



DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD
LIBRARY



Toward Environmentally Sustainable Digital Preservation

Citation

Pendergrass, Keith L., Walker Sampson, Tim Walsh, and Laura Alagna. 2019. "Toward Environmentally Sustainable Digital Preservation." *The American Archivist* 82 (1): 165–206

Published version

<https://doi.org/10.17723/0360-9081-82.1.165>

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:40741399>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Toward Environmentally Sustainable Digital Preservation

Keith L. Pendergrass, Walker Sampson,
Tim Walsh, and Laura Alagna

ABSTRACT

Digital preservation relies on technological infrastructure (information and communication technology, ICT) that has considerable negative environmental impacts, which in turn threaten the very organizations tasked with preserving digital content. While altering technology use can reduce the impact of digital preservation practices, this alone is not a strategy for sustainable practice. Moving toward environmentally sustainable digital preservation requires critically examining the motivations and assumptions that shape current practice. Building on Goldman's challenge to current practices for digital authenticity and using Ehrenfeld's sustainability framework, we propose explicitly integrating environmental sustainability into digital preservation practice by shifting cultural heritage professionals' paradigm of appraisal, permanence, and availability of digital content.

The article is organized in four parts. First, we review the literature for differing uses of the term "sustainability" in the cultural heritage field: financial, staffing, and environmental. Second, we examine the negative environmental effects of ICT throughout the full life cycle of its components to fill a gap in the cultural heritage literature, which primarily focuses on the electricity use of ICT. Next, we offer suggestions for reducing digital preservation's negative environmental impacts through altered technology use as a stopgap measure. Finally, we call for a paradigm shift in digital preservation practice in the areas of appraisal, permanence, and availability. For each area, we propose a model for sustainable practice, providing a framework for sustainable choices moving forward.

© Keith L. Pendergrass, Walker Sampson, Tim Walsh, and Laura Alagna.



KEY WORDS

Digital preservation, Sustainability,
Climate change, Appraisal, Permanence, Access

Cultural heritage organizations (CHOs)¹ rely on a stable society to fulfill their missions: from researchers who make use of collections, to funding sources, to collecting activities, CHOs' success is built on the social and physical constructions of a stable society. Climate change threatens this stability. The scientific reality of anthropogenic climate change is stark, and, as Mandy Henk notes when calling for a carbon-neutral library, "presents a very real threat to the world we have built and the world in which our libraries and the institutions they serve have thrived."² The Intergovernmental Panel on Climate Change's 2018 special report underscores the urgency: society must substantially reduce greenhouse gas emissions by 2030 to avoid the catastrophic effects of global warming above 1.5 degrees Celsius.³ The existential threat to CHOs from climate change⁴ requires an urgent switch to sustainability⁵ in all areas of practice, including digital preservation.

Although discussion of sustainability is common in the cultural heritage literature, the term most often refers to financial and staffing resources. When discussing environmental sustainability, efforts tend to focus on reducing organizations' impacts through improvements to the built environment⁶ and on adapting, or reacting, to the effects of climate change. While awareness of the negative environmental impact of digital preservation practice has increased over the last several years, the focus remains on developing sustainable financial models and staff workflows to cope with the increasing scale of digital content.⁷

Digital preservation relies on an interconnected infrastructure known as information and communication technology (ICT). This infrastructure includes items such as personal and handheld computing devices, other items with embedded microprocessors or network connectivity, data centers, cellular towers, satellites, and the networking infrastructure required to connect them all. Thomas Hecker discusses how the digital infrastructure on which academic organizations rely is vulnerable to disruptions in energy and raw material supplies,⁸ yet he does not explore the environmental impact of this infrastructure or how its use may increase the likelihood of future energy or raw material supply disruptions. With a lack of literature showing the complete environmental impact of digital preservation practices, we conducted a literature review focusing on the full life-cycle impacts of ICT components on which digital preservation relies. The review reveals that ICT components, and therefore the digital preservation practices that they enable, have a negative environmental impact throughout their life cycles, contributing to the threats to CHOs against which Henk and Hecker warn.

To move digital preservation toward environmentally sustainable practice, we draw on John Ehrenfeld's framework for creating sustainability. He argues that society needs a paradigm shift in how we think about satisfaction and

fulfilling our needs: more efficiently satisfying our needs through consumption is simply reducing unsustainability; creating true sustainability only occurs by changing the way we, as a society, fulfill our needs.⁹ In the same manner, we argue that using technology to reduce the impact of digital preservation practices, while needed as an interim measure, is only reducing unsustainability. To create environmentally sustainable digital preservation, the profession¹⁰ needs a paradigm shift in how we view and enact appraisal, permanence, and availability of digital content. Expanding on Benjamin Goldman's challenge to current practice for digital authenticity,¹¹ we propose that rethinking these three core areas of digital preservation practice provides a framework for making sustainable choices by integrating environmental sustainability into existing decision-making criteria; that environmentally sustainable digital preservation will come only from critically examining the underlying motivations and assumptions of our practices. Such a shift has been proposed in records management and big data contexts, and the cultural heritage sector as a whole would do well to work toward a similar reevaluation.¹²

While the changes proposed in this article will be a paradigm shift for many organizations and, based on current standards and best practices, for the profession as a whole,¹³ we acknowledge that, due to efforts toward financial and staffing sustainability, they will not be a paradigm shift for all. Rather, for some, they may validate that decisions made out of financial or staffing necessity are the environmentally responsible choices.

The article is organized as follows: first, we ground our discussion of sustainability by conducting a literature review of how the term is used in the cultural heritage literature; second, we fill a gap in the literature by detailing the negative environmental effects of digital preservation practice resulting from its reliance on ICT, with a focus on the full life cycle of ICT components; third, we offer practical options for reducing unsustainability with current technology; and, finally, drawing on Ehrenfeld's framework, we call for a paradigm shift in digital preservation practices for appraisal, permanence, and availability of digital content.

Literature Review

METHODS

We reviewed cultural heritage literature on the term "sustainability" and analyzed the results to determine how the term is discussed. The analysis revealed two major categories: financial and staffing, and environmental.¹⁴ For environmental sustainability, we further divided the results by main topic or argument of the work: adapting to climate change, mitigating the negative

impact of the built environment, and environmentally sustainable digital practice. We do not intend this review to be a comprehensive search of the term “sustainability,” but instead to offer a broad view of the differing ways the literature discusses sustainability. The review revealed a gap in how the impacts of digital preservation are discussed: cultural heritage literature focuses almost entirely on the electricity use of ICT, with some acknowledgment of e-waste. We sought to fill this gap by assessing the impact of ICT throughout the life cycle of its components (from raw material extraction to end of life). We reviewed literature from the fields of industrial ecology¹⁵ and computer science, as well as reports from national governments and international organizations, to establish a scope of ICT and to assemble a more complete assessment of its environmental impact than the cultural heritage literature has given thus far. The discourse on sustainability and digital preservation does not always occur in peer-reviewed works, therefore, we also cite non-peer-reviewed works throughout this article.

FINANCIAL AND STAFFING SUSTAINABILITY

In the cultural heritage literature, the term “sustainability” overwhelmingly refers to sustainability of financial and staffing resources. A lack of resources, whether in funding, equipment, or labor, has been, and remains, a common obstacle in CHOs. As a result, the profession has continually developed, debated, and implemented policies and methods designed to make the most of limited budgets, materials, and staff time; essentially, to do more with less. In her article “Archival Adaptation to Climate Change,” Eira Tansey comes to the same conclusion, noting that “[w]ithin the archives profession, much of the work on sustainability has examined processing, digitization, and digital preservation practices intended to lead to economic sustainability, given the precarious budget and staffing of most archives.”¹⁶

Much of the move toward sustainable finances and staffing in archival practice has been framed by the More Product, Less Process (MPLP) approach proposed by Mark Greene and Dennis Meissner.¹⁷ According to this model, archivists should address large backlogs of unprocessed materials by reducing the amount of arrangement, preservation, and description that they consider necessary before a collection is ready for research access. When faced with a lack of resources, archivists should “own up to the limitations we work under and accept that the golden minimum recommended here (or doing ‘good enough’ rather than insisting on perfection) is all we can realistically accomplish.”¹⁸ Greene and Meissner do not use the term “sustainability,” however, their work can be interpreted as an approach to sustainably allocating staff time and effort.¹⁹

Although the MPLP framework has faced its share of criticism,²⁰ many archivists have applied Greene and Meissner’s proposal to other areas of archival

practice. Indeed, the authors themselves later comment that MPLP “is not about specific processing actions. It is about resource management, whether on a program or enterprise level.”²¹ Greene theorizes that the basic tenets of MPLP could impact the way archivists approach appraisal, electronic records, conservation, and reference.²² Shan Sutton discusses how MPLP can apply to digitization projects, recommending that organizations “critically examine their existing practices in selection, image capture, and metadata creation in search of adjustments that can increase production without compromising the usability of the material.”²³ Similarly, Sophia Lafferty Hess and Thu-Mai Christian apply the broad principles of MPLP to research data curation in an attempt at scalability, but conclude that “building sustainable models to fund data curation will require the data archiving community to articulate the amount of skills, time, and labor that are non-negotiable when a high level of data quality is expected.”²⁴

Hess and Christian’s work reveals an issue that MPLP does not fully address: curating and preserving digital content often comes at significant financial cost. Kevin Bradley acknowledges this issue, stating that for digital preservation,

the word [sustainability] is used to mean building an economically viable infrastructure, both social and technical, for maintaining valuable data without significant loss or degradation. This includes the whole socio-technical composition of the repository, the short- and long-term value of the material, the costs of undertaking an action, and the recognition that technologies do not sustain digital objects: institutions do, using the available technology.²⁵

To address these needs, the report *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information* examines financial sustainability as a guiding principle for digital preservation. Its authors argue that digital preservation requires human and technical, as well as financial, resources and describe the challenges of digital preservation through an economic lens. The authors frame their report around answering three questions: what digital information should be preserved, who should preserve it, and who should pay. By answering these previously unanswered questions, the authors provide a set of principles for achieving financially sustainable digital preservation.²⁶ Yet, even with the detailed framework provided by *Sustainable Economics*, financial constraints remain among the most common digital preservation challenges, according to a 2017 survey of digital preservation practitioners. Although the survey focused on interoperability between local and distributed digital preservation systems, the authors note that “the survey responses and interviews clearly indicate that lack of funding and staffing are significant impediments to building a robust digital preservation program.”²⁷

More broadly than budgetary challenges, sustainable digital preservation often also refers to longevity of a digital preservation program. Many

instructional documents or guidelines on creating and managing a digital preservation program refer to sustainability in terms of maintaining actions over time. For example, the Digital Preservation Network's "Digital Preservation Workflow Curriculum" includes an entire module on sustainability, which focuses on ensuring that a digital preservation project endures beyond the project phase; that the practices developed as part of the project become systematic.²⁸ UNESCO's guidelines for selecting materials for digital preservation refer to sustainability as endurance, instructing CHOs to evaluate the sustainability of materials to assess their "capacity to preserve it for long-term access and use."²⁹ Additionally, case studies on digital preservation initiatives frequently use sustainability to mean longevity: for example, Joseph Williams and Elizabeth Berilla discuss "the implementation of a sustainable born-digital institutional archives plan"³⁰ in terms of longevity; and Adrian Brown and Christopher Fryer describe the utilization of cloud storage at the Parliamentary Archives of the United Kingdom as a move toward "a sustainable digital preservation service," by which they mean a program that can be maintained over the long term.³¹

ENVIRONMENTAL SUSTAINABILITY

Environmental sustainability has been an intermittent focus of cultural heritage professionals since the mid-twentieth century. We categorize this work as adapting to climate change, mitigating the negative impact of the built environment, and more recent work on environmentally sustainable digital practice. Perhaps the most concise review of the history of sustainability in the cultural heritage field is provided in Heidi Abbey's "The Green Archivist," which discusses environmental, economic, and social sustainability in the library and museum fields, making the case that archives have fallen behind them.³² Henk provides a longer but no less clear and direct guide to environmental sustainability in libraries. Her work is equally applicable to museums and archives, and it addresses reliance on environmentally harmful data centers, which we will discuss further in this article.³³

Adapting to Climate Change

The cultural heritage field's early engagements with sustainability include the Special Library Association's (SLA) Natural Resources Division (1968) and Environmental Information Division (1976).³⁴ The American Library Association's (ALA) Social Responsibilities Round Table created a Task Force on the Environment in 1989 and began publishing *The Green Library Journal: Environmental Topics in the Information World* in 1994; it now runs as the *Electronic Green Journal*.³⁵ The ALA Sustainability Round Table, formed in 2013, and ProjectARCC, a group founded

in 2015, have brought librarians and archivists, respectively, together to address climate change issues in the field.³⁶ For the most part, these efforts have focused on adapting to climate change, interest in which, unsurprisingly, crosses geography: Matthew Gordon-Clark and Simon Shurville explore the unique, climate-related vulnerabilities of island ecosystems in the context of Pacific islands archives;³⁷ G. E. Ebunuwele explains the global warming implications for library work in Nigeria;³⁸ and Tansey and T. Mazurczyk et al. explain how climate change will affect North American organizations.³⁹ Globally, historic sites and environments are at risk, as shown in reports and promotional materials by the Centre for Sustainable Heritage⁴⁰ and the National Trust,⁴¹ and by numerous case studies from UNESCO.⁴²

Mitigating the Negative Impact of the Built Environment

Libraries, archives, and museums have hosted research on adjusting the built environment to changing environmental needs. Rebecca Meyer, Shannon Struble, and Phyllis Catsikis review environmental practices such as green building adoption, integrated pest management, and humidity and temperature controls.⁴³ They discuss how to select and use conservation and preservation supplies, while considering factors such as minimum-impact building design, energy efficiency, fund-raising strategies, and social sustainability efforts. Exploring the effects of less stringent environmental controls, Jeremy Linden, James Reilly, and Peter Herzog demonstrate that scheduled daily shutdowns of HVAC systems can produce energy savings for academic libraries and archives without compromising collections,⁴⁴ and Tim Padfield et al. analyze relative humidity and temperature ranges resulting from passive air conditioning in architectural design.⁴⁵ Sarah Brophy and Elizabeth Wylie provide guidance for sustainable design of museum buildings and operations in *The Green Museum: A Primer on Environmental Practice*,⁴⁶ and Padfield et al. in "Low Energy Museum Storage" provide "alternative ways of building museum storage to minimise energy cost and mechanical complexity while actually increasing the durability of the collection."⁴⁷ Richard Kerschner recounts his experience improving efficiency and collection storage conditions in a diverse range of historic buildings in New England by designing preservation systems that work with, instead of against, ambient conditions.⁴⁸ These types of studies suggest that previous advice on environmental controls, such as tight ranges of allowable temperatures and relative humidity, are not necessary for many collections and spaces.

