The State of Digital Preservation in 2018

A Snapshot of Challenges and Gaps

October 29, 2018

Oya Y. Rieger

ITHAKA S+R

Ithaka S+R provides research and strategic guidance to help the academic and cultural communities serve the public good and navigate economic, demographic, and technological change. Ithaka S+R is part of ITHAKA, a not-for-profit organization that works to advance and preserve knowledge and to improve teaching and learning through the use of digital technologies. Artstor, JSTOR, and Portico are also part of ITHAKA.

Our cultural, historic, and scientific heritage is increasingly being produced and shared in digital forms. The ubiquity, pervasiveness, variability, and fluidity of such content raise a range of questions about the role of research libraries and archives in digital preservation in the face of rapid organizational and technological changes and evolving organizational priorities. Ithaka S+R is interested in exploring the current landscape of digital preservation programs and services in order to identify research and policy questions that will contribute to the advancement of strategies in support of future scholarship. To this end, during June and July 2018, I talked with 21 experts and thought leaders to hear their perspectives on the state of digital preservation.[1] The purpose of this report is to share a number of common themes that permeated through the conversations and provide an opportunity for broader community reaction and engagement, which will over time contribute to the development of an Ithaka S+R research agenda in these areas.

The digital preservation ecology is rich with many initiatives and significant contributors. [2] Rather than making broad generalizations, this paper aims to provide a snapshot of what is going through the minds of a selected group of individuals with expertise and responsibility in the related program areas. The primary goal of the interviews was to identify opportunities and needs, rather than to celebrate the successes of the community. Therefore, inevitably, the report focuses on the gaps and challenges, with more limited coverage of the significant advances that have been registered in the

---

[1] We gratefully acknowledge the generosity of the following colleagues in sharing their insights: Stephen Abrams, Associate Director, University of California Curation Center; Alan Darnell, Director, Scholars Portal Services (Canada); Kathleen Fitzpatrick, Director of Digital Humanities, Michigan State University; Mike Furlough, Executive Director, HathiTrust; Chip German, Director, APTrust; Wolfram Horstmann, Director of the Göttingen State and University Library (Germany); Leslie Johnston, Director of Digital Preservation, NARA; Anne Kenney, retired University Librarian, Cornell University Library; William Kilbride, Executive Director, Digital Preservation Coalition (UK); Steve Knight, Programme Director Preservation Research, National Library of New Zealand; Deanna Marcum, Ithaka S+R Senior Advisor; Carol Mandel, retired dean of NYU Libraries; Nancy McGovern, Director, Digital Preservation, MIT Libraries; Trevor Owens, Head, Digital Content Management, Library of Congress; Art Pasquinelli, LOCKSS Partnership Manager, Stanford University Libraries; Ben Fino Radin, Founder Small Data Industries, formerly Media Conservator at MOMA; Bernie Reilly, Executive Director, Center for Research Libraries; Kathleen Shearer, Executive Director, Confederation of Open Access Repositories & Research Associate, Canadian Association of Research Libraries; Tyler Walters, Dean, University Libraries and Professor, Virginia Tech; Kate Wittenberg, Managing Director, Portico; Kazu Yamaji, Information Scientist, National Institute of Informatics (Japan).

[2] Key terms such as archiving and preservation continue be used interchangeably, sometimes depending on the preferences of specific communities. For example, Open Archival Information System (OAIS) uses "archive" when referring to an organization that intends to preserve information for access and use by a "designated community." One of the issues the interviewees mentioned was the need to better align the roles and responsibilities of libraries and archives as the boundaries blur with the emergence of programs such as web archiving and research data management. For instance, one might argue that the materials held at archives (selected in accordance with their acquisition policies and institutional mandates) are more vulnerable due to their unique and unpublished nature.

field during the last two decades. We recognize the substantial contributions of the individuals interviewed, all the members of their organizations, and the field more broadly. Through their valuable work we see the beginning groundwork for a technical, service, and policy infrastructure. The community is getting larger and stronger and is committed to exploring how we can strengthen collaborations in this complicated landscape.

