These might include intellectual property rights that affect conditions of access, and the experiential or phenomenological aspects of digital cultural materials. The following chapter provides examples of how context-specific attributes play out in a variety of ways in the four content areas of our focus.

BOX 2.2 **Context-Specific Attributes**

Context-specific attributes shape or constrain choices in sustainable preservation strategies. The following attributes, arranged along vectors of value/incentives/roles and responsibilities, can vary according to what is being preserved. The list, while not comprehensive, highlights the range of possibilities within a given scenario.

Value

- Digital assets have value in the aggregate (and can/cannot be sampled for preservation); or as individual digital objects (with diminishing value in preserving the entire class of knowledge assets)
- Digital assets represent historical information such as cultural or natural phenomena that no longer exist; data gathered in clinical trials or other instances where ethics rule out repeatability; or data that can be easily replicated
- Patterns of demand are distributed over time with relatively intensive use in the short or medium term with use declining over time; with steady intensity over the long term; or in some combination
- User communities are the same over the short and long term (such as users of academic e-journals); or are two distinct groups, with little overlap between short-term users and long-term users (as might be the case with blogs)

Incentives

- Incentives for decision makers to act in the public interest are aligned with beneficiaries and providers; or are misaligned and at odds with beneficiaries and providers
- Digital assets are protected by copyright; must be protected for confidentiality or fiduciary reasons (clinical-trial data, communications between lawyer and client); or are in the public domain and legally non-excludable
- Costs for preservation are frontloaded, with operating costs relatively low; low at the point of ingest or acquisition but with relatively high maintenance costs; or some combination of the two
- Costs of preservation are borne by beneficiaries and users directly; by intermediaries such as libraries who represent the interests of users; or by volunteers with a mix of incentives, such as content collectors or scientists who contribute their time to develop and maintain a common resource

Roles and Responsibilities

- Key actors in preserving digital assets (the beneficiaries, the owners, and the preservers) are well-defined; diffuse; or some combination
- Roles and responsibilities of stakeholders for preservation are clear or fall into a natural allocation; unclear and/or unspecified; or clear and well defined in the present but will not be in the future

2.2.2 Choice Variables

So far, we have focused on factors that are beyond the control of decision makers—those intrinsic to a given environment, intrinsic to the process of preserving digital materials, or intrinsic to the materials themselves. The last piece of the puzzle in sustainable preservation is the one most familiar to preservation stakeholders—the choices available to decision makers in designing sustainable preservation strategies. In most cases, the appropriate strategy for economic sustainability will emerge from a combination of the conditions prevailing in the environment and the choices made by various actors.

Choice variables are not attached to any particular class of digital assets and we find many of them across a spectrum of preservation strategies. Some contexts allow more flexibility for decision makers than others. There may also be cases when choices make economic sense, but do not meet the accepted norms of a given community. Charging users fees to read or contribute to a blog is not a viable choice, for example.

Choice variables are found on both sides of the supply and demand equation. In most cases, preservation decision makers have some discretion over the following questions:

- ***Who benefits*** **from use of the preserved asset?**
- ***Who selects*** **what to preserve?**
- ***Who owns*** **the asset?**
- ***Who preserves*** **the asset?**
- ***Who pays*****?**

Who Benefits

A decision maker may have the discretion to choose between restricting the benefits to a limited group, or making them freely available to all. For example, ICPSR manages and preserves collections of social science research data. Some of the data are made available free of charge, but other data are accessible only if the researcher is affiliated with a dues-paying member institution, or pays a per-use fee. In contrast, the Worldwide Protein Data Bank makes its data collections freely available to all. Funding for the effort is provided not by the direct users of the data (either the researchers or their institutions), but by a variety of public agencies, institutes, and philanthropic organizations. If either the free-access or pay-per-use model proved to be unsustainable over time, decision makers would need to revisit their options and make a different choice.

Who Selects

Similarly, decisions must be made about what to preserve. In most cases, selection criteria are highly flexible. For example, the Sloan Digital Sky Survey (SDSS) manages a range of data types including raw data collected directly from instrumentation and

processed data representing derived forms of raw data.[16] Generally, both the raw data and the processed data are preserved, but this need not be the case. Because the processed data can usually be recreated from the raw data, it may not always be necessary to preserve the former along with the latter. Wikipedia presents another interesting selection policy: currently, the online encyclopedia chooses to preserve a complete record of all changes made to each entry, as well as community discussion related to these changes. Other preservation contexts adopt more circumscribed selection strategies. For example, the Law Library of Congress collects and preserves a set of legal blogs it believes are particularly significant; it does not attempt to comprehensively preserve the entire legal blogosphere.

Who Owns

Ownership and control are important determinants for sustainability, because the individual or group who owns or has effective control over digital materials often has ultimate power, as well as responsibility, for developing and implementing preservation strategies. Sometimes, decision making can be decentralized: that is often the case with institutional repositories. Repository services are made available to faculty members, who then decide whether or not they will deposit data sets, preprints, learning objects, and other materials. In some cases, faculty members even control the level of access permitted to the materials they deposit. In other contexts, preservation decision making is centralized within proxy organizations. For example, a research library acts on behalf of its constituent faculty and students to take the steps needed to ensure that its collections remain accessible and usable over time. In still other cases, it is society that decides preservation strategy, through public agencies that issue preservation mandates to ensure that certain public objectives are fulfilled. The Federal Aviation Administration, for example, mandates that aircraft companies preserve aircraft design data as long as a particular aircraft model is in operation.

Who Preserves

The allocation of preservation responsibilities can vary widely and is usually bound up in such issues as ownership, perceived economic value, and existence of internal preservation capacity. The increasing volume of movie studio output in digital form, which is a highly valuable corporate asset, is for the most part managed internally by technical specialists employed by the studios themselves. However, this is not always the case: many analog films have been turned over to third-party archives, such as the Library of Congress or the UCLA Film Archives, for long-term preservation in the public interest. Scholarly publishing has seen content owners, usually publishers, cede digital preservation responsibilities to third-party services such as JSTOR or Portico, while other publishers have chosen to keep this responsibility in-house.[17] The Internet Archive is an initiative explicitly set up to take on preservation responsibilities that others do not exercise.[18] It collects and archives regular snapshots of the Web, although website owners can opt out of the archive if they choose.

---

[16] See http://www.sdss.org.

[17] See http://www.jstor.org; http://www.portico.org.

[18] See http://www.archive.org.

The passage of time is not just a potential threat to digital asset longevity; it can also be a source of flexibility and discretion to decision makers when sustaining valuable materials for long periods of time

Who Pays

In some cases, the choice of who will pay is inherent in the nature of the organization vested with the responsibility to preserve. National libraries and archives are funded by society as a whole, through general taxation mechanisms. In other preservation contexts, the choice can be more explicit. JSTOR, a not-for-profit e-journal repository service, allocates the costs of its services to the intermediaries or proxy organizations—universities, colleges, museums, libraries—that act on behalf of affiliated scholars who use the journals for research and learning. ICPSR adopts a hybrid approach, charging membership fees to institutions, or charging per-use fees to unaffiliated individuals. Finally, on-demand digital archiving services such as Iron Mountain allocate the costs of preservation to those who directly benefit—those that choose to purchase their services, such as a company wishing to archive e-mail or other internal documentation.[19]

2.2.3 Summing Up

Viewing preservation through an economic lens shows that there are structural and systemic problems in providing sustainable digital preservation. These structural problems can, however, be offset by certain choices and remedies that reduce barriers to sustainability. In the next chapter, we take a closer look at those remedies, focusing in particular on the effects of the path-dependency and temporal dynamics inherent in digital preservation. Somewhat paradoxically, it is the very dynamic nature of digital assets and their value that offers the key to ensuring sustained access. The passage of time is not just a potential threat to digital asset longevity; it can also be a source of flexibility and discretion to decision makers when sustaining valuable materials for long periods of time.

[19] See http://www.ironmountain.com.

Chapter

3

Addressing Economic Risks to Sustainability

Ensuring that valuable digital assets will be available for future use is not simply a matter of finding sufficient funds. It is about mobilizing resources—human, technical, and financial—across a spectrum of stakeholders diffuse over both space and time. However, the core attributes common to all digital assets, explored in the previous chapter, create fundamental structural problems for preservation. An economic good having the core attributes of being a derived demand, a depreciable and durable asset, nonrivalrous in consumption, and temporally dynamic and path-dependent will always encounter problems aligning incentives among the beneficiaries, owners, and preservers.[20]

Ensuring that valuable digital assets will be available for future use is not simply a matter of finding sufficient funds

To determine how precisely those attributes may pose problems for sustainability, we can ask:

- ***Who benefits*** **from use of the preserved asset?**
- ***Who selects*** **what to preserve?**
- ***Who owns*** **the asset?**
- ***Who preserves*** **the asset?**
- ***Who pays*****?**

If answers to these questions are *all the same*—when those who receive the greatest benefit from access to digital assets are the ones who decide what to preserve and are in a position to ensure that the assets are preserved—then the prospects for sustainability are bright indeed. Under those circumstances, there will be close alignment of perceived value, incentives to preserve, and appropriate organizational arrangements and allocation of resources to sustain preservation. As a rule, the more closely aligned the distribution of benefits and costs, the more sustainable the activity.

[20] A good mechanism design should suggest ways to strengthen incentives where possible. For more about mechanism design, see Appendix 4. In this chapter, we take up ideas suggested by this field as they apply to preservation.

However, we have also seen that *this can never be the case with preservation* because so many benefits flow to future generations, as well as to users who are neither proximate to nor have control over the assets. Our case studies indicate that there are numerous ways to compensate for poorly articulated demand, gaps in incentives, and sustained periods of technical, organizational, and financial changes.

At first blush, the particular nature of preserved assets—especially their nonrivalrous consumption, temporal dynamism, and path-dependency—are characteristic of goods that often require public provision. And indeed, public provision is important for preserving assets with long-term societal value. But it is far from the only form of provision possible, particularly in the short run and in early stages of the digital lifecycle. Digital materials are a vital and significant resource for contemporary commerce, public affairs, entertainment, research and education, and most sectors of the economy that rely on information. In many cases, people acting to ensure that they have continued access to the assets under their control are effectively preserving materials of value to them—historical, cultural, informational, or other kinds of benefits—for at least some period of time. In this chapter, we focus on the remedies that help mitigate some of the most common systemic risks to sustainable preservation.

3.1 Systemic Problems with Demand

3.1.1 What to Do When Demand is Diffuse or Weakly Articulated

Proxy organizations broaden the base of demand and provide efficiency of effort by representing numerous stakeholders who do not have the wherewithal to take timely actions

Sustainable digital preservation needs a demand sufficient to cover the cost of supply. But demand is often spread across present and future generations, and also across diffuse communities. Who articulates the demand of future users? This is where proxy institutions, sanctioned to act on behalf of present and future stakeholders, come into play. Proxy organizations broaden the base of demand and provide efficiency of effort by representing numerous stakeholders who do not have the wherewithal to take timely actions.

Proxy Organizations

Libraries, archives, and museums are the most familiar proxy organizations representing the interests of one or more stakeholder groups, both in demand and in supply.[21] We can expect them to continue in the digital realm, and public and private funding should continue to flow to support their public missions. But they are not the only important agents. In cohesive communities with shared values, professional societies such as the American Association for the Advancement of Science and the Academy of Motion Picture Arts and Sciences can raise awareness, aggregate demand, stimulate supply, and specify solutions that meet the targeted needs of their domain. Where there are no such institutions, they should either be created within a domain, or domain experts should reach out to partner with libraries, data repositories, and other proxy organizations with which they can develop a trusted relationship.

[21] See Appendix 5 for more about proxy organizations and the representation of stakeholders.

Various stakeholder communities in science have created proxy organizations to represent their data interests, including addressing criteria for selecting materials with long-term value. The Data Preservation Alliance for the Social Sciences (Data-PASS) takes a broad view of its community to identify, acquire, and preserve high-value data.[22] The Astrophysical Research Consortium (ARC) is responsible for representing the interests of domain scientists in SDSS. An ideal and efficient stakeholder organization would articulate constituent demand, promulgate best practices, develop selection criteria, and have the authority to enter into contractual relationships with archives.

3.1.2 What to Do When Future Value is Uncertain

When the future value of an asset is uncertain, the likelihood of long-term preservation can well depend on current use cases. Preserving materials with clearly defined current uses implicitly creates the option of having the assets available for as-yet-unknown uses that may emerge in the future. Therefore, to the extent possible, value propositions should focus on the benefits generated for current users—about whom decision makers are understandably most concerned—rather than focusing too much on benefits to future generations and unknown future uses. But sometimes this is just not enough.

Option Strategy to Avoid Irreversible Loss

Given the path-dependent nature of digital assets, it is important to think of selection decisions that leave options open, but are not necessarily long-term or open-ended commitments that can deter responsible decision makers. Hedging against irreversible loss is especially important for certain classes of assets such as historical records, observational data, or clinical trial data that are simply irreplaceable. In such cases, the benefits of preservation can be compellingly articulated in terms of negative benefits—the costs that will be incurred if digital assets are not preserved.

BOX 3.1 **Option Strategy for Materials of Uncertain Long-term Value**

When future conditions are particularly uncertain—as they are for prospective future demand for many categories of digital materials—it is often economically justified to make a small current investment that in effect purchases the option to make a choice sometime in the future. For example, if the decision is made not to preserve a set of currently popular, high-traffic blogs because of uncertainty over future interest or use, this decision is irreversible; it cannot be changed later should a clear and compelling demand for those materials emerge. On the other hand, if a small investment were made in saving the blogs in secure storage, with no additional curation, decision makers could choose to intensively curate the blogs at some future date, should conditions warrant. Preservation decision makers facing uncertain future demand for at-risk digital content should consider *purchasing an option* to postpone the final preservation decision until future uncertainties are at least partially resolved.

In cases where materials may have future value but that value cannot be projected with reasonable confidence, the *option strategy* of modest initial investments in preservation

[22] See http://www.icpsr.umich.edu/DATAPASS.

can be a prudent hedge against irreversible loss. Techniques of archival appraisal have been used to assess future value of analog materials in similar cases of uncertainty. These techniques work both in cases where the underlying or intrinsic value of the data is uncertain, and when there is a body of materials with an aggregate value but resource restrictions dictate that only part of the whole can be preserved. (If we cannot save all of a given set of digital materials, how do we develop a sample that represents the whole in a meaningful and authentic way?)[23]

The purpose of the option strategy is simple—to buy time and wait until better information is available about the value of the content in question or preservation techniques have become more efficient. Each option strategy could be given a term limit, with preservation managers holding materials for a specific period of time; then, near the end of that time, they could decide to renew, hand off, or abandon their stewardship. This might be an appropriate strategy for large sets of observational data or targeted sectors of the blogosphere.

3.2 Systemic Problems with Supply

3.2.1 What to Do in the Case of Insufficient Incentives

Public policy can provide a useful nudge or necessary intervention in the absence of sufficient incentives to preserve.[24] But public policy can also be effective and efficient when it strengthens existing private or institutional interests for investing time or money to preserve valuable materials. In nonprofit endeavors, factors that motivate such investment could be pleasure, prestige and reputation, and other nonmonetary rewards. In the commercial sector, the incentives are likely to be to be monetary, though not exclusively so. Contributing to the public good and gaining prestige can be powerful motivators in the private as well the public sector. Policies should aim to align with private interests when possible, providing a stimulus for private parties to act on societal as well as private interests.

FIGURE 3.1 **Digital Preservation Risks and Remedies**
All decision makers face risks to sustainability over the course of the digital lifecycle. In each context there are usually several choices available that will reduce risks long-term. Which remedy a decision maker opts for depends on the context: what the perceived value of digital information is; which incentives will be most compelling to various stakeholders; and which organizational strategies are best suited to a particular stakeholder community.

[23] The real options theory of investment is further explained in Appendix 6.

[24] Policy frameworks for preservation, a component of sustainability as important as an economic framework, are outlined in Appendix 7.

Lowering Barriers to Preserve

Insufficient incentives can sometimes be remediated simply by lowering barriers for those willing to act. For example, people who contribute to collectively create Web content, such as open-source software, may want to have their materials and contributions preserved and reused. But they may not have the means to do so. However, if these creators would use a license (such as Creative Commons) to grant nonexclusive rights to others to preserve the content, the incentives for those who control the content could be closely aligned with those who want to preserve it.[25]

Giving Incentives to Private Parties to Preserve in the Public Interest

Private parties—commercial or noncommercial, corporate or individual—play crucial roles in preserving materials they own or have control of. In the case of analog cultural materials, for example, there have long been a number of incentives for private companies and collectors to curate and preserve cultural materials for public benefit. With analog cultural artifacts, policies aimed at strengthening incentives to preserve in the private sector and allocating preservation responsibilities among private owner and stewardship organizations relied on two mechanisms: (1) copyright incentives for owners to preserve materials for private purposes; (2) a suite of financial incentives for owners to preserve in the public interest. Both of these are taken up in greater detail in Chapter 4, where we address preservation of privately held cultural materials.

Often overlooked is the important role of private collectors in identifying valuable materials, collecting and curating them, and ensuring their stewardship at their own expense. These individuals are likely to be influential early collectors of Web-based materials, for example, by downloading and curating materials of cultural significance long before museums and archives have settled on what to collect. Individuals who donate analog collections to public institutions receive tax benefits. Owners of digital materials should enjoy the same type of tax incentives when they donate them to stewardship organizations. This would send a significant signal to society that digital cultural assets are valued and that private citizens have an important role to play in the stewardship of our shared culture.

Imposing Mandates to Preserve

Mandates to preserve can be effective in some circumstances, such as ensuring access to public records by guiding the policies of public archives, or protecting public safety by mandating preservation of design records for public structures. Mandates in other contexts can be equally effective, such as those that funding organizations place on grantees to be accountable for preserving the information produced under a grant. To be effective, however, mandates must be closely monitored and enforced, and there should be penalties for noncompliance. Monitoring compliance and imposing penalties can become a burden on funding agencies whose priority is to grant funds to advance knowledge, not police grantees.