Although several certifications exist for green building construction and operations, with the Leadership in Energy and Environmental Design (LEED) certification arguably foremost among them,⁴⁹ studies cited here demonstrate that certification is not conclusive evidence of environmentally friendly practices

and that many other measures can mitigate collection management's environmental impact.⁵⁰ Furthermore, developments in information technology capabilities and the practices that result from them may have a greater impact than whether or not a storage facility has achieved a green certification. Mark Wolfe investigates how adopting information technology in a modern recordkeeping environment can result in more records for archivists to manage.⁵¹ He applies Jevons' Paradox, an ecological economics concept that explains how greater efficiency of a process leads to increased use of the resources in that process.⁵² Wolfe notes that "paperless office" technology in fact increased the production of paper records and that, as Jevons' Paradox implies, sustainability cannot "passively emerge solely from efficiency improvements to technologies and archival practices."⁵³ New technological efficiencies in record-making practice can increase the amount of material that an organization preserves, leading to a greater net environmental impact through increased use of physical storage.⁵⁴

Environmentally Sustainable Digital Practice

Recent work addresses how digital content management and preservation impacts the environment. Maria Jankowska calls for further investigation of the impact of libraries switching from print to electronic resources,⁵⁵ and Jennifer Poggiali investigates the ethical and environmental impacts of libraries purchasing and using electronic devices.⁵⁶ These studies argue that before purchasing electronic devices or resources, cultural heritage professionals should consider factors such as proper end-of-life recycling, the impact of digital versus analog, and the production chain and labor practices involved in making the product. Taking a longer view, Hecker envisions a severe future given the status quo of libraries and other CHOs. In two articles, he argues that "we are not in a (sustainable) transition from physical formats to digital formats"⁵⁷ and that, in light of an eventual postpetroleum future coupled with more impactful near-term ecological changes, libraries will be unable to maintain their current collecting priorities and digital content.⁵⁸

Thankfully, the environmental sustainability of digital preservation practices has recently received increased attention. A 2013 National Digital Information Infrastructure and Preservation Program partners meeting featured a plenary session on environmentally sustainable digital preservation strategies,⁵⁹ and the *2015 National Agenda for Digital Stewardship* report includes a section on environmental sustainability of digital collections that notes the need for both an interdisciplinary approach to this area as well as basic research and development.⁶⁰ Presentations and panel sessions on the environmental impact of digital preservation are becoming more common. For example, Goldman discussed the energy- and carbon-intensity of digital preservation

practices at the Preservation and Archiving Special Interest Group (PASIG) Fall 2016 Meeting,⁶¹ and Linda Tadic explored the environmental impact of digitally preserving audiovisual magnetic media at the Association of Moving Image Archivists (AMIA) 2015 Annual Conference.⁶² Tadic noted the importance of responsibly recycling and reusing the physical media itself and that large-scale digitization of magnetic media in CHOs will exert massive stress on digital preservation infrastructure. Bethany Nowviskie and Dot Porter introduced the term “graceful degradation” as an appropriate response for digital humanities project preservation in light of climate change at Digital Humanities 2010,⁶³ and Sarah Demb questioned the feasibility of digital archives due to mounting pressures on CHOs from climate change, political instability, and economic inequality at Libraries and Archives in the Anthropocene: A Colloquium, and at the Society of American Archivists’ 2018 Annual Meeting.⁶⁴ Goldman further examines the carbon-intensity of digital preservation practice in his contribution to the Society of American Archivists’ *Festschrift* for Mark Greene.⁶⁵ He shows how the profession’s focus on the authenticity of digital content and trustworthiness of digital repositories has led to resource-intensive digital preservation and offers a challenge to this framework in which “we might consider evolving our theories and practices to embrace acceptable levels of mutability” of digital content. Although increased awareness is a move in the right direction, the recent work on environmentally sustainable digital preservation focuses primarily on the electricity use of technological infrastructure, which addresses only a portion of the full environmental impact.

ENVIRONMENTAL IMPACT OF INFORMATION AND COMMUNICATION TECHNOLOGY

Digital preservation relies heavily on ICT, from storage infrastructure, to networking, to end-user devices. Information and communication technology encompasses the entire communications network and all devices that connect to it, whether to support its functioning (e.g., storage and computational power in a data center) or to use it (e.g., an end-user’s mobile phone).⁶⁶ To show the pervasiveness of ICT in the modern world, Santosh Venkatraman describes it in three tiers. Tier 1 consists of the hardware that makes modern computing possible, from personal computers to data centers. Tier 2 is the telecommunications infrastructure required to network Tiers 1 and 3, and includes components such as routers, cabling, cellular towers, and satellites. Tier 3 consists of devices with embedded processors connected to the network, such as networked home appliances and vehicles with built-in GPS functionality.⁶⁷

As cloud and other networked storage systems become more prevalent, digital storage is increasingly disassociated with its physical impacts by

end-users, leading to greater, and less judicious, use.⁶⁸ This has serious consequences for environmental sustainability, due to the large support infrastructure needed for usable cloud and networked storage, and the raw materials and energy that this infrastructure requires. Cultural heritage organizations' reliance on ICT is increasing as they adopt cloud storage and distributed digital preservation services.⁶⁹ Additionally, discovery and delivery services, which fulfill digital preservation's purpose, further enmesh digital preservation in ICT infrastructure.

Assessing ICT's Environmental Impact Using a Life-Cycle Approach

The most common way to assess the full environmental impact of a product is to conduct a life-cycle assessment (LCA), a process codified by the International Organization for Standardization as ISO 14040.⁷⁰ Although other assessment methods exist,⁷¹ the LCA framework uniquely accounts for all impacts of a product, from raw material extraction through the end of the product's life, by creating life-cycle inventories that document the water, materials, and energy a product uses, as well as any emissions or other wastes released.⁷² An LCA allows one to calculate the full cost of a product based on its environmental impact, not just its purchase price or recurring energy costs. Given the pervasiveness of ICT components and their complex life cycles—each requires raw material extraction and refining, shipping at multiple points, manufacture, electricity and cooling during use, and, finally, disposal—the scope of ICT's environmental impact is staggering.

The impact begins during raw material extraction and refining, which, for many materials, is environmentally damaging and energy intensive. Klaus Fichter and Ralph Hintemann expose the materials needed to build and operate data centers in Germany. To illustrate the diversity of raw materials in ICT components, they describe a common circuit board, which contains over twenty types of raw and processed materials, from plastics to common and precious metals.⁷³ Fichter and Hintemann estimate the total weight of all the material for ICT components in German data centers in 2008 at nearly 110,300 metric tons.⁷⁴ This assessment does not include any elements of ICT outside of data centers. The material intensity of German data centers reveals the magnitude of the problem: When expanded to a global scale and including all network and end-user components that make up ICT, the materials alone have a significant negative environmental impact, before even a single component is powered up. Extracting and refining all of these raw materials results in pollution from fossil fuel use⁷⁵ and decreases in biodiversity due to habitat loss.⁷⁶

Once extracted and refined, raw materials require additional energy for shipping, manufacture, and delivery to be turned into a useful end product.

All of this work is quantified in the term “embodied energy,” which is the total energy required to produce an item, from raw material extraction until an end product is installed by a user.⁷⁷ Depending on the greenhouse gas emissions of the electricity supply used to power the device and the type of device, embodied energy can account for nearly all of a product’s carbon footprint over its life.⁷⁸

During the use phase of ICT, the two main needs are electricity and cooling. Electricity powers all “active” components, such as hard drives, circuit boards, processors, and networking equipment. Primarily needed in data centers, cooling is both energy- and water-intensive.⁷⁹ The high levels of electricity—approximately 7 percent of global use in 2012, with the potential to have exceeded 12 percent in 2017⁸⁰—and water needed to keep the global ICT network running have come under increasing scrutiny in the past decade.⁸¹ Data centers are often located where electricity is cheapest, not where it is cleanest, and, although many data service companies are making efforts to reduce the greenhouse gas emissions associated with their data centers’ energy needs, major improvements are still needed to create sustainability in this area.⁸² Additionally, as projected changes in precipitation patterns due to climate change result in fresh-water scarcity, the large amounts of fresh water needed to cool data centers will likely become more difficult to obtain in the future.⁸³

Information and communication technology components require disposal when taken out of service due to failure or obsolescence. Globally, C. P. Baldé et al. estimate that approximately 45 million metric tons of electronic waste (e-waste) are produced annually, with an expected annual growth rate of 3 to 4 percent.⁸⁴ Only 20 percent of global e-waste is documented as collected and properly recycled, while an estimated 4 percent is improperly disposed of in landfills in higher-income countries, and 76 percent globally is dumped, traded, or recycled in undocumented ways.⁸⁵ These undocumented flows of e-waste often end up in lower-income countries, where workers disassemble components to salvage raw materials, resulting in extremely toxic conditions at e-waste sites in countries such as China, Nigeria, Pakistan, and Ghana.⁸⁶ As those with higher incomes enjoy the benefits of ICT, many thousands of lower-income workers, including tens of thousands of children, suffer the negative health impacts of e-waste disposal: long-term illnesses and death.⁸⁷

Digital Preservation in an ICT Context

Cultural heritage organizations’ use of ICT infrastructure is growing as they increasingly preserve born-digital and digitized audiovisual materials, research data, personal digital archives, websites and social media, and digital records, and as users increasingly request access to these materials. Although the quantity of digital content that CHOs preserve may seem inconsequential

compared to that stored by major data companies, it becomes significant when considered in aggregate. Two recent reports, *Beyond the Repository: Integrating Local Preservation Systems with National Distribution Services* and the National Digital Stewardship Alliance's (NDSA) *2017 Fixity Survey Report*, reveal this scale.⁸⁸ Respondents to the *Beyond the Repository* survey were asked to quantify the digital content that their organizations had collected, using ranges from "Less than 1 terabyte" to "More than 100 terabytes." Converting these ranges to single values suggests that the 156 respondents to this question represent organizations that collectively preserve approximately 5.5 petabytes of unique digital content.⁸⁹ The NDSA *2017 Fixity Survey* data show this to be a greatly conservative estimate. This survey used a wider choice of value ranges, from a low of "0 GB–100 GB" to a high of "More than 500 TB: (Please enter amount)," which provided more accurate values for outlier organizations on the upper end of the range. These data reveal that the seventy-four respondents to this question represent organizations that preserve approximately eighty-one petabytes of unique digital content, or thirty-one times more per respondent than *Beyond the Repository* suggests, with the ten organizations that preserve over one petabyte each accounting for seventy-six petabytes of the total.⁹⁰ While the respondents to the two surveys may not be a representative sample of CHOs globally, mapping the average amount of content preserved per respondent organization to OCLC's estimated count of libraries and museums worldwide gives a rough sense of the global scale of digital preservation by CHOs. OCLC estimates that there are approximately 84,000 academic, national, and special libraries, and 79,500 museums globally as of 2016.⁹¹ Applying the lower per-respondent figure from the *Beyond the Repository* survey to OCLC's count results in a conservative estimate of approximately 5,750 petabytes preserved by CHOs. Yet, as designed by the survey questions, these data do not address redundancy. When one factors in digital preservation's tenet that multiple copies coupled with active management is the best loss-prevention method, our estimate multiplies at least three times,⁹² and energy intensity increases due to frequent fixity checking.⁹³ Additionally, the *Beyond the Repository* survey reveals a high growth rate of preserved content, with many respondents expecting the content that their organizations preserve to double or more in the next year.⁹⁴ While further research is needed for a more accurate estimate of digital content preserved by CHOs globally, one can compare our rough estimate to Cisco Systems' two million petabyte estimate of 2016 global data storage⁹⁵ to get a sense of the relative scale of CHOs' digital preservation activities.

Cultural Heritage Organizations' Response

While our estimate shows that CHOs' digital preservation activities account for a small percentage of global ICT use, the amount of content preserved is significant and growing rapidly, and CHOs have an obligation to critically evaluate digital preservation practice in light of its environmental impact. Arguments that CHOs do not have this obligation because the impact cannot be quantified easily, or because it is a small percentage of global ICT impact, miss the point. Cultural heritage organizations are responsible for the impact of digital preservation activities regardless of the relative magnitude of the impact or its quantifiability. No matter the relative scale, CHOs' digital preservation activities contribute to the negative effects of climate change through the use of ICT infrastructure, and this causation is grounded in rigorous scientific studies and peer-reviewed works. Therefore, CHOs should take ownership of this impact and respond to the consequences of digital preservation practice at a professional, engaged level. This response aligns with CHOs' core mission to preserve the heritage of human society for future generations.

We propose two responses to reduce the environmental impact of digital preservation practice. To address the impacts of ICT throughout the component life cycle, CHOs need to reduce the amount of digital content that they preserve while reducing the resource-intensity of its storage and delivery. To do so, cultural heritage professionals must reevaluate their basic assumptions of appraisal, permanence, and availability of digital content. Paradigm shifts involving many participants are rarely swift or easy⁹⁶—though the high stakes and urgency of climate change require decisive action⁹⁷—but strategies exist to reduce digital preservation's impact through technology that cultural heritage professionals can implement immediately as a stopgap measure. The remainder of this article starts with these immediate strategies to reduce impact and concludes with our call for a paradigm shift to create environmentally sustainable digital preservation.