Methodology

Digital preservation involves the management and maintenance of digital objects to ensure the authenticity, accuracy, and functionality of content over time in the face of technological and administrative changes. Historically, archives, national libraries, and research libraries have assumed a leadership role in the stewardship of primary resources and cultural heritage materials over generations to ensure long-term access. Therefore, we selected the interviewees for this study based on their current leadership roles in overseeing related services or their deep history in digital preservation programs. For a broader view, we included colleagues from Canada, Germany, New Zealand, and the United Kingdom, as well as the United States.[3] The overall purpose was to survey the preservation landscape within the context of evolving research workflows and the scholarly and cultural record from the perspective of these colleagues. Although the conversations were open-ended, the following questions framed the discussions:

- What seems to be working well now (in which areas have we seen significant progress)?

- What are your thoughts on how the preservation community is preparing for new content types and formats?

- Do you have any observations about new research workflows and practices (or cultural practices) and their potential impact on the future of scholarly record?

- What do you see as gaps or areas that need further attention (where the community has not made sufficient progress)? These could be organizational, technical, sociocultural, etc.

- If you were writing a new preservation research or implementation grant, what would you focus on?

---

[3] Sixteen of the interviews were conducted through individual phone interviews, five through in-person meetings.

There was a significant level of convergence in opinions related to the trends, outstanding issues, emerging needs, and gaps in the digital preservation realm.

What's Working Well

In response to the accelerating rates and increasing complexity of digital information, since the early 1990s a number of studies have investigated the trends and needs in digital preservation and have recommended research and policy agendas.[4] The interviewees agreed that a considerable amount of progress has been made in many areas since the framing of digital preservation as a critical program area for long-term accessibility of the social, economic, cultural and intellectual heritage:

Stronger community and robust sharing of practices. The digital preservation community is getting larger, representing deeper expertise around a wide range of digital content types. Through several digital preservation and repository conferences and organizations, there is a robust exchange of best practices, standards, and preservation technique. Publications such as the *Digital Preservation Coalition Technology Watch Report* series provide an advanced introduction to specific issues by identifying and tracking the development of the standards and tools that are critical to digital preservation activities.[5] We witnessed the emergence of conferences, such as the International Conference on Digital Preservation (iPres), which are entirely devoted to technical, organizational, and policy aspects of the program domain with an international scope.

Increasing knowhow of preservation strategies. There is now significant experience in implementing preservation strategies such as normalization, refreshing, migration, and emulation as the community of practitioners successfully moved these techniques from theory to practice.[6] The community appreciates that digital preservation programs should factor in a range of organizational, business, and policy issues, not limited to

---

4 Some examples include Donald Waters & John Garrett, *Preserving Digital Information, Report of the Task Force on Archiving of Digital Information*, 1996, https://www.clir.org/pubs/reports/pub63/; National Science Foundation and the Library of Congress, *It's about Time: Research Challenges in Digital Archiving and Long-Term Preservation*, 2003, http://chnm.gmu.edu/digitalhistory/links/pdf/preserving/8_4b.pdf; National Digital Stewardship Alliance, *National Agenda for Stewardship*, 2015, https://ndsa.org/documents/2015NationalAgenda.pdf; Nancy McGovern, "A Digital Decade: Where Have We Been and Where Are We Going in Digital Preservation?," *RLG DigiNews*, Apr 15, 2007, https://deepblue.lib.umich.edu/bitstream/handle/2027.42/60441/McGovern-Digital_Decade.html?sequence=4.

5 See *Digital Preservation Coalition Technology Watch Report* at https://www.dpconline.org/knowledge-base/tech-watch-reports.

6 See the "Digital Preservation Handbook" by the Digital Preservation Coalition for the descriptions of key preservation strategies: https://dpconline.org/handbook/organisational-activities/preservation-action.

technological matters. The training and professional development opportunities continue to expand with the emergence of new specialized positions, such as Digital Archivists or Digital Forensics Specialists, indicating the increasing sophistication of specialization areas.

Shared standards and frameworks. The development and adoption of shared standards, such as Open Archival Information System (OAIS), PREservation Metadata: Implementation Strategies (PREMIS), and PRONOM online file format registry, have been instrumental in facilitating the access, discovery, management, and preservation of digital resources. The OAIS Reference Model has provided a much-needed framework for establishing a common vocabulary for describing roles, processes, and functional components required for long-term preservation. In addition to general models, specific industries have become active in the development of preservation standards tailored for specific formats, for instance the "Guidelines on the Production and Preservation of Digital Audio Objects by the International Association of Sound and Audio-visual Archives."