Mandates work only to the extent that there is capacity to fulfill them

More significantly, mandates work only to the extent that there is capacity to fulfill them. There must be enough trustworthy and accountable archives, and well-defined

[25] See http://creativecommons.org.

roles and responsibilities among those who are charged with fulfilling the mandate. Funding agencies and government authorities, who are usually the ones in a position to mandate preservation, should consider specifying that the university or research institution with whom the grantees are affiliated—rather than the individual grantees—be held accountable for preserving funded research results. In addition, the mandate should be clear about which materials warrant preservation and which do not. **A mandate to preserve should never be construed as a mandate to save everything**.

3.2.2 What to Do in the Case of Misaligned or Competing Incentives

Because preservation is a derived demand—the demand is actually for access— a most effective incentive to preserve is to bundle preservation with access services. This is typical in the commercial sector where an entity is the producer and owner of an asset and represents, via derived demand, the interest of future users in the marketplace. The commercial entity will not, however, always fully represent the interests of all future users. Owners of this type of content do not necessarily take into account the societal or public implications of their private preservation decision making. When private decision making does not take into account the benefits an activity confers on society as a whole, it often leads to an under-provision of this activity from a societal standpoint. One textbook remedy to ensure adequate production of such goods would be a government subsidy of the desired activity—in this case, preservation in the public interest—to increase the provision of the activity to a level commensurate with the societal optimum. This principle could hold in all cases where preservation decision making is being done exclusively by private entities, even though public interests are at stake. The motive to protect the public interest and private assets might well be addressed through ongoing public-private partnerships, which are addressed below.

Copyright and Licensing

In analog preservation and access, copyright law provides incentives to create valuable content by giving owners the exclusive right to provide access for specified periods of time. But it also grants limited rights to qualifying public institutions to preserve the same content in the public interest. This alignment between the public and private interest does not exist in the digital realm, for a number of technical reasons. The Copyright Office and the Library of Congress have made recommendations to change the law in light of digital technology realities.[26] We urge Congress to take up this matter expeditiously. That said, most commercial digital assets are licensed, rather than purchased directly from the owner. Copyright legislation does not apply in such cases. This means that the owner has full responsibility for preserving assets, not only in their own private interest, but also for the public.

This problem has long been acknowledged with e-journals. Many current e-journal archiving initiatives have responded by explicitly transferring preservation

[26] One of the recommendations of *The Section 108 Study Group Report* by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress is to qualify museums for archival exception for the first time. See http://www.section108.gov/docs/Sec108StudyGroupReport.pdf.

responsibilities to a third party committed to long-term preservation, such as JSTOR, Portico, or the Koninklijke Bibliotheek (KB).[27] However, securing this responsibility does not guarantee sustainable preservation. The incentive to preserve now resides with an organization willing to act on it. They in turn need to secure appropriate funding. In the cases we cite, each archives has arrived at a funding strategy most suited to their constituencies—charging fees in the first two instances, and receiving public funds in the third.

Another possible way to strengthen preservation incentives is for stakeholders to aggregate and leverage their demand-side power to negotiate preservation requirements as a part of access agreements. An example of this is the NESLi2 Model License for Journals in the United Kingdom, used by the Joint Information Systems Committee (JISC) in negotiating e-journal licenses on behalf of UK higher education institutions.[28] The original suppliers of the asset—the scholars who produce the journal articles—or their institutions could require a perpetual, nonexclusive license to their works that cannot be transferred to third parties. If such practices were widely adopted, institutional repositories, libraries, or third-party preservation services would likely preserve most scholarly works. This would help restore the alignment of interests and roles that exists in the print regime. This, too, would not resolve issues about financial support, but it would resolve the matter of who owns and who pays.

Creating Public-Private Partnerships

Strong and robust mechanisms for partnering between public and private sectors are necessary to ensure the long-term public interest in private assets

Because a significant portion of digital assets with long-term value are privately created and owned, strong and robust mechanisms for partnering between public and private sectors are necessary to ensure the long-term public interest in private assets. This is an urgent need in all four of the areas we address: in scholarly discourse, where a significant sector of scholarship is owned and published by private companies; in research data, where both private and public sectors have important claims in long-term access to data; in commercially owned cultural content, much of which is privately owned for increasing periods of time and thus potentially puts at risk the long-term public interest in its preservation; and in collectively produced Web content, where many of the platforms that enable content to be freely exchanged and shared are privately owned and not transparent about their long-term intentions for their digital assets.

Creating productive and stable working relationships between the public and private sectors often requires developing a trustworthy (that is, transparent and accountable) intermediary to effectively align the benefits, incentives, and roles and responsibilities among partners who work with very different business models and are motivated by very different incentives. In such partnerships, the first step is to develop trust among partners. This is frequently done by working together to solve common problems that cannot be addressed by a single community. (Preservation offers fertile ground here.) In the research community, the Worldwide Protein Data Bank is one type of organization that manages effective partnerships between the public and private

[27] See http://www.kb.nl/index-en.html.

[28] See http://www.nesli2.ac.uk.

sectors. For e-journals, LOCKSS is a consortium of libraries collaborating to manage their own preservation.[29] The third-party archiving organization Portico was created to serve the long-term interests of scholars and scholarship by providing a mechanism for both publishers and libraries to support preservation.

In the area of multimedia production, there are grassroots efforts under way to organize the development and promotion of digital preservation best practices. One such effort, Cinegrid, is a volunteer organization that brings together media experts from many domains of knowledge and practice to develop and promote the use of technologies for media asset management that includes, but is not limited to, preservation.[30] For materials that are held in private hands for extended periods of time—as copyright allows—there is a powerful need for ongoing interaction between commercial and cultural heritage institutions throughout the digital lifecycle, to ensure a continuous chain of stewardship by preparing for an event that may trigger a handoff, sharing best practices, coordinating selection decisions, and so forth. Society cannot wait for the single handoff point when the economic value of the digital content is exhausted.

National organizations play a critical role in sponsoring long-term public-private partnerships. At the beginning of its decade-long initiative to address the preservation needs of the nation, the Library of Congress convened both public and private sector stakeholders to design their strategy.[31] Important elements of that strategy include convening public and private partners on an ongoing basis under the aegis of the library; organizing public-private working groups on contentious issues such as copyright; and supporting long-term preservation needs of the creative industries by working through trade organizations to support the development of best practices.

Clarifying Rights Status

Preserving digital orphans is a priority for public institutions

One of the critical roles that national organizations such as the Library of Congress, the Smithsonian Institution, and the British Library can play is to clarify the rights status of materials on the Web whose ownership and provenance are uncertain. As long as they are uncertain, public institutions wishing to archive them may demur. These institutions, as well as other public libraries and archives, should step into the breach to collect and preserve materials at high risk of loss because they may be privately owned but economically orphaned, their owner unwilling or unable to preserve. Digital materials can also be at risk when their ownership status is uncertain. Preserving *digital orphans* is a priority for public institutions.

Two national organizations operating under the aegis of the Library of Congress play important roles in identifying cultural assets of long-term cultural value. The National Film Preservation Board and the National Recording Preservation Board, comprising

[29] See http://lockss.stanford.edu.

[30] See http://www.cinegrid.org.

[31] See 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.* Available at http://www.digitalpreservation.gov/library/resources/pubs/docs/ndiipp_plan.pdf.

experts in the public and private sectors, convene annually to select films and sound recordings of historic significance to be preserved.[32] By building national registries of film and recorded sound treasures, they have set a precedent in declaring the public value of privately created cultural assets and have provided for their preservation. Analogous efforts on behalf of Web-based cultural assets are needed, and appropriate governance and funding of selection and preservation should be a public priority.

3.2.3 What to Do When Roles and Responsibilities among Stakeholders and Actors are Unclear

Organizations play a vital role in preservation because stewardship is responsibility that spans generations. Preservation mandates should allocate responsibilities and invest those responsibilities not in individuals but in responsible organizations. (Such would be the case if a funding agency were to make the university or research institute, rather than a principal investigator receiving funds, responsible for fulfilling preservation mandates.) Institutions should make clear and explicit allocations of responsibilities among preservation stakeholders. This means developing and promulgating preservation policies within organizations that specify such roles and responsibilities.

Sometimes the allocations of responsibility must be legally binding, as in the case of archiving services such as ICPSR or Portico. A third-party agent trusted with long-term stewardship of assets they do not own should be open and accountable to its constituents, and accountability begins with a Memorandum of Understanding (MOU) or service-level agreement specifying processes and outcomes. Less formal alliances that allow participating institutions to allocate responsibilities among themselves can also be effective. The International Internet Preservation Consortium (IIPC), an international organization of national libraries and stewardship organizations, as well as the Library of Congress-sponsored National Digital Stewardship Alliance, allow members to coordinate their interests and collecting responsibilities.[33] Such alliances are also very effective at leveraging expertise, sharing best practices, and developing economies of scale among large organizations.

3.2.4 What to Do When the Benefits from Assets No Longer Outweigh the Cost of Maintaining Them

Handoffs

Loss of perceived value is a critical moment in the life cycle of a digital asset, and it must be anticipated and planned for. Institutions that judge an asset to have lost its value should consider handing their digital materials to another preserving organization for stewardship. It may often be the case that private institutions make a sound business judgment to no longer support the maintenance of certain assets, but that public institutions will. It also happens that research libraries change their collective priorities as the interests of faculty and scholars change over time and may wish to divest themselves of a specific collection, as has happened over the decades with some

[32] See, for example, http://www.loc.gov/film/filmnfr.html or http://www.loc.gov/rr/record/nrpb/nrpb-masterlist.html.

[33] See http://www.netpreserve.org; the Alliance will be chartered in 2010.

The best practice for all institutions is to have a plan for assessing long-term value periodically, at which point they decide whether to renew, handoff, or abandon a preservation commitment

of their analog collections (for example, the case of some area studies collections and special collections.) They, too, should consider finding another home for the collection. The best practice for all institutions is to have a plan for assessing long-term value periodically, at which point they decide whether to renew, handoff, or abandon a preservation commitment. We have recommended that stakeholders have service-level agreements or MOUs with preserving institutions; a handoff clause should always be part of that agreement, as in the case of the SDSS astronomical data.

Free Riders

Organizations can help dilute the risk of free riding. A commonly used mechanism for avoiding free-rider consumption is to make access to or use of the goods exclusive to a group, usually a fee-paying group. Tiered access is a standard remedy for leveling the costs of access to institutions with different financial means without prejudicing access by the community at large. Excluding access obviously does not work for those assets that communities wish to share freely.

What to Do about the Funding Gap

Even the most compelling incentives to preserve do not generate resources. So how do we find money to pay for current and ongoing preservation? There are several models of cost recovery that may be appropriate in different situations.

Funding Models

There is no single "best" funding model for digital preservation. Selection of an appropriate model requires an in-depth knowledge of the circumstances surrounding the effort, preservation goals, the stakeholder community, and so on. In surveying the landscape of digital preservation activities, the Task Force identified five general categories of funding models used in various forms to allocate resources to digital preservation. Some organizations use combinations of models to fund their activities. Box 3.2 highlights some of the most common funding models for preservation.

BOX 3.2 **Common Funding Models for Digital Preservation**

Internal Budgeting

- An organization's preservation needs are funded out of its internal budget
- The organization views preservation as a cost of doing business and funds it accordingly
- Examples: movie studios and other owners of proprietary digital content; banks, pharmaceutical companies, and aircraft manufacturers preserving financial, research, and design data in compliance with federal policy mandates

Public funding

- Preservation of a certain class of digital materials is undertaken by a public agency on behalf of society as a whole and is paid for by public funds
- Public agencies are allocated funds collected through taxation, fees, and other methods of raising public revenue
- Examples: national archive preserving the records of government, the UK National Digital Archive of Datasets[34]

Voluntary Contribution

- Ongoing access to a digital collection or resource is funded through voluntary contributions from stakeholders, or organizations acting on behalf of stakeholders
- Resource includes contributions from individuals, philanthropic organizations, government agencies
- Examples: Wikipedia, Worldwide Protein Data Bank

Fees for Preservation

- Preservation activities are funded by fees charged to those who benefit from preservation
- The fee is a direct charge for preservation services, rather than an indirect charge based on some other service such as access (see next category)
- Examples: Portico, Chronopolis[35]

Fees for Access

- Preservation activities are funded by fees charged to those who benefit from access services made possible through preservation
- The fee represents an indirect charge for preservation, in the sense that users do not explicitly pay for preservation *per se*, but instead pay for access to preserved materials
- Examples: ICPSR, LexisNexis

[34] See http://www.ndad.nationalarchives.gov.uk.

[35] See http://chronopolis.sdsc.edu.

Reducing Costs

Of course, it is always desirable to reduce preservation costs over the long term. One way to do that is to keep materials at a relatively low curation level and provide curatorial services on demand, which could be done by charging fees on a cost-recovery basis, as libraries do for photoduplication services. Another strategy is to create economies of scale by providing services that benefit many. Bundling preservation and access services to create economies of scope can also be efficient, where appropriate. Dark archives that are dedicated to preservation without providing access have an important place in the preservation landscape, because they provide technical backup and protect materials with restricted access. That said, combining the right to preserve with the right to provide access maximizes the prospects of achieving sustainability in many circumstances, while also creating economies of scope.

Current investments in preservation must not be thought of as open-ended; nor are they wholly separate from access costs

It is not possible to generalize about ongoing and future costs. Few preservation managers are able to separate the cost of preserving materials from the cost of making them accessible. And while they can account for expenses within their institutional framework, comparisons of costs between institutions are notoriously misleading. Current investments in preservation must not be thought of as open-ended; nor are they wholly separate from access costs. Achieving cost efficiencies in preservation is complicated by uncertainties about which materials to preserve, in what state, and for how long. Those uncertainties mean that calculating return on investment is essentially impossible. It cannot be overemphasized that selection criteria—decisions about priorities for preservation investment—play a critical role in sustainable preservation. So, too, does sharing the costs as well as the benefits broadly.

Finally, preservation, like all services that are path-dependent, may find a broad and diverse base of funding to be the most reliable. Unanticipated failures in funding can be devastating. As the Arts and Humanities Data Service learned, even public provision of funding is not a guarantee of long-term support.[36]

Public Provision

Public provision of funding is nonetheless critical to sustain assets with societal value. With all the avenues to sustainability we have noted, there are still significant risks to critical classes of assets—research data, digital orphans, educational resources, data used in making public policy, and the creative output of independent artists, among many others. Trusted public institutions at the local and national levels should be funded to build the critical digital preservation capacity that currently falls far short of demand.

3.2.6 It Depends on the Context

In all preservation contexts, one must examine the aims of the various stakeholders to determine whether incentives for preservation exist, and whether it is viable to develop a funding model that covers the fixed and marginal costs of preservation and access. Where incentives for preservation exist, the problem reduces to developing the

[36] See http://ahds.ac.uk. For information on the funding discontinuation, see http://www.ahrc.ac.uk/News/Latest/Pages/AHRCreshapesitsfundingofICTresearch.aspx.

Trusted public institutions at the local and national levels should be funded to build the critical digital preservation capacity that currently falls far short of demand

appropriate funding or chargeback scheme. Where direct incentives do not exist, long-term institutions are a typical solution because they are charged with preserving cultural and scientific assets broadly.

For some content types—most notably for novel Web-based genres—a lack of awareness about potential value is the greatest risk to future accessibility. For more established content, such as e-journal literature and commercially owned cultural assets, the risks are generally in the area of incentives. For all types of content, organizational issues are troubling but can be resolved through diligent coordination of responsibilities. And with respect to funding, because of uncertainties about the costs of preservation over time, short-term strategies with the option to renew, refine, or reverse are clearly better than unrealistic and imprudent long-term ambitions.

3.2.7 Summing Up

If we use time well, it may work for us, not against us. Hedging against uncertainties, postponing decisions when possible, recognizing that benefits, demand, and users will change, anticipating better information over time—these are the habits of mind that mark responsible digital stewardship and will help husband scarce resources while creating enough flexibility for bold moves and rescue of endangered assets when that becomes necessary.

Chapter

4

Sustainable Preservation in Context

Longstanding preservation strategies for scholarly discourse have been disrupted by digital technologies.

Preservation can be understood as an economic activity, with core attributes and context-specific attributes that shape the choices facing decision makers through the entire lifecycle of an asset. But preservation strategies themselves are series of decisions made in very concrete situations in the course of engaging in many other activities—from research and development to publication and drawing up annual operating budgets—all taking place within specific social contexts. In this chapter we look at how preservation decision making occurs in a number of real-world contexts. Our analysis is based on two dozen case studies across four content areas—scholarly discourse, research data, commercially owned cultural content, and collectively produced Web content. These areas are important for science and the humanities, research and education, public policy, cultural heritage, and the creative industries. They also represent different degrees of development of digital preservation strategies.

The specific genres we investigated range from the well known—journal literature and music—to emerging types, such as the blogosphere and real-time online games. In the cases of the former, there is widely recognized value in preserving those materials, relatively well-established allocation of roles and responsibilities among stakeholders, and many decisions are shaped by traditional practices. In the latter, no such practices exist. The long-term value of interactive games or blogging is contested, and techniques for preserving such materials are not yet developed.

The case studies provided a means of applying the economic framework discussed in the previous chapters to the real-world settings in which digital preservation activities take place. They also provided an empirical context within which to assess the utility of the economic framework. In this chapter we present a synthesis of the findings from our case studies, focusing on the salient risks to sustainability found within each scenario, and propose possible remedies. We also identify an agenda of near-term actions to be taken by stakeholders to reduce those risks. A summation of action agendas for the four content areas can be found in Table 5.2. Recommendations that apply to all scenarios are discussed in the following chapter and we do not recapitulate them here.

We begin with the record of scholarly discourse—a body of knowledge with a long history and a relatively mature preservation strategy—and end with the scenario characterized by uncertainties large and small across the board—collectively produced Web content.

4.1 Scholarly Discourse

Scholarly discourse—often called scholarly communication—is the published output of scholarly inquiry: the ideas, theories, analyses of data, assessments of previous scholarship, and conclusions that collectively form the scholarly record. The audience for this literature is a relatively focused, tightly scoped, highly specialized readership that shares a general consensus about the value of the scholarly record as a whole as well as individual contributions to that record. The actors are well-established: scholars who are both creators and users; the research institutions that support them and, by and large, fund the enterprise of scholarship, scholarly publishing, and library preservation; publishers, both commercial and nonprofit, who represent the interests of scholars as authors; and libraries, who represent the interests of scholars as users and represent the diffuse demand for access from different disciplines and across generations.