REDUCING DIGITAL PRESERVATION'S ENVIRONMENTAL IMPACT THROUGH TECHNOLOGY

Common approaches to reducing environmental impact and carbon intensity are to improve efficiency by implementing new systems or changing the use of existing ones, to alter scheduling so that high-energy and high-bandwidth tasks occur at off-peak times, and to switch to clean energy sources. These strategies are relevant for ICT and for digital preservation practitioners. The suggestions for reducing impact through technology given in the following subsections are designed to be practical for all types of repositories—from those that use

local, consumer-grade equipment, to those with in-house digital repositories, to organizations that use cloud storage vendors—and to be implementable directly or in negotiation with service providers.

Efficiency

One way that cultural heritage professionals can use technology to reduce digital preservation's environmental impact is to implement practices that increase energy efficiency. This can mean prioritizing efficiency over performance when selecting new devices and systems, and using them following standard energy efficiency practices, such as shifting to standby power or powering off devices when not in use. Limiting the performance characteristics of processors on local machines and servers can lead to further savings by reducing the electricity required and heat produced, while increasing the service life of the processor. Ensuring that these energy-saving practices are in place both for local devices and those managed by an organization's IT department can reduce the amount of electricity that digital preservation activities require. However, focusing solely on efficiency can reduce performance to an extent that user needs are no longer met. One way to mitigate this concern is to configure hardware for performance under high demand and energy efficiency when demand is low.

In addition to these day-to-day actions, cultural heritage professionals can design more efficient storage systems. One example is Facebook's cold storage system for nonproduction data, which uses a variety of approaches to increase efficiency. While some of these approaches—such as removing all redundant and backup power supplies—may not be practical for, or applicable to, CHOs, others can be implemented. Only powering one hard drive per rack at a time allows for less electricity use and causes a cascade reduction in auxiliary components. Less heat demands fewer fans and cooling systems, while less electricity requires fewer power supply units, and the increased percentage of space per rack used for hard drives results in reduced rack infrastructure. Implementing these changes at Facebook produced servers that use only 25 percent of the electricity of a conventional server, while also reducing the need for, and impacts of, a substantial amount of raw material.⁹⁸ Another potential example of increased efficiency through system design is the use of self-checksumming and self-healing file systems such as ZFS, which may reduce the need for frequent, large-scale fixity checks by building in redundant levels of error detection and correction when data are read and written. We discuss the benefits and tradeoffs of error-correcting technology like ZFS further in the "Paradigm Shift: Permanence" section.

Scheduling

A second major way to use technology to reduce digital preservation's environmental impact is to schedule high-energy and high-bandwidth tasks for off-peak times. Both the electricity grid and telecommunications infrastructure are built to support peak load, meaning that the grid and the network can accommodate the highest load forecast for the year, even if this means that a percentage of assets are nearly always idle. Not contributing to peak load helps to mitigate the need for new infrastructure and avoids the environmental impact that such investment would have. Off-peak times on the electricity grid may also provide a higher percentage of emission-free generating resources, making it possible to reduce greenhouse gas emissions simply by shifting the time of electricity use.⁹⁹ As many of the energy- and bandwidth-intensive tasks of digital preservation are scheduled events, the barriers to shifting to off-peak use are fairly low. For example, instead of running fixity checks or transferring large amounts of data over the network during the day, scheduling these tasks to run overnight takes advantage of times when both electricity and bandwidth demand are low. Some tasks, though, cannot be scheduled, and cultural heritage professionals should continue to prioritize donor relations and meeting user needs when completing tasks for acquisition, such as creating initial checksums and capturing donated materials, and for access, such as rendering or visualization services.

It can also be worthwhile to schedule recurring tasks during seasonal off-peak periods on the electricity grid. For example, on the New England regional electricity grid, demand is highest on hot days during the summer and cold days in winter, when cooling and heating loads, respectively, are high. Furthermore, winter days during cold snaps often have higher emissions per kilowatt hour because a greater percentage of the region's natural gas supply is allocated for space heating, forcing the use of oil and coal to generate electricity. During the spring and fall, however, overall demand is generally low because there is little need for heating or cooling, and emission-free resources make up a larger percentage of the region's electricity generation. Therefore, long-running, high-energy tasks, such as running fixity checks on large amounts of digital content, have the least impact when run during the spring and fall. Additionally, sunny spring and fall days have the highest percentage of solar electricity generation compared to overall demand. When taken together with the high percentage of other emission-free resources during these seasons, it is least impactful to execute short, high-energy tasks during the midday hours in spring and fall in New England.¹⁰⁰ Understanding daily and seasonal variations of the regional electricity grid allows for thoughtful task scheduling that reduces environmental impact without negatively affecting needed preservation

activities and, depending on the electricity rate structure, potentially also reduces financial costs.¹⁰¹

Clean Energy

Possibly the most important way to reduce digital preservation's impact through technology is to make use of clean electricity sources. This can be done in a variety of ways: installing clean generation, such as solar or wind, on site; purchasing electricity or renewable energy certificates from clean generation sources; or selecting vendor-provided facilities and services powered by clean electricity. It is beyond the scope of this article to debate the merits of these options in detail, or to discuss carbon accounting compared to the physics of electricity delivery. However, it is important to note that, if purchasing electricity or renewable energy certificates, one should verify that the electricity or certificates are from the same regional electricity grid as the storage infrastructure and that the utility does not count them to meet its compliance mandates. Making purchases in this way promotes additionality,¹⁰² which ensures that new renewable resources will be added to the regional grid and will help to displace nonrenewable powered resources.¹⁰³ A further point to consider when using a cloud storage provider is the energy load needed for cooling. For vendors that provide a choice on data center locations, it is best to select those in colder climates that allow for ambient air or water cooling, which drastically reduces both the electricity and fresh water needs related to cooling.¹⁰⁴

DIGITAL PRESERVATION PARADIGM SHIFT

Using technology to reduce digital preservation's environmental impact is only an interim measure that will not result in environmentally sustainable digital preservation. A sustainability framework explains that greater efficiencies in the use phase do not result in sustainability.¹⁰⁵ Written by Ehrenfeld in an industrial ecology and design context, the framework is based on society's current social construction of satisfaction, achieved primarily by consumption of goods and services. Ehrenfeld argues that increasing the efficiency of consumerism only reduces unsustainability; to create sustainability, society needs a paradigm shift away from satisfaction through consumption, toward more authentic means of satisfaction that result in physical and psychological well-being.¹⁰⁶

Current digital preservation practice is similar to Ehrenfeld's satisfaction by consumption: organizations focus on the management of digital content, often tracking success based on the total storage size or file count of digital content under management and the implementation of preservation risk

reduction strategies, most commonly frequent fixity checking and redundancy. Digital preservation frameworks and standards reflect the focus on striving for optimal management. For example, the NDSA “Levels of Digital Preservation”¹⁰⁷ four-tier structure outlines a path for enhancement in five functional areas to fulfill digital stewardship responsibilities, with no guidance on when Level 1, 2, or 3 may be sufficient. This can lead to the interpretation that an organization should strive for Level 4 in all cases that resources allow.¹⁰⁸ The international standard ISO 16363—Audit and certification of trustworthy digital repositories—takes this a step further, requiring that organizations receive a rating of 4/4 on all criteria to achieve compliance.¹⁰⁹ While there has been some push-back on these standards—such as the Digital POWRR project, which addresses how to achieve “good enough” digital preservation at organizations that lack sufficient funding or staffing to achieve optimal levels¹¹⁰—most organizations continue to strive for optimal digital preservation without evaluating whether their actions help the organization fulfill its mission. Stephen Abrams notes how ISO 16363 and ISO 14721—Open archival information system (OAIS)¹¹¹—focus on the management of digital content, and how no established metrics exist for evaluating the success of digital preservation. To fill the gap, he recommends adding measures of successful use of preserved digital content to evaluations of digital preservation program efficacy.¹¹²

Building on Ehrenfeld’s sustainability framework, Abram’s reevaluation of digital preservation success, and Goldman’s challenge to current practice for digital authenticity, we propose that creating environmentally sustainable digital preservation requires a paradigm shift in appraisal, permanence, and availability of digital content. Only by reevaluating what is required for successful digital preservation, and shifting to a model where management, successful use, and environmental sustainability are explicitly integrated into decision-making criteria, can the profession create sustainable digital preservation. Our framework balances management, use, and environmental sustainability to find compromises that allow practitioners to focus on high-value materials¹¹³ through a renewed emphasis on critical *appraisal*; reduce the resource-intensity of digital storage and management by rethinking digital *permanence*; and meet user needs in different ways by challenging assumptions about the *availability* of digital content and the need for “always on” digital access infrastructure.

We recognize that every organization operates in a unique context and that there is no universal framework applicable in all situations to achieve environmentally sustainable digital preservation. Rather than a prescriptive set of rules or best practices, the proposed paradigm shift is a guide to reevaluating core digital preservation actions and, as such, is applicable across diverse organizations, regardless of size, type, or funding levels.

PARADIGM SHIFT: APPRAISAL

The importance of appraisal has long been recognized as “the first and defining responsibility of the archivist,”¹¹⁴ and the exponential proliferation of information in the digital era has made appraisal even more essential. As Geof Huth notes in “Appraising Digital Records,” appraisal is necessary to “contain the flow” of digital content to create valuable and coherent collections of historical records.¹¹⁵ Appraising digital content is often resource intensive due to the ease and frequency of replication, arrangement in often complex hierarchies, diversity of file formats, and hardware and software dependencies.

When these challenges are confronted in an environment where staff time is scarcer than digital storage, it can be tempting to appraise digital content in a cursory manner. However, to reduce the environmental impacts of digital preservation, cultural heritage professionals must be selective in their appraisal of digital content, critically examining the content they deem worthy of long-term preservation to ensure that only content with enduring value is permanently retained. Considering the environmental costs of digital preservation as part of appraisal will help CHOs prevent the preservation of truly ephemeral or duplicative material. Of course, accurately appraising future value is a difficult task, with the potential for error regardless of format. The selective appraisal of digital content presented here does not have an inherently greater risk of error than appraising analog materials, and it allows CHOs to focus resources on materials of value in accordance with their organizational missions.

Digital preservation requires that additional criteria be considered alongside existing principles for selection and appraisal.¹¹⁶ Given the environmental impacts of ICT, environmental costs should be among the additional criteria considered when appraising digital content. Cultural heritage professionals should evaluate a potential digital acquisition by carefully considering its environmental costs throughout its life cycle, from the point of transfer to preservation and access, accounting for the storage and computational resources needed for all copies. Certainly, not all digital content requires the same level of preservation. Appraisal can determine whether the content merits the most resource-intensive preservation treatment, minimal preservation actions, or a middle-of-the-road approach. Some CHOs are already using tiered approaches to preservation storage, storing more copies of certain digital content.¹¹⁷ Environmental cost should factor into these types of tiered frameworks, as any reduction in digital storage reduces the environmental impacts of digital preservation.

Cultural heritage organizations should practice selective appraisal across collections and, when possible, across organizations. For digital content, where duplication is common, it can be difficult to identify the record copy. Many

versions of documents with no or only slight differences between them are frequently found in organizational and personal digital archives.¹¹⁸ The ease of retaining all copies should not deter the practice of selective appraisal: just because a document exists in multiple locations within the collection—or across collections—does not mean that every copy is necessary. As has long been the case with analog materials, cultural heritage professionals should be comfortable determining which digital content should be retained and which should be destroyed; metadata or descriptive pointers may be sufficient placeholders for duplicate files. Deduplication of those files deemed to be unnecessary should be part of the appraisal process, further reducing the storage footprint and environmental costs of the acquisition. Additionally, organizations should increase collaboration to reduce duplication. This collaboration can be with creators, who can provide insight on what may not be necessary to preserve, or between collecting organizations, where efforts are already underway for web archiving and software preservation.¹¹⁹ Collaborative collecting may result in researchers using more queries to locate materials, thus creating additional environmental impacts, but efforts promoting cross-organizational discovery and descriptive metadata linkages can reduce this unintended consequence.

Cultural heritage organizations should also implement selective appraisal in workflows and practices throughout the digital content life cycle. For example, digital forensics practice in archives advocates capturing content from digital media via a forensic disk image in many cases,¹²⁰ which naturally leads to considerations for retaining the image. While disk images can be useful for some processing, analysis, and access scenarios,¹²¹ large disk images often contain considerable data with little archival value or privacy and security concerns (e.g., operating system files, personal information, duplicate files, and unallocated space) that can make a disk image inappropriate for preservation and access. Cases for long-term retention of full forensic disk images exist, but may be much narrower than commonly conceived. Cultural heritage professionals should critically examine the case for permanent retention of disk images in their digital forensics workflows. Are technical metadata and log files sufficient to document the capture process and context of the image, or are there strong use cases, such as preserving technical functionality or fulfilling legal requirements, for long-term retention? Instead of retaining disk images as a preservation copy by default, organizations should evaluate the environmental and financial impact of this decision and create policies and procedures for deaccession and destruction.¹²²

The growth of digital content collected by CHOs is due not only to the proliferation of born-digital content or forensic capture methods, but also to technical advancements that allow for higher-quality content creation and digitization. This is especially true for audiovisual content. For example, digitizing

a single minute of film in an uncompressed format at 2K resolution can result in a file up to 26 gigabytes in size, and at 4K resolution, the same minute of film can be more than 100 gigabytes.¹²³ This trend to higher resolution, and the accompanying increase in file size, means an exponential increase in digital storage needs, the environmental costs of which can be addressed through appraisal. Cultural heritage professionals should critically examine standards and practices for file formats of born-digital and digitized content at their organizations. Should every item be migrated or digitized to the highest quality possible? Are there true preservation benefits to high-quality digital surrogates? It is important to recognize, though, that preserving content at a lower quality can have its own environmental costs. For example, compressed formats are more susceptible to data loss and require preserving more copies for the same risk level compared to uncompressed formats. Additionally, the quality of digital surrogates may be deemed insufficient in the future, necessitating redigitization. Appraisal should take into account the costs and benefits of preserving high-resolution digital content, and cultural heritage professionals should elect the quality and preservation file formats that result in the lowest long-term environmental impact while still meeting user needs.