Availability of new preservation systems and tools. Indicative of an active market place, there are a number of open source and commercial digital preservation systems and tools available. Some examples include ArchiveSpace, Archivematica, ArchiveIt, Preservica, Memento, Portico, LOCKSS, CLOCKSS, DPN, Internet Archive, HathiTrust, Chronopolis, APTrust, DuraSpace/Cloud, Rosetta, BitCurator, MetaArchives. Each of these solutions and others like them represent a substantial investment in the time of skilled workers and the resources to support them.[7]

Demonstration of effective collaborations. There are several successful collaborative initiatives that demonstrate the value and power of broad community efforts, for example, APTrust, Scholars Portal, Texas Digital Library, Digital Preservation Coalition, Bit Curator Alliance, and the University of California Curation Center. There is an effort to create shared values to promote preservation as a service that is sustainable, affordable, practical, and available to all and the drive to accomplish these goals together rather than separately.[8]

---

[7] See the technical solutions section of the *Digital Preservation Handbook* for additional examples, https://dpconline.org/docs/digital-preservation-handbook2/1550-dp-handbook-technical-solutions-and-tools/file.

[8] The Digital Preservation Declaration of Shared Values (2017) was issued by representatives of the Academic Preservation Trust (APTrust), Chronopolis, CLOCKSS, Coalition for Networked Information (CNI), Digital Preservation Network (DPN), DuraSpace, Educopia/MetaArchive Cooperative, Stanford University - LOCKSS, Texas Digital Library (TDL), and the Council of Prairie and Pacific University Libraries (COPPUL). It is available at https://docs.google.com/document/d/1cL-g_X42J4p7d8H7O9YiuDD4-KCnRUlITC2syfXSn5s/edit.

Availability of repository architectures and implementations. There are now a range of digital repository architectures and open source collaborations, such as Fedora and Samvera, to provide open and scalable technical infrastructures specifically designed to cater to the needs of libraries and archives. Several libraries develop and maintain institutional repositories based on open standards to ensure that the curated content is easily accessible, searchable, as well as retrievable for later use. We witnessed the emergence of repositories for digital government records and documents, such as the NARA's Electronic Records Archive and the GPO's Federal Digital System to provide free online access to official Federal Government publications and manage digital content throughout its lifecycle to ensure content integrity and authenticity.

Active research and development agenda. There are many studies that investigate the increasingly complex digital ecology with a focus on web pages, software, AV, geospatial data, email, virtual reality, etc. One instance is the British Library's emerging formats project to investigate the collection management needs of works with significant software and hardware dependencies. On the software preservation front, Yale, University of Virginia, Notre Dame, Stanford, University of California at San Diego, and Carnegie Mellon are working together to ensure that the information produced on software will be accessible for future access, study, and use. [9] The University of North Texas (UNT Libraries and the UNT Computer Science and Engineering Department) is evaluating the use of machine learning algorithms to successfully identify and extract publications contained in existing Web archives. [10]

This is a rather succinct list but each one of these areas of advancement can be expanded to further illustrate the considerable accomplishments that have been instrumental in initiating a technical and organizational infrastructure. These forward-looking and specialized initiatives and services continue to bolster the community's collective ability to preserve our digital heritage.

Challenges in Need of Further Research and Action

Although the interviewees described many areas of progress, they also commented on their concerns about how to provide sufficient levels of digital preservation to meet the

---

[9] The Scaling Emulation and Software Preservation Infrastructure (EaaSI) project builds on previous work by Yale University Library and elsewhere to apply the Emulation-as-a-Service (EaaS) model for access and use of preserved software and digital objects. Information about the project available at http://www.softwarepreservationnetwork.org/eaasi/.

[10] More information about the initiative available at https://www.imls.gov/grants/awarded/LG-71-17-0202-17.

community's needs. The following section provides insights on the outstanding issues that would benefit from further research.

Organizations and Leadership

Unclear role of research libraries. As digital information is increasingly distributed and networked, academic libraries are no longer perceived as critical drivers and leaders. Nearly all the individuals we interviewed stated that the role of research libraries in digital preservation needs to be redefined and resituated. It is difficult to preserve content that is not "owned" or "controlled" by libraries. As one interviewee stated, library leaders have "shifted their attention from seeing preservation as a moral imperative to catering to the university's immediate needs." Several wondered what arguments could convince provosts and other senior university leaders to invest in digital preservation. Another expert noted that with the increasing influence of commercial and industrial actors, "the digital preservation community is becoming more diverse and the distinctive requirements of research libraries are not as dominant as they perhaps once were in the community."