Preserving scholarly discourse has been an activity of the research and education communities for centuries, and the need to preserve digital versions of scholarship is widely understood as a natural extension of the need to preserve analog antecedents. Moreover, a longstanding, if at times fragile, allocation of preservation responsibilities exists as a holdover from the print era. This balance among the stakeholders has been disrupted by digital technologies primarily in two ways. First, we see the disruption of roles and responsibilities among players, resulting from the nonrivalrous nature of digital assets, which may lead to misaligned incentives and the free-rider problem.

The second disruption comes from the emergence of new genres of scholarly discourse, such as online collaborative spaces, academic blogs, websites, e-prints, and even microblogging (referred to eponymously as Twitter). Except for fields with a tradition of grey literature and preprints—such as physics and economics—there are neither widely shared perceptions of long-term value nor well-established practices for the selection of these emerging materials for long-term retention.

For purposes of analysis we distinguish these two types of scholarly discourse as *established* and *emerging* genres. We recognize that in reality, these distinctions are difficult to make and can be misleading.

4.1.1 Value and Selection

Established forms of scholarly discourse are by nature highly curated, with digital counterparts to print conventions such as footnotes and references, indices, tables, and front and back matter specific to certain disciplines. Scholars want not just elaborate functionality within journals or monographs, but also assurance that the scholarly record will be maintained in its entirety. This record is held by a network of libraries

that has informally allocated collecting responsibilities. Because of the high standards for accuracy, reliability, and end-user functionality required by the scholarly community, the demand for scholarly discourse typically sorts into two strands, which can mean having to preserve two versions of essentially the same thing.

The first is the reader's demand for an *accessible digital object*, the version that can be readily accessed online, surrounded by all sorts of value-added reader services and tools: for example, links to additional information that the reader can find through *Science* or *Nature* online, or through Elsevier's *ScienceDirect* service.[37]

The second demand is that of the preserving institution for a *preservation copy of the digital asset*. This is what publishers, libraries, and third-party archiving services such as Portico manage. (There is another preservation service, LOCKSS, which enables libraries to manage their own archiving.) The demand for archiving services, which rely on a preservation copy, are from the journal providers themselves (either the publisher or the library) and are akin to demand for an insurance policy to protect previous investments in access services like *ScienceDirect* or JSTOR. These preservation copies can vary in significant ways from those optimized for access.

Uncertain Future Value

Many emerging forms of scholarship, such as academic blogs or SciVee videos, are at risk of loss because there are no clear articulations of their long-term value and few, if any, institutions taking responsibility for the long-term preservation of these resources.[38] However, there is much short-term use: many scholars are engaged in significant blogging activities, for example. But outside of those disciplines (chiefly in natural and social sciences) that have had longstanding traditions of preprints and grey literature, most new forms of scholarship go uncollected and unpreserved. At present, preservation interests in emerging scholarly discourse are represented only by individuals who (a) might or might not be aware of the threats to sustainability (in the case of blogs, tweets, and the like) and (b) might have little clout to effect preservation (in the case of projects that are being supported by one-off grant funds).

Libraries will assume preservation responsibilities for new titles when requested by scholars. But selecting titles one by one based on individual scholars' expressions of interest is not scalable, nor does it serve the interests of scholarship as a whole. A remedy is for professional societies to collaborate with scholars to assess disciplinary policies about the long-term value of emerging scholarly discourse. Creating several pilot collecting projects within disciplines does not necessarily impose heavy costs. Web-based scholarly materials can be harvested from the Web and do not represent the same commitment of funds as the purchase of print materials or subscription to academic literature. In this case, an option strategy to make small investments now can mitigate the risk of permanent loss and may be the appropriate action.

---

[37] See http://www.sciencemag.org; http://www.nature.com; http://www.sciencedirect.com.

[38] See http://www.scivee.tv.

Recommendation #1: Libraries, scholars, and professional societies should develop selection criteria for emerging digital genres in scholarly discourse, and prototype preservation and access strategies to support them.

4.1.2 Incentives to Preserve

Whereas in print publication there is a clear allocation of preservation responsibilities, as well as authorities to preserve on behalf of current and future users, that is no longer the case in the digital realm. Current copyright law does not provide authority for the technical preservation that digital materials require, such as the need to create more than three copies. Moreover, copyright law can be overturned by contracts. The scholar grants publishing and distribution rights to publishers, and publishers add review and editorial services and license these materials to libraries so scholars can use them. Publishers have high incentives to preserve these materials, both to secure the best authors and to keep their collective digital assets productive. Yet, publishers may not be in the best position to preserve in the public interest over the long term. The long term for scholarly discourse is at least several centuries, longer than the life of most publishing companies—but by no means all.

Misaligned or Competing Incentives

The misalignment of incentives between publishers and libraries presents a potential risk to sustainability for e-journal literature. Many of the current e-journal archiving approaches overcome this problem by explicitly transferring preservation responsibilities from the publisher to a third party committed to long-term preservation (for example, JSTOR, Portico, KB), in some cases with funding from the publisher. There has been little publisher resistance to participating in dark archive models operating on behalf of the community; such models, if economically sustainable, would provide an adequate preservation solution. But it is important to emphasize that simply securing the responsibility to preserve does not ensure sustainable preservation. While the incentive to preserve does reside with an organization willing to act on it, questions may still exist about who will pay for preservation over the long term.

Third-party archives can play an important role in mediation here, but so, too, can scholars—both in the aggregate and individually—by retaining their own rights in their work and assigning nonexclusive rights to others to preserve. Scholars can aggregate and leverage their demand-side power to negotiate preservation requirements as part of access agreements. An example of this is the NESLi2 Model License for Journals in the United Kingdom, used by JISC when negotiating e-journal licenses on behalf of UK higher education institutions. The license includes provisions requiring publishers to deposit copies of journals with appropriate archiving services. It is not clear whether this model can translate successfully into a much more extended and distributed higher education community such as that in the United States, but as a general matter demand-side stakeholders need to speak with a collective voice to articulate their preservation needs to supply-side actors.

Individual scholars—the first-order suppliers of published scholarship—can make preservation much easier by stipulating that they (and their institutions, where applicable) hold a perpetual, nonexclusive license to their works that cannot be transferred to third parties. Harvard University, MIT, the University of Kansas, and others have implemented such licenses on behalf of the scholars in their institutions. If such policies were widely enacted and such practices adopted, entities other than publishers—notably institutional repositories, libraries, or third-party preservation services—could preserve most scholarly works. The application of *collective bargaining power* would likely help in securing these rights on a widespread scale, rather than on an institution-by-institution basis, although if many institutions adopted similar or identical licenses the effect would be similar. Individual use of such licenses would also lower barriers for the preservation of emerging literature by explicitly sanctioning libraries and other stewardship institutions to preserve. Again, we underline that overcoming incentive gaps this way or any other does not address the question of securing funds for preservation. Both centralized and distributed models of digital preservation must secure an adequate flow of funds to sustain their activities, and in some cases they have successfully done so. But overcoming preservation incentive gaps ensures that those who benefit and those who have the right to preserve are one and the same.

Recommendation #2: Publishers reserving the right to preserve should partner with third-party archives or libraries to ensure long-term preservation.

Recommendation #3: Scholars should consider granting nonexclusive rights to publish and preserve, to enable decentralized and distributed preservation of emerging scholarly discourse.

4.1.3 Roles, Responsibilities, and Funding

With digital information there is little incentive for an individual institution to preserve this record and strong incentives to wait for some other institution to do it—thus is born the free-rider problem

The Free-Rider Problem

The collective interest of higher education means that the scholarly record benefits everybody. But with digital information there is little incentive for an individual institution to preserve this record and strong incentives to wait for some other institution to do it. Thus is born the free-rider problem that looms large in this community, which places high value on equitable access. For now, many research libraries continue past practices for established literature. JSTOR's delivery of products through a tiered pricing model makes access to backfiles of many journals affordable for most institutions.

In addition to JSTOR, other preservation mechanisms have emerged as the volume of electronic production has increased. With some support from publishers, research libraries fund most of the cost of LOCKSS, CLOCKSS, and Portico, which provide backstop preservation services for the formal scholarly literature.[39] We recommend that such mechanisms be explored both for e-monographs and for emerging scholarly discourse to anticipate the free-rider problems in these genres as well.

[39] See http://www.clockss.org.

Recommendation #4: Libraries should create a mechanism to organize and clarify their governance issues and responsibilities to preserve monographs and emerging scholarly discourse along lines similar to those for e-journals.

Funding Models

Preservation can be made more affordable through economies of scale. For example, accessing large corpora of materials through centralized services that serve the scholarly community as a whole (scholars, publishers, and libraries) can be more efficient than preserving a relatively small fraction of materials for relatively narrow audiences. As noted, some preservation efforts bundle access services tightly with preservation (for example, JSTOR); others segregate preservation from access (for example, Portico, KB/Elsevier). That said, all of these models exclude a significant portion of the scholarly community: smaller publishers, products of under-resourced fields, and scholarship produced by independent scholars and those in the commercial sector, for example.

A variety of funding models have been tested to support the integration of services. JSTOR implements what amounts to a two-part tariff, where a one-time archiving fee is charged to participants along with yearly access fees. Portico divides the costs of preservation between publishers and libraries, both of whom Portico perceives to be beneficiaries of its services. In the original KB/Elsevier agreement, KB essentially internalized the costs of preservation. LOCKSS is supported by the libraries that use it. All of these initiatives are making what so far have proven to be successful attempts to sustain content for extended periods of time, and all provide for handoff in the event of economic or business failure. In other words, they go a long way toward meeting the standard of sustainability for organizational responsibility and, with respect to handoffs, accountability and continuity that we advocate.

As we have seen, the supply of access services and supply of preservation services are often completely separable. This is most likely another legacy of the strong preservation traditions of the print culture, in which the copyright law grants rights to libraries and other qualifying organizations for purposes of preservation. This distinction between access services and preservation services persists in the digital world for technical reasons as well, deriving from the different demand for the preservation file and the access copy. This has important implications for the organization of markets for preservation services. We can envision two market scenarios emerging over time:

1. Preservation services bundled with access services, where preservation services are sustainable only if combined with access services (JSTOR).

2. Access-only services (*ScienceDirect* and libraries) operating in parallel with preservation-only services (Portico, CLOCKSS).

One potential risk for preservation funding, for both established and emerging genres, comes from the widely shared view that scholarship should be equally accessible to all. Some organizations—notably PubMed Central—are willing to make content freely

available.[40] But calls for open access that do not link such access policies to funding for preservation over the long term are shortsighted. Any mandate for open access must provide the resources as well as capacity for preservation.

Recommendation #5: All open-access strategies that assume the persistence of information over time must consider provisions for the funding of preservation.

4.1.4 Looking Ahead

In the case of scholarly discourse we see a relatively cohesive community of stakeholders, with a long tradition of collaborating to produce and maintain long-term access to scholarship. This community is grappling with the breakdown of existing allocations of roles and responsibilities, in large part because the consumption of digital scholarly literature is nonrival. In addition, this community faces a series of decisions about how to incorporate the emergence of new, less formal modes of scholarly discourse. These modes must be integrated into an existing ecosystem of well-established, formal publications that play a significant role in the validation and credentialing of scholars themselves. We see a wide variety of publishing and credentialing practices among disciplines and subdisciplines, just as we have in the print regime, with some fields relying on preprints and grey literature to advance the state of knowledge, and other fields having relatively long periods of gestation for producing and publishing scholarship.

BOX 4.1 **Action Agenda for Scholarly Discourse**

- **Recommendation #1:** Libraries, scholars, and professional societies should develop selection criteria for emerging genres in scholarly discourse, and prototype preservation and access strategies to support them.
- **Recommendation #2:** Publishers reserving the right to preserve should partner with third-party archives or libraries to ensure long-term preservation.
- **Recommendation #3:** Scholars should consider granting nonexclusive rights to publish and preserve, to enable decentralized and distributed preservation of emerging scholarly discourse.
- **Recommendation #4:** Libraries should create a mechanism to organize and clarify their governance issues and responsibilities to preserve monographs and emerging scholarly discourse along lines similar to those for e-journals.
- **Recommendation #5:** All open-access strategies that assume the persistence of information over time must consider provisions for the funding of preservation.

We can expect that each field will elaborate and extend disciplinary modes for its discourse within the digital realm, to arrive at a new ecosystem of production and consumption of scholarship. However, this will require stakeholders to reassess their roles and responsibilities to maintain the scholarly record, and create new allocations of

[40] See http:// www.ncbi.nlm.nih.gov/pmc.

resources that can make scholarship as accessible as possible, to both present and future generations.

4.2 Research Data

While scholarly discourse focuses on the outputs of scholarship, research data represent the primary inputs into new research, as well as the first-order results of that research. Research across virtually all knowledge domains is being transformed by digital technologies. Scientific instruments and information technologies extend powers of observation, documentation, and simulation, at the same time that we are able to discover, analyze, and visualize or otherwise represent more data than we could using only analog recording media. Some observers refer to these new modes of doing research as *data-intensive research*.[41] As a direct result, we have seen a new focus on the accessibility, integrity, and stewardship of data, and significant changes in archiving practice among nearly all academic domains.

The pace of innovation in data-intensive research is so rapid that there is always the risk stewardship practices embraced today will be superseded by new ones tomorrow. Strategies and best practices should be flexible enough to adapt rapidly to changes in technology, selection criteria, and data uses. We focus primarily on natural and social science domains of knowledge, but data-intensive research is also rapidly transforming the humanities, which increasingly depends on digital primary sources. Consequently, much of what we say here applies to the humanities disciplines as well.

4.2.1 Value and Selection

Research data vary enormously—in type and volume, as well as in use and long-term value. We considered four types of research data that exemplify particular context-specific attributes.

> *Observational data* come from telescopes, satellites, sensor networks, surveys, and other instruments that record historical information or one-time phenomena (such as astronomical data from SDSS). This category also includes social science research (such as demographic surveys hosted by ICPSR). In many cases these data cannot be replicated and should be retained.
>
> *Experimental data* may be captured from high-throughput machines (such as accelerators), through clinical trials and biomedical and pharmaceutical testing, or through other controlled experiments. Preservation is particularly important for experimental data where it is not feasible or ethical to replicate data gathering. This includes some data dealing with human subjects and endangered species.
>
> *Computational data* are generated from large-scale computational simulations. Although such data can be regenerated by re-running the simulation, there are

[41] See, for example, "Beyond the Data Deluge" by Gordon Bell, Tony Hey, and Alex Szalay, S*cience Magazine* 323, March 2009, pp. 1297-98.

two reasons why computational data may need to be preserved over the medium term (three or more years). First, the data may be used as the basis for substantive and subsequent analysis, visualization, or data mining. Second, time on a computer for additional computations may not be available within a short time frame. This is a common occurrence for very large-scale computations that run on supercomputers shared by the national community, such as those found at Department of Energy national laboratories and National Science Foundation (NSF) centers.

Reference data sets are highly curated data that are often in high demand by multiple scientific communities. Such data are created for purposes that range from mapping the human genome and documenting proteins to amassing longitudinal data on economic and social status. The Worldwide Protein Data Bank and Panel Study of Income Dynamics are such reference data sets.[42]

With all these data, there is often a need to preserve ancillary materials, such as calibrations of instruments, parameters of experiments, and lab notebooks.

While most large research data collections are produced and used by researchers, they are also valuable for public policy. Public policy needs for information about climate, seismology, oceanography, clinical trials and social science research surveys, endangered species, indigenous sites, archaeological sites, and sensitive security matters go well beyond the demands of research, and become a matter of urgent public priority. This secondary public use of research data points to a new set of constituents willing to support long-term access to these data.

Making decisions about what not to retain can be as important as making decisions about what to retain

Uncertain Future Value

Professional societies and other recognized proxy organizations can play crucial roles in resolving issues about what to preserve and for how long. The most sustainable cases we looked at were those, such as astronomical observation data and certain social science collections, that had a proxy organization or governing board (ARC and ICPSR respectively) that was empowered by the community to make assessments about selection priorities. All data-intensive fields will forge a consensus about their own selection criteria, and empowering an organization to do so on behalf of domain experts is an efficient way of creating and maintaining accountability.

Each domain must anticipate that new data, more complete data sets, and the introduction of new instrumentation may render old data obsolete. Making decisions about what *not* to retain can be as important as making decisions about what to retain. In some cases there is a compelling reason to keep data for the indefinite future, such as in humanities disciplines and in matters where ethics demands long-term retention. But in other fields—such as those relying on computational or observational data—the value of older data for research can be eclipsed by subsequent of superior data sets—for example, data from a later run, or readings from a more precise instrument.

[42] See http://psdonline.isr.umich.edu.

Recommendation #1: Each domain, through professional societies or other consensus-making bodies, should set priorities for data selection, level of curation, and length of retention.

4.2.2 Incentives to Preserve

Insufficient Incentives

Even with widespread recognition of long-term value and well-defined selection criteria, data creators may lack strong incentives to preserve.[43] A good rule of thumb is that the incentive to preserve diminishes as the decision-making unit becomes more granular—the most granular being the individual researcher. Generally, and in grant-funded research in particular, preservation is framed as a zero-sum activity. Time and money spent on preservation activities are deducted from the total budget for research. This has major impact on any preservation incentive, whether mandated or not.

Imposition of mandates is commonly recommended as a way to strengthen incentives, particularly mandates imposed by funders on grantees. For such mandates to be effective, there should be clear allocation of some funds (such as some portion of the grant) to support preservation. There also needs to be clear selection criteria in place—this should not be construed as a mandate to save everything—as well as one or more partners in preservation with whom principal investigators can collaborate. Each domain should develop partnerships with data scientists or informatics specialists with domain expertise, to ensure effective curation and preparation for deposit into an archive.

Recommendation #2: Funders should impose preservation mandates, when appropriate. When mandates are imposed, funders should also specify selection criteria, funds to be used, and responsible organizations to provide archiving.

Funders should be seeding capacity broadly within research institutions to support research data. Such efforts might begin by working with university libraries to become stewards of preserved research data. It is also important to gain an understanding of what capacity is needed. Agencies such as NSF, JISC, National Institutes of Health (NIH) and the Wellcome Trust should work with grantees and domain researchers to determine archiving needs and consider how economies of scale might be developed to take advantage of specialized use scenarios. Federal agencies could also provide startup funding for the development of archival facilities as a critical element of sustainable research infrastructure.