Finally, selective appraisal policies and practices should include regular reappraisal.¹²⁴ Initial appraisal decisions do not always stand the test of time, and, given the continual costs—environmental and otherwise—of preserving digital content, reappraisal should be standard practice. Of course, reappraisal itself generally requires accessing the data and, therefore, results in its own environmental impacts. However, these likely do not outweigh the costs of preserving a larger amount of data indefinitely. Furthermore, cultural heritage professionals can reappraise born-digital and digitized content while performing other tasks that require accessing the data—such as processing, making digital storage decisions, and migrating file formats—thus reducing the impact of reappraisal. Reappraisal should include, among other factors such as use metrics and value, an assessment of the current and ongoing environmental costs of preservation.

PARADIGM SHIFT: PERMANENCE

To maintain the integrity and authenticity of files over time through bit-level preservation, digital preservation models such as ISO 16363 and the NDSA “Levels of Digital Preservation” recommend that repositories record, verify, and periodically reverify checksums of files to ensure that file contents remain unchanged. Additionally, these best practices encourage repositories to maintain several geographically distributed, redundant copies of digital files to ensure that files can be restored from backups in the event of small- or large-scale loss.¹²⁵ Digital preservation systems that strive toward these standards

Table 1. Questions to Guide Environmentally Sustainable Appraisal

Areas for sustainable decisions	Questions	
	Born-digital content	Digitized content
Storage size	• Is the entirety of the digital content within your organization's collecting policies? Is there some content (e.g., personal or system data) that you should not acquire? • Is it necessary to capture and maintain a disk image? What are the use cases? • Do duplicates exist within the collection? If so, can you deduplicate, using metadata or descriptive pointers in the place of duplicate copies?	• Do you have a demonstrated need for digital availability of the analog materials? • What file format(s) are you generating? Would different formats result in smaller storage demands while satisfying preservation and user needs? • Should you digitize material to the highest quality possible, or is lesser quality acceptable?
Capture and analysis	• What technology resources are required to capture, analyze, or arrange the digital content?	• What technology resources are required to digitize the analog materials and conduct quality control of the digital surrogates?
Reappraisal	• How regularly should you conduct reappraisal? • What procedures are in place for deaccessioning in the event that you deem some or all of the digital content not valuable during reappraisal?	

incur environmental impacts through intensive use of ICT infrastructure, thus requiring closer investigation of professional aims and expectations.

Although the energy expense of calculating checksums varies depending on the hash algorithm and storage systems used, calculating checksums is a computationally intensive activity. Checksum verification involves a significant amount of input/output activity between processor and storage devices, and hence energy expenditure:

Checksum generation for on-disk assets requires that data be passed from the storage environment to the processor, which computes the hash based on incoming data and stores it in RAM. As such, during the checksum generation/validation process, every bit of every file that is analyzed must be passed from storage to the machine generating the hashes.¹²⁶

For tape-based storage, verifying checksums means retrieving and powering up tapes, often with the added energy expense of a continually powered tape robot.¹²⁷ Other storage solutions may be more energy efficient: Lots of Copies Keep Stuff Safe (LOCKSS) and the Auditing Control Environment (ACE) are examples of systems that can perform audits “that do not require retrieving the entire document.”¹²⁸ Nonetheless, while necessary to ensure files retain their authenticity over time, auditing files is an active process that can consume substantial resources. Cultural heritage organizations should seek to reduce the

resources used for file auditing to balance the environmental costs of auditing against the need for verifiable fixity logs of their holdings.

The cost of verifying checksums multiplies when organizations enact the best-practice recommendation of keeping multiple copies of Archival Information Packages (AIPs)¹²⁹ in geographically distributed locations. Compliance with Level 2 of the NDSA “Levels of Digital Preservation” involves keeping at least three redundant copies of AIPs; some private LOCKSS networks store as many as seven copies.¹³⁰ Although the total environmental cost of digital preservation storage depends on a number of factors,¹³¹ the cost will increase with each redundant copy of data stored. As discussed, this cost derives from additional energy used for powering and cooling storage equipment, processors, and networking infrastructure, as well as from the manufacturing, transport, and disposal of ICT components. Storage and verification costs are multiplied again in many preservation repositories through format migration at the time of ingest, which generates additional migrated preservation files stored alongside the original files in AIPs. Such overhead may be necessary to ensure ongoing usability of files and their contents as file formats and software become obsolete over time. However, this effort also has the potential to be unnecessary and unsustainable, as in the case of files in formats likely to remain accessible in the medium-to-long term.

Current digital preservation practice assumes a goal of zero change or loss in digital collections over time. This assumption is worth investigating and challenging. David Rosenthal argues that, on a practical level, perfect bit-level preservation is a “myth”: “Current storage technologies are about a million times too unreliable to keep a Petabyte intact for a century; stuff is going to get lost.”¹³² Given the inevitability of loss and the high environmental cost of digital preservation, administrators of digital preservation systems should determine acceptable levels of loss in digital preservation programs. Cultural heritage professionals preserving physical collections have long accepted an inevitable amount of loss in archives due to the effects of time and the inevitability of disaster,¹³³ and, since James O’Toole’s 1989 article, archivists have tended to prefer terminology such as “continuing” or “enduring” over “permanent” to describe ongoing value for records, rejecting “both the absolutism and the idealism that the term ‘permanent’ implies.”¹³⁴ Some degree of loss in digital collections should be accepted as well, particularly if such levels enable more environmentally and financially sustainable digital preservation programs.

As the results of the *Beyond the Repository* survey and the NDSA 2017 *Fixity Survey Report* show, archives and other repositories already apply different levels or degrees of preservation treatment to their collections, depending on the nature of the digital content, organizational priorities, and available resources. As one respondent to the 2017 *Fixity Survey* wrote, “In answering the survey I found it hard sometimes to choose an answer since not all of our digital

content is managed with the same workflows or in the same system. Some of the content is highly managed and other content not so much.”¹³⁵ The reality of budgeting, staffing, and environmental impact is worth embracing, instead of encouraging the costliest, most resource-intensive digital preservation practices for all content. In tiered solutions, the preservation and storage solutions applied to high-value, unique collections differ from those applied to lower-value assets such as digital surrogates of stable analog materials.

Decisions in storage system implementation can have a particularly strong impact on fiscal and environmental sustainability. Potential details to be evaluated include determination of acceptable loss, fixity check methods and frequency, choice of storage technologies, file format migration policies, and the number of redundant copies.

Determination of Acceptable Loss

To determine appropriate storage solutions and fixity check intervals, cultural heritage professionals must determine what level of loss is acceptable in collections under their care. If the environmental costs of perfect digital preservation were null, and repositories had unlimited resources to spend on digital preservation, it would be unnecessary to formalize such limits to resources and ambitions. However, the reality is that intensive preservation actions contribute to climate change and other environmental trends that jeopardize the ability of CHOs to fulfill their missions. Therefore, determinations of acceptable loss constitute a mindful balancing of preservation efforts with the need to limit the effects of climate change.¹³⁶ In Goldman’s words, “in adapting our practices we are not just limiting the carbon impact of our professional activities, we are also limiting the effects of climate change on our holdings and making better use of what limited resources we already have.”¹³⁷

Particular determinations of acceptable loss may depend on factors such as uniqueness, value, and even file formats of digital content in a collection. For example, what would be the effect of a random, infrequent bit flip¹³⁸ or other change in a given collection? Is that effect an acceptable risk? A bit flip in a compressed JPEG file will have a much more pronounced effect in visual rendering to an average user than a similar change in an uncompressed TIFF. Similarly, unique born-digital materials may deserve higher levels of preservation effort than digitized copies of stable analog records that can be redigitized if necessary. Repository staff should evaluate the needs and expectations of stakeholders to determine what level of loss would be acceptable within and across collections to achieve a sustainable budget line and lower energy and raw material usage.

Fixity Check Methods and Frequency

The frequency of fixity checks for stored AIPs in digital preservation systems directly affects energy expenditure. One way to reduce the environmental cost of fixity checks is by taking advantage of storage solutions that increase reliability, such as those that natively incorporate data integrity checks. This additional level of verification enables repositories to responsibly reduce the frequency, and corresponding energy expenditures, of complete fixity checks of their AIP stores. In disk-based storage solutions, we recommend investigation of the use of file systems such as ZFS, which verify checksums at the block level¹³⁹ on read/write and are capable of self-healing.¹⁴⁰ Although further study is needed, Alex Garnett, Mike Winter, and Justin Simpson suggest that, despite the multiple copies of data required by such an approach, file systems such as ZFS may offer computational efficiencies for checksumming and repairing stored data.¹⁴¹ Studies likewise show that the use of error-correcting (ECC) memory in servers can substantially increase the reliability of systems and decrease the chance of erroneous data reads/writes.¹⁴² With tape-based systems, we agree with Tadic that repositories should “perform media-level fixity checks, verifying files only when the media check has an error.”¹⁴³ While specific technologies are likely to become obsolete quickly, we recommend that practitioners stay aware of trends in storage and file system technologies that may improve energy efficiency while maintaining data integrity.

The intervals at which organizations check fixity vary considerably. Responses to the NDSA 2017 *Fixity Survey Report* range from every day, to every five years, to never, with most respondents reporting monthly and/or quarterly checks.¹⁴⁴ The wide range of practices shown in these responses, and the fact that the majority of respondents filled in multiple frequencies and answered “Other” to write in additional details or conditions for fixity checking, indicates a need for standardization and guidance in determining the frequency of checks. While further studies are needed to determine appropriate timing for fixity checks, we recommend scheduling complete fixity checks as infrequently as repository staff find responsible, except for sensitive collections whose threat model,¹⁴⁵ including security threats, warrants more careful verification. When implemented in conjunction with resilient storage systems and timed appropriately with backup schedules, infrequent (such as biannual or annual) checks may be sufficient in many digital preservation system implementations for verifying that AIPs are accounted for and unchanged, and for indicating the need to restore an AIP from backups when files have corrupted or otherwise changed. Scheduling fixity checks less frequently may increase the chance of data loss or corruption; organizations need to balance this risk against other considerations such as determinations of acceptable loss. Increased risk of corruption in the

name of more sustainable practice may be acceptable for some collections and organizations, and unacceptable for others.

Sampling can be another way to reduce the energy use of fixity calculations. If administrators can justify a degree of acceptable loss in digital collections, periodically verifying checksums of a random sample of AIPs, rather than of every AIP in storage, will drastically reduce the processing time and energy expenditure involved in auditing collections. Implementing this practice means selecting a sufficiently large and random sample to increase the likelihood that fixity checks will reveal any underlying issues, such as hardware failures, that require further auditing and investigation.

Choice of Storage Technologies

Choices in storage technologies, whether local storage media, distributed digital preservation networks, or cloud-based vendor services, will influence the total cost involved in preserving digital content. Although local storage reduces the need for networking infrastructure, distributed and cloud-based solutions benefit from efficiencies of scale. We encourage a critical examination of storage technology characteristics—including efficiency, networking requirements, electricity sources, and hardware utilization percentages—before implementing a digital preservation system. When explicitly factoring environmental impact into decision-making criteria, we encourage use of a range of solutions, as appropriate for the content and local needs.

File Format Migration Policies

While format obsolescence is a real risk to digital content, practitioners should not be overly zealous with format migrations, particularly for files in formats not at high risk. Institutions should define local file format policies based on calculated risk assessments rather than relying on the default file format policies implemented in digital preservation software systems. An evaluation by the Bentley Historical Library in 2016 found that, with its default Format Policy Registry settings, Archivematica was performing file format migrations for a number of raster image and video file formats that the Bentley considered to be preservation formats.¹⁴⁶ With many organizations working at the scale of terabytes or petabytes, such mismatches can result in a considerable amount of unnecessary data that must be created, stored, and verified, multiplying the environmental impact.

For certain types of content, practitioners should question whether it is necessary to conduct format migrations during ingest to a preservation

repository, or whether on-demand migrations at the time of access would be sufficient. The LOCKSS technology takes the latter approach to format migration:

... unlike systems that preemptively migrate formats in bulk that are not yet obsolete into formats that are presumed to be less obsolete, thereby consuming processing resources, and store both the original and the migrated copies, thereby consuming storage resources, LOCKSS migrates formats only of individual files, and only when a read's request indicates that migration of that file is necessary. The migrated version is discarded when no longer needed to save on storage. This capability was demonstrated in 2005 but has remained unused in practice because the formats of content preserved in the LOCKSS system are not going obsolete.¹⁴⁷

Not storing additional copies of files in alternative file formats will have a significant positive impact on the amount of data stored, and the environmental cost, of digital preservation. Repositories storing data in more diverse file formats than the LOCKSS system may find a great deal more format obsolescence, justifying higher rates of format migrations. However, as long as open source software utilities for conducting such file format migrations remain available on modern or emulated platforms and such conversions have been tested for their ability to retain significant properties of files, practitioners should question the necessity of format normalization at the time of ingest for all but the highest-risk formats in their care. Of course, in cases where the original file need not be maintained, format migrations that result in smaller files (such as going from uncompressed to compressed formats) may have a positive environmental impact.

Number of Redundant Copies

The environmental costs of digital preservation increase with the number of redundant copies of AIPs stored. Systems with high levels of redundancy should be reserved for unique, high-value assets. We encourage storing as many copies as possible on nearline or offline storage systems, which do not require constant power,¹⁴⁸ using durable storage media that do not need frequent replacement. Organizations should limit redundancy for digital surrogates of stable analog materials.

For all decisions related to permanence, we recommend that cultural heritage professionals review digital preservation policies, donor agreements, and other policy documents for language addressing the permanence of digital content, to ensure that such documents do not bind the organization to unsustainable practices.