> "[Library leaders have] shifted their attention from seeing preservation as a moral imperative to catering to the university's immediate needs"

Ambiguity of responsibilities. The placement of preservation responsibilities within library organizations and the associated resources dedicated to this service area are often ambiguous and opaque. [11] "Sustainable communication even within the same organization is a challenge," said one interviewee. "Expertise is increasingly fragmented as web archiving, digital curation, research data, repositories, and special collections are often placed in different library units without a common preservation mandate." Also there seems to be some disconnect between how the top leadership level (University Librarians, Associate University Librarian) perceives preservation priorities and needs

---

[11] There were several references to one of the findings of the recent National Digital Stewardship Alliance (NDSA) survey about an increasing dissatisfaction with the way the digital preservation function is organized: 46 percent of survey respondents not satisfied and 25 percent satisfied in 2017 versus 34 percent not satisfied and 42 percent satisfied in 2012. The NDSA Digital Preservation Staffing Survey is available at https://ndsa.org/2017/10/17/announcing-publication-of-the-ndsa-digital-preservation-staffing-survey-report.html.

versus curators, digital collection specialists, archivists, and other specialists. Relatedly, it is important for specialists, such as curators and archivists, to have a grounded understanding of how their specific roles and priorities fit into the overall strategies of libraries and cultural heritage institutions.

> "Expertise is increasingly fragmented as web archiving, digital curation, research data, repositories, and special collections are often placed in different library units without a common preservation mandate"

Misalignment of expectations. Many interviewees expressed anxiety about how digital preservation activities are being slowed down or impeded due to politics and conflicts both within and outside of organizations.[12] For instance, individual library units may not understand their roles and responsibilities and may even sometimes compete with each other for resources or leadership. Another disconnect is between libraries and scholars as the latter, thanks to a proliferation of tools and services, have become much more autonomous in the curation, if not preservation, of data and digital objects.

Need to attend to inclusivity, diversity, and social justice. Interviewees also expressed the need to evaluate our archiving efforts to select and preserve content through the lens of inclusivity, diversity, and social justice. For instance, there were a couple of references to the importance of considering the history, criticism, and philosophy of digital culture and its impact on society. Several interviewees stressed the historic significance of documenting hate speeches, fake news, and unearthing and preserving government websites taken down due to politicized rhetoric, such as with climate change and energy.

Preservation Services and Program Areas

Confusion about new preservation services. There is a significant level of confusion about the purpose and business models of existing and nascent preservation services.

---

[12] One of the interviewees described an impromptu straw poll at a PASIG meeting at Oxford, UK: "What stops you today from putting your digital preservation plan into action: money, technology, or politics?" More hands were raised for option C, dwarfing "money" and "technology" in comparison. Shira Peltzman, "It's difficult to solve a problem if you don't know what's wrong," November 29, 2017, https://www.dpconline.org/blog/idpd/it-s-difficult-to-solve-a-problem-if-you-don-t-know-what-s-wrong.

How do such services fit together in creating a comprehensive preservation service framework? What are the gaps and redundancies? Do they compete with each other or work towards a complementary ecosystem? How do they interact with institutional digital asset management systems? Such questions indicate that the community would benefit from a deeper understanding of the available tools vis-à-vis preservation, discovery, and access stages.[13] Also, a consolidated understanding about what's being preserved and the associated technical, organizational, and policy issues is important for effective planning and implementation of a digital preservation program.

Storage is neither simple nor cheap. Cautioning that storage does not equate to preservation, the interviewees expressed the need to better understand the current storage options and cost models (especially cloud storage).[14] As one interviewee stated, "if cloud computing is a paradigm shift in personal computing, then there needs to be an equivalent paradigm shift for digital preservation too." Another expert cautioned that we need to be careful about relying on the university IT unit for building storage as the academic data centers will "focus more comfortably on providing platforms on which to run the research algorithms on data, but they're growing less interested in providing commodity storage for it."

