Santa Clara High Technology Law Journal

Volume 33 | Issue 2 Article 1

1-1-2017

"Through A Glass, Darkly" Technical, Policy, and Financial Actions to Avert the Coming Digital Dark Ages

Richard S. Whitt

"THROUGH A GLASS, DARKLY"

TECHNICAL, POLICY, AND FINANCIAL ACTIONS TO AVERT THE COMING DIGITAL DARK AGES

Richard S. Whitt[†]

This paper explains the digital preservation challenge, examines various technical, legal, commercial, and governance elements, and recommends concrete proposals in each area. The research is based on a wide-ranging review of pertinent books, reports, studies, and articles familiar to experts in the digital preservation community.

Much of our global cultural heritage, and our own individual and social imprint, is at serious risk of disappearing. More and more of our lives is bound to the ones and zeroes of bits residing on a cloud server or a mobile device. Those bits in turn are mediated by the software and hardware implements we utilize every day. The bitstreams are unintelligible, however, without suitable data formats, computer applications, operating systems, and hardware environments to interpret them for us. As those systems are modified or replaced over time, we inevitably lose our ability to access the content. The resulting technological obsolescence can leave us trapped in a "digital dark age" as a culture that has lost its collective memory. As our reliance on data grows even more pervasive in every sector, massive technology and market trends—such as born-digital content, cloud computing, "big data," and the Internet of Things—will only accelerate the scale and scope of the problem.

[†] Mr. Whitt is currently the Corporate Director for Strategic Initiatives at Google, but this paper does not necessarily reflect the views of Google as an organization. The author sincerely thanks Vint Cerf—colleague, mentor, and friend for nearly eighteen years—for spurring his work on this topic. Like his enthusiasm for the Internet, Mr. Cerf's kind and patient brand of inspiration seemingly knows no bounds. The author also thanks summer intern Greyson Nevins-Archer, who was a thoughtful and tireless researcher, drafter, and keen questioner on numerous salient points. Much appreciation as well to Nicholas Taylor of Stanford University for his numerous excellent pointers on many topics, to Lila Bailey, Pam Samuelson, and Erik Stallman for generously sharing their considerable legal expertise on copyright issues, and to Phil Weiser and Erica Gaines for their helpful suggestions. The title of this article comes from 1 *Corinthians* 13:12 (King James) ("For now we see through a glass, darkly; but then face to face."). Other translators substitute the words "mirror" and "dimly," but the gist here remains the same: we are losing the collective ability to witness ourselves.

The digital preservation challenge is multidimensional, requiring us not just to develop and implement technical solutions—such as proposed migration and emulation techniques—but also to address the relevant public policy components. Legal frameworks, notably copyright and contract laws, pose significant hurdles to the usability and accessibility of preserved content. Moreover, the misalignment of financial incentives undermines the prospects of creating viable and economically sustainable solutions.

Despite these significant hurdles, concrete and achievable next steps are possible. In addition to highlighting technical proposals, this paper recommends ways to operate within, or change, existing laws, engage content owners, and harness the interests of ordinary end users. Policymakers also can take steps that will establish the near- and long-term value of preserved content, highlight the costs of inaction, and create new financial incentives.

An underappreciated challenge is the need to organize ongoing activities into a more persistent, broad-based, and ever-evolving process. Another novel aspect of the paper is the suggested borrowing of key learnings from the Internet governance world. For example, one can combine the existing approach of managing the natural "lifecycle" of information across time with managing the various layers of technology across space. Moreover, as with the Internet, the digital preservation community of libraries, archives and museums (LAMs)—and many others—can institute a phased stakeholder model to help govern its activities, facilitate coordination and cooperation, develop further trust and permanence, and unify pertinent stakeholders behind a common mission.

TABLE OF CONTENTS

INTRODUCTION

This paper addresses the increasingly thorny and far-reaching problem of preserving the digital records of our past, present, and future cultural heritage. The technical challenge is straightforward: "[d]ata are at risk because they are recorded on a transient medium, in a specified file format, and they need a transient coding scheme . . . to interpret them."[1] The issue thus becomes "the survival of information in an electronic environment."[2] Indeed, "in just 50 years . . . the human record of the early 21st century may be unreadable."[3] The challenge is twofold: both a failure to archive what later we want to interpret, and an inability eventually to interpret what we have managed to archive.[4] Thus, "if we are not to tolerate gaps in society's memory in the future, we must intervene in the present to secure the long-term availability of culturally significant digital materials."[5]

Over the past twenty years, the rise of the Internet and other advanced digital technology platforms have had a profound economic, cultural, and social impact.[6] The Internet itself serves as a general platform technology, leading to long-term economic growth,

1. Marilyn Deegan & Simon Tanner, *Key Issues in Digital Preservation*, *in* DIGITAL PRESERVATION 1, 12–13 (Marilyn Deegan & Simon Tanner eds., 2006).

2. Gordon B. Neavill, *Electronic Publishing, Libraries, and the Survival of Information*, 28 LIBR. RESOURCES & TECHNICAL SERVS. 76, 78 (1984).

3. Deegan & Tanner, *supra* note 1, at 7.

4. David Rosenthal, *The Half-Empty Archive*, DSHR'S BLOG (Mar. 31, 2014), http://bit.do/HalfEmptyArchive. Rosenthal believes that the long-term rate of content loss to future users from an initial failure to even collect content for preservation will be at least fifty percent—dwarfing all other causes of content loss, such as bit rot and format obsolescence.

5. Brian F. Lavoie, *The Costs of Digital Preservation*, *in* DIGITAL PRESERVATION 106, 106 (Marilyn Deegan & Simon Tanner eds., 2006).

6. For a description of these impacts, see HANDBOOK ON THE ECONOMICS OF THE INTERNET (Johannes M. Bauer & Michael Latzer eds., 2016).

substantial innovation spillovers, and other positive externalities.[7] The digital era in particular—with its ubiquitous computing, widespread online access, and growing role for mobile and environmental devices—presents unprecedented opportunities to expand and deepen the world's access to information.

At the same time, increasingly we are being called upon to fundamentally rethink our conceptions of meaning and knowledge. Three brief examples of this type of reappraisal are digital humanities,[8] knowledge commons,[9] and access to knowledge.[10] While most treatments of digital preservation challenges highlight only the potential considerable loss of knowledge and culture due to the advent of digital technologies, we must not lose sight of the considerable "upside" enabled by those very same technologies. As Rosenzweig observes, the most vexing problems of digital media are the flipside of their greatest virtues.[11] In essence, we stand to be deprived of the greatest future benefits of tools that simultaneously may rob us of our past. We risk losing both the upside benefits of unborn potential, and the downside costs of squandered actuality.

The late Jim Gray used to say, "May all your problems be technical."[12] There is no doubt that terrific technical progress is

---

7. Stephen J. Schultze & Richard S. Whitt, *The Internet as a Complex Layered System*, *in* HANDBOOK ON THE ECONOMICS OF THE INTERNET, *supra* note 6, at 55, 55-59.

8. Digital humanities are new modes of knowledge production, based on the concept of user-focused archives. Ubiquitous computing opens up new opportunities for participation, dissension, and freedom by all stakeholders. The digital humanities also provide both a key rationale for wanting to preserve knowledge in the first place, and a way of thought that helps us figure out exactly what we want to preserve, and why. The capacity to create enhanced forms of curation also brings humanistic values into play in ways that were difficult to achieve in traditional museum or library settings. This includes choices about what remains and what is eliminated, what is made accessible, how, and in what form. *See* ANNE BURDICK ET AL., DIGITAL HUMANITIES 17-19 (2012).

9. The knowledge commons has grown out of various digital era information-sharing initiatives which enable peer production, collective action, and open access to information. Nancy Kranich, *Countering Enclosure: Reclaiming the Knowledge Commons*, *in* UNDERSTANDING KNOWLEDGE AS A COMMONS 85, 93 (Charlotte Hess & Elinor Ostrom eds., 2006). Digital technologies have a huge role in the robustness—or vulnerability—of knowledge as a shared resource. Hess & Ostrom, *Introduction to* UNDERSTANDING KNOWLEDGE AS A COMMONS 1, 10.

10. So-called "A2K" movements have sought to open up the ability of all the world's citizens to interact with and utilize the world's "knowledge goods," regardless of social or economic status. *Preface to* ACCESS TO KNOWLEDGE IN THE AGE OF INTELLECTUAL PROPERTY 8, 11 (Gaelle Krikorian & Amy Kapczynski eds., 2010) (Yochai Benkler identifying the "intellectual commons" as the central concept of A2K).

11. ROY ROSENZWEIG, *Scarcity or Abundance? Preserving the Past*, *in* CLIO WIRED 3, 9 (2011).

12. EDWARD M. CORRADO & HEATHER MOULAISON, DIGITAL PRESERVATION FOR LIBRARIES, ARCHIVES, AND MUSEUMS xvi (2014).

occurring in many venues across the globe. And yet, experts generally agree that digital preservation involves recognizing and overcoming not just technical challenges, but those touching as well on the relevant policy and economic environments.[13] According to a 2008 report, there are "significant technical, financial, and legal obstacles to digital preservation."[14] A pertinent standards document notes that it would be unwise to consider the problem of technological obsolescence solely from the technical perspective, as "there are also organizational, legal, industrial, scientific, and cultural issues to be considered."[15] Some refer conceptually to a "three-legged stool," consisting of the organizational, technological, and funding questions intrinsic to a viable preservation program,[16] others to "a triad of interrelated" management, technological, and content activities.[17]

Figuring out how better to organize these disparate ongoing efforts is the higher-level challenge. In many respects, data management is a "people problem," rather than a "technical problem."[18] Indeed, two of the most intractable areas of concern, the law and the money, while ostensibly separate from the technical side, significantly affect and constrain those solutions. Nor do these activities carry much meaning without understanding how they all can be coordinated and organized into a comprehensive set of solutions. Without figuring out how technical proposals can comply with existing laws and regulations, or how to create proper financial incentives to adopt a technical solution, or how ultimately to pull it all together, the best efforts of the world's software experts will be for naught.[19]

Part I of this paper presents the scope and scale of the digital preservation challenge. In the next three Parts, the digital landscape is

---

13. Lavoie, *supra* note 5, at 107.

14. LIBR. OF CONG. NAT'L DIGITAL INFO. INFRASTRUCTURE AND PRESERVATION PROGRAM ET AL., INTERNATIONAL STUDY ON THE IMPACT OF COPYRIGHT LAW ON DIGITAL PRESERVATION 2 (July 2008) [hereinafter STUDY ON COPYRIGHT LAW].

15. CONSULTATIVE COMM. FOR SPACE DATA SYS., REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS) 1-3 (2012).

16. DIGITAL PRESERVATION COALITION, *Introduction* to DIGITAL PRESERVATION HANDBOOK, (2d ed. 2015), http://bit.do/IntroToDigitalPreservationHandbook.

17. CORRADO & MOULAISON, *supra* note 12, at xix; *see also* ROSS HARVEY, PRESERVING DIGITAL MATERIALS 3 (2d ed. 2011) (technology fixes are bound up with the organizational infrastructure, resources, and legal factors).

18. Catharine Ward & Lesley Freiman, *Making Sense: Talking Data Management with Researchers*, 6 INT'L J. OF DIGITAL CURATION 265, 272 (2011).

19. In practice, the compliance, funding, and management elements tend to bleed one into the others. The technological mechanisms are informed by the economic incentives, which in turn are shaped by the policy environment, which is influenced by the organizing process, which feeds back into the technology solutions, and so on.

analyzed from the interrelated dimensions of technical issues (software and hardware), public policy issues (law and politics), and financial issues (commercial incentives and business models). Part V lays out recommended ways of better coordinating the pertinent people and processes, by building so-called "deep infrastructure" that includes a proposed new digital life-cycle/system-layers framework and ecosystem governance mechanisms. Part VI explores potential next steps in the technical, policy, and financial realms, and summarizes an array of options for interested technology companies to consider.

I. SCOPE OF THE CHALLENGE

> We may all . . . be swimming against the tide . . . . Our society is obsessed with the present and is generally uncaring of the past and of its records . . . . The technology that allows us to interact with information itself inhibits us from preserving our interaction.[20]

A. What Is at Stake?

While perhaps it is all too easy to sound alarm bells needlessly, it is no hyperbole to suggest that some sizable portion of our combined global cultural and historical heritage is at serious risk of disappearing. This vanishing will not occur at the level of the atoms we can see and interact with physically, but rather, at the level of the ones and zeroes of bits. What will fade into oblivion is not the actual instantiation of meaning, but meaning itself. We may well become a world of societal interpreters without much to interpret. As Vint Cerf puts it, "I worry that the twenty-first century will become an informational black hole."[21]

The need to preserve our digital heritage can be seen as a virtual analog to worldwide climate change. Each constitutes a looming threat that numbs the mind by virtue of its long-term timelines, diffused stakeholders, misaligned incentives, and largely invisible nature. Left unaddressed, each also promises to bring a devastating impact to our world. In the case of digital preservation, what is required to prevent this potential catastrophe are technical, legal, financial, and organizing solutions that preserve not natural ecosystems, but human understanding.

---

20. Peter S. Graham, *Intellectual Preservation and Electronic Intellectual Property,* COALITION FOR NETWORKED INFO. (1994), http://bit.do/IntelPreservation.

21. Jill Lepore, *The Cobweb: Can the Internet be archived?* NEW YORKER (Jan. 26, 2015), http://bit.do/TheCobwebInternetArchived (quoting Vint Cerf).

The ability of a culture to survive into the future depends on the richness and acuity of its collective sense of history.[22] Indeed, "the very foundations of our civilization [are] based upon our ability to pass information and knowledge, whether technical or cultural, from one generation to the next."[23] And yet, "the greatest danger to digital materials is that we forget the meaning of them."[24] Mankind may be in the process of losing its historical dimension.[25] Ironically, the main obstacle to long-term preservation of digital documents is the relentless march of innovation.[26]

We must now grapple with the very real prospect of "an impoverished digital future,"[27] due to what Mary Feeney calls "the death of the digit."[28] The life of digital information threatens to be, in the words of Hobbes, "nasty, brutish, and short."[29] Without finding technically, legally, and commercially sustainable methods of preserving our digital heritage, our civilization could well face "the grim scenario of a culture without memory."[30]

B. What Is the Matter?

All communication, in whatever form, is mediated by technology, and therefore is "context dependent."[31] Information systems have always been shaped by available technologies that have transformed "the creation, capture, preservation, and discovery of content."[32] With analog print content, for example—information that is embedded directly into a physical artifact—the user typically can perceive the content directly. With such physical documents, "saving the physical

---

22. DONALD WATERS & JOHN GARRETT, TASK FORCE ON ARCHIVING OF DIGITAL INFO., PRESERVING DIGITAL INFORMATION 1 (1996). UNESCO's "Memory of the World" project, initiated in 1992, is one example of an attempt to guard against "collective amnesia" from a growing inability to create, preserve, and access our documentary heritage. *Memory of the World*, UNESCO, http://bit.do/MemoryWorld.

23. Deegan & Tanner, *supra* note 1, at 3.

24. Michael Lesk, *Foreword* to CORRADO & MOULAISON, *supra* note 12, at xv, xvii.

25. BORGHOFF ET AL., LONG-TERM PRESERVATION OF DIGITAL DOCUMENTS v (2006).

26. *Id.* at 11.

27. Deegan & Tanner, *supra* note 1, at 5.

28. Mary Feeney, *Digital Culture: Maximising the Nation's Investment*, 5 U.K. NAT'L PRESERVATION OFF. J. 12, 12 (1999), http://bit.do/DigitalCulture.

29. WATERS & GARRETT, *supra* note 22, at 2.

30. BORGHOFF ET AL., *supra* note 25, at vi.

31. Michael Moss*, What is the Same and What is Different*, *in* IS DIGITAL DIFFERENT? 1, 5 (Michael Moss, Barbara Endicott-Popovsky & Marc J. Dupuis eds., 2015).

32. Michael Moss & Barbara Endicott-Popovsky, *Introduction and Acknowledgements* to IS DIGITAL DIFFERENT?, *supra* note 31, at xv.

carrier saves all those attributes of the original that it is possible to save."[33]

Digital documents differ from analog documents in that they are not inextricably bound to their "containers," and therefore "preserving them is not necessarily a matter of preserving containers as it is in the analogue world."[34] In fact, it is the very nature of digital content to be incredibly vulnerable to its immediate software and hardware environment. Without that environment, the content in essence does not exist, at least so far as human beings are concerned. Unlike many analog originals such as paper or paintings, "data are very bad at self-preservation."[35] "[D]igital content can be made manifest only through the use of specific software and hardware," thus creating a "dependency on a technological intermediary."[36]

Marilyn Deegan explains that the bitstream for a word processing document is totally unintelligible without the suitable computer applications, software, operating system, and hardware environments to interpret and repackage the data into a readable form. As a result, Deegan concludes:

> Digital data are in danger, not because they are inherently fragile or flawed, but because there is a continually accelerating rate of replication, adaptation, and replacement of hardware, software, and data formats and standards, which may mean that the bitstream may not be readable, interpretable, or usable long into the future.[37]

Susan Lazinger provides a helpful taxonomy of the reasons why digital content is at such risk.[38] These reasons include the uncontrolled accumulation of data; inadvertent destruction of data; unauthorized tampering with data; lack of metadata and systems documentation; electronic data in forms that cannot be preserved; technology obsolescence, both software and hardware; and encryption-related issues.[39] Of those factors, two pose the most fundamental challenge to preserving digital content: media deterioration and technological

33. Deegan & Tanner, *supra* note 1, at 13 (quoting J. Rothenberg).

34. *Id.*

35. *Id.* at 15.

36. Robin Wendler, *The Status of Preservation Metadata in the Digital Library Community*, *in* DIGITAL PRESERVATION 60, 60 (Marilyn Deegan & Simon Tanner eds., 2006).

37. Deegan & Tanner, *supra* note 1, at 6.

38. SUSAN S. LAZINGER, DIGITAL PRESERVATION AND METADATA 5-15 (2001).

39. *Id.*

obsolescence.[40] Regardless of their form, all digital materials are threatened by both.[41]

Media deterioration—often called "data rot" or "data decay"—involves the eventual break-down of the data itself as found in digital storage materials.[42] Embodying creative works in digital form "has the unfortunate effect of potentially decreasing their usable lifespan. Digital information is ephemeral: it is easily deleted, written over, or corrupted."[43] Or as Lazinger puts it, "electronic information is fragile and evanescent."[44] Digital materials are especially vulnerable to loss and destruction because "they are stored on fragile magnetic and optical media that deteriorate rapidly and that can fail suddenly from exposure to heat, humidity, airborne contaminants, or faulty reading and writing devices."[45] Natural processes mostly driven by thermal energy are a primary culprit in destroying data over time.[46]

Separate from media deterioration is technological obsolescence, where the hardware and software required to access the bits is difficult or impossible to obtain. With digital data, "a machine needs to be interposed between the data and the human interpreter, which adds another layer of complication."[47] This involves translation from machine- to human-readable form.[48] As a result, "it is the essential character of the information object, not the way it happens to be encoded digitally, that must be preserved."[49]

"Digital objects are inherently fragile, in part due to the rapid pace of manufactured hardware and software obsolescence."[50] Access to

---

40. WATERS & GARRETT, *supra* note 22, at iii; *see also* Deegan & Tanner, *supra* note 1, at 5–6 (due to the acceleration of technology obsolescence and the instability of digital resources, the integrity and authenticity of digital resources are corrupted).

41. MARGARET HEDSTROM & SHEON MONTGOMERY, DIGITAL PRESERVATION NEEDS AND REQUIREMENTS IN RLG MEMBER INSTITUTIONS 3 (1999).

42. Jim Salter, *Bitrot and Atomic COWs*, ARS TECHNICA (Jan. 15, 2004), http://bit.do/BitrotAndAtomicCows; Nic Luo, *What "Bit Rot" Looks Like*, NICLUO.COM (May 3, 2015), http://bit.do/WhatBitrotLooksLike.

43. STUDY OF COPYRIGHT LAW, *supra* note 14, at 1.

44. LAZINGER, *supra* note 38, at 6 (quoting Neil Beagrie & Daniel Greenstein).

45. HEDSTROM & MONTGOMERY, *supra* note 41, at 1.

46. Tom Coughlin, *Storage for the Next 5,000 Years*, FORBES (Dec. 15, 2015), http://bit.do/StorageNext5kYears.

47. Deegan & Tanner, *supra* note 1, at 13.

48. HEDSTROM & MONTGOMERY, *supra* note 41, at 1.

49. Kenneth Thibodeau, *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*, *in* THE STATE OF DIGITAL PRESERVATION: AN INTERNATIONAL PERSPECTIVE 4, 5 (2002).

50. Stephen Chapman, *It's Money that Matters in Long-Term Preservation*, *in* DIGITAL PRESERVATION 133, 134 (Marilyn Deegan & Simon Tanner eds., 2006). Or as Thibodeau puts it, "any digital object maintained unchanged for any length of time will become inaccessible."

digital content requires "external dependencies"—hardware, software, or physical carriers—that may no longer be manufactured, minted, or supported.[51] Take, for example, the evolving world of data storage. The storage of data started with punched cards only some fifty years ago, and has transitioned through paper tape and magnetic tape, to magnetic disc, optical disc, and portable memory such as flash memory to the present day.[52] This obsolescence can be repeated across all dimensions of the digital content—from the storage device, to the user device, to the operating system, to the source code, to the application itself. As a result**,** while "we can read the 400 year old books printed by Gutenberg, it is often difficult to read a 15 year old computer disk."[53]

While the issue of failing to collect content upfront is very real and should not be overlooked,[54] this paper will focus primarily on the technology obsolescence challenge, as the most far-reaching and difficult issue to combat. Even twenty years ago, it already was recognized that technological obsolescence represents a far greater threat to digital information than the inherent fragility of many digital media.[55] Our focus should be on ensuring the long-term survival and accessibility of bitstreams.

C. Repeating Losses of the Past

Waters notes that "any culture depends on the quality of its record of knowledge."[56] Despite the often heroic efforts of unheralded and even unknown individuals and institutions, history is replete with examples of past lost cultural heritage, to the detriment of our collective memory. In fact, most of the records of previous historical eras have disappeared.[57] Archiving has always been what Nicholas Taylor calls "a lossy endeavor."[58] Brewster Kahle of the Internet Archive notes for example that:

Kenneth Thibodeau et al., *Persistent Object Preservation: Advanced Computing Infrastructure for Digital Preservation*, DLM FORUM (1999), http://bit.do/PersistentObjectPreservation.

51. *E.g.*, *Digital Preservation: Sustainability Factors*, LIBR. OF CONGRESS (2015), http://bit.do/SustainabilityFactors.

52. Deegan & Tanner, *supra* note 1, at 15.

53. Graham, *supra* note 20.

54. The amount of content lost due to never being preserved in the first place will vastly exceed other technical causes such as bit rot and format obsolescence. *See, e.g.*, Rosenthal, *supra* note 4.

55. WATERS & GARRETT, *supra* note 22, at 5.

56. Deegan & Tanner, *supra* note 1, at 5 (citing Don Waters).

57. ROSENZWEIG, *supra* note 11, at 8.

58. E-mail from Nicholas Taylor, Web Archiving Serv. Mgr., Stan. Univ., to author (Sept. 10, 2016) (on file with author).

> Manuscripts from the library of Alexandria in ancient Egypt disappeared in a fire. The early printed books decayed into unrecognizable shreds. [actually, rag content paper is relatively robust; cheap paper from the 1800s onward far less so]. Many of the oldest cinematic films were recycled for their silver content. Unfortunately, history may repeat itself in the evolution of the Internet . . . .[59]

More recent examples from the last years of the twentieth century are less well-known but equally compelling. NASA's original slow-scan television (SSTV) tapes from the Apollo 11 moon landing were recorded over and never found (although better images subsequently were discovered).[60] Moreover, some may not realize that World War II is a better documented event than Vietnam. The reason is that "Vietnam was a computer-era war, and many of its documents are stored on electronic tapes that can be accessed only by digital equipment that no longer exists. . . ."[61] Another example is when the US Census Bureau, in 1976, attempted unsuccessfully to access its historically-significant 1960 Census files. Those files were stored on UNIVAC tapes that had become obsolete by the mid-seventies; some records were never recovered.[62] This incident eerily echoes the fateful date of January 10, 1921, when a fire at the U.S. Department of Commerce building turned most of the 1890 Census records into "irretrievable ash."[63]

And at the more mundane level, digital "stuff goes away" because "the student graduates, the department closes, all the systems people get laid off in a downturn, the government cuts funding in the library, there's an earthquake, flood, riot, coup . . . . Information preservation requires building a structure that lasts longer than the mean time between site failure."[64] The incalculable losses of the past may well pale to those of the present and near future.

D. The Urgency Intensifies

---

59. Brewster Kahle, *Preserving the Internet*, SCIENTIFIC AM., March 1997, at 82, 82; *see* Adrienne LaFrance, *Raiders of the Lost Web*, ATLANTIC (Oct. 14, 2015) (quoting Abby Rumsey, "[m]ost of the films made in the United States between 1912 and 1929 have been lost" because "we didn't think they were valuable"), http://bit.do/Raiders.

60. *Neil Armstrong & "The Lost Apollo 11" Footage*, DAILY GALAXY (Aug. 28, 2012), http://bit.do/ArmstrongLostApollo11Footage.

61. Stephen Sottong, *Don't Power Up That E-Book Just Yet*, AM. LIBRS., May 1999, at 50, 52-53.

62. LAZINGER, *supra* note 38, at 9.

63. ELIZABETH R. LEGGETT, DIGITIZATION AND DIGITAL ARCHIVING 182 (2014). That 1921 fire led in part to creation of the National Archives. Lesk, *supra* note 24, at xvi.

64. Larry Masinter, *Stuff Goes Away*, Post to URC Discussion List (Mar. 9, 1996).

Digital preservation originally was perceived as an issue limited to library science. In the mid-1990s, libraries, archives, museums (known as "LAMs"), and other "memory organizations" began to embark on their own digital preservation programs. These efforts were seen as "both a leap in the dark and a leap of faith," involving much "learning by doing."[65] Back in 1994, Peter Graham could be considered a rare voice making the case that preserving data poses a much larger societal challenge. Two years later, an expert task force concluded that "failure to look for trusted means and methods of digital preservation will exact a stiff, long-term cultural penalty."[66]

Some two decades after that, the risk, and the potential penalty, have become far greater. Now, "software is the fabric which binds our personal, social, industrial, and digital lives."[67] And that software lives behind every strand of physical and virtual connectivity that links us together through the Internet. With the advent of wireline and wireless broadband access networks, new virtual technology platforms like search and social media, new mobile devices, and new forms of app-driven commerce, software and hardware now saturate every corner of our world.

In terms of scale, software-derived data permeates communications, entertainment, finance, health, energy, education, research, national security, transportation, and politics—not to mention our personal, familial, and social interactions. In fact, IBM estimates that "90 percent of the data in the world today has been created in the last two years alone."[68] IDC estimates that the amount of digital information in the world is doubling every eighteen months.[69] And, the amount of digital content created every year is more than all of the cloud-based data storage capacity in the world.[70] Some eighty to ninety percent of that stored material—and growing—is so-called "unstructured data," which runs the gamut from ordinary emails and other text documents to more rich data types such as photographs, music, and movies.[71]

---

65. Deegan & Tanner, *supra* note 1, at 29.

66. WATERS & GARRETT, *supra* note 22, at 4.

67. *Eclipse Foundation*, SOFTWARE HERITAGE (Jan. 18, 2015) (quoting Mike Milinkovich, Eclipse Found. Exec. Dir.), http://bit.do/EclipseFoundation.

68. CORRADO & MOULAISON, *supra* note 12, at 173; *see also* Coughlin, *supra* note 46 ("We are creating as much information annually as mankind generated from the beginning of civilization to a few years ago").

69. PAULINE SINCLAIR, PLANETS, THE DIGITAL DIVIDE: ASSESSING ORGANISATIONS' PREPARATIONS FOR DIGITAL PRESERVATION 4 (2010).

70. CORRADO & MOULAISON, *supra* note 12, at 173.

71. Structured data (SD), such as well-organized databases, is a small and declining

Together, these trends point to an increasing and accelerating dependency on digital resources to satisfy all ranges of information requirements. The broad scale and deep scope of this dependency is only heightened further by tremendous market and technology changes now sweeping the world.

1. The World Wide Web

Even beyond the massive generation of data, the Web has become the operating system of society. There are more than one billion websites in existence today.[72] And yet websites are "probably the most ephemeral category of data."[73] According to Brewster Kahle, the average lifespan of a Web page is under one hundred days.[74] Even when these pages exist, they merely reflect what is deemed important for the moment, what Adrienne LaFrance calls "a constantly changing patchwork of perpetual nowness."[75] After all, the Web was designed to be a messaging system, not a library, so that "[e]phemerality is built into the very architecture of the web . . . ."[76] And there is zero guarantee that those webpages will continue to exist. Indeed, some fifty percent of Web resources archived by the British Library had links that were unrecognizable or gone after just one year.[77] No less an authority than Tim Berners-Lee pointed out some years ago the inherently transitory nature of websites:

> Users should beware that there is no general guarantee that a URL which at one time points to a given object continues to do so, and does not even at some later time point to a different object due to the movement of objects

---

percentage of all stored data. *See, e.g.*, Robert Primmer, *Structured vs. Unstructured Data*, ROBERT PRIMMER (Oct. 6, 2016), http://bit.do/StructuredvsUnstructuredData.

72. LaFrance, *supra* note 59; *see* BURDICK ET AL., *supra* note 8, at 37 (Google reportedly has indexed over 1 *trillion* URLs).