Each funding agency should find ways to signal the importance of managing data. For example, NSF and JISC could make data under stewardship a core scientific indicator

[43] See "Incentives for Data Producers to Create 'Archive-Ready' Data: Implications for Archives and Records Management" by Margaret Hedstrom and Jinfang Niu, available at https://www.archivists.org/publications/proceedings/researchforum/2008/papers/M-HedstromJ-Niu-SAA-ResearchPaper-2008.pdf. See also "The LEADS Database at ICPSR: Identifying Important 'At Risk' Social Science Data," by Amy. M. Pienta, Myron Gutmann, Lynnette Hoelter, and Jared Lyle, for further insight into why roughly 75 percent of social science data sets funded by NSF and NIH over a 40-year period are not publicly archived. Available at: http://www.icpsr.umich.edu/files/DATAPASS/pdf/Pienta_et_al_2008.pdf.

and recognize that category as a fundamental scientific asset. As such, regular measures of how much data is produced, preserved, repurposed, and reused should be monitored and reported as a standard practice.

Recommendation #3: Funding agencies should explicitly recognize "data under stewardship" as a core indicator of scientific effort and include this information in standard reporting mechanisms.

In many cases, centralized services can be most efficient when the data require high levels of domain expertise in curation and archiving

4.2.3 Roles, Responsibilities, and Funding

Providing incentives is not the same thing as providing funding. The funds needed to support preservation are often provided through soft money, one-time grants, and voluntary commitments of time (in the case of wwPDB), and hence are fragile. Funding is relatively secure, at present, in cases where there is a subscription model, such as ICPSR, and the free-rider problem is mitigated, though the tradeoff is that some potential users are likely shut out.

There is usually a tradeoff between maximizing access and maximizing funding. In the case of SDSS, maximizing access is critical for astronomical research; consequently, the research community itself provides funds. There is potentially a free-rider problem with SDSS, but there is also a culture of sharing and of amateur and professional coexistence in astronomy. A tolerably small portion of total funding for SDSS goes toward preservation. Importantly, the project has built-in mechanisms to reassess preservation priorities in the MOU it has with its archiving partner. The institutions that collectively support SDSS are likely to continue support for the broader set of interested users notwithstanding the free-rider problem.

In many cases, centralized services can be most efficient when the data require high levels of domain expertise in curation and archiving. Personnel costs are among the highest in any stewardship system. They can be reduced by automation, but they cannot be entirely eliminated. In the case of centralized services, as in all others, there should be agreements between the data community and its archives stipulating access and outcomes. Each agreement should be up for reassessment and renewal or cancellation periodically. These agreements should include a mechanism for handoffs to a new custodian.

Recommendation #4: Preservation services should reduce curation and archiving costs by leveraging economies of scale when possible.

Recommendation #5: Agreements with third-party archives should stipulate processes, outcomes, retention periods, and handoff triggers.

4.2.4 Looking Ahead

The major tradeoffs decision makers face in preserving research data always involve negotiating between current use and creation of new data on one hand, and supporting reliable stewardship of data on the other. It is the classic tension at the core of all preservation decisions: the tradeoff between expending time and resources today

versus investing to create opportunities in the future. There is no single way to resolve this tension. But stakeholders need to take into account that data-intensive research fundamentally changes longstanding allocations of resources among competing ends to support a robust and growing research enterprise.

Over time, the infrastructure that supports research and education in the natural and social sciences may well have more in common with the infrastructure that has long supported humanities, to the extent that they both rely upon the repeated use of resources made available to current researchers through the diligent stewardship of previous generations. For once knowledge production is based on the use of historical, longitudinal, or unique data, the demands for stewardship grow. Economies of scale can be built not only within the domains of natural and social scientific knowledge, but also in partnership with that of the humanities. The former communities could benefit from long expertise in the stewardship of historical and unique materials, and the latter could benefit from the expertise of those long experienced in the use of machine-generated data.

BOX 4.2 **Action Agenda for Research Data**

- **Recommendation #1:** Each domain, through professional societies or other consensus-making bodies, should set priorities for data selection, level of curation, and length of retention.
- **Recommendation #2:** Funders should impose preservation mandates, when appropriate. When mandates are imposed, funders should also specify selection criteria, funds to be used, and responsible organizations to provide archiving.
- **Recommendation #3:** Funding agencies should explicitly recognize "data under stewardship" as a core indicator of scientific effort and include this information in standard reporting mechanisms.
- **Recommendation #4:** Preservation services should reduce curation and archiving costs by leveraging economies of scale when possible.
- **Recommendation #5:** Agreements with third-party archives should stipulate processes, outcomes, retention periods, and handoff triggers.

4.3 Commercially Owned Cultural Content

The matter of *who owns* is uniquely important in developing sustainable digital preservation strategies for commercially owned cultural content. This is because copyright law reserves to the copyright holder the right to preserve, along with some limited preservation rights to libraries and archives. As we know, these limited preservation rights ceded to qualified institutions are not enough to do the job in the digital realm. As a result, misalignments of incentives between owners and controllers of digital content on the one hand and institutions acting in the long-term public interest on the other arise in nearly every case. Analysis of our case studies therefore

focuses on how to create alignments of incentives and partnerships between public- and private- sector stakeholders.

Digital cultural content is highly diverse, ranging from relatively easily managed formats such as journals (of criticism, poetry, social commentary) to less easily managed formats, such as rich experiential media produced by advanced computer manipulations. While the former may be straightforward to preserve from a technical point of view, the latter can be both complex and expensive to preserve. Maintaining the ability to watch a stereoscopic 3-D movie a decade hence in its original presentation, for example, will require preserving and migrating a vast number of production elements. Beyond this, there is the even larger issue of preserving the fully rendered work, for which there can be 80 or more versions in various localizations or presentation formats.

While the cost may be high to preserve digital content, so, too, are the benefits and commercial return on investment. Today's digitally mastered films have the ability to gain audience and produce revenue over a longer period of time than their analog counterparts because digital technologies facilitate repurposing and remixing of older content. This enhanced capability will likely create incentives for content owners to preserve the content. But it could also delay indefinitely the point at which noncommercial entities interested in preservation from a societal perspective enter the decision making process. For all these reasons, public-private partnerships figure prominently in developing sustainable preservation strategies for commercially owned cultural content.

Public-private partnerships figure prominently in developing sustainable preservation strategies for commercially owned cultural content

4.3.1 Value and Selection

Digital content that is privately held includes many genres familiar from the analog world—moving and still images, music, online literature, television, video, and radio. The demand for such materials is well known and reasonably predictable. In addition, there are new genres—such as interactive games like the World of Warcraft—that have not been assessed for long-term value, nor for which preservation techniques have been agreed upon. Much commercially owned cultural content is thought of loosely as "entertainment," and indeed much of it is produced for the sake of entertaining, satisfying curiosity, and giving pleasure. But a good deal of it is also aesthetically, historically, and politically significant. To the extent that the content in all these forms finds audiences, has influence, and expresses something deep and abiding about the human experience, it is preservation-worthy. That is why there is a general societal interest in maintaining access to works of previous generations of creators, often called *digital cultural heritage*. But that demand can be diffuse.

Demand is Diffuse

A potentially serious problem arises from the significant growth in cultural materials that may be of high value but are not commercially distributed or do not gain wide market share. Disintermediation of cultural production—from feature films and documentaries, to photography, literature, and music—has also led to the growth of a so-called free culture. When individual creators disseminate their own materials, the

demand for reuse in remixing is too diffuse to provide a market for preservation services.

Given the risk of irreversible loss of these digital orphans, it is important to capture and preserve this content right away. We recommend that leading preservation institutions, including but not limited to the Library of Congress, the British Library, the Smithsonian Institution, the Koninklijke Bibliotheek in the Netherlands, leading audio-video archives, and others convene stakeholders to map this space and to clarify selection criteria broadly for commercially owned cultural content. Determining the boundaries of the digital cultural landscape and establishing priorities for selection of this content are important first steps in enabling grassroots and local collecting to take hold. The mapping of such materials to various preservation strategies will set the stage for implementing a series of targeted incentives to seed a distributed stewardship approach.

Recommendation #1: Leading cultural organizations should convene expert communities to address the selection and preservation needs of commercially owned cultural content and digital orphans.

4.3.2 Incentives to Preserve

There is nothing new about misaligned or competing incentives with respect to cultural content, and there is much to learn from how these gaps are closed in the analog realm. With analog cultural artifacts, policies aimed at strengthening incentives to preserve in the private sector and allocating preservation responsibilities among private owners and stewardship organizations relied on: (1) copyright incentives for owners to preserve materials for private purposes; (2) a suite of financial incentives for owners to preserve in the public interest; and (3) limited rights for qualifying stewardship organizations to preserve on the public behalf. Much attention has been given to the failures of current copyright law to provide effective rights and incentives to stewardship organizations to preserve these materials. **But another, possibly more fundamental, risk is the widespread disruption to the business models that provide the primary incentives for certain commercial owners to preserve. This is because in all cases of privately held cultural assets, it is the owner who decides what to preserve and who pays for preservation.** Among commercial cultural enterprises, periods of high business instability jeopardize preservation because they throw into question the very nature of return on investment for cultural assets.

For cultural content that has existed as "hard copy," scarcity of materials is an important component in developing a viable business model. The disintegration of such models in the era of digital content has had profound effects on the funding models for preservation. The rise of digital cultural content has also coincided with waves of mergers among content producers that result in a cultural production ecology dominated by a handful of international conglomerates, such as NBC/Universal and Sony/BMG on the one hand, and a proliferation of independent producers and creators posting content directly to the Web on the other. Sometimes a company that specializes in producing what amounts to cultural patrimony (American music or

German films, for example) is a small part of a large international conglomerate whose primary business may not even be entertainment. This, too, puts the sustainability of cultural heritage at risk because it dilutes all incentives to preserve except for short-term financial return on investment.

Strengthening Weak Incentives

Several approaches can be used to remedy problems with incentives. The first is to strengthen the rights of preserving institutions by revising copyright law. A second action is to mandate deposit of copyrighted electronic content into authorized public institutions to secure their long-term preservation. In countries with mandatory copyright deposit, that mechanism has been very effective for ensuring some public provision of preservation for privately held materials. In the United States, the moving-image record has been better and more complete since the mid-twentieth century, when the Copyright Office in the Library of Congress began demanding deposit of film. Deposit in authorized libraries and archives has preserved materials that commercial firms did not keep safe.

While there is debate about how effective mandatory deposit will be in a digital environment, it does seem advisable that digital materials submitted for copyright protection should be deposited in full, which is not now always the case. This would require that copyright authorities articulate which version of the work—the *best edition*—is registered for copyright; enforce demand deposit when necessary; and scale up copyright deposit system capacities to meet the challenges posed by the enormous scale of content creation.

Recommendation #2: Regulatory authorities should bring current requirements for mandatory copyright deposit into harmony with the demands of digital preservation and access.

Both public-spirited individuals and corporations deserve recognition for preserving digital culture for public use

A second approach is to provide incentives directly to private owners of cultural assets to preserve on the public's behalf, and to encourage handoffs—such as donations—of privately held materials to public institutions. Any private entity willing to take on the responsibility to preserve in the public interest—whether a multinational corporation or individual collector—should be given some form of incentive, benefit, or other consideration. One set of incentives could be modeled along the lines of those provided to owners of physical cultural materials, which offer financial benefits in return for filling a public preservation role. These might include deductions for charitable donations to a stewardship organization, financial considerations for those who preserve historic properties, and so forth.[44]

These financial incentives would extend to private individuals as well as to corporations, enabling grassroots efforts among private collectors to take hold. There is a long history of private collectors being in the vanguard of recognizing the value of

[44] The Library of Congress undertook a study of such incentives and proposed recommendations; the study is published as "Proposals for the Creation of the Public Policy Environment Conducive to Digital Preservation," and appears in the forthcoming *Preserving our Digital Heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report. A Collaborative Initiative of the Library of Congress*.

cultural materials, and we anticipate this will be true in the digital realm. Offering preservation incentives to private collectors will leverage their expertise and their passion.

Recommendation #3: Regulatory authorities should provide financial and other incentives to preserve privately held cultural content in the public interest.

Distributed stewardship is a potentially powerful mechanism for long-term preservation, one little understood and worthy of further study

Lowering Other Barriers to Preserve

Financial incentives are far from the only compelling reasons for private entities to preserve on behalf of the public. Both public-spirited individuals and corporations deserve recognition for preserving digital culture for public use. Discussion should begin between public and private parties about commercial sponsorship of preservation activities that would enhance corporate recognition and reputation.

An interesting set of questions arises when we think about how to take advantage of the cultural content preserved on individual computers. Consuming online music and images often means downloading such materials and preserving them on one's personal computer. Leveraging that highly dispersed capacity in the service of preservation would enable potentially significant bodies of cultural creation to be safeguarded in the near term. It is not yet clear how we would match individual computing capacity with preservation needs, or what the rights issues are. But the culture of sharing and copying among consumers of Web culture is already endemic. Distributed stewardship is a potentially powerful mechanism for long-term preservation, one little understood and worthy of further study.

4.3.3 Roles, Responsibilities, and Funding

Public-Private Partnerships

Public institutions have an important role to play in preserving culturally significant content that is privately owned. At present there are few robust mechanisms for the kinds of public-private partnerships that are needed in the digital realm. Such partnerships should be able to accomplish two things: (1) they should enable reliable and transparent allocations of roles and responsibilities among partners; and (2) they should ensure secure handoffs of significant cultural materials from private owners to public preservers.

We mention two notable examples here, but there are many more. In the United States, under the auspices of the congressionally mandated National Digital Information Infrastructure and Preservation Program (NDIIPP), the Library of Congress has forged multiple partnerships between public and private entities to develop and promote digital preservation best practices.[45] A second example is in the area of multimedia production, where there are grassroots efforts under way to organize and implement such practices. One such effort, CineGrid, is a collaborative research organization that brings together media experts from many domains of knowledge and practice to develop and promote the use of technologies for media

[45] See http://www.digitalpreservation.gov.

asset production and management that includes, but is not limited to, preservation concerns.

At some point, private owners of culturally significant materials may decide they no longer want to care for the materials. Stewardship organizations should not wait passively until key private collections are at risk. They must actively engage in partnerships, forge them early in the digital lifecycle, and arrange for secure handoff when the time comes.

Recommendation #4: Leading stewardship organizations should model and test mechanisms to ensure flexible long-term public-private partnerships that foster cooperative preservation of privately held materials in the public interest.

In addition to partnering with private entities, stewardship organizations should take the lead in identifying privately held materials of significant cultural value. There are two organizations in the United States operating under the aegis of the Library of Congress that convene annually to select films and sound recordings of historic significance to be preserved. The National Film Preservation Board and the National Recording Preservation Board each year select items to add to their national registries.[46] The selection process itself, which polls many communities for nominations, is an effective way to raise awareness of the need to preserve content, while also leveraging the expertise of multiple communities, professional and amateur. These programs have paid special attention to the needs of independent artists and the creative productions of underrepresented communities.

Funding Models

As a general rule, we see three effective funding models for this content: (1) the commercial owner preserves and pays; (2) society preserves (in the form of public cultural heritage institutions), and society as a whole pays (through current taxes or through borrowing); and (3) groups of self-organizing stakeholders preserve, and collaboratively fund their effort.

What about the owners of commercial content with low market value? Documentaries and art films produced by independent artists or small production companies are less likely than "blockbuster" or Hollywood studio films to persist for long periods after initial release. Such private owners do not always have the internal capacity (or the inclination to create the capacity) to preserve the content themselves. In these cases, third parties offering on-demand preservation services to small producers would be a good solution. If these services could be provided at scale, the cost incurred by producers to preserve would likely be lower than if they maintained small-scale internal capacity. How the demand for preservation services gets expressed may be problematic, because it is by nature diffuse. Here again, funders and stewardship organizations can take the lead by assessing the demand and then sponsoring the development of preservation services. With their leadership, eventually a marketplace will emerge to offer such services.

---

[46] See http://www.loc.gov/film/filmnfr.html; http://www.loc.gov/rr/record/nrpb/nrpb-masterlist.html.

4.3.4 Looking Ahead

At least in the near-term, commercial interests will serve the needs of preservation reasonably well, although at the price of limiting potentially valuable remix and reuse. However, there are few, if any, appropriate handoffs between private and public sectors in place. Ensuring such handoffs is high priority for action and a good candidate for the use of regulatory mechanisms (accompanied by appropriate incentives to owners). It may be possible for copyright authorities to develop a registry of works that tracks current ownership—at least for the commercial sector—to provide the information needed to secure handoffs.

In this preservation context, perhaps more than in the other three under study, the importance of productive public-private partnerships will be key to successful long-term preservation.

BOX 4.3 **Action Agenda for Commercially Owned Cultural Content**

- **Recommendation #1:** Leading cultural organizations should convene expert communities to address the selection and preservation needs of commercially owned cultural content and digital orphans.
- **Recommendation #2:** Regulatory authorities should bring current requirements for mandatory copyright deposit into harmony with the demands of digital preservation and access.
- **Recommendation #3:** Regulatory authorities should provide financial and other incentives to preserve privately held cultural content in the public interest.
- **Recommendation #4:** Leading stewardship organizations should model and test mechanisms to ensure flexible long-term public-private partnerships that foster cooperative preservation of privately held materials in the public interest.

4.4 Collectively Produced Web Content

Social networking sites, crowd-sourced sites, the blogosphere—these content types are scarcely a decade old, if that, and yet they play important social, cultural, and political roles in contemporary life. Each typifies what we call here *collectively produced Web content*: a new form of communication and knowledge production not created by a single individual or group, but by the collective efforts (coordinated or uncoordinated) of many individuals. What should be preserved of this content? Whose responsibility is it? The imperative to preserve some of this content now, to keep preservation options open and prevent irreversible losses, is clear.