Table 2. Questions to Guide Environmentally Sustainable Permanence

Areas for sustainable decisions	Questions
Determination of acceptable loss	• Have you established organizational policies around what constitutes “good enough” digital preservation, and what might constitute acceptable amounts of loss over time? • Have you implemented tiered approaches to digital preservation, where resources are allocated according to the value and uniqueness of materials being preserved? • Do your digital preservation policies and donor agreements allow enough flexibility for you to engage in sustainable digital preservation? What language is used around the effort and resources spent on preservation over time? Does this language allow for any degree of loss over time?
Fixity check methods and frequency	• How often do you run scheduled fixity checks? • Do you run fixity checks during peak or off-peak energy and network hours? • Is it necessary to run fixity checks on all AIPs, or is verifying a sample of AIPs adequate to meet organizational needs? • Are file-based checksums supplemented by other integrity checks, such as natively checksumming file systems or block/media-level hardware checks? If so, can you responsibly reduce the frequency of file-by-file fixity checks?
Storage technologies utilized	• How many copies are in online vs. nearline or offline storage? • On what media do you store digital content? Do these media need frequent replacement? What are the environmental costs of manufacturing, transporting, and disposing of these media? • Have you evaluated the environmental impact and energy sources of vendor and cloud services used?
File format migration policies	• Have you developed a local file format policy? • Have you customized the file format migration policies of digital preservation systems in use at your organization to ensure that they match your local format policy? • For the types of data being stored in your repository, is it necessary to conduct format migrations at the time of ingest? Would format migration at the time of access suffice? • Is it necessary to apply file format policies equally to all materials being preserved? Would migration at the time of ingest be appropriate for some material and not others?
Number of redundant copies	• Have you conducted a proper threat model for your AIPs? • How many stored copies of AIPs are truly necessary, considering the threats (e.g., internal attack, external attack, natural disaster, hardware failure) identified in your threat model? • Is it necessary to retain the same number of copies for all preserved digital content?

PARADIGM SHIFT: AVAILABILITY

Appraisal, acquisition, processing, and preservation efforts typically culminate in access to digital content, both born digital and digitized. As others have observed, casual users and researchers alike often expect this access to take place over the Internet (i.e., online).¹⁴⁹ As with most online resources, many users expect near-constant availability of these materials, thus many CHOs implement infrastructures and policies that can secure this “default” of instant delivery. Exacerbating this effort are unexamined claims on the value

of mass-digitization projects. Cultural heritage organizations can reduce the environmental impact of digital access and delivery by critically examining the justifications for mass digitization, implementing on-demand access strategies, adjusting storage technologies for access, and ensuring timely—but not necessarily immediate—delivery.

Mass Digitization

Given the longer history of digitization projects and the still-emerging toolset for online availability of born-digital resources, it is reasonable to assume that most digital objects available online from CHOs are digitized materials, from textual documents to photographs, audio recordings, and moving images. Over this history, the cultural heritage field has used diverse and evolving strategies to meet the demand for digitized content. Organizations have learned to manage the costs of large-scale digitization, with direct pricing and crowd-sourcing models used to offset costs.¹⁵⁰ Members of the profession have also recognized the difficulties in digitizing complete collections for online access, including labor costs associated with metadata and quality control, accurate capture of context and document markings, copyright or privacy concerns, and addressing “diseconomies of scale,” that is, when the complexity and volume of a digitization project generates higher per-item costs than a smaller project due to additional work in predigitization organization, recording context, and varying formats.¹⁵¹ Growing acknowledgment of financial costs of, and challenges in, large-scale digitization projects is a positive trend for the financial sustainability of organizations, but more effective funding models and better appreciation of project difficulties do not address the environmental impact of wholesale digitization.

The notion that CHOs should digitize collections because of a general perception of value or interest, or should strive to digitize all analog materials, ought to be abandoned. Instead, organizations should aim to digitize specific items when requested by users—or those that are at risk from a preservation perspective—while also working to bolster a collection’s discoverability. While further research is needed to compare the relative environmental impacts of physical and digital storage spaces to determine if preservation digitization of at-risk materials can reduce overall impact when the originals are discarded, focusing on improving discoverability and providing digitization and delivery on demand can greatly reduce the environmental impact of digitization programs by slowing the growth rate of digital surrogate storage.

On-Demand Access Strategies

OCLC's 2011 *Scan and Deliver* report provides a concise description of a tiered strategy for on-demand digitization, articulating three broad "tracks" for digitizing and providing access to materials.¹⁵² Simplest among these is a basic scan with no metadata sent directly to the user. The organization does not retain a copy. A middle track allows basic metadata, such as a title, and the organization may retain the digital file in a simple directory structure for future reference. A final track features the most extensive treatment, with a full archival scan, metadata profile, and placement in a content management system. Notable here is the tactic of evaluating the user's need and implementing the access method that fulfills the request, while keeping organizational commitment and resources required to a minimum. We find much to recommend in this general approach. For example, the first tier—a basic scan sent directly to the user—relieves the organization of providing any storage resources or digital preservation management, while allowing the user access to the materials requested. Even if requested several times, multiple low-resolution scans of a single item are not likely to approach the resource cost of a single archival scan stored in a digital preservation repository with full metadata and quality checking, and frequently requested items can be moved to the second tier and saved on a local drive to mitigate the cost of rescanning. An important caveat for on-demand digitization policies is that CHOs must clearly articulate them to potential researchers to spare them unnecessary trips to the reading room to view materials that could have been digitized. In these cases, the transportation-related environmental impact is likely to far outweigh the impact of digitization and storage.

Adjusting Storage Technologies for Access

Tiered storage systems support reducing the environmental impact of digital delivery. Any single item in a cultural heritage organization does not typically see frequent use, but it is vital that the resource be provided when requested.¹⁵³ The costs of not meeting this request can be high, including damage to an organization's reputation. The costs of not providing this resource *immediately*, however, are much lower. Most users of cultural heritage materials will likely abide some delay in digital delivery, provided that the delay and the reason for it are well articulated. This usage trend—infrequent but vital resource requests—can align with a lower-cost storage infrastructure. For example, a nearline storage system—one that features an automatic retrieval of a digital object, but requires some intervening time for that resource to be recalled—keeps storage media powered down when not in active use, reducing cooling and electricity

costs. While tape libraries were the first nearline storage systems, more recent approaches, such as massive array of idle disks (MAID), allow disk drives to spin down when not in use.¹⁵⁴ Such systems attempt a compromise between performance and power costs; a 2002 study found significant energy savings (15%) compared to a similar cluster of always-on redundant array of independent disk (RAID) devices.¹⁵⁵ Technologists also have tested systems that do not use spinning disks, including the “fast array of wimpy nodes”¹⁵⁶—to secure broad read access for mass data searches and computations—and “durable array of wimpy nodes”¹⁵⁷—which orients the same components toward archival storage. The up-front costs and shorter budgeting horizons of most organizations introduce difficulties in adopting these strategies, however. Nevertheless, storage systems exist that strike a balance between using spinning disk and tape media for the retention of archival copies, and CHOs could help drive additional demand for these types of storage configurations. Existing or new CHO consortia may be best positioned to use or develop these strategies, particularly in the case of smaller organizations.

Considerations for Timely Delivery

For some types of digitized materials, such as still images, audiovisual content, or textual documents, access copies can be generated when a user requests them, further alleviating storage requirements. While encoding access copies for audiovisual content may seem time consuming from an instant delivery mindset, the storage savings from foregoing permanent access copies of these file formats is also the greatest. Born-digital content, however, may require additional considerations to ensure timely delivery. Archival Information Packages of born-digital content are often composed of many small files, with a single AIP potentially containing many thousands of files. Whether or not Dissemination Information Packages (DIPs, or access copies) are created on demand, it is imperative that a digital object recalled for a user be complete. Using nearline or offline arrayed storage solutions in the recall process can be more complicated for these types of packages than recalling a single file that correlates to an item found in a finding aid or catalog. Nevertheless, this is achievable, as the packaging capabilities of modern operating systems are well equipped to bundle and deliver content.

Born-digital content may also have external dependencies, such as software libraries, emulation frameworks, or graphic sets. Depending on copyright concerns or technical requirements, fulfilling these dependencies by delivering additional files with the digital object may not be possible. An organization committed to fulfilling these dependencies through its delivery systems would

need to coordinate the use of shared organizational resources with the delayed delivery of the digital object.

Table 3. Questions to Guide Environmentally Sustainable Availability

Areas for sustainable decisions	Questions
Digitization	• Is there a specific or demonstrated need for digitization of an entire collection? If not, how can you alter the scale of the digitization project to meet user needs? • If there is a demonstrated need for digital access to particular items in a collection, what digitization and access methods will most immediately serve the user while keeping organizational commitment to a minimum? • Are you clearly articulating on-demand digitization policies to users to avoid unnecessary trips to view analog materials?
Access storage	• Is it necessary to migrate content to an access copy? If so, is it necessary to do so prior to a request for the content? • Are you retaining access copies in a storage system that facilitates media powering down when not in use?
Delivery	• If you are storing access copies in a system for delayed delivery, do you have language to indicate this to users? • If a user requests born-digital content or content with component files, can the retrieval system bundle this content appropriately? • Is there any supplemental documentation you should provide to the user with the content (e.g., how to access a particular file format)?

Conclusion

The recommendations put forward here are in the context of current scientific work on climate change and the environmental impact of ICT. While society may fully embrace true sustainability—reducing consumption and habitat degradation while rapidly transitioning to clean energy—to the levels required to avert climate disaster, it may also follow a more “business as usual” approach, leading to Hecker’s severe vision for academic institutions. We believe that CHOs should be part of the solution that leads society to the first outcome, which will require a switch to sustainability in all areas, including digital preservation.

The profession has an important role to play in both advocating for and practicing environmentally sustainable digital preservation. In this article, we expand on recent work in the cultural heritage literature that shows ICT’s negative environmental impact during the use phase, revealing that the impact is much greater when one considers the full life cycle of ICT components. The

literature on these impacts is substantial and peer reviewed, and the links between these impacts and climate change and biodiversity loss are scientifically rigorous. It is time for all cultural heritage professionals who work with digital content to engage with this urgent issue and to critically evaluate current practices in appraisal, permanence, and availability of digital content to create environmentally sustainable digital preservation.

NOTES

¹ We use “cultural heritage organizations” to include more than libraries, archives, and museums, referring to all organizations that preserve our cultural heritage.

² Mandy Henk, *Ecology, Economy, Equity: The Path to a Carbon-Neutral Library* (Chicago: ALA Editions, an imprint of the American Library Association, 2014), 4–5.

³ To keep global warming at or below 1.5 degrees Celsius, anthropogenic greenhouse gas emissions must decline by approximately 45 percent by 2030 and continue to decline to zero by 2050. See Intergovernmental Panel on Climate Change, *Global Warming of 1.5°C: An IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emissions pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* (2018).

⁴ It is well known that armed conflict has negative to disastrous effects on CHOs. Climate change will have similar destabilizing effects on society, such as resource scarcity and a renewed focus on sustenance, increased frequency and severity of natural disasters, large-scale human migrations, and land degradation leading to poor crop yields. As these effects intensify, it will become more difficult, if not impossible, for CHOs to fulfill their missions. For the effects of armed conflict, see UNESCO, “Armed Conflict and Heritage,” <http://www.unesco.org/new/en/culture/themes/armed-conflict-and-heritage>; and Simon Lambert and Cynthia Rockwell, eds., *Protecting Cultural Heritage in Times of Conflict* (Rome: International Centre for the Study of the Preservation and Restoration of Cultural Property, 2012). For the effects of climate change, see Intergovernmental Panel on Climate Change, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (New York: Cambridge University Press, 2014); and Intergovernmental Panel on Climate Change, *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (New York: Cambridge University Press, 2014).

⁵ Sustainability has become a well-known concept in modern society and a goal for many individuals and organizations. A frequent definition is that given in the United Nations report *Our Common Future*. Often simply called the Brundtland Report after the commission’s chairman, it defines the term “sustainable development” as development “that meets the needs of the present without compromising the ability of future generations to meet their own needs.” To enact sustainable development principles, many organizations are using a triple bottom line approach that measures three areas of sustainability: social, environmental, and financial. See Gro Harlem Brundtland et al., *Our Common Future, From One Earth to One World: An Overview by the World Commission on Environment and Development* (United Nations, World Commission on Environment and Development, 1987), 3.27; and John Elkington, “Enter the Triple Bottom Line,” in *The Triple Bottom Line: Does It All Add Up?*, ed. Adrian Henriques and Julie Richardson (London: Earthscan, 2004).

⁶ The built environment is any human-made or -modified structure. See United States Environmental Protection Agency, “Basic Information about the Built Environment,” <https://www.epa.gov/smm/basic-information-about-built-environment>.

⁷ We consider any digital content that receives digital preservation actions as in scope for this article. For many organizations, this includes born-digital and digitized text, image, and audiovisual

content; research data; personal digital archives; digital records; and web and social media archiving.