73. Deegan & Tanner, *supra* note 1, at 15.

74. Ester Shein, *Preserving the Internet*, COMMUNICATIONS OF THE ACM, Jan. 2016, at 27 (citing Brewster Kahle). *But see* Deegan & Tanner, *supra* note 1, at 15 (citing UK Web Archiving Consortium estimating that the average website lives about as long as a housefly–44 days); *see also* ZIMING LIU, PAPER TO DIGITAL 15 (2008) (average life of a Web link is 45 days).

75. LaFrance, *supra* note 59; *see* Lepore, *supra* note 22 (labeling the Web something that "dwells in a never-ending present. It is–elementally–ethereal, ephemeral, unstable, and unreliable").

76. LaFrance, *supra* note 59; *see* ARLENE G. TAYLOR & DANIEL N. JOUDREY, THE ORGANIZATION OF INFORMATION 15 (3d ed. 2008) (even envisioning the Internet in library-like terms, acknowledges that "all the books have been dumped on the floor and there is no catalog.").

77. Andrew N. Jackson, *Ten years of the UK Web Archives: What have we saved?* ANJACKSON Slide 26 (Apr. 27, 2015), http://bit.do/10YearsofUKWebArchives (finding that a combination of link rot—broken URLs—and content drift—changes to original content—resulted in reduced access to Web resources).

on servers.[78]

Indeed, there are no existing copies of the very first Web page from 1989; it was subsequently overwritten by other files and lost forever.[79]

The Web itself is so massive that it is claimed impossible to crawl and digest everything; as Kahle puts it, "the Web was never designed for being archived."[80] Obviously content on the Web for the most part has not been selected and edited. So "we don't understand to what extent [Web-based digital content] constitutes an essential part of the cultural record that we seek to preserve."[81] Further, at a more micro level, a Web document is not inherently fixed, but comprises many dynamic links that shift and change over time. A Web page is not actually a single page. What indeed is the one "authentic" version? Or are they all?

2. Born Digital

Moreover, in terms of scope, the "vast majority of digital content is born, lives, and dies in only digital form."[82] This means there likely is no analog artifact counterpart that will live on in absence of the digital original. Thus, "countless born-digital works are also lost every day as they are removed, replaced, superseded, or left, forgotten, in obsolete formats and media."[83]

Digital objects also contain a degree of structural complexity not found in physical materials. Such objects have features with no equivalent in the analog world: they subsume multiple formats, can be interactive, mutable, broken apart, and recombined.[84] As a result, these digital objects look and feel little like their analog antecedents.[85]

---

78. Request For Comments: 1738 on Uniform Resource Locators 20 (T. Berners-Lee, L. Masinter & M. McCahill eds., Dec. 1994), http://bit.do/UniformResourceLocators.

79. LEGGETT, *supra* note 63, at 177.

80. Shein, *supra* note 74, at 27 (quoting Brewster Kahle).

81. Clifford Lynch & Nancy Y. McGovern, *Conclusions*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 309, 318 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

82. Deegan & Tanner, *supra* note 1, at 6.

83. STUDY ON COPYRIGHT LAW, *supra* note 14, at 1–2.

84. Brian Lavoie & Lorcan Dempsey, *Thirteen Ways of Looking at...Digital Preservation*, D-LIB, July/Aug. 2004, at 1, 3.

85. An additional complication is that disseminating digital works primarily via online transmission, rather than by the distribution of copies, means there are many fewer copies in existence to protect the works from obsolescence. R. Anthony Reese, *What Copyrights Owes the Future*, 50 HOUS. L. REV. 287, 311 (2012).

In fact, most digital objects are actually complex data structures that depend on a plethora of other resources, which in turn have different interpretations.[86] Researchers have identified and begun to study "complex digital objects," or "objects at the difficult end of the digital preservation spectrum."[87] Sorted into categories as diverse as computer games and virtual worlds, software art, simulations, and visualizations, they exhibit multiple layers of scaling and detail, and pose unique problems in copyright law.[88]

3. The Cloud

The "cloud" has become the preferred means by which the average Internet user stores and accesses content, whether his/her own or produced by others. More and more people are showing significant interest in the curation of their personal data.[89] Scientists in particular increasingly rely on cloud computing to manage and analyze data collaboratively, whether as infrastructure as a service (IaaS), platform as a service (PaaS), or software as a service (SaaS).[90]

4. Big Data

Along with the rise of cloud computing comes the rapid growth of so-called "big data" applications. What separates big data from other data sets is the three Vs: volume, velocity, and variety.[91] These factors make the preservation of big datasets especially challenging, as current tools break down when dealing with very large or complex digital objects, or very large numbers of objects or heterogeneous collections.[92]

5. The Internet of Things

---

86. DAVID GIARETTA, ADVANCED DIGITAL PRESERVATION 141 (2011).

87. Neil Grindley, *Preface* to PRESERVING COMPLEX DIGITAL OBJECTS xi, xii (Janet Delve & David Anderson eds., 2014).

88. *Id.* These "objects" include networked and virtual spaces, live data streams, linking to collaborative content, and 3D visualizations. *Id.*

89. ROSS HARVEY, DIGITAL CURATION 6 (2010). At least a few voices are sounding the alarm about the dangers of leaving our personal and social legacy in the cloud. *See, e.g.*, Jason Perlow, *Your social data is doomed, and don't count on Facebook to save you*, ZDNET (Jan. 26, 2016), http://bit.do/SocialDataDoomed.

90. Hugo Hiden et al., *Developing Cloud Applications using the e-Science Central Platform*, 371 PHIL. TRANS. ROYAL SOC. A 1-2 (2013).

91. CORRADO & MOULAISON, *supra* note 12, at 183.

92. Interestingly, "small data" has its own share of challenges, as it is less likely to be preserved to begin with, or to use standardized formats. *Id.* at 184-85.

In addition, the advent of the Internet of Things (IoT) means reliance on billions of devices and sensors in the physical environment. By one estimate, there will be some twenty-four billion such devices by 2019.[93] Many of these systems would be collecting, analyzing, and storing often critical data across entire sectors of the economy. It may be misguided to expect that sensors placed in the field will continue providing useful data for their expected lives of years, even decades. Many such devices are doomed to become useless "abandonware" after supporting cloud services are altered or discontinued.[94] Even those in continuing operation risk becoming insecure.[95] Of the huge amount of data to be generated, what should be preserved, and how?

6. Virtual Realities

Not just the material environment is changing before our eyes. With machine-learning capabilities driven by ever-smarter algorithms come new ways of educating and entertaining ourselves. In mixed reality computing—what could be considered part of a larger mix of "Deep Edge" technologies—virtual reality (VR), augmented reality (AR), and artificial intelligence (AI) will reshape the ways we interact with each other, and ourselves.[96] Novel categories of data, information, and knowledge, without precedent in the analog world, will be created in this new 360-degree "Metaverse," each with its own experience-based data and artifacts needing to be preserved.[97]

E. …and the Dependencies Spread

1. Science and Medical Research

The world also is facing a looming scientific and medical disaster. "A key underpinning of scientific discourse is the ability independently to verify or refute experimental results that are presented in support of a claim."[98] As just one example, Canadian researchers concluded that

93. John Greenough, *How the 'Internet of Things' will impact consumers, businesses, and governments in 2016 and beyond*, BUS. INSIDER (July 18, 2016), http://bit.do/HowIOTImpact.

94. Perlow, *supra* note 89. Likely only a long-term, open, and extensible interoperability spec can preclude this outcome.

95. David Rosenthal, *Following up on the Emulation Report*, DSHR'S BLOG (Mar. 29, 2016), http://bit.do/FollowupEmulationReport.

96. Richard Whitt, *PTC'16 Tuesday Keynote: Living on the Deep Edge: Promoting and Protecting the Human Values in Internet Architecture*, VIMEO (Jan. 29, 2016), http://bit.do/LivingOnDeepEdge.

97. *See* JEROME P. MCDONOUGH ET AL., PRESERVING VIRTUAL WORLDS FINAL REPORT (2010), http://bit.do/PreservingVirtualWorlds.

98. Paolo Missier et al., *Provenance and Data Differencing for Workflow Reproducibility*

some 80 percent of all scientific data referenced in research articles disappeared within two decades of publication, due to reliance on old email addresses and outmoded storage devices.[99] After all,

> In many scientific fields, research depends on access to persistent stores of digital information that are built and refined continuously. Consistent with the cumulative nature of scholarly research, journals that report research findings and that make references to previous studies constitute a continuous record of research and discovery.[100]

That ability is being lost, day by day, as science moves from *in vitro* (laboratory-based) to *in silico* (computer-based) research.[101] The core challenge is to make research data discoverable, usable, assessable, intelligible, and interoperable, and sustainable for a reasonable period of time.[102] Preserving "the scientific memory" in the digital era is hampered by an inability to guarantee data reusability.[103] As elastic computational facilities enable big data, the acceleration of the production of scientific results critically depends upon the large-scale availability of datasets themselves, their sharing, and their use in collaborative settings.[104] The problem is compounded not just by the need to preserve the dataset itself, but the ability it has to deliver knowledge to a future user community.[105]

Computational science too faces a "credibility crisis," from the dissemination of non-reproducible research. In part this is due to the lack of records of computer hardware and software configurations, and lost or revised source code.[106] Irreproducibility of the results of scientific and medical research is a significant problem; among the factors is reliance on software code.[107] Thus, due to unstable operational environments, and a failure to maintain related programs

---

Analysis, 28 CONCURRENCY & COMPUTATION: PRAC. & EXPERIENCE 995, 995 (2014).

99. Timothy H. Vines et al., *The Availability of Research Data Declines Rapidly with Article Age*, 24 CURRENT BIOLOGY 94, 94 (2014) ("In the long term, research data cannot be reliably preserved by individual researchers," which "further demonstrates the urgent need for policies mandating data sharing via public archives").

100. Margaret Hedstrom, *Digital Preservation: Problems and Prospects*, 20 DIGITAL LIBR. J. 3, 3 (2001).

101. HARVEY, *supra* note 89, at 4.

102. CHRISTINE L. BORGMAN, BIG DATA, LITTLE DATA, NO DATA 272 (2015).

103. Brian Matthews et al., *Towards the Preservation of the Scientific Memory*, 10 INT'L J. DIGITAL CURATION, no. 1, 2015, at 196, 197-98.

104. Missier et al., *supra* note 98, at 1–2.

105. Esther Conway et al., *Curating Scientific Research Data for the Long Term*, INT'L J. DIGITAL CURATION Sept. 2011, at 38, 38-40.

106. INST. COMPUTATIONAL & EXPERIMENTAL RES. IN MATH., SETTING THE DEFAULT TO REPRODUCIBLE 1 (V. Stodden et al. eds., 2013).

107. *Trouble at the lab*, ECONOMIST (Oct. 19, 2013), http://bit.do/TroubleAtLab.

with mutual dependencies, which become unable to function together, "scientific applications can suffer from forms of [workflow] decay that limit their longevity, and thus their reuse and evolution in time."[108]

2. Academic and Legal Publishing

Other dependencies have become apparent in just the last few years. The integrity of academic scholarship relies in some measure on the supporting references cited in documents. In the online context, there has been uncovered the growing prevalence of so-called "reference rot," a combination of link rot (broken URLs) and content drift (links leading to changed or vanished information).[109] Some see this as a "virtual epidemic."[110]

The link rot issue has been examined closely in the context of legal scholarship. According to a study by Jonathan Zittrain, for example, some 49.9% of the hyperlinks in Supreme Court decisions no longer contain the cited material.[111] As the authors note, given the distributed nature of the Internet, both link rot and reference rot are "inevitable," posing "serious problems for scholarship."[112] As a result, "the modern Supreme Court opinion is increasingly built on sand."[113] Distributed caching solutions have been proposed, and in the case of Perma.cc even implemented,[114] but the long-term scalability appears challenging.[115]

---

108. Missier et al., *supra* note 104.

109. Lepore, *supra* note 21.

110. NPR Staff, *Stopping Link Rot: Aiming to End A Virtual Epidemic*, NPR (April 26, 2014), http://bit.do/StoppingLinkRot.

111. Jonathan Zittrain et al., *Perma: Scoping and Addressing the Problem of Link and Reference Rot in Legal Citations*, 127 HARV. L. REV. F. 176, 176 (2014); *see also* Raizel Liebler & June Liebert, *Something Rotten in the State of Legal Citation: The Life Span of a United States Supreme Court Citation Containing an Internet Link (1996-2010)*, 15 YALE J.L. & TECH. 273, 273 (2012).

112. Zittrain et al., *supra* note 111, at 189.

113. Adam Liptak, *In Supreme Court Opinions, Web Links to Nowhere*, N.Y. TIMES (Sept. 23, 2013), http://bit.do/WebLinksToNowhere. Interestingly, Andy Jackson found that the legal and academic web materials show a noticeably higher link rot rate than actual results. Jackson, *supra* note 77.

114. *See Homepage*, PERMA.CC (Oct. 6, 2016), http://bit.do/PermaCC. The archival tool allows Web links to be preserved by fashioning a new, permanent link to content using a distributed system of caches. The Supreme Court's Office of Information Technology has also begun making available Web-based content that has been cited in Court opinions since the 2005 term. James A Jacobs, *Supreme Court Website Addresses link-rot and content-drift*, FREE GOV'T INFO. (Oct. 6, 2015), http://bit.do/LinkRotContentDrift.

115. Other proposed solutions include the use of digital object identifiers, and best practices, but these rely in large part on a user community that cares about creating permanent links to content. Leighton Walter Kille, *The growing problem of Internet "link rot" and best practices for*

The scholarly record itself is also at risk. The LOCKSS program at Stanford has made this a major area of focus.[116] In particular, while open access publishing can lower the entry barriers for smaller publications to publish high-quality research, these same works face a greater risk of not being preserved than larger, well-established proprietary journals.[117]

3. Financial Systems

The financial crisis of 2008 clearly was exacerbated by a lack of transparency and financial controls. What few realized at the time, and even today, is that the core management of financial data also is deeply flawed.[118] In the words of the U.S. Office of Financial Records, "data management in most financial firms is a mess."[119] That mess is only exacerbated by the growing need to preserve crucial financial records over long periods of time. Indeed, many institutions have difficulty retrieving and accessing data as little as three to five years from the point of creation.[120] Recent technology innovations such as blockchain may or may not improve the situation.[121]

4. Journalism

This pernicious dependency extends into the news reporting sphere as well. According to one reporter, "most media companies use

---

media and online publishers, JOURNALISTSRESOURCE (July 15, 2015), http://bit.do/GrowingProblemLinkRot.

116. *Lots of Copies Keep Stuff Safe*, LOCKSS (Oct. 6, 2016), http://bit.do/LotsOfCopiesKeepStuffSafe.

117. Rosenthal, *supra* note 4.

118. Victoria Lemieux, *Records and Information Management for Financial Analysis and Risk Management*, *in* FINANCIAL ANALYSIS AND RISK MANAGEMENT: DATA GOVERNANCE, ANALYTICS AND LIFE CYCLE MANAGEMENT 1, 11 (Victoria Lemieux ed., 2013).

119. Kafui Monu et al., *Using Conceptual Models to Theorize about the Relationship Between Records and Risk in the Global Financial Crisis*, *in* FINANCIAL ANALYSIS AND RISK MANAGEMENT, 73, 78.

120. Lemieux, *supra* note 118, at 9. Little has been done to address these risk factors, particularly for "unstructured" data. *Id.* at 10.

121. Blockchain platforms are decentralized networks of computers using a common, non-reproducible database. These platforms underlie bitcoins and other non-traditional currencies and ledger-based transactional systems. Blockchain has the potential to become the OS of decentralized computing, complete with cloud storage, marketplaces and social networks. Primivera DeFilippi, *Blockchain tech: a new (r)evolution in the digital economy*, UOC Open Thoughts, (Jan. 27, 2016). For digital preservation purposes, it is an open question whether blockchain could be a useful tool, another symptomatic problem—or both. For a less optimistic viewpoint, see David Rosenthal, *A Solution to Everything?*, DSHR's Blog (Jan. 20, 2015), http://bit.do/SolutionToEverything (explaining how blockchains do not guarantee anonymity, remain decentralized, or utilize adequate mining power).

a preservation strategy that resembles Swiss cheese."[122] As she observes, the print and Web versions of content management systems can differ significantly, with the latter often far less systematic and comprehensive.[123] Amazingly, a recent survey revealed that "not one publication has a complete archive of its website."[124] Many newspapers also fail to preserve the apps that drive many interactive, multimedia projects.[125] Some have begun to tackle this challenge.[126] Among other drawbacks, this hodgepodge system must affect that hazy place where investigative journalism turns into definitive history.

5. Historical and Government Records

Official history, too, is at stake. The concept of preservation as assuring the memory of civilizations has evolved over thousands of years. This includes decisions about what to collect, and not.[127] Digital technology offers a means of gaining "a newfound vantage on the totality of passing time—the profound implications of which we are just now beginning to grasp."[128] For historians, this amounts to a questioning of the basic goals and methods of their craft.[129] As just one example, should historians in the digital world be held to the same standard of "reproducibility" results as scientists?[130]

On the other hand, in a world where historical documents and artifacts are not routinely preserved, gaps are inevitable. Historians may be witnessing a fundamental paradigm shift, from a culture of scarcity to one of abundance, even as the loss of data mounts.[131] The National Archives recently concluded that most U.S. federal agencies do a poor job of managing their digital records.[132] The swirl over the

122. Meredith Broussard, *The Irony of Writing Online About Digital Preservation*, ATLANTIC (Nov. 20, 2015), http://bit.do/IronyOfWritingOnline.

123. *Id.*

124. *Id.* (citing Kathleen Hansen & Nora Paul, *Newspaper Archives Reveal Major Gaps in Digital Age*, 36 NEWSPAPER RES. J. 290-298 (2015)).

125. Meredith Broussard, *Preserving News Apps Present Huge Challenges*, 36 NEWSPAPER RES. J. 299, 299 (2015).

126. *See Dodging the Memory Hole 2016: Saving Online News*, REYNOLDS JOURNALISM INST. (Nov. 11, 2016), http://bit.do/DodgingMemoryHole (group of conferences focused on preserving online news).

127. *See* Michèle Valerie Cloonan, *Preface* to PRESERVING OUR HERITAGE xv, xvi (Michèle Valerie Cloonan ed. 2015).

128. Jenna Wortham, *How An Archive of the Internet Could Change History*, N.Y. TIMES (June 21, 2016), http://bit.do/ArchiveInternetCloud.

129. ROSENZWEIG, *supra* note 11, at 6.

130. *Id.* at 25.

131. *Id.* at 7.

132. Thomas Luan Dang & Victoria L. Lemieux, *A Functional Framework for Evaluating*

handling of Hillary Clinton's official government emails is a recent example.[133]

This issue is even more acute where those in power seek to erase unpleasant memories of the past. Officials can, and sometimes do, alter online records when that information turns out to be factually incorrect—or embarrassingly accurate.[134] Obviously "the official version of events shouldn't always be trusted or accepted without question."[135] The Soviet Union rewrote its own history regularly "to reflect the prevailing political mores, destroying valuable evidence along the way."[136] While such massive historical fraud may not be typical on the global scale, the ability and incentive of governments and other institutions to erase and replace digital accounts of their activities is substantial. Neither the United States,[137] nor Canada,[138] have been exceptions to such "libricide."

F. Defining Our Terms

For such a complex and sprawling body of work, it would be useful to utilize some common nomenclature and conceptual models.

1. Digital Preservation, Curation, Archiving, and Access

To Deegan, the term *digital preservation* means ensuring full access and continued usability of data and digital information.[139] A group of experts declares that the term "digital preservation" refers "broadly to the series of managed activities to ensure continued access

---

Financial Visualization Products, *in* FINANCIAL ANALYSIS AND RISK MANAGEMENT, *supra* note 118, at 115, 119; *see also* NAT'L ARCHIVES AND RECORDS ADMIN., *Records Management Self-assessment Report 2010* (Feb. 22, 2011). The agencies self-reported persistent problems with preserving electronic records and emails, with too few staff and resources. Alice Lipowicz, *Agencies Admit to Bad Records Management*, FCW (Mar. 3, 2011), http://bit.do/AgenciesAdmitBadRecords.

133. *See, e.g.,* Hilary Barlow, *The Clinton Emails and Proper Digital Records Storage*, ISSUES AND ADVOCACY (Aug. 19, 2016), http://bit.do/ClintonEmailProperDigitalRecords (stating that the incident is "a "discouraging portent for the future of digital recordkeeping in the highest echelons of American government").

134. Scott Althouse & Kalev Leetaru, *Airbrushing History, American Style*, CLINE CENTER FOR DEMOCRACY (Nov. 25, 2008), http://bit.do/AirbrushingHistory.

135. Wortham, *supra* note 128.

136. Deegan & Tanner, *supra* note 1, at 5; *see also* WATERS & GARRETT, *supra* note 22, at 1.

137. Althous & Leetaru, *supra* note 134.

138. *Canadian libricide: Harper government dumps centuries of irreplaceable environmental archives*, FREE GOV'T INFO. (Jan. 4, 2014), http://bit.do/CandianLibricide.

139. DEEGAN & TANNER, *supra* note 1, at 5.

to digital materials . . . ."[140] Others note that the term refers both to "born-digital" materials that have never existed in print or analog form, and digital surrogates of analog materials.[141]

Not surprisingly for such a young field, there is some confusion over terminology. Harvey talks about *digital curation* as a more inclusive and comprehensive "lifecycle" approach to preserving digital objects than digital preservation. He explains that curation considers what comes both before preservation (how the data are created and used) and after (how the data will be used, and by whom, in the future).[142] He asserts that, because curation both maintains and adds value to data for current and future uses, and focuses on ensuring the longevity, integrity, and accessibility of digital objects, it better fits the broader theme of digital stewardship.[143] While Harvey's explanation is useful and even compelling, and tracks how Europeans tend to use the phrase, most in the field in the United States continue to employ the digital preservation terminology to cover most of the activities he describes.[144]

Similarly, Susan Lazinger suggests that, per the 1996 Task Force Report, digital *archiving* is another way of defining long-term digital preservation. Archiving also includes the concept of assured access to content, separate from its preservation. Access is seen as "continued, ongoing usability of a digital resource, retaining all qualities of authenticity, accuracy, and functionality deemed to be essential . . . ."[145] So, as with curation, one could employ the archiving term to refer to the combination of preservation, plus access. To be clear, this paper will use the term "preservation" generically and expansively to include curation, user access, and other archiving-related activities.[146]

As we walk through the technical, legal, and financial challenges, it will be especially useful to separate out the notions of preservation

---

140. STUDY ON COPYRIGHT LAW, *supra* note 14, at 5.

141. HEDSTROM & MONTGOMERY, *supra* note 41, at 3.

142. HARVEY, *supra* note 89, at xvi, 55-56.

143. *Id.* at 2-3. Harvey claims that his more expansive definition of curation covers the entire information lifecycle, adds value, includes a wide range of stakeholders, allows risk management, and utilizes good data management practices. *Id.* at 7-8. *See also* CORRADO & MOULAISON, *supra* note 12, at 58.

144. BURDICK ET AL., *supra* note 8, at 34. From the broader cultural perspective of the new digital humanities, "to curate is to filter, organize, craft, and ultimately care for, a story composed out of—even rescued from—the infinite array of potential tales, relics, and voices." *Id*.

145. Digital Preservation Handbook, *supra* note 16, at 24.

146. In the absence of clear and consistent taxonomies, digital preservation and archiving are used interchangeably, but do not include the processes required to provide continued, enduring access to digital content through various delivery methods. *See* Rieger, *supra* note 174, at 3.

and of access. While the goals are usually intertwined in the analog context, where access to material depends on their being fit for physical use, the connection is more complex in the digital world.[147] Online discovery and retrieval can raise different types of questions than the initial act of preserving some content—for example, ensuring that preserved digital content remains accessible over time.

2. Types of Digital Objects

Most in the field employ the term "digital object" to describe the thing to be preserved.[148] Digital objects (DOs) can be either simple or complex. Simple objects are discrete digital items, such as textual files, images, or sound files, along with their associated metadata. Complex objects are made by combining a number of other digital objects, such as websites.[149]

There is general consensus that Thibodeau's conceptual model is useful for digital preservation.[150] According to Thibodeau, all DOs are entities with multiple inheritance. Every digital object is a physical object (inscription of signs on a physical medium), a logical object (recognized and processed by application software), and a conceptual object (recognized and understood by a person).[151]

Peter Graham notes there are three kinds of preservation problems: (1) artifact or medium (medium preservation); (2) software

147. *Id.* at 1. Kyrtsis also makes the distinction between preserving something, which requires some form of "push" technology, and retrieving something, which requires some form of "pull" technology. Alexandros-Andreas Kyrtsis, *Coping with Messiness and Fogginess in Financial Information Management*, *in* FINANCIAL ANALYSIS AND RISK MANAGEMENT, *supra* note 118, at 155-185. The construction of records involves "messiness," while the use of records involves "fogginess." *Id*. at 161-63.

148. Rosenthal prefers the term "digital artifact" to help emphasize that the "thing" to be preserved is "an aggregation of many digital components into a form designed by humans to convey information." David S.H. Rosenthal, *Emulation and Virtualization as Preservation Strategies*, MELLON FOUNDATION 3 (2015). While "digital artifact" is a more precise and telling phrase, in part because it is increasingly difficult to determine the boundaries of a digital object, *Id.* at 19-20, this paper will utilize both phrases interchangeably.

149. Harvey, *supra* note 89, at 45; Janet Delve & David Anderson, *Introduction* to PRESERVING COMPLEX DIGITAL OBJECTS, *supra* note 87, at xxxv, xxxix-xli. *See also* GIARETTA, *supra* note 86, at 31-40 (simple objects are treated as a whole; composite or complex objects are treated as a collection of simpler parts).

150. Margaret Hedstrom, *The Digital Preservation Research Agenda*, *in* THE STATE OF DIGITAL PRESERVATION: AN INTERNATIONAL PERSPECTIVE 32, 34 (2002).

151. Thibodeau, *supra* note 49, at 6-10 (Every digital object is a physical object, a logical object, and a conceptual object); *but see* Peter Graham, *Long-Term Intellectual Preservation*, *in* GOING DIGITAL: STRATEGIES FOR ACCESS, RESEARCH, AND CONVERSION OF COLLECTIONS TO A DIGITAL FORMAT 81, 81 (Donald L. DeWitt ed., 1998) (suggesting there are two types of information: "artifactual information" is the medium, while "electronic information" is the content).

and storage formats (technology preservation); and (3) information content (intellectual preservation).[152] In order to preserve a digital object, we must be able to identify and retrieve all its digital components—the logical and physical objects that are necessary to reconstitute the conceptual object. So, digital preservation is not a simple process of preserving physical objects, but preserving the ability to render the logical and physical objects.[153]

The ideal preservation system would be a neutral communications channel for transmitting information to the future, where the messages transmitted are not corrupted or changed in any way. However, says Thibodeau, "this cannot be the case for digital objects";[154] "[t]he preservation of an information object in digital form is complete only when the object is successfully output."[155]

G. Caretakers and Appraisers: Who Should Be Responsible for Preserving What?

1. The Who

Lazinger puts it plainly: Who will become the "caretakers of our digital heritage?" She identifies a number of potential stakeholders, but admits it is more difficult to determine which of these should be responsible for archiving our electronic heritage.[156] In the category of those stakeholders with an intent to archive electronic data, she includes individuals (such as emails); corporations (such as employee records); publishers (such as books and movies); and libraries, museums, and other electronic institutions. In terms of "potential responsible agencies," Lazinger lists national libraries, businesses, government agencies, and universities and university library consortia.

The LAMs—those institutions that traditionally have assumed responsibility for preserving information—continue to face technical, organizational, resource, and legal challenges in taking on the

152. Graham, *supra* note 20.

153. Thibodeau, *supra* note 49, at 12. Other classification schemes attributable to DOs include the rendered (documents, words, images) versus the non-rendered (databases, scientific measurements and observations, software itself), the static (unchanged bit sequences) versus the dynamic (files changing over time), and the active (does something) versus the passive (something with which things are done). The tools and technologies suitable for preserving one type of DO may not work well for others. *See generally* GIARETTA, *supra* note 86, at vii, 31-40.

154. Thibodeau, *supra* note 49, at 13

155. *Id.* at 13.

156. LAZINGER, *supra* note 38, at 47.

preservation of digital holdings.[157] Nonetheless, the accumulation of standards and good practices for digital preservation has been emerging and consolidating. Release of the now-seminal 1996 Report of the Task Force on Archiving of Digital Information, *Preserving Digital Information*, began that process, setting out agreed-to concepts, requirements, and challenges. In the intervening twenty years, an international community of digital preservation practice has begun to emerge, with a basic agreement on key principles and issues.[158]

Three community documents have formalized digital preservation practice: the OAIS Reference Model (2003), the Trusted Digital Repository Report (2002), and the PAIMAS Standard (2006). In addition, "InterPARES"—the International Research on Permanent Authentic Records in Electronic Systems—was launched in 1999 as a collaborative research initiative focused on long-term preservation of authentic digital materials. These documents and institutions represent community guidance that increasingly defines prevailing practice for digital preservation.[159]

2. The What

Separate from the question of who should be involved as a stakeholder is the question of which types of digital objects should be preserved. Before the Internet, distributing knowledge required the transportation of pre-selected physical objects.[160] Now, complex digital objects composed of active links and stored in the cloud predominate.

At the heart of appraising digital records is the determination of significance.[161] Typically, memory institutions holdings have long tail characteristics, with about twenty percent of the materials receiving about eighty percent of the use.[162] It is a strategic decision of some national and international importance to have selection and retention policies for digital data, just as it it for analog[163]—particularly as "the

157. HEDSTROM & MONTGOMERY, *supra* note 41, at v.

158. Nancy Y. McGovern, *Envisioning an International Community of Practice*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 5, 5-6 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

159. *Id.* at 8-9. McGovern also has identified six core aspects of alignment of national approaches to digital preservation: Organizational (Legal and Organizational), Technological (Standards and Technical), and Resources (Economic and Education). *Id.* at 14-16.