Nearly every element decision makers need to consider in designing sustainable preservation is fraught with uncertainties. Who benefits from long-term access? At present we only know what short-term demand looks like. Who owns the content? Many of the ownership and rights issues are murky because sites often comprise

contributions from many different people with obscure rights to the materials they are contributing. Who preserves the content? A blog-hosting site, for example, typically has no transparent preservation policies: we do not know what content they do or do not preserve. And who pays? This is especially vexing in cases where the existence of the site itself relies upon volunteer efforts such as open-source software or the late GeoCities.[47]

Answers to these questions will not come all at once. A sustainable preservation strategy for collectively produced Web content may look nothing like any that exists for other digital materials. These questions can only be answered over time, iteratively—the result of learning by doing. Today, it is important to formulate the right questions to pursue, and to test and model different strategies for preservation. Some actions should be taken now, such as forging partnerships with content creators and owners, that can shed light on the nature of value, incentives, and the roles and responsibilities among those who collect and preserve these materials.

4.4.1 Value and Selection

Even though less than a decade old, the demand for this content is clear: collectively produced Web content attracts users by providing opportunities for them to create, modify, interact with, and contribute back to the content. These activities form the core of the genres that we investigated—the blogosphere, social networking sites (MySpace, Facebook, Linked In, Second Life) and crowd-sourced content (Slashdot, Wikipedia, Flickr, YouTube, SourceForge).[48] In these genres, aggregation and network effects are important; **the higher the number of contributors and users, the more valuable the resource to stakeholders and the more visits a site gets**.

Diffuse or Weakly Articulated Demand

Outside of a few marquee sites—Google Earth, Flickr, MySpace, Facebook, Wikipedia—and well-read blogs, there is not much clarity about what specific content should be collected.[49] This is in part because of the interactive and dynamic nature of the sites. What does it mean to collect the blogosphere—selecting representative blogs and sampling them? What about bloggers who do not wish to have their materials preserved, a fairly common phenomenon? What does it mean to preserve Wikipedia—saving each entry *and* its entire editorial history? And if both, how is that possible? Current archival and preservation practices rest on some rough consensus about the boundary of an information asset—where it begins and where it ends—so that a final or archival form of an asset can be preserved. This is not a model that has much meaning for content whose value resides in links into and out of multiple sites.

Despite these uncertainties, there are institutions acting now to preserve collectively produced Web content and are learning what to do by doing it. The oldest such effort,

[47] See http://www.archiveteam.org/index.php?title=Geocities.

[48] See http://www.myspace.com; http://www.facebook.com; http://www.linkedin.com; http://secondlife.com; http://slashdot.org; http://www.wikipedia.org; http://www.flickr.com; http://www.youtube.com; http://sourceforge.net.

[49] See http://earth.google.com.

the Internet Archive, began crawling the Web in 1996. Since then, many libraries and archives across the globe have been crawling portions of the Web for preservation. In many cases, they are using an option strategy, collecting a large portion of some slice of the Web and keeping those materials at relatively low levels of curation pending clarification of long-term demand. We recommend that institutions engaged in these activities provide leadership to others by convening stakeholder communities—creators, users, and collectors—to carve out near-term goals for selection. This might entail developing one or more partnerships with host sites in order to refine the selection criteria and determine technical requirements for preserving the materials.

Recommendation #1: Leading stewardship organizations should convene stakeholders and experts to address the selection and preservation needs of collectively produced Web content.

The perception of value in long-term preservation may reside with stakeholders completely distinct from those who produce, contribute to, or own the content

4.4.2 Incentives to Preserve

Weak Incentives

Motivations for contributing content to collective sites are very skewed toward short-term benefits. The producers of the resources do not necessarily see a value in long-term preservation. In fact, the perception of value in long-term preservation may reside with stakeholders completely distinct from those who produce, contribute to, or own the content.

Beyond the difficulties in aligning incentives, determining who has the right to preserve content is highly complicated. In the relatively simple case of a blog, for example, any given blog post could be a composite of linked materials that may well have compound rights within them. Bloggers will embed or link to materials over which they have no clear rights, not often seeking permission to do so. A public institution that has an incentive to preserve such blogs will nonetheless think twice about collecting blogs with such obscure rights issues, and that demands remedy. One solution might be for bloggers to use a simple license (such as Creative Commons) to clarify their intentions with respect to third-party archiving—either granting nonexclusive rights to preserve in the public interest, or forbidding third-party archiving.

Another solution would provide incentives to the host sites to preserve. These hosting services by default often end up being the sole decision makers in the long-term disposition of the content. Hosting sites may have strong incentives to preserve materials for private benefits, using the data for a number of profit-making purposes. But they may be amenable to facilitating preservation in the public interest as well. The ability to mine data would not necessarily be impeded by allowing third-party archiving of content for non-commercial uses. One possible remedy, therefore, would be for host sites to maintain a default (opt-out) license granting nonexclusive preservation rights to third parties.

Recommendation #2: Creators, contributors, and host sites could lower barriers to third-party archiving by using a default license to grant nonexclusive rights for archiving.

A second set of incentives may be in play with some large commercial host sites. The access platforms on which collectively produced Web content is created and maintained—companies such as MySpace (owned by Fox Interactive Media, a subsidiary of News Corporation) and YouTube (owned by Google)—have strong incentives to keep the content on their sites, both to provide access to the subscribers and users, and to mine the data for advertising, to support search, or whatever other purposes they have in mind. In contrast, hosting services are not likely to have an incentive to act on the vague, diffused interests of future users such as sociologists and genealogists. In this case, third-party archives such as libraries will likely have the strongest incentive to preserve. A partnership between such a site and a stewardship organization could be structured to facilitate persistence of these materials while they are still commercially viable, with a default handoff in place for materials that have long-term value.

We recommend that preservation subsidies be extended to stewardship organizations that take on these responsibilities, as well as those sites, such as Wikipedia, that may be induced to preserve on the public behalf by a financial incentive. An optimal arrangement might be to support a consortium of organizations that have sorted out collection responsibilities, achieved economies of scale, and subscribe to transparent and accountable standards of stewardship. This could take place under the auspices of nationally recognized entities such as the Library of Congress, the Smithsonian Institution, the Koninklijke Bibliotheek, or the British Library.

Recommendation #3: Regulatory authorities should create incentives, such as preservation subsidies, for host sites to preserve their own content or seek third-party archives as preservation partners.

Public institutions are best positioned to ensure the long-term preservation of high-value digital materials that have no natural steward or are otherwise at risk

By and large, contemporary creators, users, and stakeholders have indicated little concern about or even awareness of issues concerning long-term access to what is in effect current cultural heritage. This creates an urgent preservation risk, and a crucial need for public provision of preservation. Public institutions are best positioned to ensure the long-term preservation of high-value digital materials that have no natural steward or are otherwise at risk. At present, there is one outstanding barrier to action in the United States and in Europe, and that is lack of clear authority to crawl the Web for preservation purposes. In the United States, granting that authority is one of the recommendations before Congress to reform copyright law, and we urge timely action to reform legislation.

Recommendation #4: Regulatory authorities should take expeditious action to reform legislation to grant authority to stewardship institutions to preserve at-risk Web content.

4.4.3 Roles, Responsibilities, and Funding

In the absence of clear roles and responsibilities to preserve collectively produced Web content, there have been numerous grassroots efforts to preserve content at risk of disappearing. In 2009, for example, a group of volunteers stepped forward to collect content from Geocities, a Web hosting site and development company that had existed

since 1994 but faced sudden closure. The same spirit of collective action that created such sites will be effective in preserving the sites, at least for the near term. The question arises about whether there are public policies or partnerships with institutions that could support and enable such grassroots efforts. This is an area that should be explored by leaders in Web preservation.

There are many opportunities for public institutions to play vital roles in preserving collectively produced Web content. National libraries and other leading stewardship organizations have forged the International Internet Preservation Consortium (IIPC) in part to articulate roles and responsibilities among Web archiving activities (dividing those responsibilities largely along lines of national interest). We recommend that IIPC members and other stewardship organizations convene key stakeholders, particularly owners and producers of collectively produced Web content, to develop partnerships with major host sites.

Because there are far more questions than answers about what to preserve, how to preserve, and who will pay for it, it is important to reach out to major content providers and host sites to raise their awareness of the need to preserve content that they serve or own. It is also important to develop partnerships that can explore the technical, legal, and financial dimensions of preservation to ensure long-term access to collectively produced content that has value for future generations.

Recommendation #5: Leading stewardship organizations should develop partnerships with one or more major content providers to explore the technical, legal, and financial dimensions of long-term preservation.

Funding Gaps

Many host sites have uncertain funding models for business survival, and in most cases, keeping their sites alive takes precedence over developing long-term preservation strategies. Collective action—public funding, either voluntary or mandated—will be needed to secure many of these assets over the next decade, a reasonable time frame for reassessing content value. In these circumstances, public funding options include direct subsidies to institutions whose missions include preservation, or direct public funding for preserving privately held assets. Furthermore, to the extent that particular actors are receiving public support, it would seem natural to impose mandates—to condition public support on the commitment to undertake preservation activities. There is nothing new here: this is how museums and libraries fund preservation of analog cultural materials.

4.4.4 Looking Ahead

For the most part, demand (and therefore services) has only crystallized around current access. A compelling value proposition for long-term preservation has yet to be articulated. When the value proposition does arise, it is likely to be highly diffused among creators, contributors, and users. In these circumstances, preservation will probably require collective action until we see a compelling value proposition emerging and sufficient incentives for actors to take timely action, and until the community of stakeholders sorts out its roles and responsibilities. We are now in the earliest stages of

developing strategies for preservation of collectively produced Web content, the stage at which preservation is *everybody's* problem and *nobody's* problem. Yet we have already seen the emergence of pioneering grassroots stakeholders and leading preservation institutions trying out different roles, using different funding models, and taking action now to ensure that options remain open long enough for people to understand the long-term value of collectively produced Web content.

BOX 4.4 **Action Agenda for Collectively Produced Web Content**

- **Recommendation #1:** Leading stewardship organizations should convene stakeholders and experts to address the selection and preservation needs of collectively produced Web content.
- **Recommendation #2:** Creators, contributors, and host sites could lower barriers to third-party archiving by using a default license to grant nonexclusive rights for archiving.
- **Recommendation #3:** Regulatory authorities should create incentives, such as preservation subsidies, for host sites to preserve their own content or seek third-party archives as preservation partners.
- **Recommendation #4:** Regulatory authorities should take expeditious action to reform legislation to grant authority to stewardship institutions to preserve at-risk Web content.
- **Recommendation #5:** Leading stewardship organizations should develop partnerships with one or more major content providers to explore the technical, legal, and financial dimensions of long-term preservation.

4.4.5 Summing Up

This survey of the current landscape of digital preservation in the context of four content profiles underscores how diverse the uses of digital materials are, and how the motivations of actors and stakeholders vary. Regardless of content, however, preservation decision makers face common constraints because all digital materials share basic attributes. **The tradeoffs decision makers face vary depending on the context. But in all cases, they must take near-term actions to ensure future access**.

In this chapter we have focused on action agendas targeted to specific stakeholder communities. In the next chapter, we synthesize these agendas and articulate principles that support sound preservation decision making in all domains.

Chapter

5

Recommendations for Achieving Sustainability

All digital preservation decision makers—creators, owners, users, and archivers—face the same difficult decisions.

- Which digital materials should be preserved, for how long, and for what use?
- Who should be responsible for preservation?
- How will preservation activities secure necessary funding and resources over time?
- How do we measure return on investment in the short term and the long?

These questions are formidable and are made all the more difficult by the lack of sufficient information and great uncertainties about the future. But these questions are by no means unanswerable. In our concluding chapter, we review the conditions necessary to achieve economic sustainability and propose principles that can guide decision making to achieve sustainability. We identify the range of actions that decision makers can choose among to create sustainable preservation strategies. And we propose an agenda for near-term actions to ensure sustainability over long time horizons, addressed to specific stakeholder groups. These recommendations, together with those articulated in the previous chapter that are addressed to specific stakeholder communities, are presented in Tables 5.1. and 5.2.

5.1 General Principles and Actions for Sustainable Preservation

At the beginning of this report we identified the five conditions necessary for economic sustainability of digital preservation:

- **recognition of the benefits of preservation by decision makers**
- **selection of materials with long-term value**
- **incentives for decision makers to act in the public interest**
- **appropriate organization and governance of preservation activities**
- **ongoing and efficient allocation of resources to preservation**

To these conditions we add one more, which goes directly to the temporal dimension of preservation: the fact that digital preservation takes place over time. Actions taken (or not taken) in the present will shape those available to decision makers in the future, and the near-term actions need to provide flexibility in future decision making.

- **timely actions to ensure access**

The following principles apply in all preservation contexts. Which actions will effect those principles depends upon the specific context: the type of digital materials involved and at which stage of the digital lifecycle, the community of users and stakeholders for those materials, and the types of present and future value identified in the materials. We emphasize that perceived values and benefits are always context-specific, and should be construed broadly to include historical, ethical, artistic, scholarly, cultural, public policy, and institutional values, not simply values denominated in currency.

Condition for sustainability: timely actions to ensure access

Principle of action: Preservation requires a series of decisions to be made over the lifecycle of digital assets.

Action: Take preservation steps early in the digital lifecycle; create and codify contingency plans; make and implement plans for handoffs to address economic risks over the digital lifecycle.

Sustainable preservation is not achieved by a single decision or action. It requires a series of decisions made by multiple actors and stakeholders over the lifecycle of digital resources. Preservation goals are appropriately envisioned in long increments, often denominated in centuries. But in reality, decision making occurs regularly throughout the lifecycle of digital information, beginning with choosing well-documented file formats when creating data. Digital assets can be at great risk when an organization changes its preservation priorities. Contingency plans should be made in advance of such moments (often referred to as *trigger events*), and should include provisions for handing off the materials to a responsible party.

Systems must be in place to ensure that responsible parties are notified of triggering events to avoid failure through inattention. MOUs and service-level agreements between data communities and archives can be timed to prompt a regular review of preservation priorities. The MOU in place between the Astronomical Research Consortium and the Johns Hopkins Library, which must be renewed every five years, is an example of such a prompting mechanism.

Such moments of particular risk to digital materials must be negotiated without inappropriately closing options for decision makers in the future. Employing an option strategy for selection, for example, would involve a timely decision about whether to collect and preserve data, even if at a low level of curation, in order to maintain the option of deciding later.[50]

Like other societal challenges that span generations, stewardship of our knowledge base is particularly well served by organizations that represent the public trust. Libraries, archives, and museums are specifically charged to preserve valuable assets that benefit present as well as future generations. The role of these stewardship organizations in the digital realm will become only more important as information becomes primarily digital; their funding must match this vital national mission. Some entities in the private and academic sectors also serve this important mission, and they must be recognized and enabled by their supporting institutions.

Condition of sustainability: recognition of the benefits of preservation by decision makers

Principle: To make a case for preservation, make the case for use.

Actions: Articulate the value proposition in terms of use cases; identify at-risk materials whose damage or loss is unacceptable; use proxy organizations to aggregate diffuse demand; use option strategies in cases of uncertain value.

The benefits of preservation investments should always be framed in terms of the possible uses for digital assets. Cases should emphasize near-term uses where possible, but look as far ahead as possible. A compelling value proposition will help articulate demand for digital assets, which will in turn spur provision of preservation services.

Benefits can also be expressed in terms of negative benefits—the costs incurred if assets are not preserved. This could take the form of the time and effort needed to recreate the assets, or, if they cannot be recreated, the kinds of uses that would then not be possible. This would include certain classes of observational data such as longitudinal studies, historical information such as

[50] See Box 3.1, p. 37

evidence from archaeological sites, and sensitive data such as clinical trial data in which there is an ethical imperative to preserve.

In cases where the benefits of preservation are widely diffused across stakeholder communities, it may be necessary for a trusted proxy organization to represent the aggregate demand and make the case for the benefits that will flow to many different stakeholder communities.

It is especially important to represent future stakeholders, whose interests are always underrepresented. Libraries, archives, and museums are well positioned to represent the public interest of future stakeholders. In some communities, domain archives and professional organizations may fulfill that role. In the case of scholarly discourse and commercially owned cultural content, two forms of benefits—commercial and cultural, or private and public—tend to compete with one another. When that occurs, proxy organizations must step in to represent the public interest.

In cases of uncertainty about future value, decision makers can use an option strategy, designed to postpone decision making until the time when better information about the value of the data is available. This strategy could take the form of depositing materials in archives with a low level of curation.

Condition of sustainability: selection of materials with long-term value

Principle of action: Priorities must be made for selecting materials that have the greatest promise of returning value to users over time.

Action: Make decisions about priorities among collections competing for scarce resources based on projected future use.

Sustainable preservation does not mean having the ability to save everything. Just as the value proposition means making the case for use, selection means giving investment priority to those materials of greatest use to present and future stakeholders.

A decision to preserve now need not be thought of as a permanent or open-ended commitment of resources over time. In many instances, materials can be used for relatively short periods of time and then handed off or discarded when priorities shift—when higher-resolution images, newer editions of reference libraries, or more complete data sets are available, for example.

Selection decisions should reflect the interests of future as well as current stakeholders; therefore, selection processes are often best entrusted to proxy organizations to represent the interests of present and future stakeholders. In those cases, it is vital that the proxy organizations have well-articulated and transparent selection procedures, and that they make full use of domain-specific expertise to advise on the selection and preservation of materials.

Condition of sustainability: incentives for decision makers to act in the public interest

Principle of action: Incentives must be strengthened when they are weak; aligned when they diverge among different stakeholder communities; and created where there are none.

Actions: When there are insufficient incentives, use preservation mandates as appropriate. Provide financial incentives for private owners to preserve on behalf of the public. Bring copyright law and mandatory deposit requirements up to date for digital preservation. Remove barriers to creating efficient decentralized stewardship mechanisms by use of nonexclusive licenses granting preservation rights to third parties.

Weak or misaligned incentives pose a major risk to preservation, yet such gaps are endemic to preserved digital assets. Barriers to action, whether they arise from copyright law, competing benefits, the high costs of current preservation technologies, or lack of capacity, must be addressed.

Where incentives are weak, as may be the case in preserving data created in grant-funded research, mandates to preserve should be used to strengthen incentives. To avoid imposing unfunded mandates, provisions should include: selection criteria for what is to be preserved; specification of the roles and responsibilities of the parties to do the preservation; and sources of funding.

A significant misalignment of incentives occurs when stakeholders who are in the best position to preserve (by virtue of physical custody or ownership, or both) lack strong incentives to do so, while those that have strong incentives (commonly, institutions with a stewardship mission) do not have the right to preserve. Where intellectual property issues are the problem, government should move quickly to work with private and public sectors to amend legislation. We urge legislative and regulatory bodies to revise copyright code preservation rights to make them effective for digital preservation. Similar efforts are required to address the international legal frameworks for intellectual and moral rights.