- ⁸ Thomas E. Hecker, "The Post-petroleum Future of Academic Libraries," *Journal of Scholarly Publishing* 38, no. 4 (2007).
- ⁹ John R. Ehrenfeld, *Sustainability by Design: A Subversive Strategy for Transforming Our Consumer Culture* (New Haven, CT: Yale University Press, 2008); and John R. Ehrenfeld and Andrew J. Hoffman, *Flourishing: A Frank Conversation about Sustainability* (Stanford, CA: Stanford Business Books, an imprint of Stanford University Press, 2013).
- ¹⁰ We use the term "profession" to refer to those in the cultural heritage field who are involved in digital preservation or work with digital content.
- ¹¹ Benjamin Goldman, "It's Not Easy Being Green(e): Digital Preservation in the Age of Climate Change," in *Archival Values: Essays in Honor of Mark Greene* (Chicago: Society of American Archivists, forthcoming in 2019).
- ¹² Geert-Jan van Bussel, Nikki Smit, and John van de Pas, "Digital Archiving, Green IT and Environment. Deleting Data to Manage Critical Effects of the Data Deluge," *Electronic Journal of Information Systems Evaluation* 18, no. 2 (2015): 12.
- ¹³ See the "Digital Preservation Paradigm Shift" section for a discussion of how current standards push digital preservation toward more resource-intensive practices.
- ¹⁴ Compare with the three commonly described areas of sustainability: social, environmental, and financial. See Elkington, "Enter the Triple Bottom Line."
- ¹⁵ Industrial ecology is an approach to product and process design that views industrial systems in the context of the ecological systems on which they depend. This often includes a study of resource and energy flows throughout process life cycles in an attempt to optimize resource and energy use.
- ¹⁶ Eira Tansey, "Archival Adaptation to Climate Change," *Sustainability: Science, Practice, & Policy* 11, no. 2 (2015): 48, <https://doi.org/10.1080/15487733.2015.11908146>.
- ¹⁷ Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68, no. 2 (2005).
- ¹⁸ Greene and Meissner, "More Product, Less Process," 255.
- ¹⁹ Indeed, both Tansey and Goldman make this connection. Tansey notes that techniques that consider limited institutional resources "may be construed as a sustainability response, albeit from a labor and resource-allocation perspective, rather than an environmental one." Goldman summarizes Greene's considerable intellectual contributions to the archival profession: "sustainability of archival endeavor is advanced through careful mobilization, application, and management of institutional resources." See Tansey, "Archival Adaptation," 45; and Goldman, "It's Not Easy Being Green(e)," 27.
- ²⁰ See, for example, Robert S. Cox, "Maximal Processing, or, Archivist on a Pale Horse," *Journal of Archival Organization* 8, no. 2 (2010); and Carl Van Ness, "Much Ado about Paper Clips: 'More Product, Less Process' and the Modern Manuscript Repository," *American Archivist* 73, no. 1 (2010).
- ²¹ Dennis Meissner and Mark A. Greene, "More Application While Less Appreciation: The Adopters and Antagonists of MPLP," *Journal of Archival Organization* 8, nos. 3–4 (2010): 175.
- ²² Mark A. Greene, "MPLP: It's Not Just for Processing Anymore," *American Archivist* 73, no. 1 (2010).
- ²³ Shan C. Sutton, "Balancing Boutique-Level Quality and Large-Scale Production: The Impact of 'More Product, Less Process' on Digitization in Archives and Special Collections," *RBM: A Journal of Rare Books, Manuscripts, & Cultural Heritage* 13, no. 1 (2012): 62.
- ²⁴ Sophia Lafferty Hess and Thu-Mai Christian, "More Data, Less Process? The Applicability of MPLP to Research Data," *IASSIST Quarterly* 40, no. 4 (2016): 12.
- ²⁵ Kevin Bradley, "Defining Digital Sustainability," *Library Trends* 56, no. 1 (2007): 157.
- ²⁶ Blue Ribbon Task Force on Sustainable Digital Preservation and Access, *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information* (February 2010), http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf.
- ²⁷ Evviva Weinraub et al., *Beyond the Repository: Integrating Local Preservation Systems with National Distribution Systems* (2018), 34, <https://doi.org/10.21985/N28M2Z>.

- ²⁸ Digital Preservation Network, "Digital Preservation Workflow Curriculum," March 8, 2017, 21–22, <http://web.archive.org/web/https://www.dpn.org/dpn-admin/resources/digitalpreservationworkshopcriculum3.8.17.pdf>.
- ²⁹ UNESCO/PERSIST Content Task Force, *Guidelines for the Selection of Digital Heritage for Long-Term Preservation* (United Nations Educational, Scientific, and Cultural Organization, March 2016), <https://unesdoc.unesco.org/ark:/48223/pf0000244280>.
- ³⁰ Joseph A. Williams and Elizabeth M. Berilla, "Minutes, Migration, and Migraines: Establishing a Digital Archives at a Small Institution," *American Archivist* 78, no. 1 (2015): 84.
- ³¹ Adrian Brown and Christopher Fryer, "Achieving Sustainable Digital Preservation in the Cloud" (conference paper, Archives and Cultural Industries Congress, Girona, Spain, October 11–15, 2014): 1, <http://www.girona.cat/web/jca2014/ponents/textos/id87.pdf>.
- ³² Heidi N. Abbey, "The Green Archivist: A Primer for Adopting Affordable, Environmentally Sustainable, and Socially Responsible Archival Management Practices," *Archival Issues* 34, no. 2 (2012).
- ³³ Henk, *Ecology, Economy, Equity: The Path to a Carbon-Neutral Library*.
- ³⁴ Abbey, "The Green Archivist," 95–97.
- ³⁵ Beth Filar Williams, Madeleine Charney, and Bonnie Smith, "Growing Our Vision Together: Forming a Sustainability Community within the American Library Association," *Sustainability: Science, Practice & Policy* 11, no. 2 (2015), <https://doi.org/10.1080/15487733.2015.11908147>.
- ³⁶ ALA Sustainability Roundtable, <http://www.ala.org/sustainrt>; and Project_ARCC, <https://projectarcc.org>.
- ³⁷ Matthew Gordon-Clark and Simon Shurville, "To Take up Arms against a Sea of Troubles': Finding Safe Havens for the National Archives of Low-Elevation Pacific Islands and Nations Threatened by Climate Change," *Archives and Manuscripts* 38, no. 1 (2010); and Matthew Gordon-Clark, "Paradise Lost? Pacific Island Archives Threatened by Climate Change," *Archival Science* 12, no. 1 (2012), <https://doi.org/10.1007/s10502-011-9144-3>.
- ³⁸ G. E. Ebinuwale, "Global Warming: Implication for Library and Information Professionals," *International Journal of Humanities and Social Science* 5, no. 6 (2015).
- ³⁹ Tansey, "Archival Adaptation"; and T. Mazurczyk et al., "American Archives and Climate Change: Risks and Adaptation," *Climate Risk Management* 20 (2018), <https://doi.org/10.1016/j.crm.2018.03.005>.
- ⁴⁰ May Cassar, *Climate Change and the Historic Environment* (London: Centre for Sustainable Heritage, University College London, 2005).
- ⁴¹ The National Trust, *Playing Our Part* (2015), <https://www.nationaltrust.org.uk/documents/national-trust-playing-our-part.pdf>.
- ⁴² Augustin Colette and Kishore Rao, *Case Studies on Climate Change and World Heritage* (Paris: UNESCO World Heritage Centre, 2009); and United Nations Educational, Scientific, and Cultural Organization (UNESCO), *Policy Document on the Impacts of Climate Change on World Heritage Properties* (2008), <http://whc.unesco.org/uploads/activities/documents/activity-397-2.pdf>.
- ⁴³ Rebecca Meyer, Shannon Struble, and Phyllis Catsikis, "Sustainability: A Review," in *Preserving Our Heritage: Perspectives from Antiquity to the Digital Age*, ed. Michèle Valerie Cloonan (Chicago: ALA Neal-Schuman, 2015).
- ⁴⁴ Jeremy Linden, James Reilly, and Peter Herzog, "Research on Energy Savings Opportunities in University Libraries," *Library Hi Tech* 30, no. 3 (2012), <https://doi.org/10.1108/07378831211266537>.
- ⁴⁵ Tim Padfield et al., "The Climate Control of the Arnarnagnæan Archive," November 2014, <http://www.conservationphysics.org/arna/arnarnagnæan-CC16.pdf>.
- ⁴⁶ Sarah S. Brophy and Elizabeth Wylie, *The Green Museum: A Primer on Environmental Practice* (Lanham, MD: AltaMira Press, 2008).
- ⁴⁷ Tim Padfield et al., "Low Energy Museum Storage," December 2013, <http://www.conservationphysics.org/storage/low-energy-museum-storage.pdf>.
- ⁴⁸ Richard Kerschner, "Providing Safe and Practical Environments for Cultural Property in Historic Buildings—and Beyond" (contribution, Experts' Roundtable on Sustainable Climate Management Strategies, Tenerife, Spain, April 2007).
- ⁴⁹ US Green Building Council, "What Is LEED?," <http://www.usgbc.org/help/what-leed>.

- ⁵⁰ George J. Aulisio, "Green Libraries Are More than Just Buildings," *Electronic Green Journal* 1, no. 35 (2013).
- ⁵¹ Mark Wolfe, "Beyond 'Green Buildings': Exploring the Effects of Jevons' Paradox on the Sustainability of Archival Practices," *Archival Science* 12, no. 1 (2012), <https://doi.org/10.1007/s10502-011-9143-4>.
- ⁵² Jevons showed how increases in the efficiency of coal use led to decreases in the amount of coal used per unit of production, which led to decreases in unit prices, increases in consumption, and, ultimately, increases in the use of coal. See William Stanley Jevons, *The Coal Question: An Enquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-mines* (London: Macmillan, 1865).
- ⁵³ Wolfe, "Beyond 'Green Buildings,'" 39.
- ⁵⁴ Wolfe, "Beyond 'Green Buildings,'" 47.
- ⁵⁵ Maria A. Jankowska, "Practicing Sustainable Environmental Solutions: A Call for Green Policy in Academic Libraries," *Against the Grain* 22, no. 6 (2014), <https://doi.org/10.7771/2380-176X.5690>.
- ⁵⁶ Jennifer Poggiali, "Incorporating Ethical Consumption into Electronic Device Acquisition: A Proposal," *Portal: Libraries and the Academy* 16, no. 3 (2016), <https://doi.org/10.1353/pla.2016.0037>.
- ⁵⁷ Thomas Hecker, "The Twilight of Digitization Is Now," *Journal of Scholarly Publishing* 35, no. 1 (2003), <https://doi.org/10.3138/jsp.35.1.52>.
- ⁵⁸ Hecker, "Post-Petroleum Future."
- ⁵⁹ David Rosenthal, Kris Carpenter, and Krishna Kant, "Green Bytes: Sustainable Approaches to Digital Stewardship" (plenary session, National Digital Information Infrastructure and Preservation Program partners meeting, Alexandria, VA, July 24, 2013).
- ⁶⁰ National Digital Stewardship Alliance, *2015 National Agenda for Digital Stewardship* (2015), 5.3.2, <http://www.digitalpreservation.gov/documents/2015NationalAgenda.pdf>.
- ⁶¹ Benjamin Goldman, "The 14th Blackbird: Digital Preservation as an Environmentally Sustainable Activity" (lightning talk, Preservation and Archiving Special Interest Group annual meeting, New York, NY, October 2016).
- ⁶² Linda Tadic, "The Environmental Impact of Digital Preservation" (presentation, Association of Moving Image Archivists conference, Portland, OR, November 18–21, 2015), updated December 2018, <https://www.digitalbedrock.com/resources-2>.
- ⁶³ Bethany Nowviskie and Dot Porter, "The Graceful Degradation Survey: Managing Digital Humanities Projects through Times of Transition and Decline" (paper presentation, Digital Humanities, London, England, July 9, 2010). See also Bethany Nowviskie, "Digital Humanities in the Anthropocene," *Digital Scholarship in the Humanities* 30, supplement 1 (2015).
- ⁶⁴ Sarah R. Demb, "When the Lights Go Out: Digital Information and Existential Risk" (presentation, Libraries and Archives in the Anthropocene: A Colloquium, New York, NY, May 13–14, 2007); and Sarah R. Demb, "When the Lights Go Out: Digital Information and Existential Risk" (presentation) in "Dinosaur Footprints: Archiving and Preservation in the Anthropocene" (panel session, Society of American Archivists 2018 Annual Meeting, Washington, DC, August 12–18, 2018).
- ⁶⁵ Goldman, "It's Not Easy Being Green(e)."
- ⁶⁶ For examples of definitions of the scope of ICT, see Jens Malmödin et al., "Life Cycle Assessment of ICT: Carbon Footprint and Operational Electricity Use from the Operator, National, and Subscriber Perspective in Sweden," *Journal of Industrial Ecology* 18, no. 6 (2014): 830; and Klaus Fichter and Ralph Hintemann, "Beyond Energy: The Quantities of Materials Present in the Equipment of Data Centers," *Journal of Industrial Ecology* 18, no. 6 (2014): 846–47.
- ⁶⁷ Santosh S. Venkatraman, "The Dark Green Side of Information Technology," *International Journal of Business Insights and Transformation* 3, special issue 3 (2011): 97–99.
- ⁶⁸ Mél Hogan, "The Archive as Dumpster," *Pivot: A Journal of Interdisciplinary Studies and Thought* 4, no. 1 (2015).
- ⁶⁹ See for example Weinraub et al., *Beyond the Repository*.
- ⁷⁰ *Environmental Management—Life Cycle Assessment—Principles and Framework*, ISO 14040:2006 (Geneva, Switzerland: International Organization for Standardization, approved July 2006, reaffirmed 2016).