160. ABBY SMITH RUMSEY, WHEN WE ARE NO MORE: HOW DIGITAL MEMORY IS SHAPING OUR FUTURE 103 (2016).

161. HARVEY, *supra* note 89, at 132.

162. BORGMAN, *supra* note 102, at 272. The currently useful 20 percent changes continuously; the remaining usage is random and unpredictable. *Id.*

163. Deegan & Tanner, *supra* note 1, at 15-16.

rate of data production continues to outstrip the rate at which our resources for digital curation are made available."[164]

Individual research institutions and archives have differing collection and selection criteria. One example is the Internet Archive, which policy Susan Feldman has called "Preserve everything and then we'll decide."[165] Brewster Kahle himself dismisses this perspective on the Archive's work, saying that although we should "cast a very wide net," not all the Web should be archived for all time.[166] A variation on this "preserve everything" approach is to retain content now in its current form, and subsequently figure out the actual preservation technique—"save all, and preserve later."[167]

Most preservationists argue for some degree of selectivity, in keeping with the founding principles of humanistic scholarship.[168] "While it is of course necessary to propose strategies for dealing with all categories of digital data, it is not feasible to propose that all digital data should be preserved for the long term."[169] Lesk points out that "[o]ne danger in the digital world is that we will believe we can save everything, and not recognize the costs of cataloging materials so that somebody can find them again. We need to use the same kinds of principles that have been used in the past . . . ."[170] There is also considerable value derived from "the crucial tasks of forgetting, of strategically looking away, of ignoring, of letting go and even of erasure."[171]

Recently, Niu has proposed an appraisal/section framework that incorporates three sets of variables: statistical sampling (systematic and random), risk analysis (based on probabilities and consequences), and appraisal (assessing the value of materials).[172] Kastellac similarly has proposed three general models of digital object selection, on a sliding scale from most- to least-direct human involvement: the selective

---

164. HARVEY, *supra* note 89, at 133.

165. LAZINGER, *supra* note 38, at 36.

166. Shein, *supra* note 74, at 28 (quoting Kahle).

167. Lavoie & Dempsey, *supra* note 84, at 5.

168. BURDICK ET AL., *supra* note 8, at 115. "The task of cultural memory is not exhaustive, but selective, that the shape of who we are is determined as much by what does not remain as what does." *Id*.

169. Deegan & Tanner, *supra* note 1, at 15; *see also* HARVEY, *supra* note 89, at 136-37 ("Why We Can't Keep Everything").

170. Michael Lesk, *Preserving Digital Objects* § 4 (1995), http://bit.do/PreservingDigitalObjects.

171. BURDICK ET AL., *supra* note 8, at 111.

172. Jinfang Niu, *Appraisal and Selection for Digital Curation*, 9 INT'L J. DIGITAL CURATION 65, 68-69 (2014).

model (narrow sets of digital objects are selected), the thematic and collaborative models (sets of objects are selected in relatively narrow domains), and the whole domain model (everything that can be harvested is preserved).[173]

Even if we can classify selection scales, however, achieving a consensus among all stakeholders will be difficult. Whose values are taken into account and how can they be known? How do we arrive at the best criteria, and apply them correctly? Which stakeholders are involved? And is consensus even a virtue in this case? A diversity of policies can mean a wider array of bets on the value of future uses. Perhaps the better alternative to consensus may be first to decide policy, based on asking ourselves a series of pertinent questions,[174] and then articulate clearly the expected outcome in a transparent manner.[175] As Holdsworth remarks, the ultimate objective should be to make the preservation costs "so cheap that there is little reluctance to keep things that have only a small probability of being accessed in the future."[176] This suggests focusing on lowering barriers to entry and cost rather than driving to consensus.

H. Philosophical Angles and Issues

The issues surrounding digital preservation also raise profound philosophical questions. While many are beyond the scope of this paper, it is obvious that our society must find a way to wrestle with the role of digital preservation as part of our very cultural fabric.

1. Meaning and Abstraction

For starters, what is meaning? What is information? What is knowledge? As we discuss and analyze the various elements of digital preservation and access, where do they fall within the so-called "data-information-knowledge-wisdom" (DIKW) hierarchy?[177] After all, it takes a mental process to turn meaningless data into meaningful information, and from there, to knowledge and even wisdom. [178]

---

173. CORRADO & MOULAISON, *supra* note 12, at 172.

174. Oya Y. Rieger, *Preservation in the Age of Large-Scale Digitization*, COUNCIL ON LIBR. AND INFO. RESOURCES 11-15 (2008).

175. Lavoie & Dempsey, *supra* note 84, at 4.

176. David Holdsworth, *Strategies for Digital Preservation*, *in* DIGITAL PRESERVATION 32, 33 (Marilyn Deegan & Simon Tanner eds., 2006).

177. CORRADO & MOULAISON, *supra* note 12, at 111.

178. *See, e.g.,* Hans Christian von Bayer, INFORMATION: THE NEW LANGUAGE OF SCIENCE, 28-34 (2004) (information is an invented, ill-defined, and intangible concept). Despite the

Human learning is based upon the ability to "analyze, organize, and retrieve data, information, and knowledge; to recognize patterns; to compare experiences, concepts, and ideas; and to process the relationships among all of them."[179] Or, "cognition is categorization."[180] Retrieval of information is dependent upon its having been organized, and such organization also allows us to save for posterity copies of all kinds of works that result from human endeavors—"our collective knowledge reserve." [181]

So, are preservation efforts aimed at the data level, the information level, or the knowledge level? As one example, digital data can be highly complex, "and meaning derived from data can depend as much on how individual data objects are linked as on what those objects are."[182] Marshall McLuhan famously proclaimed that "the medium is the message." In the digital era, is the more accurate formulation "the message is the medium"?

What are the essential attributes of a digital object, such as a Web page? The intellectual substance contained in information objects—the content—is itself a complex idea that operates at several different layers of abstraction.[183] At the lowest level (the bit configuration), content is just ones and zeroes; at higher levels, the content reads as format and structure; and at the highest level, the content constitutes specific ideas and knowledge.[184]

Digital content is inherently mutable and therefore abstract.[185] Holdsworth believes the key is to take a view of digital data as an abstract quantity, divorced from the storage medium but associated with technical metadata that permits ready access to its intellectual content.[186] Seen this way, the concept of information—and its

---

centrality of the concept of information in our daily lives, "we are entirely lacking a clear physical account that explains how information about some abstract concept can have massive and sometimes devastating physical consequences." *See* Terrence W. Deacon, *What is Missing from Theories of Information*?, *in* INFORMATION AND THE NATURE OF REALITY 186, 189 (Paul Davies & Neils Henrik Gregersen eds., 2014).

179. TAYLOR & JOUDREY, *supra* note 76, at 2.

180. *Id.* at 1 (quoting Steve Hornad).

181. *Id.* at 2.

182. Deegan & Tanner, *supra* note 1, at 13.

183. WATERS & GARRETT, *supra* note 22, at 12.

184. *Id.* at 12-13. Another way to divide up the digital world is data (the set of facts we do not pay attention to) and "capta" (the subset of facts we pay attention to and use in order to generate information). PETER CHECKLAND & SUSAN HOLWELL, INFORMATION, SYSTEMS AND INFORMATION SYSTEMS 89 (1998).

185. Chapman, *supra* note 50, at 139.

186. Holdsworth, *supra* note 176, at 34-35. Or from another perspective, which Deacon calls the "absent content problem," "the property of something that warrants calling something

describable, recorded units as an "information resource"[187]—can survive the passage of time. So can the notion of representing everything as a sequence of bytes. We must bridge the longevity of the information concept to the certain mortality of the media on which the digital data live.[188]

2. Mediation

Just as our brains filter the external world for us, and culture provides the basic template by which we interpret the world,[189] technology acts as an intermediary between ourselves and the experiences we seek. The history of communications technology in part is a story of enhancing and expanding our limited natural senses in order to interpret the world. For what we have come to call "content," this process began with texts, readily readable from papyri to vellum to parchment to rag wood. The world of images required the camera lens, and the photographic plate. Then came sound, and video, each mediated by mechanical devices.

Now, in the age of digital content and the Web, the bits themselves that comprise the experience require software and hardware to decipher (often also with electricity to power, and connectivity to realize). In fact an emerging field of "information architects" seeks to play this "Internet librarian" mediation role on behalf of society; its practitioners determine the uses to which information will be put, and create patterns for finding "information spaces."[190]

The nature of the "immaterial" digital also allows for other forms of mediation. As information becomes a more valuable commodity, monetary value has been placed onto the intangible substance of knowledge, thoughts, and ideas, and the means of production shifted to the new domain of cyberspace and bits. While this immaterial world now offers greater modes of connectivity, creativity, and communication, at the same time the rules of the marketplace have become blurred.[191]

As bits and atoms further separate, the concept of ownership is becoming hazier. Possessing a physical book, for example, is a

---

information … is that it is something that the sign or signal conveying it is not." Deacon, *supra* note 178, at 192.

187. TAYLOR & JOUDREY, *supra* note 76, at 4.

188. Holdsworth, *supra* note 176, at 35.

189. RUMSEY, *supra* note 160, at 25.

190. TAYLOR & JOUDREY, *supra* note 76, at 20. *See also* Guy Pessach, *The Role of Libraries in A2K*, 2007 MICH. ST. L. REV. 257, 261-62 (describing libraries as "knowledge intermediaries").

191. Joasia Krysa, *Introduction* to CURATING IMMATERIALITY 7, 13-17 (Joasia Krysa ed., 2016).

relatively straightforward proposition. Typically the possessor can do countless things with it, such as read it, scrawl in it, loan it, sell it, mutilate it, destroy it, or bury it in the backyard in a time capsule. Practically the only thing forbidden, under copyright laws, is to put one's name on the book and attempt to pass it off as one's own creation. Once the contents of that book have been digitized, however, the publisher has the technical ability, and often the lawful right, to treat its transfer as a lease, rather than a sale. In so doing, the publisher can scale back or even eliminate many of those attributes of ownership. So, "owning" an e-book carries fewer actual ownership privileges than possessing its physical equivalent.[192] Similar restrictions on our supposed ownership of physical goods—phones, cars, televisions, and the like—are being created by companies seeking to "interpose a software layer" between the user and the good.[193]

Certain digital preservation technical solutions, if adopted and applied, could insert an additional element of centralized control into the Internet, further distancing the individual from the ideas and things of his or her world. Should this likelihood be resisted? Per Rosenzweig,[194] is there a way that digital preservation itself—via technical solutions, legal frameworks, or institutional arrangements—can become an instrument of direct user engagement, rather than distancing?

What, then, does it all mean, when our ability to encounter the world is so fraught? More complexity, more richness, more variety—yes, all of that, and more. But perhaps at a price. All that giving also can entail a taking away. We can become more removed from the source than ever before, with less immediacy, less graspability—and less control.[195]

3. Loss of Fixity and Authenticity

---

192. *See generally* AARON PERZANOWSKI & JASON SCHULTZ, THE END OF OWNERSHIP (2016) (explaining how the digital marketplace has shifted conceptions of ownership and private property).

193. Jason Schultz, *The Internet of Things We Don't Own?,* COMMUNICATIONS OF THE ACM, May 2016, at 36, 36-38 (2016).

194. Rosenzweig took the view that the use of democratized access promises "direct and unmediated access to the past," with the universal library bringing a "cultural disintermediation" without "cultural brokers" standing between people and the documents of the past. ROZENZWEIG, *supra* note 12, at 22.

195. Kyrtsis asserts that accessing digital records invites "fogginess" (a veil between the author and interpreter), while constructing such records involves "messiness." Kyrtsis, *supra* note **Error! Bookmark not defined.**, at 161-63.

According to Harold Innis, communications technologies favor one of two contrasting characteristics: "time binding" (related to the immutability of information) and "space binding" (related to the mobility of information).[196] Despite the social significance in framing information and giving pertinent context, John Seely Brown and Paul Duguid argue that the digital world embraces this latter space binding element, or "fluidity," and pulls against the time binding element, or "fixity," of things like documents. So even though "context shapes content,"[197] the Internet tends not to capture context particularly well.

As a result, the most threatened records in modern archives are usually not the oldest, but the newest. "The letters Paul read in Portugal may well be around in another 250 years. The files on which he recorded their text are unlikely to last twenty five."[198] Seen in this light, digital preservation is a deliberate countervailing force against the inherent fluidity of the Internet, rescuing it from the vagaries of time.

Another related issue is deciding what is the authentic information to be preserved. Rothenberg talks about the authenticity of "digital information entities,"[199] which show reliability over time.[200] However, "it is frighteningly easy to change a digital document, leaving no trace, no ghostly palimpsest to tell us what was there before."[201] Even the use of "digital signatures" tells us only that a document has been altered, not how. Graham's "taxonomy of changes"—accidental, well-meaning, and fraudulent—bears this out.[202] But who determines what is the authentic version of something?[203]

The concept of fixed elements of a digital object is fraught in the online context. "Classical" text documents are self-contained, while digital documents present more challenging line drawing between the document and its environment.[204] Rosenthal asks what it means to preserve an artifact that changes every time it is examined.[205] Masanes

196. JOHN SEELY BROWN & PAUL DUGUID, THE SOCIAL LIFE OF INFORMATION 200 (2000).

197. *Id.*, at 200-01.

198. *Id.* at 201.

199. Jeff Rothenberg, *Preserving Authentic Digital Information*, *in* AUTHENTICITY IN A DIGITAL ENVIRONMENT 51, 51 (2000).

200. George MacKenzie, *Searching for Solutions*, Managing Information, July/August 2000, at 7, 59-65.

201. Deegan & Tanner, *supra* note 1, at 8.

202. Graham, *supra* note 20.

203. Authenticity is particularly critical for government, scientific, and cultural domains. GIARETTA, *supra* note 86, at 203. In essence, we need to be able to trace back a digital object to a trusted person.

204. BORGHOFF ET AL, *supra* note 25, at 129.

205. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 20.

argues that the Internet should not be considered only as a content repository, but also as an information space system, with its specific structure, rules, and organization.[206] Among other things, this means that Web archiving consists of constructing a "local" and preserved version of certain parts or segments of this "virtual information space." This space exists only as a result of the interaction of several complex and active Web Information Systems (WIS). So, an archived segment must be collected through interactions with Web servers, and organized in a way that entails navigation.[207]

Moss points out one of many ironies of our current challenge. "It is the context of the technology that has led us to this state of affairs, but the context of the content, which was inherent in analogue practice, has vanished."[208]

4. Living for the Future

A final set of questions presents itself: what exactly are we trying to preserve, and why? What values do we employ to figure it out? How do we decide what to keep, and for how long?[209] There is an inherent paradox in digital preservation: aiming to deliver the past to the future in an unaltered, authentic state inevitably requires some alteration. This paradox can only be resolved by elaborating a basic conceptual framework for digital preservation. The OAIS reference model discussed in Part II has served this role, for example, but needs to be refined and extended to be useful for actual implementations.[210]

How long we should keep our digital stuff is also a value judgment. The Joint Information Services Committee (JISC) quantifies three different lengths of digital preservation: long-term (indefinite access into the future), medium-term (access extends beyond changes in technology), and short-term (access does not extend beyond changes in technology).[211] Each time period brings with it different preservation and access requirements.

---

206. Julien Masanes, *Web Archiving*, *in* DIGITAL PRESERVATION 78, 79 (Marilyn Deegan & Simon Tanner eds., 2006).

207. *Id.* at 82. As navigation paths in the Web are embedded and actionable in the documents themselves, the archive must be built in a way that will enable this mechanism to work.

208. Moss, *supra* note 31, at 7.

209. As a rule of thumb, archivists would claim only to keep some five percent of content in the analog world, selecting only documents deemed relevant. *Id.* at 8. That number may rise to roughly twenty percent in the digital environment. *Id.* at 8-9.

210. Thibodeau, *supra* note 49, at 28-29.

211. CORRADO & MOULAISON, *supra* note 12, at 6.

Finally, data preservation is about enabling decisions in the future. Per Hedstrom, "[w]e may know when we have failed, but we may not be alive to know whether we have succeeded."[212] At the more mundane level, "this problem requires some challenging thinking about success measures and evaluation criteria."[213] It also requires trust. But more deeply, this point raises some profound questions about our motivations in preserving our digital heritage, and the lengths we are willing to go—or not—to make it happen. We are making decisions now on behalf of generations not yet born, lacking a current day voice or vote of their own.

Inevitably we will touch on some of these questions at various junctures below. But the larger point is that we need the professional and armchair philosophers on board as well, to help guide us into productive ways of thinking about preserving our digital heritage.

II. THE TECHNICAL CHALLENGES

> Wisdom enough to teach us of our ill
> Is daily spun, but there exists no loom
> To weave it into fabric . . . .[214]

In a very real sense, digital documents exist only by virtue of software that understands how to access and display them. "In practice we need to remember that every bit stream relies on the appropriate software to give it significance."[215]

In digital preservation, the objective is to preserve access to the digital content, rather than the physical object or medium. According to Lazinger, there are three types of technological obsolescence: hardware issues (no compatible device); software issues (no backward compatibility); and media deterioration (deterioration of physical storage medium). And he believes that "the biggest enemy of long-term data storage is obsolescence."[216] As software and hardware technology continues to evolve, obsolescence is an inevitable outcome. It goes to not the presence of the disk, but the disk reader.

---

212. Hedstrom, *supra* note **Error! Bookmark not defined.**, at 35.

213. *Id.*; *see also* CORRADO & MOULAISON, *supra* note 12, at 95 (observing that successfully preserving digital information cannot be proven until years into the future).

214. Edna Millay, *Huntsman, What Quarry? in* COLLECTED SONNETS (1959).

215. FILIP BOUDREZ, *The Digital Recordkeeping System*, DIGITAL ARCHIVING IN FLEMISH INST. AND ADMIN. (DAVID) 12 (2001).

216. LAZINGER, *supra* note 38, at 76. Of course, the initial challenge is preserving the bitstreams in the first place, which is more an economic matter than a technical one.

Ironically, in the analog world, records survive for a very long time, quite often in less than ideal conditions, "provided they do not get wet or eaten by rodents."[217] By contrast, bit patterns are notoriously logically unstable; every time they are opened, their logical structure changes, along with some of the ambient metadata.[218] The paradox of digital data is that they are fundamentally simple, being made up of only two electrical states. But, when using computer programming techniques, those states can be configured into patterns so complex "that a limitless number of different documents and other artifacts can be represented."[219]

While there is still no universally accepted technology for preserving our digital heritage, below we will examine the primary technical mechanisms and standards that have been put forward to date to preserve digital objects.

A. First Step: Defining What Should Be Preserved

The initial job in preserving any material is archival scope: identifying which material merits the effort of preservation. The question comes down not only to what types of digital objects should be preserved, but how much of each digital object—including contextual information—should be preserved.

The 1996 Task Force on Archiving of Digital Information enunciates that the central goal of preservation must be to preserve "information integrity," regardless of whether the digital object is textual, numeric, image, video, sound, multimedia, or simulation.[220] This means defining and preserving those features of an "information object" (an inapt term?) that distinguish it as a whole and singular work. The Task Force defines five features that constitute the integrity of a digital document:

- **content**: intellectual substance contained in information objects. This includes, for example, character set, layout, and structure issues in print text; at the highest level of abstraction, content is defined in terms of the knowledge or ideas the object contains. These ideas "transcend the limits of the hardware and software systems needed for reading and interpreting the bits of an information object."

217. Moss, *supra* note 31, at 13.
218. See footnotes 43-46, *infra*.
219. Deegan & Tanner, *supra* note 1, at 16.
220. WATERS & GARRETT, *supra* note 22, at 12.

- **fixity**: content fixed in a discrete object. Digital text is "infinitely variable," and can include concurrent multiple versions and continuously updated databases.
- **reference**: reliable systems for locating and citing. This includes a number of identifying systems, such as Uniform Resource Name (URN), Uniform Resource Locator (URL), and Digital Object Identifier (DOI).
- **provenance**: record of the document's origin and chain of custody. The presumed authenticity can be contained in provenance metadata.
- **context**: document's interaction with elements in the wider digital environment. This includes the technical context of hardware and software; "digital objects are sometimes highly dependent on a specific technology configuration, such as a particular word processing program that runs on a particular computer with a particular operating system."[221] Other contextual dimensions include linkages between digital objects, mode of communication or distribution, and the wider social environment.

The technical context of a digital object can be further broken down, into type of material (e.g., spreadsheet); type of file formats (e.g., compressed graphics format); type of media (e.g., portable CD optical media); and type of platform/operating system (e.g., Windows NT). [222] All these attributes together create the digital object as it is experienced by the end user.

B. The Role of Metadata

Metadata is "data about data," or more precisely, standardized information about data sets.[223] Metadata can aid in the identification, description, and location of networked electronic resources. One primary function of metadata is resource discovery; another is control of the electronic resource.[224]

Metadata can be generated automatically or manually, located outside or inside the electronic items, and either established at the time

---

221. LAZINGER, *supra* note 38, at 25.

222. *Id.* at 30-31; *see also* John Bennett, *JISC/NPO Studies on the Preservation of Electronic Materials*, BRITISH LIBR. RES. INNOVATION REP. 13-20 (1997) 30-31.

223. TAYLOR & JOUDREY, *supra* note 76, at 89.

224. LAZINGER, *supra* note 38, at 139-40 (statements from Hudgins, Agnew, and Brown). Or as Higgins puts it, metadata improves both accessibility and discoverability. Sarah Higgins, *Digital Curation: The Emergence of a New Discipline*, INT'L J. DIGITAL CURATION, Sept. 2011, at 78, 79.

of the creation of a digital object, or added later.[225] Categories of metadata include administrative (managing the resource), descriptive (identifying the resource), preservation (preserving the resource), technical (systemic information), and use (using the resource).[226] Levels of metadata are simple (data extracted from the source), structured (formal element sets created for the user), and rich (comprehensive, detailed descriptions for information professionals).[227]

Metadata has an important role to play in all the current approaches to electronic data archiving, including refreshing, migrating, and emulation.[228] Higgins asserts that metadata "is the backbone of digital curation. Without it a digital resource may be irretrievable, unidentifiable or unusable."[229] And of course metadata itself must be preserved, along with the additional digital artifacts that it enables.[230]

Perhaps the best known and utilized metadata standard is PREMIS (PREservation Metadata: Implementation Strategies), founded by an international working group seeking to "define implementable, core preservation metadata, with guidelines/recommendations."[231] Other metadata standards applied to digital preservation, all of which relate to XML in some fashion, include:

- **MARC** (Machine-Readable Cataloging): a common metadata scheme, created in the 1960s to establish bibliographic records stored in library catalogs, and which to some has become somewhat outdated and unintuitive.[232]
- **RDF** (Resource Description Framework): metadata specification developed in 1999 by the World Wide Web Consortium (W3C) to promote structural interoperability.

---

225. CORRADO & MOULAISON, *supra* note 12, at 112.

226. LAZINGER, *supra* note 38, at 143-44. Others identify the categories as administrative, structural, and descriptive. TAYLOR & JOUDREY, *supra* note 76, at 91.

227. TAYLOR & JOUDREY, *supra* note 76, at 91. Metadata operates at various levels of granularity (from single documents, to individual webpages, to entire websites). *Id.* at 93. Its ability to meet the user's needs depends on its interoperability, flexibility, and adaptability. *Id.* at 95-96.

228. LAZINGER, *supra* note 38, at 168.

229. Sarah Higgins, *What are Metadata Standards?*, DIGITAL CURATION CENTER (Feb. 2007).

230. GIARETTA, *supra* note 86, at 4. Preservation metadata is the information necessary to ensure the survival of digital objects over long periods. Wendler has identified five key functions of preservation metadata: viability, renderability, understandability, authenticity, and identification. Wendler, *supra* note 36, at 61.

231. Wendler, *supra* note 36, at 61.

232. TAYLOR & JOUDREY, *supra* note 76, at 129.

- **SGML** (Standard Generalized Markup Language): describes how to construct and use markup languages, including metadata.
- **METS** (Metadata Encoding and Transmission Standard): metadata schema for complex digital objects, based on XML.
- **MODS** (Metadata Object Description Schema): developed by the U.S. Library of Congress for encoding information resources.
- **MADS** (Metadata Authority Description Schema): used to provide authority control for names of people, organizations, events, and terms.
- **Dublin Core**: a metadata element set that began at a workshop held in Dublin, Ohio (ISO 15836:2009) as a means of cataloging webpages using just fifteen categories of information.[233]

C. Proposed Techniques for Preservation

The best technical solution is "to fix . . . a document so that a user can be sure of the unaltered text when it is needed."[234] Actual digital preservation techniques include a range of options. Harvey identifies three main families of digital preservation techniques: technology preservation, technology emulation, and information migration.[235] Other options, such as digital archaeology, are more limited or reactive and hence less interesting, but still worth mentioning.

1. Technology Preservation

Technology preservation is the maintenance of the hardware and software platforms which support a digital resource. Because this method would require a regular cycle of media refreshing, involving a large number of computers and programs over a long period of time, it is relatively impractical and financially unfeasible.[236]

2. Refreshing

---

233. LEGGETT, *supra* note 63, at 150-54.

234. Graham, *supra* note 20.

235. HARVEY, *supra* note 89, at 162. Others employ different buckets. Lazinger states for example that there are three main categories of potential solutions to combat obsolescence, perhaps in combination: refreshing, migrating, and emulating. LAZINGER, *supra* note 38, at 76-88. Deegan mentions a range of options that includes technology preservation, refreshing, migration and reformatting, emulation, data archaeology, and output to analog media. Deegan & Tanner, *supra* note 1, at 17.

236. Deegan & Tanner, *supra* note 1, at 17.

Refreshing is copying the bitstream from one storage media to another to prevent media obsolescence, with no change to any of the underlying data. The refreshing process would need to be carried out, whatever other preservation strategies are adopted.[237] This option is viable primarily for digital files in a non-proprietary format, and formats not protected by digital rights management (DRM). While it fails to solve the technological obsolescence problem, refreshing could serve as a stopgap measure while more viable technologies are being developed.

3. Migration

Migration also involves copying the bitstream from one storage media to another. Unlike refreshing, however, migration changes the configuration of the underlying data, even as their intellectual content remains unaltered. The entire digital environment—the hardware/software configuration—is transferred, not just the physical storage medium. Migration will generally involve some reformatting, which can be costly and labor-intensive, and more complicated with complexly-linked artifacts such as websites.

Migration strategies can include transferring digital information from less stable to more stable media; from more complex (highly software-dependent) formats, to simplest possible (less software-intensive) formats, and from a multiplicity of formats, to a smaller number of common formats. Some forms involve bit sequence changes (Repacking and Transformation), while others do not (Refreshment and Replication).[238] Other migration approaches are to develop different standards (such as metadata standards), backward compatibility paths, and process centers.[239] **Encapsulation**—making details of how to interpret a digital object part of its encapsulated information—is considered another type of migration.[240] Impact on integrity and cost are the two key factors in determining which migration strategy to employ.

4. Emulation

---

237. *Id.* at 18.

238. GIARETTA, *supra* note 86, 198.

239. Thibodeau classifies different migration approaches: simple version, format standardization, Typed Object Model (TOM) conversion, Rosetta Stones translation, and object interchange format. Thibodeau, *supra* note 49, at 23-26.

240. Kyong-Ho Lee et al., *The State of the Art and Practice in Digital Preservation*, 107 J. RES. NAT'L STAND. TECH. 93, 98, (2001). The Universal Preservation Format (UPF) is based on the encapsulation approach, while the Digital Rosetta Stone stores the representation information separate from the encapsulation. *Id.*

While migration focuses on the digital object itself, **emulation** is the process of recreating the hardware and software environment required to access a resource. The emulation essentially mimics (performs the functions of) the obsolete hardware and other software. Emulators are pieces of software and/or hardware that transparently run applications and operating systems on non-native platforms. Digital documents are stored in their original forms, along with the original software in which they were created. Additional software and/or hardware is created to permit a more advanced computer at some future time to mimic, or virtualize, the obsolete hardware.[241]

Emulation was described in depth in 1998 by Jeff Rothenberg at RAND Corporation. Rothenberg's concept is to enable the emulation of obsolete hardware systems on future hardware of an unknown nature, so that a document's original software can be run despite being obsolete. His three-pronged approach is to (1) develop generalizable techniques for specifying emulators to run on future computers, (2) develop techniques for saving in human-readable form the metadata necessary to find, access, and recreate digital objects, and (3) develop techniques for encapsulating documents, their metadata, software, and emulator specifications. According to Rothenberg, the information that needs to be encapsulated is composed of the document and its software environment, the specification of an emulator, and metadata and other explanatory material and documentation. [242]

Developments over the last 20 years, including the evolution of digital formats and interconnectedness of digital artifacts, have significantly challenged some of the assumptions underlying Rothenberg's original model.[243] State of the art emulation frameworks, such as bwFLA (University of Freiberg), Olive (Carnegie Mellon University), and the Internet Archive's "Emularity" program, have revealed shared concerns about a lack of adequate technical support for emulation, the need for both technical and bibliographic metadata, and questionable fidelity to the original artifact.[244]

a. Emulation Versus Migration?

241. For a more thorough technical treatment, see GIARETTA, *supra* note 86, at 123-138.

242. Emulation is founded on the principle that all computers are Turing machines, and that any command that can run on one Turing machine can run on any other. Thibodeau, *supra* note 49, at 20.

243. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 19-23.

244. *Id.* at 5-9, 13-16.