Decentralized stewardship strategies should be encouraged by granting a nonexclusive perpetual license to a responsible agency serving in the public interest to preserve and make accessible the various forms of digital materials. With respect to openly accessible materials—for example, emerging scholarly discourse or collectively produced Web materials—authors should make clear their intentions for accessibility and stewardship.

Condition of sustainability: appropriate organization and governance of preservation activities

Principle of action: Roles and responsibilities among stakeholders must be clear, transparent, and well integrated; and handoffs between responsible parties must be ensured at key moments of risk in the digital lifecycle.

Actions: Create effective governance mechanisms to aggregate and rationalize collective preservation interests and costs. Create mechanisms to address free-rider problems in the provision of preservation.

Preservation organizations should be transparent and accountable. In the case of third-party archiving services, all parties should have a service-level agreement or MOU that stipulates roles, responsibilities, and outcomes that is revisited or open for renewal or cancellation at reasonable intervals (such is the case with Chronopolis, Portico, and CLOCKSS).

The need for clear allocation of responsibilities is just as strong in organizations that provide their own preservation services. Each organization that creates data should have clear policies that specify roles, responsibilities, and procedures.

While there is a sense of collective responsibility for stewardship of the digital scholarly and cultural record, there is often no clear allocation of preservation responsibilities and costs across stakeholders. In these situations, when preservation actions taken by one party can in theory benefit all parties, the free-rider problem arises and may threaten the long-term stability of preservation funding. Effective governance mechanisms are needed to aggregate the collective interest into an effective preservation strategy and, equally important, ensure that the effort and cost of carrying out preservation are appropriately apportioned. Such mechanisms can range from third-party archives that provide access through a tiered access system, such as ICPSR or Portico, to the international community initiative LOCKSS.

Condition of sustainability: ongoing and efficient allocation of resources to preservation

Principle of action: Funding models must reflect community norms, be flexible to adjust to disruptions over time, and leverage economies of scale and scope as appropriate.

Actions: Choose funding models according to norms and expectations of anticipated users; leverage economies of scale and scope; lower costs of preservation overall.

Digital assets do not need to be treated as a public good in all cases. Market channels are often the most efficient means of allocating resources for preserving many types of digital content. It is critical that those digital assets

provided as market goods or otherwise privately held have some provision for handoff to another trustworthy steward when the private owner no longer deems the content worth keeping but it may still be of value to society. For materials that are not amenable to private provision, such as certain types of research, Web-based materials, and digital orphans, public provision is necessary.

Bundling the provision of preservation and access can be an efficient means of creating economies of scope. Creating economies of scale among archives when possible is always desirable, and may be critical when the materials under stewardship require particular kinds of expertise that are scarce. This is the case for much scientific data and many cultural assets.

Finally, as the rate of digital information production continues to escalate, it is vitally important to reduce the cost of preservation for all types of digital assets. Reducing the cost of storing materials, developing sustainable sources of energy to power preservation systems, and, in particular, engineering ways to lower the cost of preserving, curating, and providing access to rich media collections and complex data sets are all important. In some cases, providing a menu of preservation quality options at different prices can align user needs and resources with appropriate levels of stewardship. Instituting more efficient preservation using a host of approaches will increase the likelihood that the most valuable digital collections will be accessible to future generations.

5.2 Agenda for Further Action

Our action agendas are necessarily broad in scope and implication. They lay a foundation for sustainable preservation by helping to create the requisite conditions. Yet there remain significant gaps in knowledge that require further investigation. Much work has yet to be done to model and test promising economic strategies for digital preservation.

Areas of priority for near-term action include the following:

Organizational Action

- developing public-private partnerships
- ensuring that organizations have access to skilled personnel, from domain experts to legal and business specialists
- creating and sustaining secure chains of stewardship between organizations over time
- achieving economies of scale and scope
- addressing the free-rider problem

Technical Action

- building capacity to support stewardship in all areas
- lowering the cost of preservation overall
- determining the optimal level of technical curation needed to operationalize an option strategy for all types of digital material

Public Policy Action

- modifying copyright laws to enable digital preservation
- creating financial incentives and handoff requirements for private entities to preserve on behalf of the public
- sponsoring public-private partnerships
- clarifying rights issues associated with Web-based materials
- empowering stewardship organizations to protect digital orphans from unacceptable loss.

Education and Public Outreach Action

- promoting education and training for 21st century digital preservation—domain-specific skills; curatorial best practices; and core competencies in relevant science, technology, engineering, and mathematics (STEM)
- raising awareness of the urgency to take timely preservation actions

We address these recommendations to the groups of actors with lead responsibilities for implementing them (see Table 5.1).

5.3 Conclusion

Preservation decisions are always made under conditions of uncertainty: technologies, policy environments, investment priorities, and societal concerns will change over the course of the digital lifecycle. But we can develop practices that anticipate or resolve uncertainties, that leverage resources among stakeholders, and above all, that leave options open for decision makers in the future. Sustainable preservation strategies will find ways to turn the uncertainties of time and resources into opportunities for flexibility, adjustments in response to changing priorities, and redirection of resources where they are most needed.

Commitments made today are not commitments for all time. But actions must be taken today to ensure flexibility in the future. There is an urgent need in all sectors of digital creation—public and private, cultural and scientific—for support in the near-term to model and test robust preservation strategies. All stakeholder communities must provide leadership and accept responsibility for the development of a common digital preservation infrastructure that is sustainable for generations to come.

Like other societal challenges, such as climate change and sustainable energy, preservation is a balancing act, weighing the needs and desires of the present day with those of the future. Preservation depends on the cooperation of generations to steward our most precious asset over time—knowledge. Society has long valued, supported, and protected the cultural heritage deeded to us by previous generations. Isaac Newton acknowledged that "If I have seen further it is by standing on the shoulders of giants." By attending to the value of digital information, providing incentives to preserve these digital assets, and ensuring allocation of roles and responsibilities among stakeholders that share a common interest in valuable digital assets, we can continue to build high the shared body of knowledge that will enable all of us to see farther.

Tables 5.1 and 5.2 summarize the recommendations proposed by the Task Force for economic sustainability: Table 5.1 lists actions to be taken by leading actors and organizations; and Table 5.2 list actions by content domain.

TABLE 5.1 **Action Agenda for Leading Actors and Organizations**

National and International Agencies
Trusted international, national, and public institutions—libraries, archives, museums, research institutes, consortia, regulatory agencies

1. Create mechanisms for public-private partnerships to align or reconcile benefits that accrue to commercial and cultural entities. These agencies can play a critical role in convening stakeholders, sponsoring cooperation and collaboration, and ensuring representation of all stakeholders.
2. Convene expert communities to address the selection and preservation needs of materials of particular interest to the public for which there is no stewardship (Web materials, digital orphans).
3. Act expeditiously to reform national and international copyright legislation to address digital preservation needs.
4. Create financial incentives to encourage private entities to preserve digital materials on the public behalf.

Funders and Sponsors of Data Creation
Private and public agencies and foundations

1. Create preservation mandates when possible, ensuring that they adhere to community selection criteria, and specifying roles and responsibilities of individuals and organizations.
2. Invest in building capacity throughout the system. The Library of Congress, the National Archives and Records Administration, the National Science Foundation, and JISC have set important precedents for supporting capacity building within specific communities of practice. Seeding stewardship capacity and developing sustainable funding models should, however, be a high priority for all funders.
3. Provide leadership in training and education for 21st century preservation, including domain expertise and core competencies in STEM. Such organizations as the National Archives, Library of Congress, National Library of Medicine, National Agricultural Library, National Science Foundation, Smithsonian Institution, Institute of Museum and Library Services, National Endowment for the Arts, and National Endowment for the Humanities in the United States; and the British Library, National Archives, JISC, Digital Curation Centre, and Digital Preservation Coalition in the United Kingdom each have a remit for promoting digital preservation skills.
4. Fund the modeling and testing of domain-specific preservation strategies. This would entail developing domain-specific requirements for lifecycle management to create a timeline of predictable risks, strategies to meet them, and triggering mechanisms to address them.

TABLE 5.1 **Action Agenda for Leading Actors and Organizations**

Stakeholder Organizations
Universities, research institutions, private companies, third-party archives, professional societies, trade organizations

1. Secure preservation of high-value institutional materials by making explicit roles and responsibilities across organizational boundaries.
2. Develop preservation strategies that assign responsibilities for achieving outcomes. Service-level agreements and MOUs with third-party archives should include contingency plans for handoffs and clauses for putting internal monitoring systems in place.
3. Leverage resources; create economies of scope and economies of scale by partnering with related organizations and industry professional associations.
4. Work with domain and preservation experts to ensure that personnel are fully equipped with the technical skills needed for selecting, curating, and preserving materials.
5. Fund internal preservation and access activities as core infrastructure.

Individuals
Principal investigators, data creators, individual authors, creators, and scholars

1. Provide nonexclusive rights to preserve content they create and to distribute this content through publicly accessible venues.
2. Partner with preservation experts early in the lifecycle of one's own digital data, to ensure that data are ready to hand off to an archive in forms that will be useful over the long term.
3. Actively participate in professional societies and relevant organizations in developing stewardship best practices and selection priorities.

TABLE 5.2 **Action Agenda by Content Domain**

Action Agenda for Scholarly Discourse

1. Libraries, scholars, and professional societies should develop selection criteria for emerging genres in scholarly discourse, and prototype preservation and access strategies to support them.
2. Publishers reserving the right to preserve should partner with third-party archives or libraries to ensure long-term preservation.
3. Scholars should consider granting nonexclusive rights to publish and preserve, to enable decentralized and distributed preservation of emerging scholarly discourse.
4. Libraries should create a mechanism to organize and clarify their governance issues and responsibilities to preserve monographs and emerging scholarly discourse along lines similar to those for e-journals.
5. All open-access strategies that assume the persistence of information over time must consider provisions for the funding of preservation.

Action Agenda for Research Data

1. Each domain, through professional societies or other consensus-making bodies, should set priorities for data selection, level of curation, and length of retention.
2. Funders should impose preservation mandates, when appropriate. When mandates are imposed, funders should also specify selection criteria, funds to be used, and responsible organizations to provide archiving.
3. Funding agencies should explicitly recognize "data under stewardship" as a core indicator of scientific effort and include this information in standard reporting mechanisms.
4. Preservation services should reduce curation and archiving costs by leveraging economies of scale when possible.
5. Agreements with third-party archives should stipulate processes, outcomes, retention periods, and handoff triggers.

Action Agenda for Commercially Owned Cultural Content

1. Leading cultural organizations should convene expert communities to address the selection and preservation needs of commercially owned cultural content and digital orphans.
2. Regulatory authorities should bring current requirements for mandatory copyright deposit into harmony with the demands of digital preservation and access.
3. Regulatory authorities should provide financial and other incentives to preserve privately held cultural content in the public interest.
4. Leading stewardship organizations should model and test mechanisms to ensure flexible long-term public-private partnerships that foster cooperative preservation of privately held materials in the public interest.

TABLE 5.2 **Action Agenda by Content Domain**

Action Agenda for Collectively Produced Web Content
1. Leading stewardship organizations should convene stakeholders and experts to address the selection and preservation needs of collectively produced Web content.
2. Creators, contributors, and host sites could lower barriers to third-party archiving by using a default license to grant nonexclusive rights for archiving.
3. Regulatory authorities should create incentives, such as preservation subsidies, for host sites to preserve their own content or seek third-party archives as preservation partners.
4. Regulatory authorities should take expeditious action to reform legislation to grant authority to stewardship institutions to preserve at-risk Web content.
5. Leading stewardship organizations should develop partnerships with one or more major content providers to explore the technical, legal, and financial dimensions of long-term preservation.

Appendix 1. Characteristics of Sustainability in Public and Corporate Records

In the most formal sense, corporate and public records are simply the byproduct of operational activity, presenting a record of an organization's official business. Beyond this narrow definition, however, records encompass information that can contribute to the well-being of society. One example might be the American Time Use Survey, a vital social science research tool produced by the U.S. Department of Labor.[51] Similarly, the design and production records of commercial airplanes, kept by aeronautical companies such as The Boeing Company, are an important public safety protection. These records are increasingly produced in digital form, forcing corporate and public records programs to continually adapt to the challenges of emerging technologies. Long-term preservation of corporate and public records is not discussed in the body of this report; nonetheless it is worthwhile to consider these programs in the framework of the economic sustainability definition.

Recognition of the benefits of preservation on the part of decision makers: In the short term, digital preservation is crucial for both governments and corporations; records, and increasingly digital records, are an integral part of ongoing operations and official business. At the same time, corporate and national archives are extremely effective tools for ensuring accountability—government to its citizens, and businesses to their shareholders and employees. In the longer term, these preserved digital records have enormous potential to advance academic research, personal and avocational inquiries, and public safety.

This ongoing societal value has been established for government records at all levels; however, for corporate records it is not always appreciated within the corporations themselves. While wider recognition of corporate records' importance is also still uncertain, researchers and cultural repositories are beginning to call attention to their significance. David Kirsch, at the University of Maryland, has investigated the public interest in private records, while the Minnesota Historical Society recruits new additions to its robust archiving program for local corporate records.[52]

Selection of materials with long-term value: Records retention policies form the backbone of selection for preservation within these archives. In identifying the qualifications for an official versus an unofficial record, a National Archives and Records Administration (NARA) specialist works with departmental records-keepers to determine the long-term value of different classes of documents.[53] Correspondingly, corporations use in-house or consultant legal experts to define their policies, often using preservation mandates from the government as the foundation of the document. Policies must be

---

[51] See http://www.bls.gov/tus.

[52] See David Kirsch, "The Record of Business and the Future of Business History: Establishing a Public Interest in Private Business Records," in *Library Trends* vol. 57, Number 3, Winter 2009; available at http://muse.jhu.edu/journals/library_trends/v057/57.3.kirsch.html.

[53] See http://www.archives.gov.

flexible enough to accommodate the discretionary addition of historically or culturally significant materials.

Whether working with public or corporate records, the most problematic aspect of this process is that once a policy is in place, it is up to the individual employee's discretion whether and how to place a document within the retention framework. With paper records, a practiced archivist can sort through all files to determine a document's relevance ex post. Not only does the complexity of digital files make such classification much more difficult, it also requires that the document's creator—a non-expert—make the first and potentially most vital decision about the value of the document. Often this leads to the unfortunate practice of keeping all files, rather than putting the effort into selection. Fully implementing a smooth selection process for digital documents remains a significant challenge for both public and corporate records programs.

Incentives for decision makers to act in the public interest: Beyond the described benefits, the maintenance of national and corporate archives is often motivated by legislative mandate. NARA was chartered specifically to document the government of the United States; incentives to preserve are intrinsic to the organization. Other agencies are legally required to give NARA their digital and physical records, and if they do not, legal actions could result. The State Department, for example, is required to transfer its central cable files to NARA; these files have been digital since 1973.

Similarly, many industries must follow government regulations regarding the records they generate. Pharmaceutical companies, for instance, must keep all records from medical trials in order to receive FDA approval for their drugs.[54] Though rarely included in legislative mandates, positive incentives that motivate corporations through rewards rather than punishment allow corporations to realize the extended value of their archives. Beyond the public relations boost of touting one's corporate responsibility, archives provide valuable marketing materials. In 2008, Coca-Cola began to engage with Web 2.0 marketing and launched a blog showcasing numerous digitized images from its 120-year-old archive.[55] As these efforts expand with a Facebook account and Twitter profile, digital records will increasingly be a part of the company's outreach strategy.

Appropriate organization and governance of digital preservation activities: NARA is a standalone federal agency that acts on behalf of society at large. It has full authority over the records-keeping practices of all agencies of the federal government, but whether this translates into full compliance is an open question. To directly address the challenges of the new digital environment, NARA created the Electronic Records Archive (ERA); as of June 2008 it had become directly involved in records retention policymaking. The persistence of physical files, however, means that NARA will require a complex and expensive hybrid strategy for the foreseeable future.

---

[54] See 21 CFR 312.58 - Inspection of sponsor's records and reports.

[55] See http://www.coca-colaconversations.com.

Corporate archives are usually positioned as internal departments; even when a records retention policy is in place, corresponding records management practices are sometimes not followed in all other departments. Still further, outside cultural repositories rarely have the ability to actively advise how corporate records are created or maintained until they are at the end of their lifecycle. Since the preservation status of digital records must often be determined from the moment of creation, this can produce insurmountable recovery or restoration problems.

Mechanisms to secure an ongoing, efficient allocation of resources: Public records programs are funded for the most part directly through the government, with varying levels of additional private donations. NARA must appear before Congress to justify a yearly budget line-by-line, but specific projects, such as exhibitions or public programs, often rely on private funds. The Smithsonian Institution has a trust that accounts for 30 percent of its budget, with the remainder provided by the government. Private foundations fund and build presidential libraries, but once complete the library becomes a part of the National Archives.

In contrast, a corporation is almost always the sole funder of its internal archive. If a corporation gives its records to another organization, however, usually the accepting organization will take responsibility for long-term funding. One new and promising funding model is being pioneered by the Minnesota Historical Society (MHS); by offering its archiving services, MHS builds relationships with corporations that include both paid, project-based work and large monetary donations.[56] Minnesota-based corporations such as the 3M Company are often pleased to outsource their archiving needs to better concentrate on their own products.

The five elements of economic sustainability are as useful for assessing public and corporate records programs as they are for the other data types discussed in this report. These programs have a long and successful history, and are confronting the new challenges of the digital era. However, an integral part of that adaptation must be a re-examination of their sustainability models.

[56] This model is discussed in James Fogerty's "Competing for Relevance: Archives in a Multiprogram Organization," a chapter in *Leading and Managing Archives and Records Programs*. Bruce W. Dearstyne, Ed. New York: Neal-Schuman Publishers, Inc. 2008.