- ⁷¹ Standard assessment methods other than LCA include material flow accounting, material intensity per unit service, and estimating impact without conducting an LCA. See Mari Tuusjärvi, Saku Vuori, and Ilmo Mäenpää, "Metal Mining and Environmental Assessments," *Journal of Industrial Ecology* 16, no. 5 (2012): 735–36, <https://doi.org/10.1111/j.1530-9290.2012.00469.x>; and Daniel Collado-Ruiz and Hesamedin Ostad-Ahmad-Ghorabi, "Estimating Environmental Behavior without Performing a Life Cycle Assessment," *Journal of Industrial Ecology* 17, no. 1 (2013), <https://doi.org/10.1111/j.1530-9290.2012.00510.x>.
- ⁷² Tuusjärvi, Vuori, and Mäenpää, "Metal Mining," 735–37.
- ⁷³ Manufacture of a circuit board requires epoxy, polyvinyl chloride, and other synthetic materials; glass fiber, silicon monoxide, ceramics, and silicon; iron, copper, aluminum, nickel, tin, zinc, and lead; and gold, platinum, palladium, silver, ruthenium, and other precious metals. Fichter and Hintemann, "Beyond Energy," 849.
- ⁷⁴ Fichter and Hintemann, "Beyond Energy," 852–54. Their assessment of the weight of ICT components in data centers includes power infrastructure, racks and containment, and cooling infrastructure in addition to IT equipment.
- ⁷⁵ BSC, Incorporated, *Mining Industry Bandwidth Study* (United States Department of Energy, 2007), 7, 21–24, 31–35.
- ⁷⁶ Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, *Summary for Policymakers of the Assessment Report on Land Degradation and Restoration of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* (Bonn, Germany: IPBES Secretariat, 2018).
- ⁷⁷ Iztok Humar et al., "Embodied Energy of Communication Devices: Modeling Embodied Energy for Communication Devices," in *Green Communications: Principles, Concepts and Practice*, ed. Konstantinos Samdanis et al. (Chichester, UK: John Wiley & Sons, 2015), 57–61.
- ⁷⁸ Malmmodin et al., "Life Cycle Assessment," 839; and James Suckling and Jacquetta Lee, "Redefining Scope: The True Environmental Impact of Smartphones?," *International Journal of Life Cycle Assessment* 20, no. 8 (2015): 1183–85.
- ⁷⁹ While industry leaders such as Google report PUE (power use effectiveness, a ratio of the amount of overall power used compared to that used for computing; for example, a ratio of 1.0 means all power is used for computing while a ratio of 2.0 means that for each watt used for computing, an additional watt is used for activities such as cooling and lighting) of 1.10 fleetwide during the fourth quarter of 2018, the reported industry average as of 2018 is only 1.58. See Google, "Google Data Centers, Efficiency: How We Do It, Measuring Efficiency," <https://www.google.com/about/datacenters/efficiency/internal>; and Uptime Institute, *Global Data Center Survey* (2018), <https://uptimeinstitute.com/2018-data-center-industry-survey-results>. See also Simon Forge, "Powering Down: Remedies for Unsustainable ICT," *Foresight* 9, no. 4 (2007): 10–12. On water use, see Mél Hogan, "Data Flows and Water Woes: The Utah Data Center," *Big Data & Society* 2, no. 2 (2015): 4–7, <https://doi.org/10.1177/2053951715592429>.
- ⁸⁰ Peter M. Corcoran and Anders Andrae, *Emerging Trends in Electricity Consumption for Consumer ICT* (2013), 44–48. These figures are corroborated by Esther Müller et al., "Material and Energy Flows and Environmental Impacts of the Internet in Switzerland," *Journal of Industrial Ecology* 17, no. 6 (2013): 820, <https://doi.org/10.1111/jie.12056>, which estimates 7 percent of Swiss electricity use in 2009; and Fichter and Hintemann, "Beyond Energy," 846, which estimates 10.5 percent of German electricity use in 2007. For an estimate with a narrower scope of ICT, and for comparison with additional studies, see Ward Van Heddeghem et al., "Trends in Worldwide ICT Electricity Consumption from 2007 to 2012," *Computer Communications* 50 (2014), 72–75.
- ⁸¹ See for example Gary Cook et al., *Clicking Clean: Who Is Winning the Race to Build a Green Internet?* (Washington, DC: Greenpeace, Inc., January 2017); Gary Cook and David Pomerantz, *Clicking Clean: A Guide to Building the Green Internet* (Washington, DC: Greenpeace, Inc., May 2015); Gary Cook et al., *Clicking Clean: How Companies Are Creating the Green Internet* (Washington, DC: Greenpeace, Inc., April 2014); Gary Cook, *How Clean Is Your Cloud?* (The Netherlands: Greenpeace International, April 2012); Gary Cook and Jodie Van Horn, *How Dirty Is Your Data? A Look at the Energy Choices that Power Cloud Computing* (The Netherlands: Greenpeace International, 2011); Hogan, "Data Flows"; Malmmodin, "Life Cycle Assessment"; Müller, "Material and Energy Flows"; and Venkatraman, "Dark Green Side."
- ⁸² Cook, *Clicking Clean: Who Is Winning?*, 8–13.

- ⁸³ Intergovernmental Panel on Climate Change, *Climate Change 2014. Part A*; Intergovernmental Panel on Climate Change, *Climate Change 2014. Part B*; and Hogan, "Data Flows," 7–9.
- ⁸⁴ C. P. Baldé et al., *The Global E-waste Monitor 2017: Quantities, Flows, and Resources* (Bonn/Geneva/Vienna: United National University, International Telecommunication Union, and International Solid Waste Association, 2017), 38.
- ⁸⁵ Baldé et al., *The Global E-waste Monitor 2017*, 39.
- ⁸⁶ Venkatraman, "Dark Green Side," 100; Ben Bridgens et al., "Closing the Loop on E-Waste: A Multidisciplinary Perspective," *Journal of Industrial Ecology*, online only (2017): 2, <https://doi.org/10.1111/jiec.12645>; Baldé et al., *Global E-waste Monitor*, 44–45; and Forge, "Powering Down," 7–9.
- ⁸⁷ Baldé et al., *Global E-waste Monitor*, 60–78; Bridgens et al., "Closing the Loop," 2; and Kristen Grant, "Health Consequences of Exposure to E-waste: A Systematic Review," *Lancet Global Health* 1, no. 6 (2013), [http://dx.doi.org/10.1016/S2214-109X\(13\)70101-3](http://dx.doi.org/10.1016/S2214-109X(13)70101-3).
- ⁸⁸ Weinraub et al., *Beyond the Repository*; and Sarah Barsness et al., *2017 Fixity Survey Report* (National Digital Stewardship Alliance, 2018), <https://doi.org/10.17605/OSF.IO/SNJBV>.
- ⁸⁹ Analysis is based on Weinraub et al., *Beyond the Repository* survey question 3: "Approximately, how many terabytes of unique content has your institution collected?," https://arch.library.northwestern.edu/concern/parent/00000009g/file_sets/pg15bd956. Storage ranges were converted to single values in the following manner: <1 TB = 0.5 TB; 1-10 TB = 5 TB; 11-50 TB = 25 TB; 51-100 TB = 75 TB; >100 TB = 100 TB.
- ⁹⁰ Analysis is based on Barsness et al., *Fixity Survey Report* question 10: "How much total content (preservation copies that are managed for long-term preservation only) are you running fixity on? Please provide your answer in total number of TB;" and question 10_7_text: "More than 500 TB: (Please enter amount)," <https://osf.io/5ftha>. Storage ranges were converted to single values in the following manner: 0-100 GB = 0.05 TB; 100-500 GB = 0.25 TB; 500 GB-1 TB = 0.75 TB; 1-10 TB = 5 TB; 11-50 TB = 25 TB; 51-100 TB = 75 TB; 101-500 TB = 250 TB; >500 TB = write-in values (the write-in value "many thousands of TB" was converted to 5,000 TB; write-in values qualifying a value by using "over," "close to," or "about" were converted using the exact value given).
- ⁹¹ OCLC, "Global Library Statistics," <https://www.oclc.org/en/global-library-statistics.html>; OCLC, "Global Library Statistics Country Data," https://www.oclc.org/content/dam/oclc/globallibrarystats/globalstats_countrydataset_oclcweb.xlsx. We did not include OCLC's count of public or school libraries, focusing on those types of organizations that we felt are most likely to engage in digital preservation. Furthermore, these data exclude some organizations engaged in digital preservation, such as corporate archives and other CHOs outside of libraries or museums.
- ⁹² Responses to the *Beyond the Repository* survey indicate that repositories preserve from three to seven copies of content. See Weinraub et al., *Beyond the Repository* survey question 8: "You indicated that your institution keeps multiple copies of digital files in multiple places. How many copies (including the original file) are kept?"
- ⁹³ Fixity checking is "the practice of algorithmically reviewing digital content to insure that it has not changed over time." This process results in checksums, also known as cryptographic hashes, which can be compared over time to determine if a file has been altered. See Barsness et al., *Fixity Survey Report*, 4. On frequency, see Barsness et al., *Fixity Survey Report* question 8: "Do you check fixity information at regular intervals—please specify the intervals that your organization uses. Select all that apply:"
- ⁹⁴ See Weinraub et al., *Beyond the Repository* survey question 4: "In the next year, how much additional data do you expect your institution to collect?"
- ⁹⁵ Cisco Systems, Inc., *Cisco Global Cloud Index: Forecast and Methodology, 2016–2021* (2018), <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf?dtdid=ossdc000283>. We do not intend this estimate to be rigorous, rather to provide a sense of the scale at which CHOs are preserving digital content. Further research is needed to quantify this scale and the direct and indirect environmental impacts.
- ⁹⁶ Donella H. Meadows, *Leverage Points: Places to Intervene in a System* (Hartland, VT: The Sustainability Institute, 1999), 18.
- ⁹⁷ See Intergovernmental Panel on Climate Change, *Global Warming of 1.5°C*.

⁹⁸ Krish Bandaru and Kestutis Patiejunas, "Under the Hood: Facebook's Cold Storage System," *Facebook Code Engineer Blog* (blog), May 4, 2015, <https://code.facebook.com/posts/1433093613662262/-under-the-hood-facebook-s-cold-storage-system>.

⁹⁹ S. Hoedl, *Location-Based ISONE Pollutant Intensity Analysis for Harvard University's Office for Sustainability*, August 30, 2016, https://green.harvard.edu/sites/green.harvard.edu/files/Hoedl_GHGIIntensity_Main_Text.pdf.

¹⁰⁰ ISO New England, *CELT Report: 2018–2027 Forecast Report of Capacity, Energy, Loads, and Transmission* (May 1, 2018), <https://www.iso-ne.com/system-planning/system-plans-studies/celt>; ISO New England, "Key Grid and Market Stats," <https://www.iso-ne.com/about/key-stats>; ISO New England, "Solar Power in New England: Concentration and Impact," <https://www.iso-ne.com/about/what-we-do/in-depth/solar-power-in-new-england-locations-and-impact>; and ISO New England, "Daily Generation by Fuel Type," <https://www.iso-ne.com/isoexpress/web/reports/operations/-/tree/daily-gen-fuel-type>.

¹⁰¹ For information on relevant load characteristics, see the regional transmission organization or wholesale electricity market for the area in which your storage infrastructure is located. In the continental United States, these are California Independent System Operator (CAISO), Electric Reliability Council of Texas (ERCOT), Midcontinent Independent System Operator (MISO), ISO New England (ISO-NE), New York Independent System Operator (NYISO), PJM Interconnection (PJM), Southwest Power Pool (SPP), Southeast electricity market, Southwest electricity market, and Northwest electricity market.

¹⁰² "Additionality" of renewable electricity generating resources is used to designate resources added to the electricity grid beyond what is already built, or what would have been built with normal market forces. A good example is wind turbines in the Midwest and Great Plains regions of the United States versus in New England. In the former, wind turbines are economically viable without renewable energy certificates (RECs) and will be built whether or not third parties purchase these certificates. In the latter, wind turbines are only economically viable with the sale of RECs. Purchasing RECs from the former does not promote additionality because the wind turbines will be built with or without the purchase, whereas purchasing RECs from the latter does promote additionality because the wind turbines will only be built with the purchase. For more information on additionality, see for example Michael Gillenwater, *What Is Additionality? Part 1: A Long Standing Problem* (Greenhouse Gas Management Institute, 2012), http://ghginstitute.org/wp-content/uploads/2015/04/AdditionalityPaper_Part-1ver3FINAL.pdf.

¹⁰³ Cook, *Clicking Clean: Who Is Winning?*, 37–41.

¹⁰⁴ Google, "Google Data Centers, Efficiency: How We Do It, Water and Cooling," <https://www.google.com/about/datacenters/efficiency/internal/#water-and-cooling>.

¹⁰⁵ Compare Jevons' Paradox cited earlier in this paper.

¹⁰⁶ Ehrenfeld, *Sustainability by Design*, 5–9, 17–21, 35–47, 108–32. See also Ehrenfeld and Hoffman, *Flourishing*. This approach agrees with an earlier work in the field of systems analysis by Meadows in which she argues that a paradigm shift is one of the most powerful leverage points for effecting change when intervening in a system. See Meadows, *Leverage Points*.

¹⁰⁷ National Digital Stewardship Alliance (NDSA), "Levels of Digital Preservation," <http://nds.org/activities/levels-of-digital-preservation>. We refer to Version 1 of the NDSA "Levels of Digital Preservation." The Levels of Preservation Working Group was actively engaged in updating the levels at the time of writing.

¹⁰⁸ The United States Geological Survey Fundamental Science Practices Advisory Committee Data Preservation Subcommittee preservation recommendations provided by NDSA as an example of how to utilize the NDSA levels guidelines include a description of each level. The description of Level 4 is "Best. USGS should plan to provide repositories that meet these criteria for all long-term USGS records." It is clear that USGS's interpretation of the NDSA levels is that Level 4 is optimal for all records requiring long-term preservation, regardless of anticipated use, value, inherent characteristics, or original format. See United States Geological Survey, *USGS Guidelines for the Preservation of Digital Scientific Data* (April 2014).

¹⁰⁹ *Space Data and Information Transfer Systems—Audit and Certification of Trustworthy Digital Repositories*, ISO 16363:2012 (Geneva, Switzerland: International Organization for Standardization, approved February 2012, reaffirmed 2017).