Back in 2001, Lazinger could report that "opinions vary on the efficacy of emulation as a practical technique for electronic data archiving."[245] Emulation decidedly has come of age as a suitable digital preservation strategy to tackle complex digital objects.[246] Emulation protects the authenticity of the document, and invokes relatively small maintenance efforts.[247] Conversely, each step in migration requires considerable processing effort, with deterioration of document authenticity almost unavoidable.[248]

While in theory emulation is superior to migration in many ways, as of 2015 migration remains strongly favored over emulation.[249] Barriers to adopting emulation include (1) the need to create Web-based emulators, (2) inadequate tools for creating preserved system images, (3) greater expense of emulation over migration, and (4) legal barriers for creating and providing access to collections of preserved system images.[250] David Bearman observes that saving proprietary software and hardware specifications and documentation raises potential intellectual property issues.[251] Indeed, various patents, licenses, and copyright restrictions may cover the systems being emulated.[252] Others point out that emulation tends to limit the ability to utilize more modern applications.[253]

Despite what many say, it is inadequate to sum up the available strategies as "emulate or migrate."[254] An increasing variety of methods useful for long-term digital preservation "do not fit nicely into the simple bifurcation of emulation versus migration."[255] Indeed, "neither emulation nor migration is a panacea for current digital artifacts."[256]

---

245. LAZINGER, *supra* note 38, at 31.

246. Delve & Anderson, *supra* note **Error! Bookmark not defined.**, at xliv.

247. BORGHOFF ET AL., *supra* note 25, at 131.

248. *Id.*

249. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 1.

250. *Id.*

251. LAZINGER, *supra* note 38, at 86-87.

252. CORRADO & MOULAISON, *supra* note 12, at 51.

253. GIARETTA, *supra* note 86, at 123-24.

254. *Id.* at 123-24.

255. Thibodeau, *supra* note 49, at 19. Both migration and emulation involve "active maintenance," with life-cycle management of digital information from the point of creation through storage, migration, and providing access. Active maintenance requires a shift in thinking from the preservation of traditional materials. First, it is important to preserve *access*, rather than the object or medium. Preservation of the integrity of the intellectual content is key, even as we discard the original storage medium, software, and hardware. Second, obsolescence of the *carrier* of electronic information must be solved, which can be accomplished by refreshing the medium or carrier. LAZINGER, *supra* note 38, at 78-79.

256. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 26. In 2001, migration and

This may explain why neither technique provides a sufficient, general solution. Rather than debate the merits in the abstract, we should recognize that "different kinds of information captured in different ways for long-term preservation will need various kinds of support."[257] Where "look and feel" are important, for example, with complex linkages and interactive computer programs, emulation is the better option.[258] On the other hand, Boudrez claims that "the ideal situation from an archival point of view would be to cut hardware and software dependence to a minimum and involve software as little as possible in the preservation process. The migration route, in which files are transferred to a standard format, comes closer to this ideal than emulation . . . ."[259]

Interestingly, as Rosenthal explains, one outcome of the advent of the Web is the massive reduction in the rate at which formats have become obsolete, due to the standardization of formats to support Web-published content.[260] This greatly-reduced rate of format obsolescence reduces the need for either migration or emulation as a way of interpreting individual documents. However, the fact that digital artifacts and their associated infrastructure continue to evolve beyond a state of individual static documents raises a new set of technical problems. For example, it is unclear that emulators can handle commonplace digital artifacts such as Google Maps, or important scientific computations such as climate models.[261]

5. Universal Virtual Computer

In 2000, a project in IBM Research proposed using a UVC (**universal virtual computer**), a general purpose computer that would specify a process to be executed on an unknown machine of the future. The program would be written for the UVC, so that in the future only an emulator of the UVC would be required to run the program and

---

emulation were the two most cited strategies for preserving digital content, but with no real consensus over the usefulness and advantages of each approach. BOUDREZ, *supra* note 215, at 7, 12-13.

257. Donald Waters, *Good Archives Make Good Scholars*, *in* THE STATE OF DIGITAL PRESERVATION: AN INTERNATIONAL PERSPECTIVE 78, 80 (2002).

258. CORRADO & MOULAISON, *supra* note 17, at 51.

259. BOUDREZ, *supra* note 215, at 12. Note, however, that semantics may not be preserved, in documents such as spreadsheets. And something as seemingly insignificant as missing or incorrectly substituted word fonts in legacy documents can mean a significant risk of information loss. Broom & Woods, *Born Broken: Fonts and Informational Loss in Legacy Digital Documents*, INT'L J. DIGITAL CURATION Mar. 2011, at 5.

260. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 19.

261. *Id.* at 23-26.

return all data in a logical view.[262] So the UVC would provide essential functionality for an unlimited variety of data types. The UVC's downside is that it can only provide a limited set of basic functions, and is not optimized for any specific software.[263]

While a UVC is similar to emulation in some respects, unlike emulation it specifies processes that have to run in the future. As a result, its supporters claim, only a single, reasonably simple UVC emulator needs to be available to decipher future programs, as compared to creating and maintaining emulators of various real hardware and software-based machines of the past.[264] In essence, emulators could be written in the language of a UVC, rather than on demand. A fully functioning UVC has been used by IBM, and resides at the National Library of the Netherlands.

6. Other Options

Data archaeology describes a number of techniques for rescuing a digital resource which has not been migrated.[265] As a digital preservation strategy, because no migration would be performed or programs preserved to be emulated at a later stage, data archaeology would leave the data structures and connections to be puzzled out in the future.

Output to analog media entails creating a high-quality surrogate as an analog version of the data file. One example is "computer output to microfilm" (COM).

Another technique that keeps old software running takes the opposite approach from emulation: it relies on a special kind of hardware, a **configurable chip**, rather than software emulators. Such chips seem like a simpler approach than emulation.[266]

The cornerstone of **persistent archives** is to articulate the essential characteristics of the objects to be preserved, and to preserve in a manner that is independent of any specific hardware or software. This articulation is expressed at the data level by tags that identify every byte sequence that must be controlled to ensure preservation.[267]

---

262. Raymond Lorie & Raymond J. van Diessen, *A Universal Virtual Computer (UVC) for Long-Term Preservation of Digital Objects*, IBM RES. REP. 1 (Feb. 4, 2005); *see also* J.R. van der Hoeven, R.J. van Diessen, & K. van der Meer, *Development of a Universal Virtual Computer (UVC) for Long-Term Preservation of Digital Objects*, 31 J. INFO. SCI. 196, 196 (2005); Chapman, *supra* note 50, at 134.

263. Thibodeau, *supra* note 49, at 22.

264. Lorie & van Diessen, *supra* note 262, at 2.

265. GIARETTA, *supra* note 86, at 171.

266. Thibodeau, *supra* note 49, at 20. The IBM 360 computer included such micro-coded capacity.

267. *Id.* at 26.

D. Proposed Techniques for Authentication

Graham describes a "taxonomy of changes" that can affect the fixity and authenticity of an electronic document, including accidental changes during copying, well-meaning changes during updating or restructuring, and fraudulent changes from changing or damaging one's own work. One way to combat these changes is to use an electronic technique to fix the document in some manner. The most common techniques for authentication of digital objects are encryption, hashing, and digital time-stamping.[268]

Encryption depends on "mathematical transformation of a document using an algorithm requiring a particular number, or key, as the basis of the computation."[269] The key is used to decode the resulting encrypted text. While useful for security and for authenticating the identity of the user, encryption as a means of authenticating the document can be problematic because either the keys are widely available enough to invite abuse, or they are so limited in availability that their loss would be devastating.

Hashing involves using an algorithm to assign arbitrary values to each portion of a document, yielding specific computational values, or "hashes." The resulting hash number is a series of characters unique to the document. Any subsequent changes to the document would yield a different hash, helping prove the existence of an original version, or detect alterations.[270]

Digital time-stamping authenticates both the document and its existence at a specific time, much like rubber stamping incoming mail with the date and time it was received. A hash is created and combined with a hash derived from the current time and date, resulting in a certificate. Cryptography is necessary to secure the hash from alterations.[271]

Digital signatures combine all three techniques to authenticate both the document and the creator. A hash is created and encrypted, usually with a public key, and the creator's identity certified through digital IDs issued by a third party. A digital signature currently is a common technique for ensuring the authenticity of documents,[272] although its chief drawback is its inability to specify how a particular document may have been altered.

---

268. Graham, *supra* note 20. *See also* LAZINGER, *supra* note 38, at 89-91.

269. LAZINGER, *supra* note 38, at 91.

270. *Id.* at 92-93.

271. *Id.* at 93-94

272. Holdsworth, *supra* note 176, at 48-49.

E. Standards for Digital Preservation

Standards "embody the outcomes of negotiations that are simultaneously technical, social, and political in character."[273] Digital preservation relies on interoperability between computer systems, and thus is dependent on standards. It is estimated that at least two hundred standards are related in some way to preservation and digital curation.[274] However, standards have not yet been developed for all aspects of digital curation.[275]

Jeff Rothenberg decidedly is in the minority in warning that reliance on standards, like proprietary formats, will become obsolete over time, and thus should play a minor role in a long-term digital preservation solution.[276] Other experts continue to point to standards as crucial to the digital preservation process.

1. Open Archival Information System (OAIS)

The first international standard to describe a digital archive system was the OAIS Reference Model. Considered the "ur-standard" for many others to follow, it is of prime importance in the digital storage field.[277] In 2003, OAIS became ISO Standard 14721; the newest version is known as Magenta Book 2. It is a generic, context-neutral standard that utilizes a "lifecycle" approach to lay down the principles and style of operation of digital preservation, without specifying the detail of data formats or hardware technology.[278]

The OAIS Reference Model comprises four basic components: producers, consumers, management, and the archive. The Model also provides four attributes: a uniform vocabulary, an information model, a recommended functional model, and a set of responsibilities for an archive. The Model describes how digital objects should be preserved for a certain group of users (the Designated Community), from the point where the objects are deposited into the system to the point where they are disseminated.

---

273. Paul N. Edwards, *"A Vast Machine": Standards as Social Technology*, 304 SCIENCE 827, 827 (2004).

274. Raivo Ruusalepp et al., *Standards Alignment*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 115, 117 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

275. HARVEY, *supra* note 89, at 99.

276. Ruusalepp et al., *supra* note 274, at 118.

277. *Id.*; Holdsworth, *supra* note 176, at 36.

278. Holdsworth, *supra* note 176, at 37. Some point out that many in the digital preservation community continue to conflate this reference function with a system design. Michael Seadle et al., *Technical Alignment*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 167, 178 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

The Model's information system consists of five layers: Physical (the storage medium), Stream/Binary (delimited byte streams and file systems), Structure (primitive data types and logic), Object (data, container, and description objects), and Application (analysis and display programs).[279] The foundational conceptual structure is the Information Package, which includes both the digital object and the necessary associated metadata. The Functional Model includes six main functional entities: Ingest, Archival Storage, Data Management, Administration, Preservation Planning, and Access. OAIS also includes an Archive that has accepted six responsibilities to preserve information long term, and make it available for the Designated Community.[280]

The OAIS reference model is supplemented by the Producer-Archive Interface Methodology Abstract Standard, or PAIMAS (ISO 20652: 2006), which describes the workflow of negotiating and coordinating the submission and transfer of objects to an archive. OAIS also has influenced various preservation metadata standards, with PREMIS (cited above) being the most widely adopted.[281]

2. COP

The InterPARES Project's Chain of Preservation (COP) Model, although not a formal standard, has also been quite influential. It was adopted in 2007 as part of the InterPARES "Phase 2" process.[282] This model represents the activities of making, keeping, appraising, and preserving digital records during their entire lifecycle, and thus encompasses two more modules than the OAIS model.[283]

3. DLRM

The Digital Library Reference Model (DLRM), originally created through the auspices of the DELOS archiving project, provides a conceptual framework describing the characteristics of a digital library

279. CONSULTATIVE COMM. FOR SPACE DATA SYS., *supra* note 15, at E-1.

280. *Id.* at 1-1. Preserving Descriptive Information constitutes Fixity (authenticity), Reference (identification), Context (environment), Provenance (history), and Access Rights Management (pernmissibility). GIARETTA, *supra* note 86, at 185.

281. Ruusalepp et al., *supra* note 274, at 126.

282. Since launching in 1993, InterPARES has completed three separate phases. The fourth phase (2013-2018) is focused on digital records entrusted to the Internet. *See also* Sherry Li Xie, *Preserving Digital Records: InterPARES Findings and Developments*, *in* FINANCIAL ANALYSIS AND RISK MANAGEMENT, *supra* note 118, at 187, 187-204 (summarizing InterPARES third phase findings that the status of digital records preservation remains challenging due to weak or lack of preservation foundation).

283. *Id.* at 202-03.

management system. Classified domains are Organization, Content, User, Functionality, Policy, Quality, and Architecture.[284]

4. WARC

WARC (Web ARCHive) is an international file standard that can be used to combine different digital resources into an aggregate archival file. The WARC format is a version of the Internet Archive's ARC File Format, used to co-package networked objects and their context. In some cases, this means storing "web crawls" as sequences of content blocks harvested from websites.[285]

5. Audit Methods

Key challenges in the standards world include establishing trust, conforming to preservation metadata standards, and determining the appropriate scope for such standards.[286] In recent years the use of audits has become an important component in building trust in the operation of digital repositories. For example the TRAC (Trustworthy Repositories Audit and Certification) checklist presents nearly ninety organizational, technological, and digital object management criteria for digital repositories.[287]

6. Standards for Structural Interoperability

Disparate data systems must interoperate, and so too there are standards to assist in creating interoperability. Some include: Apple's Bento container; the Universal Preservation Format (UPF) (self-describing wrappers, or containers, containing both the data and the metadata); Open Media Framework (OMF) Interchange; and the Warwick Framework, which is developing the RDF (Resource Description Framework).

7. Open Source Software

David Rosenthal argues that digital preservation should utilize open-source software, claiming that closed-source preservation has "the same fatal 'just trust me' aspect that closed-source encryption (and

284. *Outputs*, DIGITAL LIBRARY INTEROPERABILITY (Nov. 11, 2016), http://bit.do/Outputs.

285. *Sustainability of Digital Formats*, DIGITAL FORMATS (Nov. 11, 2016), http://bit.do/SustainabilityDigitalFormats.

286. Ruusalepp et al., *supra* note 274, at 140-146.

287. CTR. FOR COMPUT. LIBR. ET AL., *Trusted Repositories Audit & Certification: Criteria and Checklist* (2007).

cloud storage) suffer from."[288] Others agree.[289] A compelling case can be made for adopting an open source license,[290] including the fact that open source software is less encumbered by legal constraints for archiving purposes. An open source approach also would provide full details of architecture, make available all code used, use open standards, and encourage collaborative development process.[291] Nonetheless, open source still has a versioning challenge which cannot easily be avoided.

F. Source Code and Other Software Preservation

While the focus thus far has been on the reproducibility of executable software programs—"digital vellum" or "digital artifacts" or just "content"—another area of deep concern is the preservation of software more generally. Examples of software can include system software (operating system or device driver), programming software (compiler or debugger) and application software (web browser or graphic design program). [292] The preservation of software is a sub-field yet to be thoroughly explored.[293] Even computer viruses and other malware threats should be saved for future analysis.[294] Only limited consideration has been given to software preservation as a digital object in its own right; as a result, some have begun exploring creation of a conceptual framework necessary to express "a rigorous approach" to software preservation.[295]

---

288. David Rosenthal, *Economic Sustainability of Digital Preservation*, SLIDESHARE Slide 23 (Sept. 28, 2014), http://bit.do/EconomicSustainabilityDigitalPreservation.

289. *See* Cal Lee, *Open Source Software: A Promising Piece of the Digital Preservation Puzzle,* 29 MIDWEST ARCHIVES CONF. 26, 26-28 (2001), http://bit.do/PromisingPieceDigital PresPuzzle.

290. Michel Castagne, *Consider the Source: The Value of Source Code to Digital Preservation Strategies*, 2 SAN JOSE ST. UNIV. SCHOOL STUDENT RES. J. 1, 7 (2012).

291. *Approach*, SOFTWARE HERITAGE (Nov. 11, 2016), http://bit.do/ApproachSoftware Heritage.

292. Castagne, *supra* note 290, at 2.

293. *Id.* at 1.

294. Howard Besser & Jonathan Farbowitz, *Why save a computer virus?*, CONVERSATION (Aug. 9, 2016). Some 82,000 new malware threats are created every day, and have become a pervasive feature of the Internet. Nonetheless, few "malware archives" are being created to preserve their technical, social, and historical value. *Id.* This can be especially problematic given the significant risk of ingesting compromised software as part of an emulation technique. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 25-26.

295. Brian Matthews et al., *A Framework for Software Preservation*, INT'L J. DIGITAL CURATION, July 2010, at 91, 92. UNESCO's PERSIST program seeks to preserve software as a matter of cultural heritage. *PERSIST: UNESCO Digital Strategy for Information Sustainability*, UNSECO (Jan. 30, 2014), http://bit.do/UNSECODigitalStrategy. The IMLS-funded Software Preservation Network is dedicated to collecting and preserving executable software. *About*,SPN,

Of particular concern is source code, which is deemed "integral to durable software preservation."[296] In essence source code is the flipside of the concern with digital objects, distinct yet inherently related to content data. Source code goes through several steps before it becomes an executable program, first passing through a compiler which creates object code that then passes through a linker, which combines the various modules to create machine code.[297]

III. THE PUBLIC POLICY CHALLENGES

> Let us save what remains: not by vaults and locks which fence them off from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident.[298]

A. Public Policy Frameworks

Public policy is "the art of determining a mix and dosage of instruments that can achieve the desired objectives."[299] Broadly defined, policy encompasses a range of institutions and organizations, engaged in what Bromley calls "an exercise in practical inference," aimed at achieving preferred future outcomes.[300] Social policy preferences are displayed in common, statutory, treaty, and constitutional law, as well as in agreements between parties who avail themselves of contract law.[301] A sustainable policy environment is one where, as Cherry and Bauer put it, the outcomes are both "adoptable and achievable."[302]

Digital preservation obviously takes place within broader public policy frameworks, including private law, public law, and criminal

---

http://bit.do/AboutSPN.

296. Castagne, *supra* note 290, at 1.

297. Cal Lee, *supra* note 289, at 2. Software Heritage has taken a leading role in preserving source code, with a mission to collect, preserve, and share all software that is publicly available in source code form. Glyn Moody, *Software Heritage, the 'Library of Alexandria of Software,' Launches Today*, ARS TECHNICA (June 30, 2016), http://bit.do/SoftwareHeritage.

298. Thomas Jefferson, *From Thomas Jefferson to Ebenzer Haward, 18 February 1971*, 19 THE PAPERS OF THOMAS JEFFERSON 287 (1974). Jefferson's reference to physical "locks" is echoed in the modern day practice of the LOCKSS program ("lots of copies keeps stuff safe").

299. Richard S. Whitt, *Adaptive Policymaking*, 61 FED. COMM. L.J. 485, 496 (2009) (quoting Cherry & Bauer)

300. DANIEL W. BRAMLEY, SUFFICIENT REASON: VOLITIONAL PRAGMATISM AND THE MEANING OF ECONOMIC INSTITUTIONS 14 (2006).

301. Whitt, *Adaptive Policymaking*, *supra* note 299, at 496-97.

302. *Id.* at 497 (quoting Cherry & Bauer).

law.[303] Nonetheless, no unifying legal framework exists today for digital preservation activities. Indeed, "a holistic legal understanding of digital long-term preservation is missing."[304] Instead, the governing law in this area is sprawling and complex.[305] This situation increasingly is problematic. Myriad laws and regulations can profoundly affect both the initial preservation, and subsequent re-use of, and access to, documents, data, metadata, and software.[306] Often these laws and regulations are adopted in complete ignorance of their potential impact on digital preservation. The inadvertent impact is only heightened as the laws change, regulations are revised, and licenses expire. Among other drawbacks, this makes it more difficult to effect alignment between national responses to the legal issues arising from digital preservation.[307]

In the United States, most of the discussion about digital preservation has been taking place in the shadow of the nation's copyright law.[308] While somewhat understandable, given a carve-out in the copyright statute for the work of libraries and archives, the end result has been a crimped conversation about the broader public interest in digital preservation. If it is true that "preserving creative works isn't only, or even primarily, a task that copyright law can accomplish,"[309] we should be casting a wider net in order to devise a holistic policy framework for digital preservation. For now, we will follow the prevailing scholarship and industry conversation, which focuses almost exclusively on copyright law. Additional relevant policy considerations will be raised in Part VI.

B. Copyright Law

Back in 2002, Donald Waters could say that there is "considerable confusion" among US policymakers about how the nation's intellectual property laws apply to digital preservation.[310] Not much has changed in the intervening years.

---

303. THOMAS HOEREN ET AL., LEGAL ASPECTS OF DIGITAL PRESERVATION 3 (2013).

304. *Id.*

305. *See generally id.*

306. *Id.* at xiv.

307. Adrienne Muir, Dwayne Butler, & Wilma Mossink, *Legal Alignment*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 43, 45 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

308. Alyssa N. Knutson, *Proceed With Caution: How Digital Archives Have Been Left in the Dark*, 24 BERKELEY TECH. L.J. 437, 450 (2009) (copyright law dominates the digital archiving discussion).

309. Reese, *supra* note 85, at 311.

310. Waters, *supra* note 257, at 90.

Originally, copyright laws were "envisaged as providing a limited degree of economic protection to a relatively small group of creators and content-producing industries, for a short and clearly delimited period of time."[311] More recently, these laws have been extended in terms of the covered creations and the time of protection. In 1998, the Digital Millennium Copyright Act (DMCA) was enacted, ostensibly to extend the copyright laws into the digital sphere.[312] Unfortunately, the concept of digital preservation was still new enough at the time that members of Congress failed to appreciate its significance. Instead, the relevant portions of the statute focused on making it illegal to circumvent security measures in digital items.

Opinions vary as to the relative impact of copyright and other intellectual property laws on digital preservation and access. Some commentators have cited copyright in particular as inimicable to the effective preservation of digital works for future use and re-use.[313] The 1996 Task Force put it bluntly: "The biggest problem for preserving digital information . . . isn't technology, it's intellectual property rights."[314] Liberating the content of a document from its medium has "unsettling consequences for the protection of IP in digital form."[315] This is because IP law and practice have been predicated on the context of physical artifacts, on "the familiar properties of information closely bound to a physical substrate."[316]

The actual legal analysis is straightforward enough. The unauthorized exercise of the rights in a work may result in infringement of copyright law, unless (1) the material is not protected by copyright (in the public domain); (2) digital preservation is undertaken by the owner of copyright in the work, or with the permission of the owner; or (3) the material is permitted under an exception in the copyright law.[317]

---

311. Andrew Charlesworth, *Intellectual Property Rights for Digital Preservation*, DPC TECH. WATCH REP. 12-02 3 (2012).

312. MARYBETH PETERS, A REPORT OF THE REGISTER OF COPYRIGHTS PURSUANT TO §104 OF THE DIGITAL MILLENNIUM COPYRIGHT ACT v (2001).

313. Charlesworth, *supra* note 311, at 1.

314. WATERS & GARRETT, *supra* note 22. Digital archives today continue to face many legal barriers, including "practically perpetual copyright terms in the material they include, an uncertain fair use doctrine, a chaotic licensing scheme, and a proliferation of online contracts that threaten archivists' efforts to construct comprehensive digital libraries." Knutson, *supra* note 308, at 437.

315. NATIONAL RESEARCH COUNCIL ET AL., THE DIGITAL DILEMMA: INTELLECTUAL PROPERTY IN THE DIGITAL AGE 33 (2000).

316. *Id.*

317. STUDY ON COPYRIGHT LAW, *supra* note 14, at 6.

There is no doubt that digital technologies create tension between long term preservation needs and copyright laws. Indeed, digital preservation can trigger copyright concerns in a way that analog preservation does not.[318] In the digital environment, nearly every action taken on a digital object may be classed as making a copy, broadcast, or performance.[319] So, the crux of the problem becomes: How can one preserve something that one does not own? Under what circumstances does the preserving organization have the right or permission to ingest the protected content into the preservation system?[320] And then, to provide user access on the other end of the process? Does one need a presumptive authorization to preserve something?

1. Exclusive Rights

The Berne Convention provides the foundation for governance of copyright law internationally. Various treaties provide the modern updates to the Berne Convention, such as the World Intellectual Property Organization (WIPO) Copyright Treaty (WCT). Neither the Convention nor the WCT mandates any exceptions or limitations specific to preservation activities or institutions.[321] However, the Berne Convention, carried through in the WIPO treaty, allows exceptions to the right of reproduction under the "three step test," where the reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author.[322]

Copyright applies to work that is recorded in some way. Rights exist for musical and dramatic work, as well as films, sound recordings, and literary, artistic, or typographic arrangements. In the United States, the United Kingdom, and much of the Western world, a person's work is automatically copyrighted. As a result, there is rarely such a thing as copyright-free material. However, some material may have had its rights waived, or the rights may have expired.[323]

US copyright law provides a copyright owner with the following exclusive rights: the right to reproduce the work (make copies); the right to create adaptations (derivative works); the right to distribute copies of the work to the public, as limited by the "first sale doctrine"

318. *Id.*
319. Deegan & Tanner, *supra* note 1, at 23.
320. *Id.* at 24.
321. STUDY ON COPYRIGHT LAW, *supra* note 14, at 7-8.
322. *Id.*
323. Deegan & Tanner, *supra* note 1, at 23.

(which does not apply to digital works); the right to perform the work publicly; and the right to display the work publicly.[324] One can be liable under US copyright law for infringement, even if one is not a direct infringer. Secondary liability includes vicarious and contributory (monetary) liability.[325]

A few highlights as applied to digital preservation:

- **Reproduction**: Copyright law is concerned with making copies. And reproduction is a fundamental activity of digital preservation. It is impossible to make any use of a copyrighted work in digital form without a computer also making a number of temporary copies.[326] Visiting a Web page technically means a server sends a copy of the data to recreate the Web page via a browser program. All that data is temporarily stored on the user's computer. Technically, then, a copy of the copyrighted material has been made.[327]

- **Performance Rights**: As indicated above, the owner of copyrighted works have a right of public performance Experts agree that this right of "making available," or of public performance or display, may be implicated by digital preservation.[328] State law may also provide some protection for unfixed performances.[329]

- **Database Rights**: A database may be protected by copyright as a compilation if there is originality in the selection, coordination, or arrangement of the contents.[330]

- **Moral Rights**: Section 106A of the U.S. Copyright Act provides the moral rights of attribution and integrity to authors of certain types of visual works.[331] Further, "whatever

---

324. 17 U.S.C. §§ 102, 109.

325. STUDY ON COPYRIGHT LAW, *supra* note 14, at 120.

326. NATIONAL RESEARCH COUNCIL ET AL., *supra* note 315, at 50.

327. ELIZABETH R. LEGGETT, DIGITIZATION AND DIGITAL ARCHIVING: A PRACTICAL GUIDE FOR LIBRARIANS 166 (2014). Some suggest that making a copy should not be a relevant factor, since computers by their nature make copies that are ephemeral and insignificant, and thus should not be considered infringing. NATIONAL RESEARCH COUNCIL ET AL., *supra* note 315, at 141-144.

328. STUDY ON COPYRIGHT LAW, *supra* note 14, at 6.

329. *Id.* at 122-23.

330. *Id.* at 122.

331. 17 U.S.C. § 106A; STUDY ON COPYRIGHT LAW, *supra* note 14, at 121.

> the copyright circumstances, creators have moral rights under IPR that relate to paternity and integrity."[332] The moral right to paternity is the right to be identified as the author of the work. Unlike copyright, this moral right would persist for as long as the digital item persists. The moral right to integrity is the right to not have the work altered in a derogatory manner. A person's image is also her or his property.[333] It is conceivable that some authors would deem the reformatting, migration, or other techniques used to preserve a digital object, to be potentially derogatory.[334]

2. Relevant Exceptions and Limitations

While the U.S. copyright statute contemplates a number of exceptions and limitations to the enumerated set of rights, those exceptions do not necessarily accommodate all the actions required for digital preservation.[335] Two key provisions are Section 108 and Section 107. The interplay between these two provisions provides much of the discussion, and confusion, over the extent to which the U.S. copyright laws allow digital preservation-related activities.

- **Section 108—Libraries and Archives**: Outside the United States, many other countries have recognized the global significance of copying and preservation exceptions for libraries and archives. Some 156 World Intellectual Property Organization (WIPO) member states have at least one statutory library exception.[336] The issue comes down to whether and how those exceptions encompass the types of activities necessary to preserve digital objects.

  Section 108 of the U.S. Copyright Act contains the "safe harbor" for libraries and archives. Enacted in 1996, and modified slightly by the DMCA in 1998, this provision establishes the ground rules for libraries and archives to preserve certain types of information content without violating the rights of copyright holders. The provision in

332. Deegan & Tanner, *supra* note 1, at 25.

333. *Id.* at 26.

334. *Id.*

335. STUDY ON COPYRIGHT LAW, *supra* note 14, at 2.

336. Notice of Inquiry, Section 108: Draft Revision of the Library and Archives Exception in U.S. Copyright Law, 81 Fed. Reg. 36594, 36597 (June 7, 2016).

particular allows making and distributing full-text preservation copies of unpublished works, making full-text replacement copies of published works, and making and distributing copies of excerpts or full texts of works for distribution to patrons.[337]

Importantly, Section 108 on its face does not mention digital preservation or curation. Nor does it refer to the types of techniques or processes involved in preserving or accessing digital objects.[338] As the Copyright Register explained in 2015, Section 108 "is replete with references to analog works and fails to address the ways in which libraries really function in the digital era . . . ."[339] This even extends to the limitation on copies to be made in preserving an analog object (three). The provision on its face also does not apply to museums or other memory institutions. One former publisher testified before Congress in 2014 that Section 108 "is so outdated and inadequate as to no longer serve its function."[340] The so-called Section 108 Study Group agreed, and in 2008 recommended specific fixes to help bring the provision up to date.[341] To date however no congressional action has led to any changes to Section 108.