Appendix 2. Analog and Digital Preservation Strategies

Preservation of analog content—print-on-paper books, maps, sheet music, manuscripts; or formats such as film, analog tape, sound recordings on wax cylinders, wires, and lacquer discs—differs in significant ways from preservation of digital content. For all practical purposes, preserving information that has been recorded on analog formats is as simple (and as complicated) as preserving the physical object itself. The content and the physical carrier are inseparable. Generally, physical formats are prone to deterioration and thus require preservation action to remain usable over time. Copying the content of a book from the book itself, whether by microfilming or photocopying, almost always produces inferior copies and thus unauthorized and illegal copying is readily detected. These characteristics of analog artifacts have implications for all aspects of the economic models that support their preservation and access: the scale of production, the modes of distribution and consumption, the ease of reproduction, and the protection of copyright claims in a given work. To see how this plays out in the real world, let us take the case of books.

First, the production of books is costly, requiring paper, ink, and binding materials, as well as the labor and materials involved in editorial and physical production and managing sales and distribution (books are bulky and heavy and easily damaged in shipping and handling). Because of the labor and resource costs of producing books, many manuscripts are never published, to the great frustration of their authors.

Preservation, too, demands a concentration of resources. Hardbound books can be remarkably sturdy, but are prone to accidental damage and, if not printed on high-quality paper, will eventually become brittle. This is seldom a short-term hazard that presents problems for individuals (though one cannot say that about other analog media, such as cassette tapes or Super 8 movies). But ensuring the usability of books over generations requires dedicated resources and a continuity of stewardship that only institutions such as libraries typically provide. Crucially, the interest that an academic library has in making books available to its faculty and students can be realized only by preserving the books in physical form. In a world of print books, the ability of libraries to deliver information to their constituents depends on the size and quality of their local collections. Thus there is rivalry in the quality of library access.

Because libraries preserve on behalf of the larger and long-term interests of society, they are granted legal and financial considerations that make their work possible. Section 108 of the U.S. copyright law grants libraries and archives limited rights to replace or reformat a book or other artifact no longer fit for use. Moreover, under the doctrine of first sale the copyright law allows that once a book is purchased, the owner of that book has the right to dispose of the physical object as he or she wishes, such as giving it to someone or reselling it. This enables the library to buy a single copy of a book and make it maximally accessible—that is, serially; the access is limited only by the fact that one person can use a book at a time.

Book owners can dispose of the physical objects as they see fit, but they have no such claim to the intellectual property held in the book. This protects the rights of the owners of the intellectual property, either the author or, more frequently, the publisher who represents the author by proxy, to continue to benefit from consumer demand for the book.

Thus there is an acceptable alignment of the interests of the stakeholders: the creators, producers, distributors, and the libraries and users. Most important, libraries acquire and preserve on behalf of their users. The reliability of library preservation is so strong that many publishers do not keep full archives of their own publications. Just like the rest of the public, publishers count on libraries to hold the stock of publishers that have long gone out of business, in the public interest of users present and future.

Compare this to the digital preservation scenario. The scale of production is immensely larger, because the marginal cost of publishing is virtually nil. With direct, unmediated access to the Web, many creators post directly online, without going through publishing houses, recorded sound or film studios, journals, magazines, or newspapers.

The doctrine of first sale is inapplicable to an increasing portion of library materials, with disastrous consequences for preservation. Because consumption of digital content is nonrival, most digital producers of content with commercial potential license their materials. Libraries have no right to preserve licensed materials; nor do they have strong incentive, as they did when they owned books. Publishers themselves are unused to preserving content for the benefit of the public or of future generations of users, thus putting at risk large corpora of culturally and historically significant materials. And finally, Section 108 of the copyright law has not been amended to account for the technical demands of redundant copying for the purposes of ensuring long-term usability and rendering of content.

Digital materials are subject to two types of deterioration that are quite different from many analog materials, including books. The first is immediate physical degradation. Put simply, a CD is much less reliable as an archival medium than acid-free paper, even in the short term. The second type of deterioration is loss of usability when there are changes in the formats used. The CD requires an appropriately configured drive to be used. The book requires much less. If this problem seems far-fetched, consider floppies written on ten years ago using programs that are now defunct. The time for migrating the information on these media has already passed or is close to it.

Appendix 3. When Markets Do Not Work

In an idealized, textbook-perfect competitive market, with many well-informed buyers and many well-informed profit-maximizing sellers of goods or services, there is a single price for each item being sold. That price provides valuable information to both buyers and sellers; it assures that (1) no buyer will pay more for an item than it is worth; (2) no seller will sell an item at a loss; and (3) all buyers who value an item more than its cost of production will have the opportunity to purchase it.

Many markets reasonably approximate perfectly competitive markets. However, many other markets do not and they fail. There are at least five distinct sources of market failure.[57] First, some businesses have organizational structures that support goals different from or more complex than simple profit maximization. For example, a cooperative is an organization that is jointly owned by many individuals or businesses for their mutual benefit. However, because members of the cooperative are likely to pursue their individual interests, rather than the profits of the cooperative as a whole, we cannot be sure that profit maximization will be achieved. In other organizations, managers might pursue short-term growth so as to increase their own marketability, while deemphasizing the longer run interests of the company. Recent experiences with Wall Street are consistent with this form of market failure.

Second, either buyers or sellers might have substantial *market power*, i.e., an ability to affect price. Thus, if there are very few sellers of a good or service, those sellers are likely to find it profitable to sell at a price that not only covers their cost, but that allows them to enjoy a supra (larger than)-competitive profit.

Third, either buyers or sellers might be poorly informed about the choices they face. In the most typical instance, certain buyers might lack adequate information about the products and services they are considering. This *asymmetric information* will almost certainly cause markets to fail to allocate goods and services efficiently, and in some cases markets may not exist at all. Asymmetric information problems can be resolved only if both buyers and sellers are given accurate signals about product prices and quality.

Fourth, the products or services at issue may generate *externalities*, whereby the effects of production and consumption activities are not directly reflected in the market. When externalities are harmful and significant, the goods and/or services at issue are likely to be oversupplied. On the other hand, when externalities are beneficial, markets will not sufficiently supply the good.

Fifth, and final, certain goods might be what economists call *public goods*—goods that benefit many consumers, but are undersupplied or not supplied at all by markets. Conditions that lead to public goods include *non-rivalry in consumption* (one person's use

[57] This discussion draws from Robert S. Pindyck and Daniel L. Rubinfeld, *Microeconomics*, 7th Edition, Prentice-Hall, 2009.

of a good does not diminish the ability of another person to use the good) and *non-excludability* (it is impractical or impossible to prevent anyone from using the good—for example, those who do not pay for it—once it is produced). Both of these conditions diminish the ability of markets to supply an optimal amount of the good in question, and as a result, public goods are usually supplied through non-market mechanisms such as government agencies.

Appendix 4. Mechanism Design

Digital preservation poses a set of problems whose solutions depend on the interaction of many individuals with different interests, backgrounds, and knowledge. Mechanism design is a subfield of microeconomics that uses game-theoretic concepts to offer abstract characterizations of solutions to problems involving multiple self-interested individuals or *agents*, each of whom has an individual perspective and private information. As such, the mechanism design literature has the potential to offer insights into the difficult issues associated with digital preservation.

Most economists prefer to rely on the market to solve most problems. When buyers and sellers are well informed and markets are reasonably competitive, markets offer the most efficient mechanisms for arranging the sale and purchase of goods and services. However, markets fail to operate efficiently when the goods and services are nonrival. In such instances, it is impossible to stop individuals from *free riding*—that is, from enjoying the benefits of a good or service without paying. Market failures can also arise when there is incomplete or asymmetric information—when economic actors have access to different information. Nevertheless, even when markets are likely to fail on their own, the market mechanism may well be preferable to an alternative in which the government (which can itself fail) is actively involved.

The literature on mechanism design offers some answers to a core question: Can one design institutions that create incentives for private individuals, acting in their own interests, to make choices and take actions that achieve the desired public purpose—in this case, the socially desirable degree of digital preservation? The mechanism-design literature entails a search for public-private institutions that solve the problems associated with providing goods that are nonrival and would consequently be unlikely to be provided adequately by the market. In some cases the role for government will be narrow; for example, government will set up the rules of the road, by placing some limits on private choices and by using carrots (subsidies) or sticks (penalties) that encourage appropriate private choices. In other cases, government will play a broader role—perhaps by setting up incentives for public agents or agencies to provide nonrival goods and services, such as digital preservation, efficiently.

Questions surrounding mechanism design have been asked for generations. However, the economics literature has formalized design problems, adding some useful terminology, and in the process has developed a number of innovative ideas. Indeed, the 2007 Nobel Prize in Economics was given jointly to Leonid Hurwicz, Eric Maskin, and Roger Myerson for their work on *incentive-compatible mechanisms*, also known as *mechanism design*.

In the most basic of mechanism design frameworks, the *principal* (a social planner—perhaps a member of the executive or legislative branch of government charged with achieving certain preservation goals) employs an *agent* or *agents*, such as a member of an organization with preservation capabilities, to perform tasks. Agents are assumed to act in their own self-interest; the effort that they put forward, if any, will depend on the

incentives that are created by the principal. The principal wants to encourage the agent to work hard to achieve the principal's goals. To make the design worthwhile for the principal, the cost associated with providing incentives to the agent must not exceed the benefit to be gained. For the mechanism to encourage the agent to work hard to achieve the desired purpose, the reward to the agent from doing so must outweigh the benefits associated with alternative choices.

There are a variety of instruments that are potentially available to a principal that wants to achieve a desirable social goal, such as digital preservation. Conduct rules might specify technological standards, while pricing rules might spell out subsidies for preservation of particular digital elements. Other rules might require the provision of socially valuable information that might otherwise be kept private. All of these rules build on the belief that with an appropriate regulatory overview and a well-designed set of incentives, the market can function effectively. In some cases, however, the divergence between private and public benefits may be so great that the best mechanism is government ownership, either partial or complete. The issue then becomes one of setting appropriate incentives for the public entity to strive to achieve the desirable social goal.

Because markets tend to encourage efficient (cost-reducing) behavior, government ownership is usually seen as the solution of last resort to the mechanism design problem. Most basic mechanism designs rely on markets; they involve the use of bonuses, commissions, and other financial incentives to induce agents to make the appropriate work effort. An agent is offered a higher reward for a greater effort and a lesser reward for a lesser effort. While it is easy to see that simple financial incentives of this type have the potential to stimulate effort, it is more difficult see how one can design a mechanism that will ensure that the agent honors the implicit contract (to work hard) between the agent and the principal. Something stronger than trust is likely to be needed. It is important for the principal to monitor the agent's effort, but if undertaken by a third party, the monitoring itself is likely to be costly and difficult to enforce. Therefore, the mechanism-design literature also entails a search for indirect measures of effort that can be easily measured and monitored.

The theory of mechanism design is often tied closely to the theory of non-cooperative games. In the "game" the principal is seen as offering a contract to an agent. The agent either accepts the contract or not, and if she accepts it, she decides on the appropriate action (to exert effort or not). The principal and the agent each face a series of payoffs that depend on the action that the other party takes (the principal offers a contract or not; the agent accepts the contract or not). The social objective is to design the contract between the principal and the agent that achieves the socially desirable outcome. The most desirable contract can, in principle, change over time in response, for example, to changes in technology.

It is natural to think of mechanism design solutions to digital preservation problems because beneficiaries of digital preservation often do not have control over the assets to be preserved when important decisions are made. These beneficiaries, in the language of the theory described here, are principals, and those with responsibility

(legal, physical, or both) for the digital assets are agents. It is not surprising that digital preservation issues can be complex, especially when the principals themselves are many and varied, and the preserved assets are not likely to be adequately provided by the market.

Appendix 5. Representing Stakeholder Interests and the Role of Proxy Organizations

Stakeholder communities attached to a particular set of digital materials can assume a variety of forms. They can be large or small; their boundaries can be clearly demarcated or only vaguely defined. The nature of the benefits the preserved digital materials are perceived to confer may be homogeneous across the stakeholder community, or vary considerably. The stakeholders themselves may be real people whose interests and views on preservation can be solicited directly, or they may be an unborn generation of future users whose interests and views must be anticipated by others.

In any given digital preservation context, it is important that the interests of all stakeholders be represented in the decision making process. In some cases, this is relatively straightforward: for example, when stakeholders constitute a small, well-defined group with a shared set of preservation interests. In other cases, however, representing the interests of all stakeholders becomes much more challenging: for example, when the stakeholders are a large, vaguely defined group widely diffused across both space and time, representing a plethora of varied and sometimes competing preservation interests.

In the latter case, it is often expedient and beneficial to aggregate the collective interests of a stakeholder group associated with a particular set of digital materials within a single *proxy organization* that represents these collective interests in the preservation decision-making process. Proxy organizations often are sanctioned by society, through law or custom, to act on behalf of present and future stakeholders; because they are perceived to act in the public interest, they are often granted special privileges, such as tax considerations or limited rights in copyright law to enable long-term preservation and access. Libraries, archives, museums, historical societies, and other collecting institutions are familiar examples of proxy organizations that act on behalf of large, amorphous groups of current and future stakeholders with diverse interests in preserved materials. These organizations are entrusted to represent the interests of stakeholders in preservation decision making, such as selection, and often carry out the preservation process itself on behalf of stakeholders.

The merits of centralizing preservation decision making in the hands of a proxy organization, rather than attempting to coordinate the participation of the individual members of the constituency in the decision-making process, are readily apparent. Yet the use of proxy organizations introduces other problems for economic sustainability. In particular, there is the question of who funds the proxy organization. Just as it is difficult to coordinate individual stakeholders in the decision making process, it may be equally if not more difficult to coordinate a process by which those who benefit from preservation make appropriate contributions to defray costs. To overcome this problem, proxy organizations responsible for preservation are often embedded within a larger funding context—for example, a national archives might be funded from general tax revenues, or a research library might be funded from the overall university budget.

Proxy organizations have worked well in the print world, as clear divisions of labor have emerged over time. Publishers, for example, keep as many books on their backlist as is economically advisable, but they are not obligated to keep more for the indefinite future. They know that libraries will preserve the books and make them accessible to the public, regardless of commercial value. Commercial film studios have also relied on a small number of archives that specialize in film collecting and preservation to ensure access to old films over time. When the redistribution market grew in the 1980s, film studios began to take a new interest in preserving old movies so they could put them back on the market. Many of them had to turn to these archives for the best available prints for republishing.

Some new forms of digital content, such as collectively produced Web content, exhibit characteristics that suggest a possible role for proxy organizations in their preservation. The stakeholders are scattered and uncoordinated with vaguely defined future interests. However, if proxy organizations are to assume the primary responsibility for preserving content of this kind, they must promptly be granted an appropriately defined writ to preserve, along with adequate resources to achieve their preservation goals. A key point of failure in preserving new genres is that their future value is often not appreciated until it is too late and much of the content has disappeared. The canonical example of this is the fate of silent film. Less than 20 percent of silent films still exist: studios did not assume responsibility for preserving them, and libraries and archives delayed action because of the novelty of the genre and a lack of consensus on the cultural or historical value of these movies. We are in a similar situation with many emerging digital genres; proxy organizations have a useful role to play in preserving these materials, but first we must remove the barriers to their ability to act.

Appendix 6. Flexibility in Preservation Decision Making

Economic models often represent the act of investing as analogous to placing a bet. An investor stakes a sum of money on an investment that is expected—but not guaranteed—to have a profitable outcome at some future time. Once the investment is made, the investor waits passively until the outcome of the investment bet is revealed: success and a positive return on investment, or failure and a loss of the original investment capital.

Although this simple view of investing is sufficient in most cases, in practice the process of managing an investment is much more nuanced and controlled than simply placing a bet. Once an initial investment is made, the investor usually does not become a passive spectator, but instead becomes an active manager, monitoring the investment through a continuous process of evaluating progress, reassessing the probability of success, making incremental course corrections, or even abandoning the investment. For example, a company that has invested in the development of a new technology will usually monitor R&D activities closely, periodically refining expected probabilities of successful commercialization and projected rates of return, and assessing these in the context of costs sunk into the project to date and expected cost to completion. Should the prospects for success take a turn for the worse mid-stream, the company may choose not to see the R&D through to completion, but to abandon the project altogether and avoid incurring additional costs.

Economists recognized the need to provide richer descriptions of the investment process in their models, and have begun to develop theoretical frameworks for this purpose. For example, *the real options theory of investment* adapts the theory of financial options as a foundation for understanding how firms can optimize strategies for managing investment opportunities. This includes the decision to exercise the option to invest, and once exercised, managing the investment activity continuously over time in response to shifting circumstances in the uncertain investment environment. In general, these new theories of investment depart from traditional thinking by focusing not just on the *outcome* of the investment, be it success or failure, but also on the *process* of investment.

Digital preservation is a form of investment, in the sense of allocating resources now in hope of achieving benefits later. Indeed, the Task Force's interim report emphasizes that decision making about digital preservation activities is analogous to decision making about investment. However, as with traditional economic modeling, decision makers attached to digital preservation often focus all of their attention on investment outcomes, to the neglect of investment process. In particular, the decision whether to invest in digital preservation by allocating resources to preservation activities is often perceived in the over-simplified form of a once-and-for-all decision to incur potentially vast upfront and ongoing costs, and to do so in the name of long-term preservation goals that may be defined in terms of centuries. The implicit assumption is that once set in motion, the structure and associated economic commitments of the digital

preservation investment will remain virtually unchanged for the duration of the activity. Not surprisingly, the opportunity to invest in an activity of this kind often seems unattractive, if not daunting, to funders and other decision makers.

Digital preservation is indeed a significant commitment, but to present that commitment as a once-and-for-all, all-or-nothing decision may exaggerate the economic challenge of initiating and sustaining digital preservation activities. This becomes apparent if we focus more attention on the investment process, and less on the long-term outcome. If we do, we can see that investment in digital preservation is not necessarily once-and-for-all, or all-or-nothing.

- *Not once-and-for-all:* If digital preservation investment is perceived as a long-term commitment of resources to extend for centuries, it becomes an exceedingly difficult commitment to make. The long-term costs of preservation are potentially significant, and in many cases, the long-term benefits are uncertain and difficult to quantify. Making such a costly bet on a future with uncertain rewards is not only daunting, but may also seem downright imprudent. It would be more tractable if we were to see such investments as a series of shorter investments, each spanning a time horizon of five to 10 years. That way, decision makers have the opportunity and flexibility to modify or even stop the preservation activity in light of changing circumstances, including the emergence of new information about costs and benefits. A sustainability plan need not extend over 100 years. Planning for five to 10 years, with an option to continue if circumstances warrant, is often the strategy that will have the best outcomes in the short and long terms. Perceiving preservation investment in this light makes it a much less fateful and risky proposition for decision makers.