- ¹¹⁰ Jaime Schumacher et al., *From Theory to Action: "Good Enough" Digital Preservation Solutions for Under-Resourced Cultural Heritage Institutions* (Digital POWRR, August 2014).
- ¹¹¹ *Space Data and Information Transfer Systems—Open Archival Information System (OAIS)—Reference Model*, ISO 14721:2012 (Geneva, Switzerland: International Organization for Standardization, approved September 2012, reaffirmed 2018).
- ¹¹² Stephen Abrams, "Nothing Succeeds Like Success: An Approach for Evaluating Digital Preservation Efficacy," in *Proceedings of the 15th International Conference on Digital Preservation* (2018), <https://doi.org/10.17605/OSF.IO/Z78R6>.
- ¹¹³ As with all materials, digital content can have many types of values, including administrative, enduring, historical, informational, evidential, and monetary. For a more detailed explanation of value, see <https://www2.archivists.org/glossary/terms/v/value>.
- ¹¹⁴ Terry Cook, "'We Are What We Keep; We Keep What We Are': Archival Appraisal Past, Present and Future," *Journal of the Society of Archivists* 32, no. 2 (2011): 175, <https://doi.org/10.1080/00379816.2011.619688>.
- ¹¹⁵ Geof Huth, "Appraising Digital Records," in *Appraisal and Acquisitions Strategies*, ed. Michael Shallcross and Christopher J. Prom (Chicago: Society of American Archivists, 2016), 14.
- ¹¹⁶ See for example Huth's detailed description of the appraisal process for digital content in "Appraising Digital Records," or Tallman and Work's framework for collaborative appraisal by curators and digital preservationists: Nathan Tallman and Lauren Work, "Approaching Appraisal: Guidelines and Criteria to Select for Digital Preservation," in *Proceedings of the 15th International Conference on Digital Preservation* (2018), <https://doi.org/10.17605/OSF.IO/8Y6DC>.
- ¹¹⁷ Weinraub et al. demonstrate that many CHOs are electing to store some portion of their digital content in distributed digital preservation services, the selection of which depends on various criteria. See Weinraub et al., *Beyond the Repository*, 25. Other CHOs have integrated a tiered approach to digital preservation into their policies, though without specifically addressing environmental impacts. See for example "Rockefeller Archive Center Digital Preservation Policy," The Rockefeller Archive Center, <http://rockarch.org/programs/digital/DigPresPolicy.php>; and "Digital Preservation Policy," Northwestern University Libraries, <http://www.library.northwestern.edu/about/administration/policies/digital-preservation-policy.html>.
- ¹¹⁸ For example, Creighton Barrett demonstrated in a 2017 presentation that digital forensics tools could identify millions of duplicate, software, and system files within a 3.5-terabyte born-digital archive (an accession of over 26.5 million files was reduced to approximately 100,000 files after analysis and filtering). See Creighton Barrett, "Tools for Identifying Duplicate Files and Known Software Files" (presentation, BitCurator Users Forum, Evanston, IL, April 28, 2017).
- ¹¹⁹ This is germane to personal digital archives, which can be duplicated and donated to multiple organizations, but is most relevant for web archiving and software preservation, where collaborative collecting is already underway. See for example the Ivy Plus Libraries Web Resources Collection Program, a collaborative web archiving program among thirteen universities: Columbia University Libraries, "Ivy Plus Libraries Web Resources Collection Program," http://library.columbia.edu/find/web-archives/Ivy_Plus_Libraries.html. Collaborations in the software preservation field, such as the Fostering a Community of Practice: Software Preservation in Libraries and Archives (FCoP) project, are currently underway as well. See Software Preservation Network, "FCoP: Software Preservation in Libraries and Archives," <http://www.softwarepreservationnetwork.org/fcop>.
- ¹²⁰ See for example Sam Meister and Alexandra Chassanoff, "Integrating Digital Forensics Techniques into Curatorial Tasks: A Case Study" *International Journal of Digital Curation* 9, no. 2 (2014); Christopher A. Lee et al., *From Bitstreams to Heritage: Putting Digital Forensics into Practice in Collecting Institutions* (BitCurator Project, September 30, 2013); and Kam Woods and Christopher A. Lee, "Acquisition and Processing of Disk Images to Further Archival Goals," in *Proceedings of Archiving 2012* (Springfield, VA: Society for Imaging Science and Technology, 2012), 147–52.
- ¹²¹ See for example, Kam Woods, Christopher A. Lee, and Sunitha Misra, "Automated Analysis and Visualization of Disk Images and File Systems for Preservation," in *Proceedings of Archiving 2013* (Springfield, VA: Society for Imaging Science and Technology, 2013): 239–44; Christopher A. Lee and Kam Woods, "Automated Redaction of Private and Personal Data in Collections: Toward Responsible Stewardship of Digital Heritage," in *Proceedings of Memory of the World in the Digital Age: Digitization and Preservation: An International Conference on Permanent Access to Digital Documentary*

Heritage, ed. Luciana Duranti and Elizabeth Shaffer (United Nations Educational, Scientific, and Cultural Organization, 2013), 298–313; and Kam Woods et al., “Functional Access to Forensic Disk Images in a Web Service,” in *Proceedings of the 12th International Conference on Digital Preservation*, ed. Christopher A. Lee et al. (Chapel Hill: University of North Carolina, School of Information and Library Science, 2015).

¹²²For example, Baker Library Special Collections at Harvard Business School has developed a disk image retention policy that takes into account ethical issues, risk tolerance, and environmental sustainability. See Keith Pendergrass, “One Does Not Simply Keep Disk Images: Ethical, Risk Tolerance, and Sustainability Issues with Forensic Disk Image Retention” (presentation, BitCurator Users Forum, Evanston, IL, April 28, 2017).

¹²³File sizes given are for 16 Bit DPX. Gamma Ray Digital, “Just How Big Are Those Files, Anyway?” (blog), December 17, 2014, <https://www.gammaraydigital.com/blog/just-how-big-are-those-files-anyway>.

¹²⁴Society of American Archivists, “Guidelines for Reappraisal and Deaccessioning,” 2017, https://www2.archivists.org/sites/all/files/GuidelinesForReappraisalDeaccessioning_2017.pdf.

¹²⁵Megan Phillips et al., “The NDSA Levels of Digital Preservation: An Explanation and Uses,” *Proceedings of the Archiving (IS&T) Conference* (April 2013).

¹²⁶Alex Duryee, *What Is the Real Impact of SHA-256? A Comparison of Checksum Algorithms* (New York: AVPreserve, October 2014), 2, https://www.avpreserve.com/wp-content/uploads/2017/07/ChecksumComparisons_102014.pdf.

¹²⁷David S. H. Rosenthal, “Why Is Green Preservation Hard?” (presentation) in “Green Bytes: Sustainable Approaches to Digital Stewardship” (plenary session, National Digital Information Infrastructure and Preservation Program partners meeting, Alexandria, VA, July 24, 2013), <http://digitalpreservation.gov/meetings/documents/ndiipp13/Rosenthal.pdf>.

¹²⁸David S. H. Rosenthal, “How Few Copies?,” *DSHR’s Blog* (blog), April 19, 2016, <https://blog.dshr.org/2016/04/how-few-copies.html>.

¹²⁹The Open Archival Information System (OAIS) reference model (see note 111) defines three information packages: the Submission Information Package (SIP), the Archival Information Package (AIP), and the Dissemination Information Package (DIP). For an accessible introduction to the OAIS reference model, see Brian Lavoie, *The Open Archival Information System (OAIS) Reference Model: Introductory Guide*, 2nd ed. (Great Britain: Digital Preservation Coalition, October 2014), <http://dx.doi.org/10.7207/twr14-02>.

¹³⁰MetaArchive, “How DDP Works,” <https://metaarchive.org/how-ddp-works>.

¹³¹Such as storage media, hardware setup, auditing protocols, and the electricity source of the servers and storage components.

¹³²David S. H. Rosenthal, “The Half-Empty Archive,” *DSHR’s Blog* (blog), March 31, 2014, <https://blog.dshr.org/2014/03/the-half-empty-archive.html>.

¹³³James O’Toole, “On the Idea of Permanence,” *American Archivist* 52, no. 1 (1989), <https://doi.org/10.17723/aarc.52.1.3x85283576r43387>.

¹³⁴Margaret Hedstrom, “Digital Preservation: A Time Bomb for Digital Libraries,” *Computers and the Humanities* 31 (1998): 190.

¹³⁵Barsness et al., *Fixity Survey Report*, 32.

¹³⁶In this section, we build on Goldman’s challenge to practices of digital authenticity and Nowviskie and Porter’s concept of graceful degradation. See Goldman, “It’s Not Easy Being Green(e);” Nowviskie, “Digital Humanities in the Anthropocene;” and Nowviskie and Porter, “The Graceful Degradation Survey.”

¹³⁷Goldman, “It’s Not Easy Being Green(e),” 291.

¹³⁸A *bit flip* is said to have happened when a binary bit of data unexpectedly changes state from a 0 to a 1 or vice versa.

¹³⁹In many storage systems, such as hard drives, Storage Area Networks (SAN), Network Attached Storage (NAS), and some cloud storage systems, *blocks* refer to consistently sized sequences of bytes that the file system interacts with as discrete chunks of data. Blocks are managed independently of logical files; the data in a file may span one or more blocks.

- ¹⁴⁰Yupu Zhang et al., "End-to-End Data Integrity for File Systems: A ZFS Case Study," <http://research.cs.wisc.edu/wind/Publications/zfs-corruption-fast10.pdf>. Note that self-healing in ZFS requires a second source of the data from which to heal, such as that in a mirror or RAID, and so does not eliminate the need for redundant copies of data.
- ¹⁴¹Alex Garnett, Mike Winter, and Justin Simpson, "Checksums on Modern Filesystems, or: On the Virtuous Consumption of CPU Cycles," in *Proceedings of the 15th International Conference on Digital Preservation* (2018), <https://doi.org/10.17605/OSF.IO/Y4Z3E>.
- ¹⁴²Bianca Schroeder, Eduardo Pinaheiro, and Wolf-Dietrich Weber, "DRAM Errors in the Wild: A Large-Scale Field Study," in *SIGMETRICS/Performance '09—Proceedings of the 11th International Joint Conference on Measurement and Modeling of Computer Systems* (2009), <http://www.cs.toronto.edu/~bianca/papers/sigmetrics09.pdf>.
- ¹⁴³Tadic, "Environmental Impact," 80.
- ¹⁴⁴Barsness et al., *Fixity Survey Report*, question 8: "Do you check fixity information at regular intervals—please specify the intervals that your organization uses. Select all that apply:"
- ¹⁴⁵Threat modeling is a process in the computer security domain during which specific potential security vulnerabilities and risks are identified so that they can be addressed in a targeted manner.
- ¹⁴⁶Max Eckard, "Customizing Archivematica's Format Migration Strategies with the Format Policy Registry (FPR)," *Bentley Historical Library Curation Team* (blog), October 18, 2016, <https://archival-integration.blogspot.com/2016/10/customizing-archivematicas-format.html>.
- ¹⁴⁷CLOCKSS, "LOCKSS: Basic Concepts," https://documents.clockss.org/index.php?title=LOCKSS:_Basic_Concepts.
- ¹⁴⁸Offline storage consists of data written to media that are not powered continuously and are accessed infrequently (e.g., LTO tapes stored on a shelf). Nearline storage describes an intermediate solution between offline and online storage, where storage is not immediately available but can be moved online quickly without human intervention (e.g., LTO tapes stored in a tape library with a robot, or disk drives that are powered down when not in use).
- ¹⁴⁹James Bantin and Leah Agne, "Digitizing for Value: A User-Based Strategy for University Archives," *Journal of Archival Organization* 8, nos. 3–4 (2010), <https://doi.org/10.1080/15332748.2010.550791>; and Larisa Miller, "All Text Considered: A Perspective on Mass Digitizing and Archival Processing," *American Archivist* 76, no. 2 (2013), <https://doi.org/10.17723/aarc.76.2.6q005254035w2076>.
- ¹⁵⁰See for example Stanford Libraries, "Pricing," <https://library.stanford.edu/research/digitization-services/services/pricing>; and Reveal Digital, <http://revealdigital.com>.
- ¹⁵¹Samantha Thompson, "Why Don't Archivists Digitize Everything?," *Archives @ PAMA* (blog), May 31, 2017, <https://peelarchivesblog.com/2017/05/31/why-dont-archivists-digitize-everything>; Bruce R. Kingma, "The Costs of Print, Fiche, and Digital Access: The Early Canadiana Online Project," *D-Lib Magazine* 6, no. 2 (2000), <https://doi.org/10.1045/february2000-kingma>; and Miller, "All Text Considered."
- ¹⁵²Jennifer Schaffner, Francine Snyder, and Shannon Supple, *Scan and Deliver: Managing User-Initiated Digitization in Special Collections and Archives* (Dublin, OH: OCLC Research, 2011), <http://www.oclc.org/research/publications/library/2011/2011-05.pdf>.
- ¹⁵³David S. H. Rosenthal et al., "Requirements for Digital Preservation Systems: A Bottom-Up Approach," *D-Lib Magazine* 11, no. 11 (2005), <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>.
- ¹⁵⁴TechTarget, "MAID (massive array of idle disks)," <https://searchstorage.techtarget.com/definition/MAID>.
- ¹⁵⁵D. Colarelli and D. Grunwald, "Massive Arrays of Idle Disks for Storage Archives," *SC '02: Proceedings of the 2002 ACM/IEEE Conference on Supercomputing* (2002), <https://doi.org/10.1109/SC.2002.10058>.
- ¹⁵⁶David G. Andersen et al., "FAWN: A Fast Array of Wimpy Nodes" (paper, Association for Computing Machinery Symposium on Operating System Principles, Big Sky, MT, October 11–14, 2009), 14.
- ¹⁵⁷Ian F. Adams, Ethan L. Miller, and David S. H. Rosenthal, *Using Storage Class Memory for Archives with DAWN, a Durable Array of Wimpy Nodes* (Storage Systems Research Center Technical Report, October 2011), 8.

ABOUT THE AUTHORS



Keith L. Pendergrass is the digital archivist for Baker Library Special Collections at Harvard Business School, where he develops and oversees born-digital content workflows. His research interests include integration of sustainability principles into digital archives standard practice, systems thinking, energy efficiency, and clean energy and transportation. He holds an MSLIS from Simmons College and a BA from Amherst College.



Walker Sampson is an assistant professor and digital archivist at the University of Colorado Boulder, where he leads the management of the Special Collections and Archives' born-digital accessions. His research interests include digital object preservation, storage, and curation, as well as the technical history of computing and gaming. He holds an MSIS from the School of Information at the University of Texas at Austin and a BA from the University of Mississippi.



Tim Walsh is the digital preservation librarian at Concordia University, where he leads the development of the library's digital preservation infrastructure and activities. His research interests include automation of digital processing workflows, software preservation, management of private and sensitive information, and digital design records. He holds an MSLIS from Simmons College and a BA from the University of Florida.



Laura Alagna is the digital preservation librarian at Northwestern University, where she develops and implements policies and workflows for preserving born-digital and digitized content. Her research interests include repository interoperability, sustainability in digital libraries, and preservation of mobile environments. She holds an MLIS from the University of Illinois School of Information Sciences and a BA from Amherst College.