- **Section 107—Fair Use**: Where the preservation activity does not otherwise fall within Section 108, the exclusive right of the copyright holder to reproduction can be waived by application of the fair use doctrine.[342] Fair use permits reproduction of limited amounts of copyrighted material for restricted purposes, such as review, analysis, commentary, and parody. The courts normally consider four factors (sometimes referred to as "PNAM") to determine whether or not a particular practice is fair use:

---

337. 17 U.S.C. § 108(b), (d), (e).

338. Laura N. Gasaway, *America's Cultural Record: A Thing of the Past?*, 40 HOUSTON L. REV. 643, 645 (2003); Knutson, *supra* note 308, at 452-53.

339. *The Register's Perspective on Copyright Review: Hearing before the H. Comm. on the Judiciary*, 114th Cong. 14 (2015) (Statement of Maria A. Pallante, Director, U.S. Copyright Off.).

340. *Preservation and Reuse of Copyrighted Works: Hearing before the Subcomm. on Courts, Intellectual Prop., & the Internet of the H. Comm. on the Judiciary*, 113th Cong. 28,30 (2014) (Statement of Richard S. Rudick, Co-Chair, Section 108 Study Group).

341. *Id. See* Section 108 Study Group, THE SECTION 108 STUDY GROUP REPORT 28 (2008), http://bit.do/Section108.

342. 17 U.S.C. § 107.

- the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit education purpose;
- the nature of the copyrighted work;
- the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- the effect of the use upon the potential market for or value of the copyrighted work.[343]

Section 108(f)(4) makes clear that nothing in that section affects the ability of libraries and archives to rely on fair use.[344] Further, the legislative history of the 1976 Copyright Act suggests that certain preservation activities may qualify as fair use.[345] Thus, the fair use standard in Section 107 acts as a potentially important bulwark where Section 108 is unclear as to its applicability to a particular digital preservation activity.[346]

- **Legal Deposit and Registration**: The "legal deposit" requirement is a statutory provision that obliges publishers to deposit copies of their publications in libraries in the country in which they are published.[347] The aim of legal deposit is the preservation of a country's published output for posterity.[348] Traditionally the requirement has applied to print publications. In the United States, the legal deposit requirement is contained in the copyright statute. Legal deposit in the U.S. covers all types of non-print publications.[349] However, the Copyright Office currently does not require digital objects to be deposited because it lacks the

343. *See* Mary Minow, *How I Learned to Love FAIR USE. . . .*, STANFORD COPYRIGHT AND FAIR USE CTR. (July 6, 2003), http://bit.do/LoveFairUse.

344. 17 U.S.C. § 108(f)(4). *See also Register's Perspective on Copyright Review*, *supra* note 339, at 15 (Section 108 has always had a savings clause for fair use).

345. STUDY ON COPYRIGHT LAW, *supra* note 14, at 123 n.245. *See also* Peter Hirtle, *Digital Preservation and Copyright*, STANFORD UNIV. LIBRS. (Nov. 10, 2003), http://bit.do/HirtleDigitalPreservation (Senate has concluded that making duplicate copies for preservation purposes falls within the fair use provision).

346. Libraries as a practical matter rely heavily on Section 107, "particularly with respect to the use of digital works, for which there is currently little clear legislative guidance." MARY ROSENBERG & CHRIS WESTON, OVERVIEW OF THE LIBRARIES AND ARCHIVES EXCEPTION IN THE COPYRIGHT ACT 31 (2005).

347. LAZINGER, *supra* note 38, at 59.

348. Muir et al., *supra* note 307, at 46.

349. STUDY ON COPYRIGHT LAW, *supra* note 14, at 125-26.

necessary infrastructure for holding and securing them. Under current U.S. and European law there is no requirement to register copyrights, or copyrighted materials.

- **Technological Protection Measures Provisions**: U.S. law prohibits circumventing technology measures intended to protect against content piracy. For example, DVD players incorporate circuits supporting technologies designed for the express purpose of preventing copying.[350] This provision has been read to prohibit legally bypassing the access control mechanism, even to preserve the content.[351]

3. Licensing and Contracts

Contract law has an interesting intersection with copyright law. In the United States, contracts normally are governed by state laws, defined in turn by their adoption of the uniform commercial code (UCC) concerning business dealings. However, the copyright owner can utilize contracts to give third parties the permission to do what otherwise would constitute an infringement of statutory rights.[352] So, the copyright law operates by default, and can be limited by voluntary agreements between parties.[353] Similarly, licensing agreements can override the terms of Sections 108 and 107.[354] So, "private" contract law can incorporate, or trump, "public" copyright law.

Permission to utilize a copyrighted work normally is granted by a third party through a license. Licenses can apply to the entire bulk of exclusive rights, or just to some. They can be limited or absolute, exclusive or non-exclusive, paid or free, voluntary or compulsory (mandated by law), and negotiated individually or collectively.[355] These licenses also can extend beyond copyright, to include other terms

350. LAZINGER, *supra* note 38, at 10-11.

351. Hirtle, *supra* note 345. Reese observes that these very same control measures intrude another layer of software and hardware that can become obsolete and itself must be preserved in some way. Reese, *supra* note 85, at 310.

352. Knutson, *supra* note 308, at 466-67.

353. 17 U.S.C. § 108(f)(4). One example is the Uniform Computer Information Transactions Act (UCITA), adopted on a state-by-state basis as part of the governing commercial code. *E.g.*, *What is UCITA*? UCITA ONLINE, http://bit.do/WhatIsUTICA.

354. Minow, *supra* note 343; *see also* Knutson, *supra* note 308, at 470 (federal copyright law does not preempt state contractual rights).

355. PETER B. HIRTLE ET AL., COPYRIGHT AND CULTURAL INSTITUTIONS 130 (2009). One trend against exclusive licensing is the growing use of free, standardized copyright licenses through Creative Commons. *What We Do*, CREATIVE COMMONS (Nov. 11, 2016), http://bit.do/CreativeCommons.

of service (TOS). In the online context, website owners often will seek to employ so-called "click-through" or "click-wrap" licenses, essentially nonnegotiable contracts that require users to give their assent to specific stipulated uses of the digital material.[356]

Licensing is more extensively used for digital works than in the print world, where retail sales are the primary means of distribution. While a sale normally would involve the transfer of ownership rights in the copy,[357] licensing constitutes a more limited transfer of rights to use under terms and conditions governed by contract law and determined by the publisher of the information.[358] Online publishers typically see such transfers as conditional leases to remotely access and utilize a resource, rather than outright ownership. Ironically, then, content in the cloud already consists of a more limited bundle of user rights than what otherwise would be found in the analog context.

C. Other Laws

1. Patent and Trademarks Laws

Another area of the law that potentially affects digital preservation activities is software patents. Patent restrictions need to be taken into account when choosing a preservation format.[359] As with copyright laws, there is no explicit digital "preservation" or "archiving" use exception in the patent laws of the United States, or elsewhere. Separately, digital objects could also represent the text of patents or images associated with trademarks.[360]

2. Bankruptcy Laws

Bankruptcy laws typically treat tangible assets of a firm or individual as private property. This would include, for example, the

---

356. HIRTLE ET AL., *supra* note 355, at 147-48. The practice has its roots in the "shrinkwrap" licenses commonly found with purchased software. Clickwraps have since been extended (controversially) to "browserwrap" licenses that only require viewing a website as grounds for creating a binding agreement. *Id.* at 148.

357. This element of the copyright law's "first sale" doctrine has come to be known as "exhaustion," and refers to the notion that transferring a copy of a work to a new owner diminishes the original owner's rights against the new owner. *See generally* Aaron K. Perzanowski & Jason Schultz, *Digital Exhaustion*, 58 UCLA L. REV. 889, 889 (2011).

358. LAZINGER, *supra* note 38, at 97-98. For example, the one common factor in every digital book is device-specific licensing. *Id.* at 11. This means that every electronic text typically is exclusively licensed for a specific e-book device.

359. Gary McGath, *T Minus 9 Days*, FILES THAT LAST (Jan. 6, 2013), http://bit.do/FilesThatLast.

360. CORRADO & MOULAISON, *supra* note 12, at 24.

software code, hardware, and other elements of an online business. When an entity files for bankruptcy, those assets would be subject to claims by creditors. The same arguably would be true of the third party digital materials stored by a data repository or cloud services provider. Without an explicit agreement in place that says otherwise, the courts may treat the data as part of the estate, or corporate assets, and thus not eligible to be returned to the content "owner."

3. Privacy and Data Protection

Laws relating to confidentiality and privacy, such as data protection acts and freedom of information acts, may restrict an entity's ability to access, use, and reuse data.[361] While the United States is generally recognized as having less stringent data protection laws than places like the European Union, financial and health records are accorded special treatment.

Depending on national law, informed consent often is required when gathering data from or about individual persons. In the EU in particular, privacy is considered a human right, and violation of data protection requirements can bring stiff penalties.[362] The EU's data protection directive, adopted in 1995, governs personal and sensitive data.

4. Content Liability

Preservation bodies may also face liability for the content they acquire and maintain; content may be libelous, offensive, or obscene, or fall afoul of blasphemy or anti-terrorism laws. Providing access to such material may expose the preservation institution to liability for the material, not only in its own country but also in other jurisdictions.[363]

361. HARVEY, *supra* note 89, at 206.

362. *Information Society, Privacy and Data Protection*, EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS (Nov. 12, 2016), http://bit.do/InformationSociety.

363. Muir et al., *supra* note 307, at 44.

IV. THE FINANCIAL CHALLENGES

"It's money that matters in preservation."[364]

Who precisely is responsible, both morally and financially, for the long-term stewardship of digital content? After all, "economic sustainability—generating and allocating the resources necessary to support long-term preservation activities—is fundamental for the success of long-term digital preservation programs . . . . And yet, this fundamental point has not received the attention or the analysis it deserves."[365] Compared to the substantial literature on the technical and policy aspects of digital preservation, the economic aspects until recently have been "relatively neglected."[366]

Uncertainty about the funding question "creates a significant barrier to the coherent, systematic preservation of digital information."[367] More to the point, "lack of capital will hasten the obsolescence of digital works—whether they are managed within safe repositories or live in more threatening domains—because at some point people forget, they stop paying attention, or they lack time, expertise, documentation, rights, or tools."[368] Or as Gioretta succinctly puts it, the one real foolproof solution for digital preservation is "MONEY"—enough of it, and for an indefinite period.[369]

Public policy also lags behind the need to motivate creators and distributors "to save the national patrimony for future generations."[370] There is little to no budget for digital preservation at the federal, state, and local levels, even though most of their documents, legislation, land maps, and other vital records now are in digital form.[371] Because "long-term benefits hardly ever convince politicians and other power-holders," more direct benefits must be shown.[372]

---

364. Chapman, *supra* note 50, at 146.

365. Lunghi et al., *Economic Alignment*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 195, 195 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

366. *Id.*

367. WATERS & GARRETT, *supra* note 22, at 37.

368. Chapman, *supra* note 50, at 145-46.

369. GIARETTA, *supra* note 86, at 7-9.

370. Laura Campbell, *A Tale of Two Countries: Part I*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATIONS 17, 27 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

371. *Id*.

372. Inge Angevaare et al., *Organizational Alignment*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATIONS 89, 94 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

There is also an important linkage between the technical and the economic. In the words of one set of researchers, "Simplicity contributes to economic sustainability; complexity undermines it."[373] Even as the proposed technical solutions for digital preservation continue to evolve, we have yet to see a maturing of the economic basis for this activity.[374] The next three sections briefly explore the incentive structures, costs and benefits, and potential business models involved in preserving digital materials.

A. The Financial Incentives

Economics is primarily about incentives, and all ways that humans interact via the exchange of resources. Not all incentives are monetary. For example, for many people sharing Web content stems from an interest in sharing life experiences and ideas. Nonetheless, getting people to agree to spend money on a certain activity is both a science and an art.

1. Preservation As a Public Good (And More)

Broadly speaking, a public good is a shared benefit at a societal level. Because one would assume that cultural memory is deemed a public good, insuring against its loss also would be a public good. To many observers, preserving the past is a shared benefit that is conferred upon both current and future generations.[375]

As a matter of economic theory, the archiving of digital information has special properties as a public good.[376] Digital preservation exhibits attributes of a public good—such as national defense and public parks—for which there are no practical means to exclude those who do not contribute towards the cost of provision.[377] In economic terms, the argument would be that preserved digital materials are both non-excludable and non-rivalrous.[378] As a result, the

373. Lunghi et al., *supra* note 365, at 209.

374. Deegan & Tanner, *supra* note 1, at 26.

375. Lavoie, *supra* note 5, at 107; ROSENZWEIG, *supra* note 12, at 7 (preservation of the past is a public responsibility).

376. Waters, *supra* note 257, at 83. Waters sees cultural memory, and its preservation in digital form, as a public good.

377. Lavoie, *supra* note 5, at 110.

378. Richard S. Whitt, *A Deference to Protocol*, 31 CARDOZO ART & ENT. 689, 721 (2013). Brett Frischmann, noted authority on physical infrastructure, has discussed the concept of intellectual infrastructure. He has labeled the Google Books program a "mixed" infrastructure with substantial spillovers—a non-rival digital collection, but with partially non-rival facilities and services. BRETT M. FRISCHMANN, INFRASTRUCTURE, 260 n.6 (2012). Similarly, digital preservation could be seen as basic infrastructure, as it is difficult to value and markets arguably

incentives are significantly reduced for any institution to be the one to take on the effort and expense of preserving the materials.[379] "It is far better, from an economic perspective, to wait for another institution to assume responsibility for this task; the benefits from preservation then can be enjoyed by all."[380] This free-rider problem—a collective attitude of "wait for someone else to do it"—must be taken seriously.[381]

Digital preservation raises the classic problem of the political economy of public goods—the incentives for individuals and institutions to participate in the provision of a good from which others cannot be readily excluded. As one example, "an institution that preserves the last copy of a resource has performed a service of potentially incalculable value to the public. In these circumstances, the benefits from preservation are widely distributed; unfortunately, the costs of preservation are not."[382]

Moreover, it could be argued that not all digital materials qualify for preservation as a public good. The 2010 Blue Ribbon Task Force specified four types of content they concluded were in the public interest: scholarly discourse, research data, commercially-owned copyrighted cultural content, and collectively produced Web content.[383] On the other hand, it may be difficult if not impossible to determine ahead of time what information may be relevant and even critical to future generations.

2. Stakeholders and Incentive Structures

"We have to be realistic about the viability of any of our institutions and their abilities to invest at the level of funding needed for the creation of digital archives on a comprehensive scale."[384] In determining who regards the content as valuable, it is useful to know the designated roles of the creator, the principal keeper, and the

---

are failing to efficiently supply it. *Id*. at 262, 378.

379. There is a counter-argument that preserved digital resources could be treated as private goods, excluding those who do not pay for their preservation and access. Whitt, *A Deference to Protocol*, *supra* note 378, at 722. However, it would be difficult to exclude those who do not contribute to the provision of the good from enjoying its benefit. Waters, *supra* note 257, at 83. Also, limiting access largely defeats the purpose of preserving content for the future.

380. Lavoie, *supra* note 5, at 110-11.

381. *Id*. at 111.

382. Lavoie & Dempsey, *supra* note 84, at 12.

383. BLUE RIBBON TASK FORCE ON SUSTAINABLE DIGITAL PRESERVATION AND ACCESS, SUSTAINABLE ECONOMICS FOR A DIGITAL PLANET: ENSURING LONG-TERM ACCESS TO DIGITAL INFORMATION 49 (2010).

384. Meredith A. Butler, *Issues and Challenges of Archiving and Storing Digital Information*, 24 J. LIBR. ADMIN., no. 4, 1997, at 61, 76.

principal user.[385] How do we create incentives for entities to voluntarily adopt proffered technical solutions, and/or seek legal changes?

LAMs and related institutions directly own, and have physical custody of, one or more copies of the analog materials in their collections. They also control public access. The institutions therefore are uniquely placed to undertake the preservation of these materials, and this enhances the incentives to preserve. However, "these incentives to preserve analog materials—physical custody and limited opportunities for sharing—break down in the digital world."[386] Thus, "in the absence of a formal preservation mandate, incentives to preserve digital materials, without compensation for the benefit of society as a whole may be weak indeed."[387]

Outside the library context, there may be some hope:

> We are fortunate that electronic preservation is of some interest to other communities for the mundane commercial reasons. The financial, publishing, and other business communities have a stake in the authenticity of their electronic communications. The business and computing communities wish to protect against the undesired loss of data in the short term. The governmental and business communities profess an interest in the security of systems.[388]

Theoretically, however, some government and company officials may have perverse incentives not to preserve certain information, because it creates public accountability (a cost to some) without other benefits.[389] Further, the "Right to be Forgotten" in Europe now allows users to have unwelcome content about themselves removed from search indexes based on their names.[390]

B. The Economic Costs and Benefits

1. Large but Often-Unknown Costs

While scarcity of funds is hardly a new problem in preservation (often commanding less than three percent of a typical library's total budget), the economic requirements of digital preservation, "fueled by the immediacy and scale of the problem, will exacerbate this familiar

385. Lunghi et al., *supra* note 365, at 219.

386. Lavoie & Dempsey, *supra* note 84, at 13.

387. *Id.*

388. Graham, *supra* note 20, at 11.

389. *See* Althous & Leetaru, *supra* note 134.

390. *Factsheet Data Protection*, EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS (2015), http://bit.do/FactsheetDataProtection.

problem to a degree heretofore unseen."[391] In fact, the question of technical feasibility of digital preservation is overshadowed by the question of its economic feasibility. "Even the most elegant technical solution is no solution at all if it is not economically sustainable."[392]

Evaluating digital archiving is impossible without concrete measures of costs, benefits, and values of digital objects.[393] Unfortunately there are a limited number of viable cost models available, and costing assumptions vary greatly between different institutions.[394] By most measures, however, the most significant factor in the costs of digital preservation is staffing; in some cases seventy percent of all costs goes to human resources.[395] One bit of good news is that the costs of keeping digital data continues to fall over time, a very special and convenient property.[396] Storage costs in particular are "dropping by 50% every 18 months."[397]

The costs involved in electronic data archiving include the costs for converting information into digital form, and costs for maintaining digital information.[398] Three components of long-term preservation costs related to a repository are: (1) capital costs (upfront expenditures to set up repository systems and processes); (2) direct operating costs (expenditures needed to sustain the repository's operations over time); and (3) indirect operating costs (the repository's overheads).[399] The "LIFE" model for example lists six different cost components in the preservation lifecycle.[400]

---

391. Lavoie, *supra* note 5, at 107.

392. *Id*.

393. *Id*. at 125 (quoting Hedstrom).

394. CORRADO & MOULAISON, *supra* note 12, at 77-78; HARVEY, *supra* note 89, at 90; Deegan & Tanner, *supra* note 1, at 26. Some, such as the EU's "4C" group, have sought to address this knowledge gap. *See Overview*, 4C PROJECT (Nov. 11, 2016), http://bit.do/4CProject (describing sustainable "Collaborations to Clarify the Costs of Curation"). This project since has evolved into the CCEx (Curation Costs Exchange). *About*, CURATION COSTS EXCHANGE (Nov. 11, 2016), http://bit.do/CurationCostsExchange.

395. CORRADO & MOULAISON, *supra* note 12, at 27. However much costs appear to come down, it will be much more expensive and require specialized staff with the necessary technical skills to ensure not just that a digital object is preserved and what it purports to be, but also that it is held as securely as in the analog world. Moss, *supra* note 31, at 13.

396. Holdsworth, *supra* note 176, at 57.

397. Richard Wright et al., *The Significance of Storage in the "Cost of Risk" of Digital Preservation*, 4 INT'L J. OF DIGITAL CURATION 104, 119 (2009).

398. LAZINGER, *supra* note 38, at 111.

399. Lavoie, *supra* note 5, at 109. (citing Sannett).

400. CORRADO & MOULAISON, *supra* note 12, at 28. The California Digital Library also has developed a total cost of preservation (TCP) model. *Id*.

Because most digital preservation initiatives are still in their infancy, they have yet to produce a record of the complete cost trajectory underlying stewardship of digital resources through successive shifts in prevailing hardware and software environments.[401] It can be particularly difficult to separate out digital preservation costs from costs associated with producing, acquiring, disseminating, and rendering digital objects.[402] Nonetheless, Lavoie maintains that there are some certainties about cost. First, recovering costs will require a substantial ongoing commitment, through the allocation of ongoing, budgeted funds. Second, costs will be incurred throughout the entire lifecycle. Third, digital preservation costs increasingly will become inseparable from other aspects of digital collection management. Finally, intangible costs must be reckoned with as well, including "the costs of changing the mindset of all stakeholders in digital materials," so that they perceive preservation to be an immediate, ongoing issue for which all bear responsibility.[403]

2. Ongoing Commitments

A significant commitment in effort and resources is required to preserve digital materials. This is due to recognizing that preserving digital materials requires active—and costly—intervention throughout "the information lifecycle."[404] Smaller institutions have a particularly difficult time grappling with "feeling overwhelmed and under-resourced," even as tools and services are "developed, updated, and occasionally abandoned at alarming speed."[405]

David Rosenthal has examined the costs of digital preservation in the context of the information lifecycle, and concluded that about one-half are due to ingest, one-third for preservation and one-sixth for access.[406] Resource allocations tend to be driven towards "low hanging fruit"—content from larger publishers at low risk of loss (due to relative ease of discovery, migration, and collection)—and away from content at higher risk of loss. Perversely, as Rosenthal finds, the more difficult it is to find, collect, and migrate content, the less likely it will be funded.[407]

---

401. Lavoie, *supra* note 5, at 111.

402. Chapman, *supra* note 50, at 135.

403. Lavoie, *supra* note 5, at 115.

404. *Id.* at 108.

405. JAIME SCHUMACHER ET AL., FROM THEORY TO ACTION: GOOD ENOUGH DIGITAL PRESERVATION FOR UNDER-RESOURCED CULTURAL HERITAGE INSTITUTIONS 3 (2014).

406. Rosenthal, *supra* note 4.

407. *Id*.

Any dedication to preserving digital resources into an indefinite future "presupposes a parallel commitment to marshal, on an ongoing basis, the resources—funds, infrastructure, expertise, etc.—necessary to do so." For many this is an "often neglected" point, with far more time and effort ordinarily devoted to the technical issues than to the economic issues.[408]

3. Technical Choices Are Key

The choice of digital preservation strategy will figure heavily in the associated long-term costs.[409] The costs become more considerable as one moves from the objective of mere bit preservation, to preservation of intellectual content, to preservation of original form and functionality.[410] The cost implications associated with the relative simplicity or complexity of a digital preservation strategy can be linked to the scope of what is chosen to preserve.

Lavoie explains that the scope of preservation can be interpreted in both a horizontal and a vertical sense. The vertical interpretation comprises the layers of technology that sit between the user and the digital content: storage technologies, network and computing resources, operating systems, application programs, etc. Technology preservation aims to preserve the entire vertical stack environment, which of course adds to cost. The horizontal interpretation addresses the variety or range of environments, or portions of environments, that are chosen for preservation. As the range expands, so will the costs.[411]

As one example, while emulation best preserves the original form and functionality, supporting that most ambitious preservation objective consequently makes it the most expensive option. Emulation will require developing and maintaining a library of emulators, preserving the software and hardware environments along with the digital object, and maintaining a constant supply of new emulators as new environments emerge.[412]

Future costs are so uncertain because of the current lack of consensus on best practice to carry out digital preservation, presumably including format migration and/or emulation. In addition, preserving digital content for decades or even centuries in the future increases the variance attached to any forecast of long-term costs—the more distant

408. Lavoie, *supra* note 5, at 108.
409. *Id*. at 115.
410. *Id*. at 117.
411. *Id*. at 123.
412. *Id*. at 120-21.

the time horizon, the greater the uncertainty surrounding the effort and expense necessary. Further, the goals of digital preservation themselves are uncertain, and—depending on evolving user expectations—can range from preserving only the intellectual content, to all aspects of the original look, feel, and functionality.[413]

C. Funding Mechanisms and Business Models

While a cost model provides a framework for recording and allocating costs, a business model shows how the service can sustain itself financially. This begs the question: who will pay to support digital preservation, and how?

It is obvious that "there remain large challenges and gaps in both defining the business case and the business models for preservation."[414] The short-term perspective (no longer than three to five years out) is "probably the single most critical reason that making a business case or economic argument for preservation is a difficult proposition."[415] Making that case is complicated by long time horizons, misaligned or weak incentives, and stakeholders who are diffused and lack clarity about roles and responsibilities.[416] Many digital preservation initiatives are funded by "soft money," such as grants, one-time donations, or other one-time expenditures.[417] In the United States, funding dedicated to digital preservation "has traditionally lagged behind that available in the European and British contexts in particular."[418] And so far, nonprofit organizations have not demonstrated an ability to earn enough money to operate independently.[419]

Nonetheless, interesting work has emerged in recent years to begin classifying and examining possible options.[420] Multiple funding streams have been posited, including government funding, philanthropy, private markets (a self-sustaining business activity), and

---

413. *Id*. at 112-13. Lavoie goes on to argue that preservation costs likely will emerge as a byproduct of negotiation and compromise between principal stakeholders and decision-makers. Lavoie, *supra* note 5, at 123.

414. Lunghi et al., *supra* note 365, at 215.

415. *Id*. at 196.

416. *Id*. at 213.

417. CORRADO & MOULAISON, *supra* note 12, at 79.

418. Christopher Jordan et al., *Encouraging Cyberinfrastructure Collaboration for Digital Preservation*, TEXAS ADVANCED COMPUTING CENTER (2008), http://bit.do/Encouraging Cyberinfrastructure.

419. Nancy L. Maron et al., *Sustaining Digital Resources: An On-the-Ground View of Projects Today, in* ITHAKA CASE STUDIES IN SUSTAINABILITY 10 (2009), http://bit.do/Sustaining DigitalResources.

420. Lunghi et al., *supra* note 364, at 215.

publisher-based archives. Factors to influence commercial sustainability also have been identified, including dedicated and entrepreneurial leadership, a clear and compelling value proposition, minimizing direct costs, developing diverse sources of revenue, and clear accountability and metrics for success.[421] On the commercial side, sustainable pricing models must be developed, whether as one-time, upfront computational archiving fees, or annual fees.[422] Direct charging for use will be acceptable to some communities, such as corporate content creators and professional associations.[423] A subscription software model would entail monthly use of software and the right to use in an emulation environment. One-time payment models, on the other hand, are not likely to be sustainable.

Hedstrom and Montgomery maintain that:

> Considerable research is needed to develop funding and business models for repositories that assume preservation responsibilities. Repositories may be expected to preserve digital resources even though their utility may not become apparent until well into the future and even though the future users are not yet born. Over the long term, new communities of users will emerge with needs and expectations that differ from those of the communities that created the digital content.[424]

User expectations also must be factored into the equation. For example, is one paying for a certain guaranteed outcome associated with the preservation process, or only a guaranteed process?[425]

The digital repositories concept is still relatively new. Professionally managed repositories represent digital preservation as a business, where long-term use of stored objects will result from *quid pro quo* transactions between object owners and preservation professionals. The object owners would pay fees to repository managers. Indeed, "one can imagine making an actuarial calculation of the lifetime cost of preserving a digital information object, finding creators/providers/owners with an economic interest in paying to preserve their information, and constructing an archival service that functions much like a safety deposit system for digital information objects."[426] Digital repositories also can engage in fee-for-service

---

421. Maron et al., *supra* note 419, at 11. *See also* BLUE RIBBON TASK FORCE, *supra* note 382, at 12 (Conditions for sustainability also include selecting digital objects with long-term value, and establishing appropriate governance of digital preservation activities).

422. Lavoie & Dempsey, *supra* note 84, at 11.

423. WATERS & GARRETT, *supra* note 22, at 38.

424. HEDSTROM & MONTGOMERY, *supra* note 41, at 5.

425. Lavoie & Dempsey, *supra* note 84, at 11.

426. WATERS & GARRETT, *supra* note 22, at 38.

pricing, including one-time infrastructure development fee, annual support fee, storage fee, ingest per-batch fee, and fees for transformation or other interventions.[427]

V. HARNESSING "DEEP INFRASTRUCTURE" TO MOVE FORWARD

"We've GOT to get organized!"[428]

Hopefully, the foregoing discussion leads to but one inevitable conclusion: the world must get better organized, and quickly, to preserve our digital present and future. As Vint Cerf puts it, we need the digital equivalent of vellum—an enduring virtual platform and process for preserving our current and future bits.[429] The technology and laws and finances must come together, across time and space, to render our digital heritage. Not as a one-time silver bullet solution, however, but as a persistent, ever-evolving process.

To be clear, much is happening already. Thousands of dedicated researchers and archivists and academics and government officials and volunteers are doing what they can. Institutions large and small are allocating scarce resources to study the intertwined problems and implement changes. And some of the results so far are most impressive.[430]

But through no one's fault, the current pace and scale and scope is not sufficient to meet all the recognized challenges. And outside the relatively small and tight-knit community of interest, despite the urgency of the situation, the cause of digital preservation is "generally met with indifference."[431] At its root, digital preservation should become a broad-based social cause, not limited to any particular set of players. In particular, the digital preservation challenges confound national boundaries, and require international alignment across legal, organizational, standards, economic, and educational lines.[432] More

---

427. Chapman, *supra* note 50, at 135.

428. THE RUSSIANS ARE COMING, THE RUSSIANS ARE COMING, THE MIRISCH CORP. (1966).

429. *See, e.g.,* Vint Cerf, *Digital Vellum*, YOUTUBE (Aug. 7, 2015), http://bit.do/DigitalVellum (describing need for digital vellum).