- *Not all-or-nothing:* Another misleading perception about digital preservation investments is that the technical and curatorial choices are binary: either we implement intensive preservation and curatorial techniques such as format migration or emulation immediately and forever; or we do nothing. This is not a new issue; it is the substance of a longstanding debate in technical quarters. What are the economic implications of implementing a minimalist preservation strategy, something akin to just saving the bits, and maintaining the assets in a condition that preserves the option of implementing more elaborate preservation techniques later, if and when conditions warrant? Recognizing this implicit option may help reduce the perceived magnitude of the economic commitment to investing in digital preservation. A relatively small investment may be enough to preserve the option of making larger commitments in the future if needed.

In practice, digital preservation investments are not once-and-for-all, or all-or-nothing. Decision makers will make course corrections as time passes, and if the activity is no longer economically feasible, it will end. So the implicit flexibility in the investment process—to reduce the time horizon or the current commitment—will likely be

exercised sometime over the life of the investment, whether or not it is explicitly recognized at the beginning. Making clear that this flexibility exists when an investment is first considered may make the scope of the perceived economic commitment less daunting, and correspondingly increase decision makers' willingness to make this commitment.

Appendix 7. Policy Frameworks for Digital Preservation

Digital preservation activities confer benefits on society. The output of such activities—preserved digital materials—can be used to create value in a variety of areas, including research, learning, cultural heritage, and creative expression.

Often, the level of investment in preservation by private decision makers to meet their own needs falls short of what is optimal for society. When this happens, policies that intervene in the private decision-making process can help increase the amount of preservation activity. Such policies can include public policy made by government agencies, as well as policy made by other entities that have influence over preservation decision makers: professional societies, universities, and other organizations that have long-term planning horizons and some responsibility for acting in the public interest.

Policy interventions can be broadly categorized as those that target preservation output, and those that target preservation prices. In other words, interventions can be seen as either setting certain output requirements for preservation, or increasing the benefit/cost ratio of preservation investment perceived by decision makers.

- *Output-focused interventions* change the amount of preservation activity undertaken by removing the decision to preserve from private decision makers. Examples include policy mandates to preserve (interventions that compel decision makers to preserve) and situations where public agencies circumvent private decision makers by undertaking the preservation activity themselves (such as the National Archives and Records Administration and the Library of Congress). Output-focused policy interventions mandate that private decision makers either preserve their content or relinquish their preservation responsibility to another organization.

- *Price-focused interventions* leave preservation decision making in the hands of private decision makers, but attempt to change the incentive structure of the decision, either by increasing the perceived benefits from preservation relative to cost (such as making preservation of important research data an activity relevant for tenure decisions), or lowering the perceived cost of preservation relative to benefits (such as subsidizing preservation activities through matching funds programs). Price-focused policy interventions encourage, but do not force, private decision makers to undertake more preservation activity.

The selection of an appropriate policy mechanism will depend on the circumstances or context in which the policy will be implemented.

Output-Focused Interventions

Many current U.S. federal policies take the form of output-focused mandates, in most cases using the threat of penalty as the motivating factor. In some cases, policies

invoking mandated preservation of digital records are an extension of traditional analog record retention policies. For example, the U.S. Food and Drug Administration has for years imposed recordkeeping requirements for all companies that present new drugs for approval. In many cases, to gain approval a company must preserve records from its drug trials for a number of years past the death of the last participant. Violating this policy results in fines and the drug being pulled from the market. For most of these records, the types of information to be preserved map closely to those in paper records used in years past. While fulfilling digital preservation requirements is costly, such costs are fairly easily integrated into the existing business model.

In other cases, the evolution of the industry has produced records that pose a significant additional burden. The U.S. Federal Aviation Administration, for instance, requires that aeronautics companies keep design records of every aircraft currently in use. While previous aircraft design records had consisted of two-dimensional blueprints, computer drafting software has revolutionized the way aeronautical engineers operate. Aircraft design records now consist of three-dimensional renderings created in proprietary software that may or may not still be in use. Not only are the files enormous, they require a complex software-migration strategy if they are to remain accessible.

In the preceding examples, the U.S. government instituted preservation mandates without offering any preservation capacity or other resources to aid the process; however, this is not always the case. In 2007, the NIH announced an open-access policy for all research that it funds, with the rationale that if taxpayers pay for a project, they should be able to read the results. Rather than leaving the mechanics of this access to the researchers, NIH mandates that the information be deposited into PubMed Central Digital Library, a database of scientific literature sponsored by the National Library of Medicine.

In another type of preservation-focused mandate, public policy does not require the owner of digital content to preserve, but instead empowers other agencies to undertake preservation on society's behalf. Traditionally, the U.S. government has addressed the issue of preservation support by giving libraries, archives, and other organizations with a preservation mission the right to preserve others' material, through Section 108 in the U.S. Copyright Code. These organizations are willing and able to accept material, but it is left to the private decision maker to actually transfer protection.

Unfortunately, in the digital age, U.S. copyright law sometimes hinders preservation. Libraries and archives have the right to make up to three preservation copies of physical files, regardless of who owns the copyright; however, the U.S. Copyright Office has specifically declined to extend this right to digital files. Since the mechanics of computer storage legally constitutes copying, libraries and archives are put in a potentially precarious situation if they attempt to preserve the files. Recently, the Section 108 Study Group looked at the issue of preservation and intellectual property

rights in the digital world.[58] It recommended several policy tools that the U.S. government could use to encourage preservation.

Price-Focused Interventions

In addition to output-focused policies, there are other mechanisms by which policy makers can encourage desired preservation behaviors. A useful framework for thinking about such strategies is provided by Thaler and Sunstein in their book *Nudge*.[59] Introducing the concept of *choice architecture*, the authors describe how policies can be designed to make doing the hoped-for activity easier than not doing it. In the context of digital preservation, choice architecture is more easily assigned to policies that affect the price of preservation, rather than the output. Each of the six categories of interventions described in *Nudge* could be appropriate for encouraging preservation.

Incentives: The most basic element of choice architecture is the incentive, which gives individuals positive motivation to act. A classic government incentive is the tax credit, wherein a government rewards an activity by lowering the taxes of those who perform it. The United States rewards preservation of historic architecture in this manner, with the Historic Preservation Tax Incentives program, administered by the National Park Service. The Library of Congress has investigated how these policies could be applied to preservation of digital information.[60]

The corollary to instituting incentives is getting rid of de-motivating factors, or disincentives. One of the largest disincentives for scientific researchers to deposit their data in digital repositories is the threat of others using their data to further themselves without crediting the researcher. Scholarly publishing provides one antidote to this by allowing researchers to publish their data sets so that future users must include a formal citation to the original researcher. Many researchers in the biological sciences are particularly focused on this issue. The *Journal of Nucleic Acids Research*, for example, publishes an annual data set issue, which acts as a citation focus. This indirectly increases the incentive to preserve valuable research data by removing a powerful disincentive for doing so.

Mapping: Many times a decision maker will not choose a desired behavior simply because it is difficult to truly understand the differences between options. For example, it is often difficult to decide between cell-phone plans; hidden costs and confusing terms make comparing the ultimate monthly payment nearly impossible. *Mapping* is the term Thaler and Sunstein have given to the process of giving decision makers equivalent information for disparate choices.

---

[58] See *The Section 108 Study Group Report* by the United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress. Available at http://www.section108.gov/docs/Sec108StudyGroupReport.pdf.

[59] See Richard Thaler and Cass Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Yale University Press, 2008.

[60] See "Proposals for Creation of a Public Policy Environment Conducive to Digital Preservation," Appendix E of *Preserving Our Digital Heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report. A Collaborative Initiative of the Library of Congress,* forthcoming from the Library of Congress.

Costs for one digital preservation program often do not directly map to costs in other programs, making it extremely difficult for decision makers to create an accurate budget for preservation. The United Kingdom's Joint Information Systems Committee (JISC) sponsored a study to address this disincentive, resulting in the Beagrie, Chruszcz, Lavoie (BCL) cost model and publication *Keeping Research Data Safe*.[61] This project attempted to standardize the elements that go into preserving digital scholarly discourse and to provide cost estimates for each of those elements. The hope is that decision makers can then use these numbers as the basis of their own budgeting.

Defaults: One extremely efficient method of guiding a decision maker's actions is to set defaults to the preferred option. This method has been embraced by the business community, as seen by anyone who has registered with an online service but forgot to deselect the option to receive notifications of special discounts and one-day sales. Using *opt-out* rather than *opt-in* policies, policy makers turn inertia to their own advantage.

Academic policy makers have begun to use defaults to promote preservation within their own institutions. Many universities have created digital repositories, such as the Digital Access to Scholarship at Harvard (DASH) repository, for works created by members of their faculty and staff. Traditionally, the contract between a faculty member and his or her university includes a copyright policy that gives all rights to the faculty author, including the right to decide whether or not to provide open access to the work. To encourage preservation, however, the Harvard University Faculty of Arts and Sciences in 2008 unanimously approved a motion to require each faculty member to give an electronic copy of every scholarly article he or she publishes to the Harvard Provost's Office. The provost then deposits the article in DASH, which is freely open on the Web. While a faculty member can request a waiver, the default process requires preservation.

Feedback: Sometimes simply reminding people of the potential for disaster can be a powerful motivating force. There are numerous examples of landmark digital materials being lost, and alerting the public to this fact can spur action. The rumor that several tapes of the moon landing, a universal cultural milestone, had been lost was spread widely despite its apocryphal nature; the strength of this rumor is an indication of the powerful response such a loss would elicit. An advertising campaign focused on other digital losses, preferably ones with equal cultural weight, would put pressure on decision makers to preserve. This type of strategy, however, should be used sparingly to prevent desensitization to the message.

Another type of feedback that can have an impact on behavior is a public comparison between the decision maker's activities and those of his or her peers. Simply knowing how well your behavior compares with others' can encourage good behavior, or inspire improvement when your behavior doesn't measure up. In the world of digital preservation, many institutions currently advertise the number of files they store and

[61] See http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf.

the number of users who access their information; if this type of metric were more widely shared, it might encourage preservation efforts across institutions.

Expecting error: Making the correct decision often requires more thought and effort than making the incorrect decision. In cases where it seems that a decision maker has made a bad choice, it may be useful to assume that the choice was made in error. For example, it is easy to forget to save your work on a document before closing it, which is why software designers have incorporated the "would you like to save?" text box into so many exit mechanisms.

At the level of individual data creators, preservation can be easily ignored. Anyone who has experienced a hard-drive failure without a backup in place knows this fact intimately. Private companies know this as well, which is why most employers over a certain size institute automatic backup of all employees' computers. A solid disaster strategy requires that even these backups be mirrored in separate locations. This is the thinking behind the Lots of Copies Keeps Stuff Safe (LOCKSS) program. In the LOCKSS network, libraries serve as each others' back-ups; if something happens to their own files, the library can rely on the many copies stored throughout the world to replace them.

Choice structuring: A last element of choice architecture involves attempting to simplify difficult decisions by easily eliminating options. It is easy to become overwhelmed when presented with a barrage of choices, and the ability to exclude options often makes the decision more manageable.

The proteomics data repository Tranche has built this thinking into its system through its metadata requirements.[62] Metadata, the crucial descriptive information that accompanies research data, is extremely complex, and requirements are often voluminous. Tranche allows researchers to assign whole categories of metadata to other team members, effectively eliminating the categories that they themselves do not have the expertise to address. Through structuring of complex choices, Tranche has solved the problem of researchers becoming overwhelmed with metadata requirements and ignoring them entirely, or even worse, refusing to deposit their data.

Both output-oriented and price-oriented policies can be used to encourage digital preservation. The framework discussed here represents a way to organize these options into distinct categories, each of which will work better or worse depending on the circumstances in which it is applied. Policy makers must therefore take careful stock of the preservation landscape if they are to choose the appropriate mechanism to promote the desired behavior.

[62] See https://trancheproject.org/.

Glossary

Frequently Used Terms and Acronyms

actors: any individual, group, or organization that plays a role in a preservation strategy.

analog preservation: activities focused on maintaining usability of traditional (non-digital) information resources, including print materials, film and video, and a variety of music carriers such as vinyl discs, magnetic tape, and glass cylinders.

ARC: Astrophysical Research Consortium

archiving: activities that enable long-term retention of digital materials. Together with *curation*, often referred to as *stewardship*.

authenticity: in preservation, refers to the perfect identity of a current object with its original state.

born digital: something that originates in digital form, as opposed to being converted from analog to digital format.

choice variables: the choices available to decision makers in designing sustainable preservation strategies. Choice variables are not attached to any particular class of digital assets and are found across a spectrum of preservation strategies.

CLOCKSS: Controlled LOCKSS (see LOCKSS below)

collectively produced Web content: content that is created interactively on the Web, the result of collaboration and contributions by consumers. Because of its collective nature, the ownership of such content is often ambiguous.

commercially owned cultural content: digital assets that are owned by private entities and are under copyright protection. This includes journalism, music, film, games, and an array of creative expression and popular culture.

content: digital information; the term is often used interchangeably with *data* or *digital assets*.

context-specific attributes: features of digital information that vary according to data type or user community and that shape or constrain choices among preservation strategies.

core attributes of digital assets: elements that are common to all preserved digital assets as economic goods. The attributes are: preservation is a derived demand; the materials are depreciable durable assets; the materials are nonrival in consumption; and the digital preservation process is temporally dynamic and path dependent.

curation: activities that enable use and long-term accessibility. In digital preservation, the curation and archiving together comprise *stewardship*.

dark archive: a digital repository that is not publicly accessible; often used for secure storage and back-up, and for materials embargoed for one reason or another.

data: digital information; the term is often used interchangeably with *content* or *digital assets*.

decision maker: an individual or group whose actions can determine preservation outcomes.

depreciable durable asset: something created to produce a flow of value over time, with the quality and quantity of the flow declining over time if actions are not taken to maintain the viability or productivity of the asset.

derived demand: demand for a good or service that is produced in the service of something else that is valued. Preservation is a derived demand; it is valued because it enables access to information over time.

digital assets: digital information; the term is often used interchangeably with *data* or *content*. The term does not refer to monetary or financial aspects in the asset unless specifically mentioned.

digital orphans: information assets whose ownership and provenance are uncertain, or whose owner is unwilling or unable to preserve it. These assets are at especially high risk of loss over time.

digital repository: a place where digital assets are deposited and stored.

disintermediation: the removal of an intermediary in a process. Digital technologies enable individuals and small groups to bypass publishers, for example, in creating and distributing content.

economically sustainable preservation: a means of keeping information accessible and usable over time by ensuring the ongoing and efficient allocation of resources to its maintenance.

emerging literature: new genres of scholarly discourse, such as online collaborative spaces, academic blogs, websites, e-prints.

free-rider problem: a situation arising when goods are nonrival in consumption, when benefits accrue to those who don't pay for them. For example, the costs of preserving digital assets may be borne by one organization, but the benefits accrue to many.

handoff: mechanism for transferring custody and/or preservation responsibility from one party to another.

ICPSR: Inter-university Consortium for Political and Social Research

IIPC: International Internet Preservation Consortium

irreversibility: in preservation, the fact that actions taken at one point in the digital lifecycle can predetermine others downstream. For example, a decision not to preserve for one reason or another may mean that the data are forever lost.

JISC: Joint Information Systems Committee

KB: Koninklijke Bibliotheek [Dutch National Library]

layered demand: when a good or service is of interest to those beyond the immediate constituency for which it was created.

lifecycle: a series of stages through which something, in this case digital information, passes during its lifetime. The lifecycle for digital information includes creation, use and reuse, migration or emulation, and storage.

LOCKSS: Lots of Copies Keep Stuff Safe

on the margin: investments are made "on the margin" when they are seen as incremental to existing expenditures or processes.

misalignment of incentives: a situation in which the factors that motivate one stakeholder to take action differ from those that motivate another stakeholder. An example of misaligned incentives in preservation is when an entity that owns information no longer sees value in keeping it and has no motivation to hand it off, yet another entity sees value in preserving the resource, but does not have the right to.

MOU: memorandum of understanding

NARA: National Records and Archives Administration

NDIIPP: National Digital Information Infrastructure and Preservation Program of the Library of Congress

negative benefit: the undesirable effects of a decision or action (or failure to decide or act).

nonrivalrous consumption: a situation in which one person's use of a good does not impede or detract from another person's use of the same good. Digital assets are nonrivalrous goods because the cost of providing access to additional users is close to zero.

NSF: National Science Foundation

option strategy: the strategy of making modest investments in preservation to hedge against irreversible loss. By making a small current investment—for example, in secure storage, with no additional curation—decision makers effectively purchase the option of postponing their choice until they have better information.

path-dependency: when decisions one faces for any given circumstance are limited by the decisions one has made in the past, even though past circumstances may no longer be relevant.

PSID: Panel Study of Income Dynamics

positive externality: an economic good produced by one party that has benefits that accrue to others. Such goods tend to be under produced in the marketplace.

preservation: activities that enable the use and long-term accessibility of information; often used interchangeably with *stewardship*.

preservation strategy: the series of decisions taken over the course of the digital lifecycle to ensure long-term accessibility and usability, and to reduce outstanding risks to loss and degradation of the materials.

proxy organization: an entity that is sanctioned to act on behalf of present and future stakeholders. In preservation, a proxy organization might articulate constituent demand, promulgate best practices, develop selection criteria, and have the authority to enter into contractual relationships with archives. Examples of proxy organizations include libraries, archives, museums, some scientific consortia, and some professional societies.

research data: primary inputs into scientific, humanities and other research as well as the first-order results of that research.

scholarly discourse: the published output of scholarly inquiry; often called scholarly communication.

SDSS: Sloan Digital Sky Survey

stakeholder: any individual, group, or entity that benefits from access to and use of preserved information, or who support or fund those who do.

STEM: science, technology, engineering, and mathematics

stewardship: activities that enable use and long-term accessibility of information; often used interchangeably with *preservation*.

temporally dynamic: referring to an activity that takes place over time and may change during that time.

value proposition: a statement about or shared understanding of the benefits of a particular good or service.

wwPDB: Worldwide Protein Data Bank