430. "Electronic records, digital preservation, and digital curation programs have multiplied at an alarmingly fast rate, producing a useful bewildering array of theoretical frameworks, diagrams, software, and services." Christopher Prom, *Making Digital Curation a Systematic Institutional Function*, THE INT'L J. OF DIGITAL CURATION Mar. 2011, at 139, 142. Nonetheless, it is not clear that LAMs have made progress in implementing the procedures, tools, and services to actually preserve digital records. *Id.*

431. BORGHOFF ET AL., *supra* note 25, at v.

432. McGovern, *supra* note 158, 5-16. "At least some key aspects of the digital present …

can be done, to better supplement, and organize, and fund, and publicize what is already happening. And ultimately, "keeping knowledge, rather than objects, is an organizational problem."[433]

This Part focuses on moving forward with a greater sense of urgency and focus, pulling together useful strands of learnings and making concrete organizational and institutional recommendations. The discussion draws in part on prior work by the author and others in the parallel universe of Internet policymaking. In brief, the overarching message is that "form (and forum) should follow function."[434]

A. Our Goal: Implementing A "Deep Infrastructure" Process

1. Unifying Framework, Diverse Solutions

Even greater than the need for a technology fix is the need for institutional will. The often prescient 1996 Task Force Report described the need to develop what it called "deep infrastructure." Rather than a narrow exercise of fine tuning technical variables, or a clear problem of restoring crumbling books, preserving digital information:

> is a grander problem of organizing ourselves over time and as a society to maneuver effectively in a digital landscape. It is a problem of building—almost from scratch—the various systematic supports, or deep infrastructure, that will enable us to tame anxieties and move our cultural records naturally and confidently into the future.[435]

Some twenty years later, unfortunately, that task remains largely undone.

Hedstrom and others have noted that, in addition to the technical, legal, and economic challenges to digital preservation, numerous organizational barriers stand in the way as well.[436] In fact, Lesk maintains that other problems are "insignificant compared to the organizational issues."[437] Some even refer to the "Digital Preservation Triad" of Management, Technology and Content.[438]

---

do not follow national boundaries and, indeed, erode them." ROSENZWEIG, *supra* note 12, at 18.

433. Michael Lesk, *supra* note 24, at xvi.

434. Vincent Cerf et al., *A Perspective from the Private Sector: Ensuring that Forum Follows Function*, *in* BEYOND NETMUNDIAL 31, 33 (William J. Drake & Monroe Price eds., 2014).

435. WATERS & GARRETT, *supra* note 22, at 7.

436. HEDSTROM & MONTGOMERY, *supra* note 41, at 24.

437. Lesk, *supra* note 24, at xv.

438. CORRADO & MOULAISON, *supra* note 12, at 17-37.

As the 1996 Task Force Report makes plain, the process of building deep infrastructure necessarily means bringing together into a single conceptual framework all the multiple but interrelated activity streams discussed earlier. We need an ordering mechanism to encompass what Hedstrom calls a "spectrum of solutions" in terms of scale, format types, and institutional responsibilities.[439]

2. Engaging in Life-Cycles and System-Layers

Like other human activities, effective data preservation is based on the use of conceptual models. Data preservationists often speak approvingly of managing the natural "lifecycle" of digital information objects.[440] With digital objects this progression runs the gamut—from creating, editing, and describing them, to disseminating, acquiring, using, and revising them, and finally to retaining them for future use. A complex, interwoven community of creators and others participates in this lifecycle.[441] In essence, preservation must be closely integrated with the creative process itself.

A number of data lifecycle models exist today.[442] In 2008, the Digital Curation Centre released its "Curation Lifecycle Model," intended to describe the complete digital preservation process. The Model consists of four Full Lifecycle Actions: Curate and Preserve; Description and Representation Information; Preservation Planning; and Community Watch and Participation.[443] In turn, the Lifecyle is comprised of eight Sequential Actions: Conceptualize, Create or Receive, Appraise and Select, Ingest, Preservation Action, Store, Access, Use, and Reuse, and Transform.[444]

During their many iterations, digital information objects acquire the qualities of content, fixity, references, provenance, and context, raising unique issues and stakeholders at each stage.[445] The lifecycle approach helpfully can be seen as a continuum in time, to identify

---

439. Hedstrom, *supra* note **Error! Bookmark not defined.**, at 34.

440. *See generally*, DIGITAL PRESERVATION COALITION, *supra* note 16. Some trace the origins of the "records lifecycle model" to the 1940s. Xie, *supra* note 282, at 188-89.

441. WATERS & GARRETT, *supra* note 22, at 11.

442. *See, e.g.,* Line Pouchard, *Revisiting the Data Lifecycle with Big Data Curation*, 10 INT'L J. OF DIGITAL CURATION, no. 2, 2015, at 176, 180 (describing the different lifecycle models).

443. HARVEY, *supra* note 89, at xvi.

444. *Id*. at xvii. Harvey maintains that while the Curation Lifecycle Model follows closely the OAIS Reference Model, it includes activities that take place outside the archival system, and so fully encompasses the more narrow OAIS approach. *Id*. at 33. *See also* Niu, *supra* note 172, at 66 (OAIS Reference Model does not explicitly include appraisal as a function).

445. WATERS & GARRETT, *supra* note 22, at 19.

dependencies, barriers, and collaborative solutions in various infrastructure components at the appropriate stage(s).[446] Per Rieger, it also helps us give equal emphasis to the later stages of discovery, access, and delivery of digital content.[447]

In parallel, the technology, policy, and financial layers can be mapped out as well. One way to simplify this approach, as demonstrated in the Internet policymaking context,[448] is to separate out the three dimensions of Code (the target activity), Rules (the institutional tools), and Players (the organizational entities).[449] Or, Code is what we should be doing, Rules is how we should be doing it, and Players is who should be doing it.

These dimensions translate nicely into a conceptual framework that aligns with the modularity of the Internet itself. Data systems have been viewed as a composite of several interrelated information layers, which helps inform and direct specific preservation strategies.[450] Others have suggested a similar layered approach, cutting along the lines of information systems. Boudrez, for example, observes that when dealing with digital preservation "it is best to start from the information system itself."[451] To that end, he created a decision model for a digital preservation system based on four questions: WHAT do we archive, WHO manages the archive, HOW do we preserve the archive, and WHEN does the transfer take place. These questions in turn correspond to different information layers, as articulated by the five-layer OAIS digital preservation model.[452]

As is true of the Internet more generally, no solution to the digital preservation challenge can be appropriately comprehensive or effective without addressing all the affected systems. Different approaches to digital preservation suggest a natural layering of functions and interfaces.[453] For example, Lavoie suggests deconstructing a digital preservation system into four functional layers: infrastructure (hardware, software, and network); metadata; preservation mechanism;

---

446. DIGITAL PRESERVATION COALITION, *supra* note 16, at 20, 26. *See also* GIARETTA, *supra* note 86, at 431-33 (describing the digital preservation infrastructure components, including those serving the Data Producer, Data Curator, and Data Consumer).

447. Rieger, *supra* note 174, at 36.

448. Whitt, *A Deference to Protocol*, *supra* note 378, at 732-34.

449. Whitt, *Adaptive Policymaking*, *supra* note 299, at 511-36.

450. BOUDREZ, *supra* note 215, at 6, 15; *see also* GIARETTA, *supra* note 86, at 73 (describing the OAIS layered information model in terms of Media Layer and Application Layer).

451. BOUDREZ, *supra* note 215, at 27.

452. *Id*. at 14 n.18.

453. Waters, *supra* note 257, at 89.

and access.[454] Further, "the concept of organizing the digital preservation challenge into a series of components, or layers, in a model architecture provides a basis for distributing responsibility among various types of institutions."[455] This paper largely retains the OAIS model as a useful way to conceptualize the dimensions of the digital preservation challenge.

The Open System Interconnection (OSI) Reference Model was first articulated in the 1970s.[456] While never adopted officially to govern the Internet's development, the modular OSI model has remained influential in engineering circles, in part because it comports well with the virtual ecosystem of players that has sprung up at the heart of the Internet.[457] OSI lays out seven different interdependent, software-derived layers, from physical networks at the bottom to applications and content at the top.

Two additional layers are brought here into the discussion. The late Evi Nemeth semi-facetiously published a version of the OSI stack which included two layers at the top: the Financial as Layer 8, and the Political as Layer 9.[458] Nemeth's insight is a keen one—analyzing Internet-related activities should take into account the larger context.

Here then is one way to flesh out the "Code" element of our analysis, using for illustrative purposes the OSI stack as modified by Nemeth:

454. Lavoie & Dempsey, *supra* note 84, at 8.

455. Hedstrom, *supra* note **Error! Bookmark not defined.**, at 34.

456. Whitt, *A Deference to Protocol*, *supra* note 378, at 732-33.

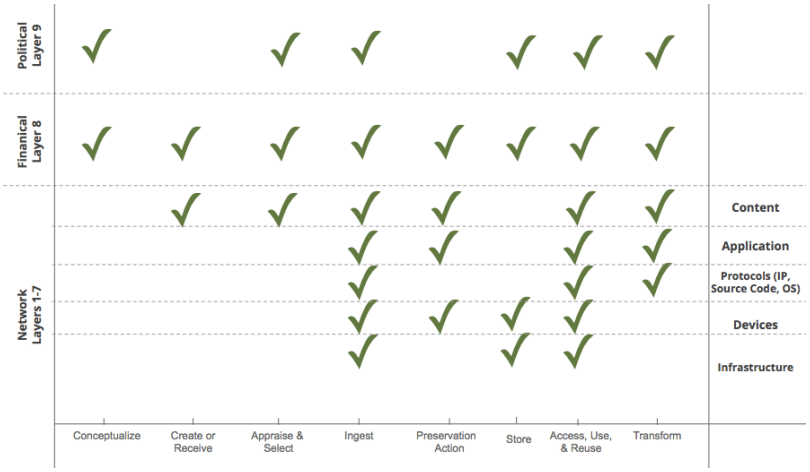
457. *Id*. at 733.

458. *9-Layer T-Shirt OSI Model per Evi Nemeth*, INTERNET SYSTEMS CONSORTIUM (Dec. 15, 2015), http://bit.do/OSIModel.

- *Layers 1-7 (The Internet itself):*
  - Software (user content and applications) (Upper Layers)
  - Software (Internet Protocol, source code, and operating systems) (Middle Layers)
  - Physical Hardware (devices and network equipment) (Lower Layers)
- *Layer 8: Finance (Business/Commercial/Financial systems)*
  - Funding Context
- *Layer 9: Politics (Legal/Regulatory/Political systems)*
  - Legal and Political Context

In this case, OSI layers 1-7 match up nicely with Lavoie's vertical "layers of technology," and Gioretta's "islands of capabilities" that sit between the user and the digital content.[459]

Plainly there are different techniques for different types of digital objects, for different software and hardware elements supporting the objects, and for different phases in the information flow. Further, each of the system layers—from the various forms of hardware, to various forms of software—is prone to errors that can destroy the meaning of the content.[460] Thus, combining digital life-cycles and system-layers helps us uncover the full complexity, so as to better understand and work effectively with it:



459. Lavoie, *supra* note 5, at 122.

460. HOEREN ET AL., *supra* note 303, at 1; GIARETTA, *supra* note 86, at 457-59 (laying out the network layers, storage and computer layers, repository layers, and preservation layers). Giaretta regards the present and future as two "islands of capabilities" which are joined by a number of infrastructure components. *Id*. at 459.

The illustrative digital life-cycles/system-layers mapping shown above is an example of taking a Code approach to digital preservation. Together these related movements in time and space constitute the digital landscape for all preservation and access activities.[461]

Creating and implementing deep infrastructure for digital preservation should encompass both the digital life-cycles and the system-layers elements, as a way of opening up new insights and solutions. A brief discussion of the remaining dimensions of institutional tools and organizational entities (Players and Rules) follows below.

3. Emphasizing Programmatic Process

The problem of digital preservation is not static; any solution must be inherently evolutionary.[462] We must look at data preservation "not just as a mechanism for ensuring bit sequences created today are renderable tomorrow, but as a process operating in concert with the full range of services supporting digital information environments, as well as the overarching economic, legal, and social contexts."[463] In other words, digital preservation is less like an event occurring at discrete intervals, and more like a process proceeding continually over time.

Much as the focus should be on process rather than outcome, we should move from project-driven activities to a fundamental program of core activity worldwide.[464] This approach will not be easy. Many digital preservation activities consist of short-term research projects, and/or institution-specific focus, and/or genre-specific focus. We also need to accept a long learning curve. As Clifford Lynch notes, "we need to acknowledge that we don't really know how to do long-term digital preservation."[465] Perhaps instead "in a hundred years the community will really know about preserving over long periods of time."[466] And while "we have never preserved everything; we need to

461. Relatedly, the EU-funded TIMBUS project has been developing a "Legalities Lifecycle Management" (LLM) tool as way to capture and account for all the relevant features of the legal landscape as they apply to digital preservation activities. *Legalities Lifecycle Management*, TIMBUS PROJECT (Feb. 9, 2015), http://bit.do/LegalitiesLifecycleManagement.

462. Thibodeau, *supra* note 49, at 28.

463. Lavoie & Dempsey, *supra* note 84, at 2.

464. Nancy Y. McGovern, *Opportunities for Alignment*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 320, 321 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

465. Lynch & McGovern, supra note 81, at 313.

466. *Id*. at 324.

start preserving something."[467] Hence the necessary focus on a process, and not an outcome.

4. Utilizing Conceptual Tools

As Owen Jones reminds us, "Reality is notoriously impervious to taxonomy."[468] Most of us think about the world largely in conceptual models that have significant consequences for the content of our thoughts.[469] For a subject as complex and abstract as digital preservation, it would be useful for the community of interest to adopt and apply some conceptual frames to make the task of understanding somewhat easier.[470] We must be conscious of course not to let our models override the messiness of reality.[471] That said, a few salient examples include:

- **imagined futures**: The stakeholder community can utilize long-range planning to help reduce uncertainty that can surround current day decisions. This would include "the art of the long view," scenario planning (going forward in time from today), backcasting (going backward from a future end goal),[472] and "accelerated lifetime" testing (envisioning potential changes).[473]
- **viewpoint analysis**: Borrowing from the software community, it may be helpful to identify the various stakeholders so as to view the system from each of their viewpoints.[474]
- **mental models**: As described further below, we should look as well to utilize ideas borrowed from the Internet governance space, in terms of process and stakeholders, as well as fundamental design principles.

---

467. ROSENZWEIG, *supra* note 12, at 20.

468. Owen D. Jones, *On the Nature of Norms*, 98 MICH. L. REV. 2072, 2072 (2000).

469. Whitt, *Adaptive Policymaking*, *supra* note 299, at 537.

470. For further discussion of these concepts, see *id.* at 536-41, 558-67.

471. *See* Pouchard, *supra* note 442, at 180 ("Models tend to report data-related activities in an orderly and linear fashion, which is rarely the case in reality"); They also tend to overlook the diversity of approved practices, reflect the biases of the sponsoring organization, and not be readily adaptable. *Id*. at 471.

472. One example would be the "theoretical ideal" situation ten years in the future. HARVEY, *supra* note 89, at 171.

473. GIARETTA, *supra* note 86, at 267-68.

474. William Y. Arms, *Strength in Numbers*, *in* WHOLE DIGITAL HANDBOOK 108, 108-11 (Diane Kresh ed., 2007). For example, a library is an organization, its technology, and its users. *Id.*

5. Utilizing Metaphors and Analogies

Concrete metaphors and analogies may be useful in helping us grasp the virtual environment of digital information. One potential analogy compares software code with genetic code, and postulates that preserving "society's genome" requires making genetic diversity the cornerstone of a robust digital preservation strategy.[475] Another potential analogy is the role of the Rosetta Stone, discovered in 1799, in deciphering ancient Egyptian texts. Deegan observes that:

> There is a direct analogy with the decipherment of ancient scripts where the knowledge of the language used and the system of coding of the written scripts is lost and must be recreated from scraps of knowledge, intuition, research and other language fragments that may be stems of the ancient script.[476]

We must be careful not to take such an analogy too far, however. Unlike the fixed texts of print data, electronic texts are subject to inadvertent destruction of both the physical medium on which they exist and the intellectual content of their information.[477] Microfilm is predicted to last 500 years.[478] By contrast, all the software from our earliest digital machines is already lost.[479]

It may also be helpful to employ an Internet mentality in thinking metaphorically about digital preservation. For example, the Internet Protocol enables data packets to flow freely between disparate network configurations of software and hardware. What we need is the digital object version of IP; rather than engendering a "network of networks," we should want a thin layer of interoperability to facilitate a "data stream of data streams," or even an "archive of archives."

B. Getting (Better) Organized: Polycentric Governance

As indicated above, the chief management challenge is not to supplant what activity is taking place today, but to help coordinate and expand and deepen those efforts. We also must put digital preservation on a sound footing. This means that the technical, legal, financial, and

---

475. NATHAN THOMPSON, BOB CONE & JOHN KRANZ, SOCIETY'S GENOME: GENETIC DIVERSITY'S ROLE IN DIGITAL PRESERVATION (2016). On the other hand, as Rumsey points out, human culture is far more efficient, flexible, and adaptable, than biology. Rumsey, *supra* note 160, at 23.

476. Deegan & Tanner, *supra* note 1, at 6-7.

477. LAZINGER, *supra* note 38, at 6.

478. Deegan & Tanner, *supra* note 1, at 12.

479. *Id*. at 3-4. Similarly, the life span of electronic information is far shorter than the 50-100 year life span of acid paper, which most libraries refuse to collect. LAZINGER, *supra* note 38, at 10. So, the "slow fire" of acidic paper actually compares favorably with the "fast fire of digital obsolescence." Deegan & Tanner, *supra* note 1, at 3-4.

operational/organizational elements all mix, so that any technical solution answers to the others. The Code, Rules, and Players must work together.

Here are but two examples of the need for a multidisciplinary, multistakeholder approach. Kirchhoff has listed the key elements of a successful digital preservation program: an independent organization with a mission to carry out preservation; a sustainable economic model to support preservation activities over targeted timeframes; clear legal rights to preserve content; relationships with the content owners, and the content users; a preservation strategy and supporting technological infrastructure; and transparency about the key decisions.[480] Gaining general agreement about these core attributes would be a good start.

Further, Thibodeau believes that four criteria should apply to any method for preserving authentic digital information objects: it must be feasible (software and hardware can implement the method), sustainable (applied indefinitely into the future), practicable (within reasonable limits of difficulty and expense), and appropriate (depends on the types of objects to be preserved and the specific objectives of preservation).[481] Our chief task should be translate these elements into an actionable framework.

1. The Multistakeholder Model

What increasingly has come to be called "governance" is just a blend of different institutions and organizations matched to achieve particular objectives.[482] A gamut of choices in these Rules and Players differs by degree of coercion, flexibility, accountability, trust, and formality.[483] Waters has concluded that:

> Government control and private interest, however, are unlikely to be sufficient, or even appropriate in many cases, for preserving the public good in digital archiving." Instead, "groups of people with a common interest in a shared resource will draw on trust, reciprocity, and reputation to devise and agree upon rules for and the means of financing the preservation of the resource.[484]

Or, put differently, we need a third way that does not simply accept the too-easy dichotomy of government and market solutions.

---

480. Amy J. Kirchhoff, *Digital Preservation: Challenges and Implementation*, 21 LEARNED PUB. 285, 288-89 (2008).

481. Thibodeau, *supra* note 49, at 15-16.

482. Whitt, *A Deference to Protocol*, *supra* note 378, at 743.

483. *Id.* at 744.

484. Waters, *supra* note 257, at 85-86.

In recent years there has been much discussion of a different manner of dealing with difficult societal concerns. The multistakeholder model (MSM) extends broadly to a community of interest—governments, civil society, and the private sector—with shared responsibility to wrestle with common issues.[485] The Berkman Center has produced several studies exploring the fundamental elements to successful multistakeholder groups. These include: (1) establishing clear success criteria, both external and internal; (2) setting initial framework conditions of inclusiveness, transparency, accountability, legitimacy, and effectiveness—and (3) adjusting the first two elements continually, based on evolving contextual factors.[486] The authors show that no single model fits all circumstances, and that strong facilitators are key.[487]

Accepting the notion that a greater degree of governance would be useful to better coordinate and collaborate across many institutions and communities of interest, some obvious parallels can be found with the Internet itself. As with digital preservation, the software-derived protocols, standards, and best practices that make up the core of the Internet's operation need continual innovation, revision, and promulgation. For more than forty years, an assortment of technical standards bodies, volunteer organizations, policymaking institutions, and influencers—like the Internet Engineering Task Force (IETF), the Internet Corporation for Assigned Names and Numbers (ICANN) and the World Wide Web Consortium (W3C)—evolved to take on this mission. Because these groups have open participation, employ "bottom-up," transparent processes, and rely on consensus-based approaches to decision-making, they are recognized as "multistakeholder" institutions that act as good stewards of the Internet.[488]

As a general matter, the incorporation of representatives from multiple groups in discussions and decision making can facilitate global governance by bringing in a range of resources and competencies to address common problems.[489] Moreover,

> the fact that technical experts from diverse backgrounds are making decisions about something as vast and complex as the Internet in an open

485. *See* Whitt, *A Deference to Protocol*, *supra* note 378, at 745-748.

486. Ryan Budish et al., *Designing Successful Governance Groups*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y (2015).

487. *Id.* at 3; *see also* Urs Gasser et al, *Multistakeholderism as Governance Groups*, BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y 90-91 (2015) (synthesizing case studies).

488. Cerf et al, *supra* note 434, at 32.

489. Gasser et al., *supra* note 486, at 2.

> and cooperative manner helps to preserve its overall utility. In addition to these questions of technical standards, governance can include conventions for behavioral norms, legal standards of practices, and the protection of users from harm.[490]

While the list of concerns in digital preservation is somewhat different, the concept is the same: governance from the bottom up, involving a richer and fuller set of players.

It should be emphasized that the suggestion is not simply to add a new top-down superstructure to decide, and then dictate, digital preservation practice to the world. That model simply would not work.[491] Rather, the hope is that the preservation community will recognize that it always has been employing a form of multistakeholderism in its various activities, even if it didn't know quite what to call it. That recognition can allow the community to begin taking advantage of the learnings and experiences from other fields, including Internet governance. It is to better harness the bottom-up activities already occurring, and funnel them in more productive ways. As with Internet policymaking, "it is clear that form (and forum) should follow function, not the other way around."[492]

In the Internet space, the IETF is where Internet-related software protocols and standards are introduced and debated. On a parallel track, the Internet Governance Forum (IGF) serves the multistakeholder role of coordination, collaboration, and dissemination of best practices. As one suggestion, perhaps a new, all-inclusive "Digital Preservation Forum" could serve a similar lightweight but effective clearinghouse function for preserving our world digital heritage.[493]

2. Utilizing a Polycentric Approach

A variation on the multistakeholder model is polycentric governance. This is an arrangement to organize political matters in a way that involves, local, national, regional, and international agents and institutions on equal footing.[494] These participants develop and apply shared technical principles, social norms, and rules and practices intended to reach decisions to evolve and use certain shared resources.

---

490. Cerf et al., *supra* note 434, at 32

491. Ward and Freiman note, for example, that top-down, policy-driven, or centralized solutions are unlikely to provide as effective as clear, appropriate, and practical support for the data research community. Ward & Freiman, *supra* note 18, at 266.

492. Cerf et al., *supra* note 434, at 33.

493. Conversely, the issue of digital preservation could be brought to the agenda of future IGF sessions.

494. Whitt, *A Deference to Protocol*, *supra* note 378, at 747.

Nobel Prize-winning economist Elinor Ostrom has championed the concept of using polycentric governance for common pool resources (CPRs). While there are no magic formulas for solving collective-action problems with public goods such as digital preservation, polycentric governance to Ostrom is a more effective way to manage CPRs than what she terms the "monocentric hierarchies" of centralized, top-down government organizations and the competitive market for private goods.[495] Among other favorable attributes, the "philosophy of process" of such resource systems tends to be "decentralized, open, transparent, consensual, and peer-reviewed."[496]

No single organization can adequately archive, preserve, and provide access to all digital materials.[497] And yet, enacting a new governance model based on massive cooperation is no sure thing. After all, "writing and speaking about cooperation are viewed as forms of leadership, while the act of cooperating is not."[498] Applying polycentric governance principles to the digital preservation community yields the following recommended, and doable, action plans.

a. Unite Behind an Expansive Concept of Digital Stewardship

Stakeholders in the digital preservation space share a common objective: ensuring that the future has a rich and robust picture of the past. Some have referred to this broad theme as *digital stewardship*.[499] Among other implications, this means that digital preservation is not merely an isolated process, "but instead, one component of a broad aggregation of interconnected services, policies, and stakeholders which together constitute a digital information environment."[500] Using this wide-angle lens can help broaden and deepen the roster of stakeholders who care about preserving digital materials.

b. Coordinate the Players, Pieces, and Processes

By 2012, it was becoming apparent that distributed preservation, infrastructure, and architectures are emerging as the predominant model for preserving digital materials.[501]

---

495. *Id.*
496. *Id.* at 749.
497. HARVEY, *supra* note 89, at 96-97.
498. Rieger, *supra* note 174, at 45 (quoting Ross Atkinson).
499. Lavoie & Dempsey, *supra* note 84.
500. *Id*.
501. Muir et al., *supra* note 307, at 52.

So, the technical objective should be establishing an interoperable archives, distributed interoperably—much like the Internet famously as "a network of networks." At the same time, the optimal way to preserve information is via community-based efforts.[502] Obviously this presents an organizational challenge to organizing and disseminating information about digital preservation practices.

Coordination is key. Because the imperative to take action to preserve digital materials reappears over and over as the information lifecycle unfolds, Lavoie maintains that "digital preservation cannot be postponed until materials pass into the custody of cultural heritage institutions." Consequently, it is likely that preservation responsibilities will not reside with a single library or archive, but instead will be diffused across many organizations and institutions. This creates a need for "cradle-to-grave" stewardship by a diverse array of stakeholders.[503]

While institutions can take action on their own, even twenty years ago there was "a strong consensus that coordinated strategies and shared resources are essential to achieve broader solutions to digital preservation and enhancing the success of local efforts."[504] Given "the fact that digital preservation is expensive, funding is scarce, and preservation responsibilities are diffuse suggests that data preservation activities would benefit from cooperation."[505]

One worthwhile proposal is for each nation to establish its own high-level administrative point of contact, to coordinate all digital preservation initiatives on behalf of that country.[506] Another is to establish a cooperative project by existing digital archives to preserve specific types of objects (such as government records) from a specific time period (say, pre-1990).[507]

c. Foster Converging Principles and Collaborative Practice

There are some recent hopeful signs that convergence is beginning to occur at the governance level, with LAMs joining together

---

502. Donald J. Waters, *Preserving the Knowledge Commons*, *in* UNDERSTANDING KNOWLEDGE AS A COMMONS: FROM THEORY TO PRACTICE 156-57 (Charlotte Hess & Elinor Ostrom eds., 2011).

503. Lavoie, *supra* note *5*, at 129-30. *See also* Waters, *supra* note 500, at 158-59 (arguing that coordination problems could be alleviated by creating two-sided markets).

504. HEDSTROM & MONTGOMERY, *supra* note 41, at 1.

505. Lavoie & Dempsey, *supra* note 84.

506. WATERS & GARRETT, *supra* note 22, at 40-44.

507. *Id.*

as single cultural entities within each country.[508] But more is required. Democratic self-governance of the "knowledge commons" will require an unprecedented degree of collaboration.[509]

Relatedly, some of this should occur at the level of key common principles, or higher-level fundamental norms. A survey of notable converged principles includes elements like: longevity, minimal interventions, choice, quality, integrity, access, long-term stewardship, scalability, risk management, focus at creation stage, and understanding structures.[510] These principles provide a "meeting place" for a holistic understanding of digital preservation management.[511]

In a field of rapid change and limited common understandings, keeping apprised of the latest developments and engaging with other stakeholders collaboratively are essential.[512] Lynch too articulates the compelling need for a more deliberate collaborative exchange of emerging tools, workflows, and technologies, across national and institutional boundaries.[513]

> Knowledge transfer and technology transfer remain significant challenges. Researchers can do wonderful things in the lab or the test-bed environment, but there is often a huge gap in translating that research into products, services, best practices, and guidelines. Use-inspired research, combined with practitioners' willingness to test research results and implement effective strategies from the research lab, will benefit all of us involved in the challenges and rewards of digital preservation research.[514]

Another proposal is to commission follow-on case studies to identify current best practices in creation, management, storage, and migration paths of digital information.[515]

---

508. Heather Brown, *Convergence and Preservation in the Digital Age*, *in* INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES 683, 683 (2013).

509. Kranich, *supra* note 9, at 106. Such an approach would be based on a common new mission and goals, organizational structure, more comprehensive planning, additional levels of communication, new kinds of authority strictures with dispersed leadership, and shared and mutual control. *Id*. Of course, determining who ultimately is responsible for preserving the knowledge commons is a tricky enterprise. Waters, *supra* note 502, at 145, 147.

510. Brown, *supra* note 508, at 686-87. Per this paper, other important considerations should include interoperability, automated mechanisms, financial viability, and legal compliance.

511. *Id.* at 687.

512. HARVEY, *supra* note 89, at 93-98. Interestingly, the last full action in the Curation Lifecycle Model is "Community Watch and Participation."

513. McGovern, *supra* note 464, at 321.

514. Hedstrom, *supra* note **Error! Bookmark not defined.**, at 37.

515. WATERS & GARRETT, *supra* note 22, at 40-44. As but one example, Miksa has observed that currently there is no common way of documenting and managing scientific research data, creating a serious risk of non-reproducible results. He has proposed adopting data-centric Process Management Plans (PMPs). Tomas Miksa, Stephan Strodl, & Andreas Rauber, *Process*

d. Develop Trust

Perhaps the most important attribute of a digital archive is trust: the notion that it is what it says it is, and that the information stored there is safe for the long term.[516] Authentication of preserved objects also is a matter of trust. Ultimately one must "trust some person, some organization, some system or method that exercises control over the transmission of information over space, time, or technological boundaries."[517] Three levels of trust of digital repositories are the trust of their designated communities of users, of third party providers, and user trust in the documents provided to them by a repository.[518] One suggested approach is to adapt the "institutional guarantee" concept from the world of monetary currency, to increase confidence and trust in the integrity and accessibility of digital information.[519] Another idea is to create a process for certifying digital archives as a means of elevating user trust.[520] Repositories themselves have begun developing systems for audits and certification.[521]

e. Create Permanence

Along with trust, however, is the need for some form of resilience. "The world needs to be wary of depending on institutions whose continued existence cannot be guaranteed."[522] Further, we should not be satisfied with a program that limits itself to "bitstream" preservation, without also tackling enduring access to content in all its context. The top principle for a successful digital preservation program is a long-

---

Management Plans, 9 INT'L J. DIGITAL CURATION 83, 84 (2014).

516. WATERS & GARRETT, *supra* note 22, at 1-2.

517. Thibodeau, *supra* note 49, at 14.

518. *Trusted Digital Repository*, RES. LIBRS. GRP. 1, 9 (2002). Seadle sees a useful tension between trust and distrust in the technical aspects of digital preservation. Seadle et al., *supra* note 278, at 182-83 ("Distrust presents itself as a safer basis than trust for designing systems, as long as that distrust means building in sufficient redundancy to make reasonable allowance for error, accident, external attack, or deliberate internal damage—all of which are known problems."). In such an environment, a useful proposal to engender trust is to create a digital archives certification program. WATERS & GARRETT, *supra* note 22, at 40-44. Audits, certification standards, and even a "Data Seal of Approval" all can help build trust among stakeholders. CORRADO & MOULAISON, *supra* note 12, at 96-107.

519. LIU, *supra* note 74, at 43-48. Such a guarantee necessarily requires the participation of a multitude of stakeholders. *Id.* at 48.

520. GIARETTA, *supra* note 86, at 461-80.

521. For example, the CLOCKSS Archive program at Stanford University has undergone an audit process and certification pursuant to ISO 16363, from which Rosenthal has derived some useful lessons learned. David Rosenthal, *TRAC Audit: Lessons*, DSHR'S BLOG (Aug. 12, 2014), http://bit.do/TRACAuditLessons.

522. Holdsworth, *supra* note 176, at 48.

term commitment that includes continuous lifecycle management.[523] "New ways of organizing the work are needed at a scale never seen before and which, indeed, stretches the boundaries of imagination."[524]

As Hedstrom explains:

> One unique aspect of preservation is its concern with the long term, where 'long term' does not necessarily mean generations or centuries. It may simply mean long enough to be concerned about the obsolescence of technology. In this area, preservation requirements may exceed what information technology vendors typically provide. When long-term preservation spans several decades, generations, or centuries, the threat of interrupted management of digital objects becomes critical. Digital objects cannot be left in an obsolete format and then turned over to a repository after a long period of neglect. This challenge is as much a social and institutional problem as it is a technical one, because for long-term preservation, we rely on institutions that go through changes in direction, purpose, management, and funding.[525]

f. Establish Lofty Goals: Universal Information, Universal Access

Finally, seeking to correct our flawed digital systems may lead us to embrace the full potential of preservation in the twenty-first century. Given the growing attention worldwide, preservation now is on the cusp of becoming a universally endorsed activity. "There is no small irony in the fact that digital technologies forced this hand and may well prove to be preservation's salvation."[526] We can transform our challenges into lofty goals.

One interesting observation is that the analog world has bequeathed us "the haphazard historical gerrymandering of knowledge into institutional collections belonging to communities."[527] One side benefit of the need to archive digital data is the concomitant opportunity not just to preserve information, but to actually enhance the prospects for universal information archives, and universal access. Perhaps a long-lasting societal virtue can be made out of a looming technical necessity. Or, as Michael Lesk puts it, "Digital preservation is not a problem; it is an opportunity."[528]

---

523. Angevaare et al, *supra* note 372, at 92.

524. *Id.* at 90.

525. Hedstrom, *supra* note **Error! Bookmark not defined.**, at 35.

526. Paul Conway, *Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas*, 80 LIBR. QUARTERLY 61, 76 (2010).

527. Clifford Lynch, *Colliding with the Real World*, *in* DIGITAL LIBRARY USE 196, 196 (Barbara P. Buttenfield et al. eds., 2003).

528. Lesk, *supra* note 24, at xv. *See also* Brown, *supra* note 508, at 683 (metaphysical level convergence is providing an opportunity for a new "meeting place" across the "keepers of

VI. SOME POTENTIAL NEXT STEPS

"We can only see a short distance ahead, but we can see plenty there that needs to be done."[529] What follows is a compendium of plausible initiatives to implement in the technical, policy, and financial spheres, as well as possible "deep infrastructure"-enhancing actions by entities in the online tech community. Some of these proposals are more obvious than others; more than a few may be both novel and achievable. All can be mapped usefully along the digital life-cycles/system-layers axes described above.

A. Technical Initiatives (Layers 1-7)

The wide variety of digital formats and applications makes it impossible to select a one-size-fits-all solution for preservation. Indeed, Hedstrom warns us that "the search for the Holy Grail of digital archiving is premature, unrealistic, and possible counter-productive."[530]

Preservation strategies should not be seen as competing with each other, however, but instead are best viewed as different techniques working together. Different kinds of information need various kinds of support, whether via emulation or migration.[531] Perhaps migration fits best for some data sets, emulation for others.[532] The choice of strategy can depend on factors such as the type of digital object, its essential characteristics, the user's requirements, and the institutional priorities.[533] Under one suggested combined approach, for example, one could save the bitstream and metadata, maintain the document's original state using emulation or a virtual machine, and migrate the metadata.[534]

To a certain extent, the focus on researching the various preservation techniques also may be misplaced. It may well be that only

---

collections" and "places of knowledge").

529. Alan M. Turing, *Computing Machinery and Intelligence,* 49 MIND 433, 460 (1950).

530. Margaret Hedstrom, *Digital Preservation: Matching Problems, Requirements and Solutions*, NSF WORKSHOP ON DATA ARCHIVING AND INFORMATION PRESERVATION (Mar. 1999), http://bit.do/MatchingProblems.

531. Waters, *supra* note 257, at 80. To a supporter of emulation, the remaining technical issues are "relatively trivial," while the continuing evolution of digital artifacts makes them less suitable for migration. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 29-30.

532. HARVEY, *supra* note 89, at 163

533. Jeffrey van der Hoeven et al, *Emulation for Digital Preservation in Practice*, INT'L J. OF DIGITAL CURATION, Dec. 2007, at 123, 124.

534. Borghoff *et al, supra* note 25, at 132. The authors claim that the combined approach takes advantage of how the two techniques complement each other well. *Id.* at 131-32.

by actually using preservation strategies for a number of years, can we conclude which ones might work best in practice.[535] Under this approach, documents should be stored now in the most promising formats, and then immediately begin testing preservation strategies systematically.[536]

1. Research Generally

Potential research methodologies cover a spectrum—from theory building to exploratory research, simulations, and experiments. One difference between digital preservation research and research on preserving physical objects is that we can make copies of bits or objects and experiment with them. We can run digital objects through a number of processes and get observable and measurable results. Such experiments would allow researchers to compare the results of different preservation strategies in terms of effectiveness, cost, and user acceptance.[537]

The maturing digital preservation field needs to create greater international alignment on infrastructure and testing.[538] Broadly defined, infrastructure includes all the hardware and software elements necessary to manage digital archiving systems. Potential useful next steps could include designing and implementing common infrastructure testing practices, and initiating benchmarking strategies.[539] Another avenue is creating test beds where researchers create a prototype environment with metrics making it possible to measure the effectiveness of different strategies. This can involve a feedback loop that includes the people managing collections, and the people designing test beds. It is important that these test beds be realistic, including threat model analysis similar to what is utilized in designing a security system.

2. Technical Elements

The community should develop a series of experiments comparing emulation and migration. Researchers could conclude that, for a particular type of digital object, an emulation approach preserves

535. *Id.* at 124.

536. *Id.* at 128.

537. Nicholas Taylor notes the unfortunate chicken-and-egg problem: "We don't know what tools to build, because no research has been done, but the research hasn't been done because we haven't built any tools." Lepore, *supra* note 21 (quoting Nicholas Taylor).

538. Seadle et al., *supra* note 278, at 167-68.

539. *Id.* at 185-90.

specific properties, with particular complications, and costing this amount of money. By contrast, a migration approach to the same material over three format conversions has specific consequences and costs this much. More concrete evidence is necessary, along with an empirical basis for evaluating different preservation strategies, and for deciding which strategy is most appropriate for particular types of resources.[540]

Other proposals include:

- Develop evaluation protocols and benchmarking via pilots and test beds.[541]
- Develop a commodity bit storage locker.[542]
- Focus on more advanced technologies for more complex, non-rendered digital holdings (such as databases, scientific data, and software itself).
- Investigate the concept of emulation as a service (EaaS) in a distributed, cloud-based environment utilizing remote access.[543]
- Define minimum digital preservation requirements necessary to ensure the persistence of digital materials.[544]
- Explore newer virtualization techniques, such as CASPAR.[545]
- Develop an auto-archiving website process (to further the Internet Archive's current remapping of the content of webpages), perhaps through an X-Prize-like mechanism.[546]
- Foster demonstration projects on specific technical challenges, such as emulation algorithms.[547]
- Develop a matrix or hierarchy of selection criteria—a "triage chart"—for digital preservation.[548]
- Establish a "technology watch" to monitor changes in technology indicating when hardware and software are in danger of becoming obsolete.[549]

---

540. Rosenthal for one posits that emulation can become a major digital preservation strategy only after the cost per preserved system image is greatly reduced. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 27-28.

541. McGovern, *supra* note 464, at 322.

542. Lynch & McGovern, *supra* note 81, at 313, 324.

543. Dirk von Suchodoletz et al., *Towards Emulation-as-a-Service*, 8 INT'L J. DIGITAL CURATION, no. 1, 2003, at 131, 131 (2013).

544. Rieger, *supra* note 174, at 41.

545. GIARETTA, *supra* note 86, at 112

546. *E.g.*, *About Us*, ARCHIVE-IT (Nov. 11, 2016), http://bit.do/ArchiveItAboutUs.

547. WATERS & GARRETT, *supra* note 22, at 40-44.

548. Lunghi et al., *supra* note 365, at 222-226; BORGHOFF ET AL., *supra* note 25, at 129.

549. HARVEY, *supra* note 89, at 100-101. As just one close-to-home example, think of

3. Standards Elements

The community also could develop potential OAIS-related standards work. This can focus on interactions with digital archives, such as creating interfaces between archives, ingest methodology used by an archive, ingest of digital data sources to the archive, delivery of digital sources from the archive, submitting digital metadata about digital or physical sources, identifying digital sources to the archive, migrating data across media and formats, and recommended archival practices.[550]

Interestingly, some in the Internet technical community have recognized the need to develop protocols for accessing long-term archiving services.[551] It is not clear whether and how such protocol development has taken place, but sharing insights between the two communities should be encouraged.

Other possible initiatives include:

- Delineate and support interoperability standards.[552]
- Express platform-agnostic digital preservation requirements that move away from the repository-centric worldview.
- Monitor external standards relevant to the digital preservation community.
- Develop defenses against attacks to the security and integrity of cultural heritage institutions.[553]
- Work within the science and medical communities to ensure that software and data should be "open by default," and the complete software/hardware environment represented in scientific and medical work.[554]
- Establish widely distributed caching networks (like Perma.cc) that allow Web links to be captured permanently.

Apple's decision to no longer install CD-ROM players in MacBooks.

550. CONSULTATIVE COMM. FOR SPACE DATA SYS., *supra* note 15, at 1-5. Some have suggested moving beyond OAIS due to its incomplete nature. Giaretta for one believes that conformance with OAIS is a necessary but not sufficient condition because it does not cover all aspects of digital preservation. GIARETTA, *supra* note 86, at 49.

551. *See* C. Wallace, V. Pardesch, & R. Brandner, RFC 4810, *Long-Term Archive Service Requirements*, IETF (2007) (discussing technical requirements and financial and operational considerations of protocols to protect the future integrity of data).

552. McGovern, *supra* note 464, at 321-24.

553. Lynch, *Closing Thoughts*, *in* ALIGNING NATIONAL APPROACHES TO DIGITAL PRESERVATION 309, 313 (Nancy Y. McGovern & Katherine Skinner eds., 2012).

554. *See* Limor Peer *et al*, *Committing to Data Quality Review*, 9 INT'L J. DIGITAL CURATION, no. 1, 2014, at 263, 281 (data curation practices should encompass verifying published research results).

B. Public Policy Initiatives (Layer 9)

1. Wanted: A Comprehensive Policy Framework

How one grapples with possible public policy issues depends largely on whether or not the legal status quo is seen as impeding digital preservation strategies. As discussed above, copyright laws have been the traditional focal point of concerns about digital preservation. Many in the digital preservation community appear to believe that some activities might violate current copyright laws. Rightly or wrongly, that belief can create an environment of inaction, where those same laws will never be directly challenged.[555] Indeed, in some circles, legal uncertainty shrouds preservation activities today, creating a "chilling environment" that threatens current efforts.[556]

Ideally, a comprehensive, unifying public policy framework governing digital preservation would be forged, adopted, and implemented, one that takes into account the nuances of the various challenges. For example, as discussed above, contract law, bankruptcy law, property law, and privacy and data protection laws, all have some relevancy for digital preservation activities.[557] Further, defining the rights and privileges of ownership of digital objects—including forms of access to such objects, and the information and knowledge they contain—should be on that public policy agenda as well.[558] These and other considerations should be analyzed holistically and made part of a broader digital era solution set.[559]

Political acumen suggests, however, that this ideal decidedly is far from becoming reality. Instead, we must content ourselves for now with a more incremental, scatter-shot approach that attempts to address the worst infirmities of our present situation by shoe-horning improvements into existing statutory and case law. Comprehensive frameworks must await another day.

---

555. *See* Knutson, *supra* note 308, at 438 (archivists operate over cautiously to avoid copyright litigation).

556. *Id.*

557. Knutson notes that archivists typically face legal claims based on racketeering (RICO), wire fraud, trespass, conversion, and computer fraud and abuse. *Id.* at 450.

558. Rosenthal insists it is "essential" to clarify the legal basis for building and providing access to preserved digital objects. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 29.

559. As just one example, the Society of American Archivists has identified non-negotiated agreements, such as end user licenses and online terms of service, to pose a large enough problem to digital preservation that no efforts to amend Section 108 should be undertaken without addressing them. *See Statement on U.S. Copyright Office*, SAA, 1, 6 (2016).

2. Dealing with Copyright Law

At least some commentators believe that techniques such as emulation and migration violate one or more exclusive rights under U.S. copyright law. By way of example, Hoeren and others conclude that, if one is conserving data substance, migrating infringes reproduction rights, while converting file formats infringes reproduction rights and adaptation (alteration) rights. If one is conserving services, processes, or program, then migrating software, content, or data infringes adaptation rights, while porting software infringes reproduction rights, and perhaps adaptation rights as well.[560] They state that "As the idea of emulation is to alter the environment and not the preserved data, it is questionable how the process of emulation should be classified with regard to copyright law."[561] If the preserved data is not altered in any way, re-use may have no impact under copyright laws.

That view is not universal, however. Others see more positive trend lines in both the United States and Europe, particularly with regard to the fair use doctrine. The Hathi Trust litigation and related caselaw in the U.S. suggests that some elements of digital preservation qualify as fair use.[562] The European Union has authorized member states to include preservation privileges in their national copyright laws; the *Eugen Elmer v. Darmstadt* case makes it now acceptable for European libraries to make digital copies of books.[563] Nonetheless, enough uncertainty remains that a public policy approach involving a mix of legal and political strategies seems most suitable.

a. Work with Rightsholders

First, how can the digital preservation community offer to work with the so-called "content community"? Despite Rudick's claim that "this is a family quarrel" between two communities that need each other,[564] convincing copyright owners to work with the digital preservation community may take some effective messaging and

560. HOEREN ET AL., *supra* note 303, at 24.

561. *Id*.

562. Authors Guild v. HathiTrust, 755 F.3d 87, 105 (2d Cir. 2014) (library's digitization of books for purpose of long-term preservation is fair use). *See* Lila Bailey, *How Copyright Law is Promoting Cultural Amnesia*, 20 COPYRIGHT AND NEW MEDIA L. 1, 3 (2016).

563. The Court of Justice of the European Union stated that "[a] Member State may authorize libraries to digitize, without the consent of the rightholders, books they hold in their collection so as to make them available at electronic reading points," Press Release, Court of Justice of the European Union, Judgement in Case C-117/13 (Sept. 11, 2014).

564. Rudick, House 2014 Testimony, *supra* note 340.

outreach. In particular, some in the content community may perceive digital obsolescence not as a flaw to be fixed, but a feature to be embraced. After all, selling a single copy of content that theoretically could live on forever in a variety of futuristic incarnations does not appear quite as financially renumerative as leasing a copy of content that must be replaced, over and over, as technological innovation marches on. Further, "many publishers simply do not have the financial incentive, or the institutional stability, to preserve digital materials for decades, let alone centuries."[565] So the incentives for working with digital preservationists may not be suitably aligned.

Still, one must start somewhere. The curators community can create a new mantra: "no value in copyright without effective preservation."[566] Digital preservation can be recast as an engine for innovation, "whether in terms of outputs from reuse of copyright works, or in terms of copyright holders developing new business models, new approaches to marketing their works, and new ways to ensure that their rights are respected."[567] Digital preservation should be seen not as a commercial threat, but as a new marketplace opportunity, and even advantage. Some voluntary options include persuading content owners to (1) preserve the materials in their custody, (2) cede the rights to preserve to another entity; and/or (3) be willing to assume responsibility for preservation, through "escrow repositories" or "archives of last resort."[568]

- **Safe Harbors**: It would be useful to explore establishing "safe harbor" principles about IP rights, which could form the basis for digital archiving agreements among interested parties.[569] This could parallel the safe harbor agreements that have governed the privacy-related actions of US companies operating in the European Union. Menell has suggested a similar safe harbor concept for private companies that work under certain guidelines to preserve and disseminate digital information.[570]

---

565. *Id.* at 124 (statement of James G. Neal). *See also* Reese, *supra* note 85, at 292 (financial incentive to preserve an author's work is not always enough).

566. Charlesworth, *supra* note 311, at 7. Or, more to the point, "Save the Bits!"

567. *Id.*

568. Lavoie & Dempsey, *supra* note 84. Of course, non-voluntary options include enjoining the content owners, directly or indirectly, to preserve the materials in their custody. *Id.*

569. Waters, *supra* note 257, at 91.

570. Peter S. Menell, *Copyright in Context: Knowledge Accessibility and Preservation Policy for the Digital Age*, 44 HOUS. L. REV. 1013, 1064-65 (2007). *See also* Knutson, *supra* note

- **DRM Metadata**: Another avenue is to bless newly incorporated digital rights management (DRM) metadata. Under this approach, "[a] digital archive will have to collect and store any relevant rights management information, which could be stored as part of the descriptive metadata."[571]

- **Licensing Agreements**: Libraries typically rely heavily on commercially-produced content, which, in turn, is licensed with provisions barring its preservation by libraries.[572] Because individual preservation institutions may lack the bargaining power to greatly influence license agreements, model license agreements containing archiving provisions can be useful.[573] Lynch argues for pursuing cooperative agreements as a means of governing and implementing preservation activities.[574] Collective and cross-border licensing also should be investigated, given the national nature of copyright laws.[575] Such collective licensing also is a potentially useful way of dealing in particular with orphan works, where supposed rightsholders cannot be identified.[576] Finally, the industry should examine "clear and concise copyright licensing options," like the Free Software Foundation's general public licenses (GPL), open source software, and Creative Commons.[577]

- **Access to Material**: There is little point in preserving digital material without also giving access to it at some point. "The only real justification for preservation is to provide access."[578] But access creates its own separate legal and public policy questions, however.[579] For example, under what

308, at 472-73 (citing Menell with approval).

571. Michael Day, *Issues and Approaches to Preservation Metadata*, NAT'L PRESERVATION OFF. (1998), http://bit.do/PreservationMetadata.

572. *Preservation and Reuse of Copyrighted Works: Hearing before the Subcomm. on Courts, Intellectual Prop., & the Internet of the H. Comm. on the Judiciary*, 113th Cong. 25 (2014) (Statement of James G. Neal).

573. Muir et al., *supra* note 307, at 52.

574. McGovern, *supra* note 464, at 320-21.

575. Muir et al., *supra* note 307, at 70.

576. McGovern, *supra* note 464, at 321.

577. LEGGETT, *supra* note 63, at 172-73; Charlesworth, *supra* note 311, at 18.

578. Rosenthal, *supra* note **Error! Bookmark not defined.**, at 29.

579. Muir et al., *supra* note 307, at 57-58.

> circumstances would making content accessible be defined under copyright law as publication, performance, or broadcasting.[580] The digital domain creates this new problem "because of the naturally one-to-many relationship of digital content and networked access."[581]

Placing software and hardware in "escrow" would allow preservationists to have access, even if the public cannot. Under an escrow agreement, the software developer agrees to deposit the software's source code for the benefit of the user. If a triggering event occurs—such as insolvency, or lack of continuing support for the software,—the deposited material would be made available to the user.[582] These so-called "dark archives" may be able to address this issue of acceptable archiving, but they also may conflict with the primary purpose of preservation: to provide future access.[583]

b. Operate Under Existing Law

Should content providers be unwilling to work directly with the digital preservation community, another path is to attempt to comply with existing laws, so far as they can be understood.

- **Case-By-Case Analysis**: Each preservation technique triggers a different analysis. One plausibly can argue, for example, that emulation does not violate copyright law. An emulator is simply an interface that allows software and hardware to function together. Reverse engineering for the purpose of emulation, therefore, should be permitted under copyright law.[584] This view is far from universal.[585] Some also argue that the copyright owner has granted an implicit license for webpages to be copied to, and displayed by, a local machine.[586]

  Others conclude that many preservation activities such as copying would meet the four-part "PNAM" test for fair use protection. However, this question cannot be resolved with any

580. Deegan & Tanner, *supra* note 1, at 23.

581. *Id.* at 24.

582. HOEREN ET AL., *supra* note 303, at 171.

583. Deegan & Tanner, *supra* note 1, at 24.

584. Charlesworth, *supra* note 311, at 20-21.

585. *See, e.g.,* Rosenthal, *supra* note **Error! Bookmark not defined.**, at 16-18 (emulation appears to trigger copyright and end user license agreement concerns).

586. Hirtle, *supra* note 345.

certainty outside the fact-specific context of each individual case.[587]

- **Section 117: An increasingly viable option?** Additional elements of copyright law might well be useful. Section 117 is another statutory exception to exclusive rights in the United States, one that permits the user to make an additional copy of a legally purchased copyrighted software program for archival purposes.[588] This particular provision appears to have been largely overlooked in the ongoing disputes over the reach of Sections 107 and 108. Hirtle for one points out that Section 117 allows a copy of legally-purchased software to be made for archival purposes, so long as the file is not shared.[589] He posits that digital information is like a computer program, which can be adapted to run on a new platform without compensating the copyright owner.[590]

  Network evolution may lend further credence to that argument. Rosenthal observes that the Web is evolving from a set of hyperlinked documents to being a distributed programming environment, from HTML to Javascript.[591] As the Web becomes one giant executable software platform, this software-focused provision could be applied to a growing number of online activities.

- **Opt-Out**: Another possible approach is to take an "opt out" stance regarding material from the Web and elsewhere. Archivists essentially would archive and make available content unless and until a rights holder objects.[592] The Internet Archive employs this tactic today. Flipping the default from opt-in to opt-out "dramatically reduces transaction costs and correspondingly expands the scope of the undertaking."[593] On the other hand, utilizing what amounts to a "publish and

587. *See id.*
588. 17 U.S.C §117.
589. Hirtle, *supra* note 345.
590. *Id.*
591. Rosenthal, *supra* note 4.
592. Muir et al., *supra* note 307, at 66.
593. FRISCHMANN, *supra* note 378, at 359.

takedown" stance "diminishes the utility of the archives as a whole by excluding important parts of the Web."[594]

c. Change Existing Law

As "our copyright laws . . . are based on the marketplace and technologies of the eighteenth century,"[595] one solution to the digital preservation challenge is a legislative fix. In particular, this means expanding the reach of the copyright laws to more clearly include all forms of digital objects, as well as digital preservation activities (techniques, functions, and processes) and institutions.

Lobbying for changes to existing copyright law, however, even if a constructive exercise, must be accepted as only a medium to long-term goal. Political reality in the United States suggests that any significant near-term changes are unlikely.[596] Inertia is a primary obstacle to a legislative solution, particularly because "digital preservation is not a legislative priority."[597] To many politicians, long-term issues like digital preservation "are effectively someone else's problem."[598] At minimum, creating and advocating a more robust concept of the "public interest" probably will be necessary to underpin exemptions from copyright.[599]

- **Digital Preservation Exception**: Efforts to recalibrate Section 108 have been ongoing since 2005, with zero results thus far. Despite the fact the Section 108 Study Group Report yielded some concrete (if modest) proposals, with the unanimous support of a diverse group of stakeholders, no legislative changes have been adopted. Obviously consensus "has proved elusive over the past decade," and is "extremely unlikely" going forward.[600]

---

594. Minow, *supra* note 343.

595. WILLIAM PATRY, HOW TO FIX COPYRIGHT 37-38 (2012).

596. *Preservation and Reuse of Copyrighted Works: Hearing before the Subcomm. on Courts, Intellectual Prop., & the Internet of the H. Comm. on the Judiciary*, 113th Cong. 48-49 (2014) (Statement of James G. Neal) (Due to content owner opposition, "any legislative process concerning orphan works, mass digitization, or Section 108 is bound to fail.").

597. Charlesworth, *supra* note 311, at 5.

598. *Id.*

599. *Id.* at 75. Charlesworth claims that the international intellectual property regime has been "captured" by special interests, resulting in a "hollowing out" of the concept of public interest as applied to issues like digital preservation. *Id.* at 5. To the extent this is a true statement, seeking changes to copyright law via the political process may invite an erosion of existing preservation rights under Sections 108 and 107, rather than strengthening them.

600. Letter from University of Virginia Library to The Honorable Bob Goodlatte, US

Some commentators agree that Section 108 requires revisions that expressly include digital preservation.[601] The current version of Section 108 has failed to provide "a useful, clear, and unambiguous exception that practicing librarians can employ . . . ."[602] U.S. Register of Copyrights Maria Pallante goes so far as to insist that Section 108 "must be completely overhauled."[603] In June 2016, the U.S. Copyright Office sought public input on proposed changes to the provision, referring to it as outmoded and "stuck in time."[604] The Copyright Office pointed out "fundamental problems with organization and clarity" in the current Section 108 language, and suggested a thorough redraft based in part on the Section 108 Study Report. [605]

However, this perspective has been far from universal, even in the LAMs community.[606] Perhaps this is because there have been only a handful of cases in the courts citing Section 108 at all, and no instances of a nonprofit library unsuccessfully asserting a Section 108 defense.[607] Interestingly, major groups of librarians and archivist have actively opposed efforts to revisit Section 108, arguing that it is working fine in conjunction with the Section 107 fair use standard.[608] The Internet Archive too questions the need for reform of Section 108.[609] The Copyright Office's seemingly less-than-transparent approach to this inquiry also has sparked controversy.[610]

Congress 1, 4 (Aug. 16, 2016).

601. *See, e.g.,* Knutson, *supra* note 308, at 453-54.

602. David R. Hansen, *Copyright Reform Principles for Libraries, Archives, and Memory Institutions*, 29 BERKELEY TECH. L. J. 1559, 1559 (2014).

603. Pallante, *supra* note 339, at 15.

604. Section 108 Notice of Inquiry, *supra* note 336, at 36595.

605. *Id*. at 36598.

606. *Compare* Rudick, House 2014 Testimony, *supra* note 340 (Section 108 must be revised) *with* Neal, House 2014 House Testimony, *supra* note 572 (Section 108 needs no revision).

607. David R. Hansen, *Digitizing Orphan Works: Legal Strategies to Reduce Risks for Open Access to Copyrighted Orphan Works*, HARVARD LIBR. (Aug. 2016), http://bit.do/OrphanWorksReport.

608. Statement of the Library Copyright Alliance on the Copyright Office's Notice of Inquiry Concerning Section 108 of the Copyright Act, (June 16, 2016) ("LCA specifically urges Congress *not* to address Section 108. . . ."); Society of American Archivists, Statement on U.S. Copyright Office Draft Revision of Section 108: Library and Archives Exceptions in U.S. Copyright Law (Docket No. 2016-4) (July 11, 2016) ("SAA does not consider Section 108 to be obsolete or in need of reform").

609. Lila Bailey, *The Copyright Office is trying to redefine libraries, but libraries don't want it—Who is it for?*, INTERNET ARCHIVE BLOGS (July 27, 2016), http://bit.do/CopyrightOfficeRedefine ("Now is not the time for changes to Section 108.").

610. *See, e.g.,* LCA Statement, U.S. Copyright Office NOI ("LCA is concerned about the

Some modest changes to Section 108 to expand the coverage of digital preservation activities do appear warranted, so long as they are not accompanied with a scaling back in other ways. This means, at minimum, adopting some of the Section 108 Report proposals;[611] these include: (1) covering museums as well as libraries and archives; (2) modifying the three-copies limit; and (3) allowing authorities outside the LAMs to perform at least some outsourced preservation functions.[612]

Other proposals in the Section 108 Study Report would constrict, rather than expand, digital preservation rights. The proposed redefinition of libraries and archives would actually limit the type of players that could participate in digital archiving activities.[613] Basing the rights on a functional or activities-based approach would be far preferable. The Report also proposes allowing LAMs to capture, preserve, and redistribute publicly available online content (also known as "web harvesting"). While on its face this appears to be a reasonable expansion of the existing statute, the limiting language suggests that the new provision would actual curtail existing activities. Fair use may already provide a sufficient basis for website archiving.[614]

Alternatively, the existing language could be left as is (perhaps with minor clarifying modifications) and a new digital preservation/access provision created from scratch within Section 108. This provision would be applicable to all types of digital objects, as well as supporting software and hardware, and the preservation entity's techniques, functions, and processes.

- **Fair Use Standard**: Libraries and archives have come to rely heavily on Section 107, in part because of the inadequacies of Section 108 in the digital era.[615] Another solution posed by

---

lack of transparency relating to this inquiry."); Carrier Russell, *Top Secret, Hush Hush*, THE ALA WASHINGTON OFFICE DISTRICT DISPATCH (July 7, 2016), http://bit.do/TopSecretHushHush ("We have never heard of an instance where a government agency seeking public comments does not provide public access to the comments."); Bailey Blog ("So why is the Copyright Office holding 'hush hush' meetings to 'answer their last questions' before going to Congress with a proposed rewrite . . . ?"). Of course, a closed-door, non-transparent process for discussing and formulating national policy affecting digital preservation does not well represent a robustly multistakeholder approach. *Infra* Part V, above.

611. Pallante, *supra* note 339, at 15; Section 108 Notice of Inquiry, *supra* note 335, at 36598; Reese, *supra* note 85, at 313-314.

612. Section 108 Study Group Report, *supra* note 341, at iii-v.

613. SAA Statement, *supra* note 559, at 2-3.

614. LCA Statement, *supra* note 608, at 1.

615. Rudick, House 2014 Testimony, *supra* note 340, at 25.

> many is to revise the fair use standard in Section 107 so that it more clearly governs digital objects, and preservation activities and institutions.[616]

Many commentators point out that the fair use doctrine is unpredictable, and not always easy to determine.[617] The standard is fact intensive and can be difficult to prove, which makes for an uncertain application to the copying and display of digital archives.[618] "A provision so dependent on balancing and analyzing individual facts and circumstances in specific situations is not well suited to the major [digital preservation] projects . . . ."[619] Some LAMs have attempted to remedy this zone of uncertainty by devising a code of best practices in fair use,[620] which—as if to reinforce the point—was then criticized as overly broad and one-sided by members of the content community.[621]

Nonetheless, Pallante and others suggest leaving Section 107 to the courts to decide. Further codification of the fair use standard is "ill-advised," given the existing rich and comprehensive court-derived jurisprudence.[622] Rudick too proposes using the court-derived fair use doctrine to fill in the gaps where Section 108 does not provide adequate support, in part because the courts have been adopting relatively expansive views of fair use.[623] Archivists then are left with a conundrum: relying on fair use as the best defense for their digital preservation activities, even as that very same provision may act as an unstable legal basis for long-term archiving programs.[624] The quandary may be unavoidable, at least for now.

Absent a thorough and inclusive review leading to a future-proof copyright regime for digital preservation activities, this suggests that

---

616. Muir et al., *supra* note 307, at 48. *See also* Reese, *supra* note 85, at 314 n.191 (relying on fair use alone is uncertain and time-consuming and expensive).

617. Section 108 Notice of Inquiry, *supra* note 336, at 36598.

618. Knutson, *supra* note 307, at 454; Rudick, House 2014 Testimony, *supra* note 340, at 26 (noting that reliance on Section 107 invites "expensive litigation with uncertain results").

619. Rudick, House 2014 Testimony, *supra* note 340, at 26.

620. *Code of Best Practices in Fair Use for Academic and Research Libraries*, ASSOCIATION OF RESEARCH LIBRARIES (Jan. 2012), http://bit.do/FairUseAcademic.

621. In recent congressional testimony, a representative from the publishing industry argued that such a code represents a "one-sided… wish list" that "perpetuates unreasonably broad assertions of fair use." *Hearing on "The Scope of Fair Use,"* ASS'N OF AM. PUBLISHERS (Jan. 28, 2014), http://bit.do/ScopeOfFairUse.

622. Pallante, *supra* note 339, at 28-32.

623. Rudick, House 2014 Testimony, *supra* note 340, at 25-26. On the other hand, proposing legislative changes to Section 107 to make it more digital preservation-friendly likely is a politically risky move.

624. Knutson, *supra* note 308, at 454.

the most sound approach going forward is to adopt a "as best as possible of both worlds" stance. Preservationists would continue relying on case-by-case, flexible, and adaptable Section 107 fair use caselaw, as a backstop to a slightly modified and updated Section 108 provision.[625]

- **Legal Deposit and Registration Requirement**: While legal deposit may have had its origins in the control of intellectual output, today its primary purpose is to preserve the national published heritage.[626] Today's provision is out of date and requires attention.[627] Imposing a mandatory deposit obligation—either initially, or whenever the owner issues the work in a new format[628]—would be the most effective and efficient way to ensure that material is preserved. To that end, case studies could be identified, articulated, and disseminated demonstrating the benefits and impact of legal deposit to different stakeholder groups.[629]

- **Other Legislative Proposals**: Other potential pro-preservation revisions to the U.S. copyright laws include:
    - requiring that content owners authorize reproduction of copyrighted works, as part of a compulsory licensing scheme, for the purpose of digital preservation and access;[630]
    - establishing a legislative "rescue right" premised on eminent domain and abandonment property law concepts, which libraries and archives could invoke to prevent the disappearance of digital assets;[631]

---

625. *See, e.g.,* Section 108 Notice of Inquiry, *supra* note 336, at 36598 (underscoring "the advisability of allowing section 108 and section 107 to co-exist, while ensuring that each provision is positioned for the future, free from the analog restrictions of a bygone era"). Perhaps it also makes sense to separate out the preservation and access functions for purposes of compliance with copyright law. To be more precise, Section 108 could be looked to as providing the digital preservation/curation standard, while Section 107 could be looked to as the digital distribution/access standard.

626. Muir et al., *supra* note 307, at 75-76.

627. Pallante, *supra* note 339, at 31.

628. Reese, *supra* note 85, at 312-303.

629. McGovern, *supra* note 464, at 320. *See also* Menell, *supra* note 570, at 1066-1067 (endorsing a digital deposit mandate).

630. Knutson, *supra* note 308, at 472-473.

631. Alicia Ryan, *Contract, Copyright, and the Future of Digital Preservation*, 10 BOSTON UNIV. J. OF SCIENCE AND TECH. L. 152, 160 (2004).

- clarifying that Section 108 trumps contractual limitations, especially in non-negotiated situations such as online terms of service;
- dealing with the so-called "orphan works" provisions;[632]
- limiting monetary damages for LAMs undertaking digital preservation activities;
- modifying Sections 109 and 117 of the Act to apply the full meaning of exhaustion limitations to digital work;[633] and
- allowing repository contents to be replicated in different geographic locations, much as the Internet's Domain Name Service (DNS) relies on dispersed root servers.

As with many federal legislative proposals, political reality suggests these would be difficult to enact into law, at least in the near term.

3. Give End Users Tools

An interesting take on the legal challenges posed by digital preservation is to assume the end user's perspective. The notion that end users cannot rely on the Web for the long-term care of their content, for example, is slowly seeping into popular consciousness.[634] The fact that users now receive a lesser bundle of rights in the digital than in the analog context is also coming under greater scrutiny.[635] Educating end users about the inherent limitations of possessing and storing digital content could help exert useful demand-side pressure on the public policy arena, and pave the way to rewrite the social contract for digital objects.

One idea is that users could assert product liability concerns about their content residing in the cloud. For example, can an end user sue an

---

632. As Reese points out, the longer a work is preserved, the older it becomes, and the more likely it will be difficult to identify the current rights holder. Reese, *supra* note 85, at 314. Metadata obviously could assist in tracking down the owner.

633. Aaron Perzanowski & Jason Schultz, *Reconciling Intellectual and Personal Property*, 90 NOTRE DAME L. REV. 1211, 1211 (Feb. 1, 2015).

634. *See, e.g.,* Kevin Skobac, *Preventing Digital Memory Loss*, TECHCRUNCH (Jan. 10, 2015), http://bit.do/PreventingDigitalMemoryLoss (users must take back control by backing up content in trusted places); Lafrance, *supra* note 59 ("You can't count on the web, okay? It's unstable. You have to know this."); Broussard, *supra* note 122 ("There is no guarantee that we will be able to read today's news on tomorrow's computers.").

635. PERZANOWSKI & SCHULTZ, THE END OF OWNERSHIP, *supra* note 192 (explaining how the digital marketplace has shifted conceptions of ownership and private property).

entity because its digital objects are not well enough manufactured to survive into the future? Or if the bits themselves are corrupted in some manner?[636] Is "bit rot" a basic flaw in the software, actionable at the Federal Trade Commission? May users seek compensation for not receiving the desired outcome of a long-term sustainable digital object?[637]

Another idea is for users to utilize the common law standard of permissive bailment. At common law, those who held and transported goods on behalf of another entered into an implied contract and engagement regarding those goods. This "holding out" involved a duty to exercise due care when handling the bailor's property.[638] A similar duty could be held to apply in the case of digital content that fails to include a sustainable preservation capability.

4. Undertake Outreach and Advocacy

Reaching out to acquaint current and potential stakeholders with the value of digital preservation is an important but often overlooked activity.[639] It is not clear whether the preservation community is talking to the public about the benefits of preservation and what it means to them.[640] Lynch believes the community must actively engage in a concerted public campaign about digital preservation.[641]

On the advocacy front, Campbell suggests establishing an international preservation body or association that would focus on public policy aspects of digital preservation.[642] Another proposal is to create a national policy to develop "National Information Infrastructure."[643] A third idea is to convince Congress to make a modest appropriation to fund a limited program of symbolic importance, such as preserving all digital records from the 9/11 terror attacks.

A related activity would introduce the digital preservation issue in international trade talks. For example, the World Economic Forum's

---

636. Conversely, having an effective digital preservation program may make it easier for an entity to defend a product liability lawsuit. Hoeven et al., *supra* note 262, at 4-5.

637. Or, alternatively, should entities entrusted with creating or accessing digital content be required to issue a disclaimer that the content is not guaranteed to be readily accessible?

638. Richard Whitt, *Evolving Broadband Policy: Taking Adaptive Stances to Foster Optimal Internet Platforms*, 17 J. OF COMM. L. AND POL'Y 417, 494-95 (2009).

639. CORRADO & MOULAISON, *supra* note 12, at 29.

640. Muir et al., *supra* note 307, at 66.

641. McGovern, *supra* note 464, at 322.

642. Campbell, *supra* note 370, at 29.

643. WATERS & GARRETT, *supra* note 22, at 40-44.

E15 Experts Group on the Digital Economy recently issued a report calling for digital goods and services to be included in future trade negotiations.[644] Digital preservation can and should be on that list of items. The World Trade Organization's Trade-Related Aspects of Intellectual Property Rights (TRIPS) would be another international trade-related advocacy option.

5. Harness the Sway of Government

A final set of steps that bleed over into the Financial Layer is to utilize the government's power of the purse to engineer change.

- **Power of Procurement**: As discussed previously, U.S. Government agencies are struggling to manage the transition to a digital world. This includes in particular electronic records and emails.[645] By some measure, the U.S. Government utilizes some $60 billion annually simply to maintain its legacy computer systems.[646] Those systems can be modified or rebuilt to specifications that encompass the best thinking about digital preservation practices. The U.S. Government could direct, for example, that all responses to RFPs worth at least $100 million must include a viable digital preservation plan.

- **Power of Funding**: The U.S. Government also can help play a role in the long-term storage and access to public data. Currently several federal agencies require that data preservation be included as a component of grant-funding research applications. The National Science Foundation, an early champion for stimulating research on digital archiving,[647] mandates that research grant proposals include Data Management Plans (DMPs).[648] These Plans require the

644. *See* Joshua P. Meltzer, *A New Digital Trade Agenda*, E15 INITIATIVE (Aug. 2015), http://bit.do/NewDigitalTradeAgenda.

645. *See* Jeffrey Zients, *Managing Government Records Directive (M-12-18)*, NATIONAL ARCHIVES (Aug. 24, 2012), http://bit.do/ManagingGovernmentsRecords (summarizing Directive goals and responsibilities).

646. *Federal Agencies Need to Address Aging Legacy Systems*, U.S. GEN. ACCOUNTABILITY OFFICE (May 2016), http://bit.do/FedAgenciesAddressAgingLegacy.

647. *Digital Archiving and Long-Term Preservation (DIGARCH)*, NATIONAL SCIENCE FOUNDATION (Sept. 14, 2004), http://bit.do/DIGARCH.

648. CORRADO & MOULAISON, *supra* note 12, at 71-72. Beginning January 18, 2011, for example, all proposal to NSF must include a data management plan. Kathleen Fear, *You make it, you take care of it*, 6 INT'L J. OF DIGITAL CURATION 53, 55 (2011).

> applicants to attest to, among other things, the period of research data retention, the specific data formats, access and sharing policies, and physical and cyber resources used to preserve and store the data.[649]

The White House also has begun requiring that federal agencies investing in R&D must have "clear and coordinated policies" for providing access to research and scientific data, including storage for long-term preservation.[650] These types of obligations could be extended to all funding for programs over a certain dollar amount. Any such plans also could be required to comply with a comprehensive set of criteria for an effective digital preservation program, such as Process Management Plans (PMPs), or even acquire a data seal of approval.[651] The requirement can extend to the access component as well.[652]

C. Financial Initiatives (Layer 8)

A chief objective of the digital preservation community is to create economic sustainability for its activities, to "support the indefinite persistence of digital preservation systems, enabling access to and use of the information assets into the long-term future."[653] Unfortunately, the current reliance on short-term, project-based funding from governmental bodies "does not support good digital curation practice."[654] Other approaches to the money issue must be explored.

Cost is an obvious impediment. David Rosenthal believes we need nothing less than "a radical re-think of our entire set of digital preservation techniques with the aim of vastly reducing their cost" by

---

649. *See, e.g.,* Data Management for NSF Engineering Directorate Propsals and Awards, National Science Foundation (2014), http://bit.do/NSFDataManagement (describing details of DMPs).

650. Memorandum from John P. Holdgren on Increasing Access to the Results of Federally Funded Scientific Research to Heads of Exec. Dep't. and Agencies (Feb. 22, 2013) (on file with author).

651. Tomasz Miksa, *Process Management Plans*, 9 INT'L J. OF DIG. CURATION 83, 84 (2014).

652. *See* Katherine Bricceno, *Finding a home for scientific data*, AAAS (Sept. 5, 2013), http://bit.do/FindingHomeScientificData (debate on the topic of access to scientific data continues). *See also* Francine Berman & Vinton Cerf*, Who Will Pay for Public Access to Research Data?*, AAAS (Aug. 9, 2013), http://bit.do/PublicAccessResearchData (arguing for a shift in scientific culture to accepting shared data costs, with private companies, academics, and corporate labs investing in data centers and storage systems).

653. CORRADO & MOULAISON, *supra* note 12, at 67.

654. HARVEY, *supra* note 17, at 11.

some four to ten times.[655] He proposes some ideas for reducing ingest costs (preserve content in place), preservation costs (accept it as a losing process, and utilize the cheapest processes), and access costs (combine computation and storage).[656] At bottom, however, the chief challenge is to convince those who hold the purse strings that digital preservation and access are worth the financial investment.

1. The Importance of Establishing Value

Lunghi emphasizes that the clearer the arguments are for the value of digital materials, the easier it will be to win the argument about funding.[657] And the most important concept to argue is that some digital information has implicit enduring value, and/or can be used to create entities that will have such value. "Content gains value by being put to use, regardless of which revenue models support it."[658] It is this long-term value proposition that underpins all other arguments and evidence for engagement in this area.[659]

- **Articulating Additionalities and Externalities**: Non-market valuation techniques in economics estimate "the value or benefit of goods and services conferred on society without the intermediation of markets."[660] These techniques focus on goods and services that are provided through mechanisms other than the usual price-based voluntary market transaction. The common theme is estimating the value of preserving a societal asset, whether intellectual, cultural, or natural, or in analog, digital, or physical form. These techniques include contingent valuation (surveying people about willingness to pay), travel cost models, and hedonic pricing.[661]

  Recent economic literature highlights the importance of developing and applying impact metrics. Traditional financial metrics do not account for the so-called economic "additionalities" that collectively benefit users overall whenever a certain action occurs in a particular market.

655. David Rosenthal, *Talk "Costs: Why Do We Care?"*, DSHR'S BLOG (Jan. 18, 2014), http://bit.do/CostsWhyDoWeCare.

656. *Id*.

657. Lunghi et al., *supra* note 365, at 215.

658. Maron et al., *supra* note 419, at 28.

659. *Id.* at 196-97.

660. Lavoie, *supra* note 5, at 127.

661. *Id*. at 127-28.

> Additionalities can help capture that market impact, whether as inputs (the resources invested), outputs (the improved technical capabilities), or behavioral (responsive actions of other market players).[662] Further, online networks such as the Internet generate enormous positive externalities. These "spillovers" stem from a user's new or improved ability to interact with the rest of the world.[663] Preserving for future access digital artifacts should have a similar spillover effect.

2. The Cost of Inaction

The need to marshal solid evidence speaking to the benefits of expensive new activities such as digital preservation programs has never been greater. Traditional cost/benefit analysis is difficult to carry out, due to the challenge of valuing the long-term preservation of cultural artifacts.[664] Nonetheless, there are ways around this constraint.

Traditional economic theory includes the concept of opportunity cost. Lavoie argues that choosing not to allocate funds for digital preservation will forego major future benefits. Further, while there has been much discussion of the costs of preserving digital material, "relatively little attention [has been] paid to the reverse side of the problem," namely "the costs of not preserving digital materials."[665] Indeed, "failure to look for trusted means and methods of digital preservation will exact a stiff, long-term cultural penalty."[666] As one example, AVPreserve has developed the "Cost of Inaction Calculator" to quantify the financial and cost cost of failing to digitize legacy physical audiovisual media.[667]

---

662. *See, e.g.*, Laurent Bach and Mirielle Matt, *From Economic Foundations to S&T Policy Tools: A Comparative Analysis of the Dominant Paradigms*, *in* INNOVATION POLICY IN A KNOWLEDGE-BASED ECONOMY: THEORY AND PRACTICE (Patrick Llerena and Mirielle Matt, eds., 2005).

663. *See* Richard S. Whitt and Stephen J. Schultze, *The New "Emergence Economics" of Innovation and Growth, and What It Means For Communications Policy*, 61 J. ON TELECOMM. & HIGH TECH L. 217, 278-84 (2009).

664. Lavoie, *supra* note 5, at 123.

665. *Id.* at 107.

666. WATERS & GARRETT, *supra* note 22, at 4. The community also could assemble and make available case studies of digital preservation costs, with benchmark figures and real-life cost scenarios.

667. Chris Lacinak, *The Cost of Inaction: A New Model and Application for Quantifying the Financial and Intellectual Implications of Decisions Regarding Digitization of Physical Audiovisual Media Holdings*, AVPRESERVE 7 (July 2014), http://bit.do/CostOfInaction.

- **Invoking Risk Management**: Couching digital preservation as a proactive exercise in risk management is another way to attract a broader base of financial supporters. Preserving information assets can lower the risk associated with degraded performance and future "unfortunate events."[668]

  A new field is developing that demands more rigorous financial analysis of data management techniques.[669] The 2008 financial crisis demonstrated weaknesses in the quality and management of financial records, information, and data, which led to operational risks that hindered effective risk management.[670] Further, financial bubbles and collapses are in part organizational problems from the way we develop and apply financial technologies.[671] The financial world desperately needs stronger risk management techniques that include viable programs for preserving and accessing data.

  Moreover, various legal obligations to preserve data invite the creation of uniform approaches, including both general rules and sector-specific rules. One potential area to explore is the discovery rules for judicial bodies. Implicitly, entities are required to maintain records in accessible formats to comply with those rules. If so, one can imagine preservation being deemed a part of that requirement.[672] Further, showing that a corporate officer or employee failed to exert adequate control over process or people regarding data could yield the basis for a breach of duty and negligence of the corporation.[673]

- **Invoking Legacy Costs**: What also often is overlooked is the immense waste of present day corporate and government

668. CORRADO & MOULAISON, *supra* note 12, at 68; HARVEY, *supra* note 89, at 83-86.

669. *See generally* Lemieux, *supra* note 118.

670. Lemieux, supra note 118, at 1-2.

671. Alexandros-Andreas Kyrtsis, *Coping with Messiness and Fogginess*, *in* FINANCIAL ANALYSIS AND RISK MANAGEMENT: DATA GOVERNANCE, ANALYTICS AND LIFE CYCLE MANAGEMENT, 155, 156 (Victoria Lemieux ed., 2013).

672. By way of example, the Sedona Conference Working Group on Electronic Document Retention and Production has addressed "Information Governance" for purposes of legal compliance and risk management, including maintaining "the integrity and availability of long-term information assets throughout their intended useful life." The Sedona Conference, *The Sedona Conference Commentary on Information Governance*, 15 SEDONA CONF. J. 125, 129 (2014).

673. HOEREN ET AL., *supra* note 303, at 11.

resources utilized to maintain legacy systems. "At the moment, 60 to 90 percent of IT budget is being spent to keep legacy applications alive simply to provide access to the content and records they contain."[674] Moving these records to a sustainably-preserved environment can avoid many of these needless costs.

- **Invoking the Web's Declining Utility**: Along different lines, there is a viable argument that the total value of the Internet, and the World Wide Web in particular, declines significantly in a world without digital preservation. The open dissemination of and access to information through the Internet plays a crucial role in innovation, economic growth, and countless non-economic benefits like human flourishing.[675] Because of non-rivalry, and increasing returns on ideas, growth in the world's stock of knowledge drives the rate of growth in every country. Ideas create growth and all its emergent benefits. But as that pool of information slowly dries up from incomplete human access, those benefits begin to fade. What does the Internet become without the ability to connect people to meaning?

3. Creating Incentives

There are a variety of ways that the current incentive system can be modified to improve the prospects for wide adoption of digital preservation programs. New incentives for change can be put in place for individual users and what they create, and the content community and what they create. Government bodies in particular can:

- Enact changes to the tax code and accounting rules to favor the preservation of digital information, much like investment in long-term, capital stock.[676]
- Create compliance requirements via public funding bodies.
- Fund a competition for proposals to advance the use of digital archives, focused on removing economic barriers.[677]
- Explore past successful (and unsuccessful) cases where governments attempted to mandate/induce adoption of technology; examples include IPv6 and Adobe PDF-A.

---

674. Jon Tilbury, *Digital preservation in 2016: 5 predictions*, ITPROPORTAL (Dec. 15, 2015), http://bit.do/DigitalPreservationIn2016.

675. Whitt, *Evolving Broadband Policy, supra* note 638, at 436.

676. WATERS & GARRETT, *supra* note 22, at 37-38.

677. *Id*. at 40-44.

4. Exploring New Business Models

The community should study how demand is altered based on various fee arrangements. Applicable services could include storage, software platforms, digitization-on-demand, value-added applications for end users, content enrichment, access, and consulting services.[678] Other approaches would be to:

- Study and promote community-owned solutions.[679]
- Explore opportunities for public-private partnerships.[680]
- Bundle specific and limited forms of access services with the longer and primary responsibility for preservation.[681]
- Organize professional societies to create and finance digital archives.[682]
- Investigate a self-funding set of add-on services (such as Registry of RepInfo, Knowledge Gap manager, provider of authenticity tools, license tool dark archive, brokerage system, and certification system).[683]
- Utilize an insurance model that insures against loss of digital content.[684]
- Aggregate various collections into "union archives," maintained and funded as a shared community resource, which helps distribute the costs of long-term maintenance over a larger stakeholder community.[685]
- Make the case for why at least some digital preservation initiatives constitute a public good, which should be sustained through general public funds.[686]

D. Building Deep Infrastructure at All Layers: A Tech Company Perspective

Many share the view that the primary responsibility for ensuring the long-term preservation of the human record in digital form ought

678. CORRADO & MOULAISON, *supra* note 12, at 79-81.

679. Lunghi et al., *supra* note 365, at 222-226.

680. *Id.*

681. Waters, *supra* note 502, at 162. Of course, this business model runs counter to the broad availability of information as articulated by proponents of A2K and open access, as well as the core mission of most libraries and archives.

682. WATERS & GARRETT, *supra* note 22, at 40-44.

683. GIARETTA, *supra* note 86, at 436.

684. Waters, *supra* note 502, at 149-150.

685. McGovern, *supra* note 464, at 325.

686. *Id.*

to rest with public institutions.[687] That said, "there may be room to explore the topic with some major commercial players in the digitization field . . . ."[688] From the preceding discussion it should be plain that technology companies have a number of potential roles to play in the digital life-cycle/system-layers space.[689] These include:

- A technical support role (such as offering cloud resources for researchers);
- A financial support role (such as funding various research initiatives);
- A certification role (such as crafting and implementing a data "seal of approval" to certify compliance with trust-building principles and guidelines);
- An anchor tenant role (such as becoming customer of a third party's curation, preservation, storage, and/or access services);
- A "best practices" role (such as defining industry-leading standards for building archiving into online platforms and services);
- A convening role (such as sponsoring symposia and other gatherings of experts);
- A lobbying role (such as pushing for changes to copyright laws);
- A public education and outreach role (such as including digital preservation messages in advertising); and/or
- A clearinghouse role (such as serving as informational focal point for cross-border preservation activities).

Another option for tech companies to consider is to narrow the focus to a certain subset of all digital materials, and drill deeply in terms of providing various forms and levels of support.[690]

In addition to carrying out the various functions suggested above—researching, funding, advocating, convening, outreaching—

687. Rosenthal for one questions whether the market would support an emulation offering from a major vendor such as Amazon or Google. Rosenthal, *Emulation*, *supra* note **Error! Bookmark not defined.**, at 27.

688. Lunghi et al., *supra* note 365, at 214.

689. As one example, IBM launched its Long Term Digital Preservation (LTDP) program to help "preserve large amounts of heterogeneous data for very long periods of time or even hundreds of years." *Long Term Digital Preservation (LTDP)*, IBM, http://bit.do/IBMLongTermDigitalPreservation.

690. The category of born digital Web content, for example, is both less well-understood and more difficult in terms of actual technical solutions. Companies born of the Web could decide to devote resources in that particular direction. Complex digital objects also pose some unique technical challenges. Alternatively, tech companies could preserve the stories that often don't get told: of today's small minority groups, yesterday's disappearing cultural heritage. This approach could selectively and deliberately boost the very voices than otherwise might fade into the online mists of time.

companies could take part directly in the digital preservation process itself. This could include: creating new search engine parameters for preserved content; undertaking curation activities; acting as a repository (perhaps with disaster recovery and integrity checking functions); and/or serving as an access "gateway" to preserved digital data. In whatever of these roles the tech company decides to take on, its hallmark could be to step up to the position of a trusted source, pledging to users the benefits of sustainability, integrity, and authenticity of data.

As part of that mission, the tech company could consider a variety of business models, and then develop and implement the ones that prove most promising. For example, if serving as a trusted repository of preserved digital information, the company could assess one-time or annual fees on individuals, companies, and other institutions. Alternatively, the entity could assess the fees on the access end of the process. Or the entity could create a common open source platform for preservation and access, upon which others could create a new ecosystem by building novel applications and services. Perhaps the new program would not even need to rely on fees assessed directly on creators and/or users. The entity certainly could focus on non-traditional metrics like maximizing additionality—the overall positive impact on the digital environment. Ideally, however, the entity would develop one or more novel and sustainable business models.

CONCLUSION: START MAKING SENSE

> "Digital objects last forever—or five years, whichever comes first."[691]

Thanks to the advent of an impressive array of digital technologies, our civilization stands at the brink of a golden era of near-universal access to human knowledge and culture. And yet, thanks as well to those very same innovative technologies, we run the risk of erasing not just that shining future, but the seemingly solid ground beneath our feet. So our looming collective loss extends far and wide.

Twenty years ago, as the Web was just beginning to take over the Wide World, a small band of visionaries delivered a sizable warning shot. The 1996 Report of the Task Force on Archiving of Digital Information was prescient in so many ways. What it lacked, however, was an audience ready to appreciate the severity of the situation, and a

691. Jeff Rothenberg, *Digital Preservation in Perspective: How far have we come, and what's next*?, LINKEDIN SLIDESHARE (Mar. 26, 2012), http://bit.do/RothenbergDigitalPreservations.

path forward that draws its best lessons from collaborative practice. Perhaps we are prepared, finally, to heed its call to action.

Lavoie and Dempsey put it well:

> Preserving our digital heritage is more than just a technical process of perpetuating digital signals over long periods of time. It is also a social and cultural process, in the sense of selecting what materials should be preserved, and in what form; it is an economic process, in the sense of matching limited means with ambitious objectives; it is a legal process, in the sense of defining what rights and privileges are needed to support maintenance of a permanent scholarship and cultural record . . . . And perhaps most importantly, it is an ongoing, long-term commitment, often shared, and cooperatively met, by many stakeholders.[692]

Enough said. The time to act was yesterday.

692. Lavoie & Dempsey, *supra* note 84, at 13.