Unresolved legacy preservation problems. Some pointed out that as we worry about falling behind in dealing with emergent content formats and software, we tend to forget that we have not yet addressed the problems associated with ejournals, ebooks, and Electronic Theses and Dissertations (ETDs). For instance, one of the interviewees reminded us that a 2016 study estimated that only 30 percent of ejournals had preservation strategies and questioned how the research library community has been following up on such findings.[15]

Vulnerability of open access content. Many remarked on the current and future role of institutional repositories in the discovery, access, and archiving chain. Several expressed concerns about the insufficient emphasis on the long-term sustainability and preservation of open access content, noting that freely accessible e-journals comprise the

---

[13] Several interviewees underscored the need for service providers to articulate their distinctive value proposition more clearly, especially vis-à-vis other related tools and services.

[14] One of the reviewers of a draft version of this paper pointed out that there needs to be more discussion at the international level about what constitutes "good enough" file format quality, for instance JPEG2000 (JP2) with regard to preservation risks and the practicalities of preserving data in this format.

[15] Shannon Regan, Joyce McDonough, Bob Wolven, Oya Y. Rieger, "Strategies for Expanding e-Journal Preservation," May 2016, https://confluence.cornell.edu/display/culpublic/Strategies+for+Expanding+E-Journal+Preservation.

largest, most diverse, and most problematic category for preservation as they vary widely in publication methods and publisher types.[16]

Questions about the use of web archives. There is a lack of understanding about how archived websites are discovered, used, and referenced. "Researchers prefer to cite the original live-web as it is easier and shorter," pointed out one of the experts. "There is limited awareness of the existence of web archives and lack of community consensus on how to treat them in scholarly work. The problems are not about technology any more, it is about usability, awareness, and scholarly practices." The interviewee referred to a recent CRL study based on an analysis of referrals to archived content from papers that concluded that the citations were mainly to articles about web archiving projects.[17]

"There is limited awareness of the existence of web archives and lack of community consensus on how to treat them in scholarly work. The problems are not about technology any more, it is about usability, awareness, and scholarly practices."

Concerns about usability of archived research data. Several interviewees pointed out that research data initiatives tend to focus on initial identification, ingest, and description stages without a sufficient emphasis on how the archived content will be discovered, accessed, and used by scholars at the point of need in a usable and meaningful way: "The harder problem to understand is the relationship between data and software. It is difficult to justify collecting and preserving things if they aren't providing value to your stakeholders."

---

[16] PKP, LOCKSS, Portico, and Directory of Open Access Journals have initiatives that try to address certain aspects of open access journal archiving. The Andrew W. Mellon Foundation recently awarded a research and development grant to the Internet Archive to address the critical need to preserve the "long tail" of open access scholarly communications. The announcement is available at https://mellon.org/grants/grants-database/grants/internet-archive/1710-04980/.

[17] Jeffrey Garrett, "Harvesting Web Content and Data from Emerging Regions: a Status Report," *CRL Global Resources Collections Forum*, 2018, https://www.crl.edu/events/crl-global-resources-collections-forum-2018.

"It is difficult to justify collecting and preserving things if they aren't providing value to your stakeholders"

Insufficient focus on primary materials and analysis tools. Contemporary source materials like news, blogs, and online discussion forums are vital to original scholarly research in the humanities and social sciences as they capture viewpoints and new trends and reflect how scientific discourses evolve. How do we preserve the content, context, and tools to be able to reproduce "evidence"? Interviewees described several examples to illustrate how researchers are relying on proprietary or experimental tools in mining web content and conducting network analysis.

Assessment, Evaluation, and Risk Management

Questions about value of certification and self-audit. Another occurring theme was the need to assess the implementation and utility of trustworthiness, certification, and self-auditing processes and to link the use of such tools with preservation metrics so that we have a systematic and recurrent way of assessing progress and gaps.[18] The Center for Research Libraries has applied their Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC) metrics to evaluate commercial and publisher digital asset management systems and platforms to assess if there are safeguards in place so that the source continues to be functional and useful to the community. "The goals and process for certification need to be re-examined but don't throw the baby with the bath water," cautioned one expert.

Inadequate risk assessment and management strategies. It was repeated over and over that the age of comprehensiveness in preservation (similar to collection building) is over: "It is time to shift our focus to risk assessment beyond failure and understand how to document loss and the associated implications for interactivity."[19] Such a strategy will

---

[18] There are several accreditation, trustworthiness, and self-assessment tools such as TRAC, NESTOR SEAL, DRAMBORA, ISO 16363, and CireTrustSeal.

[19] A reviewer of this paper suggested this related source as an example of risk mitigation strategies: José Barateiro, Gonçalo Antunes, Filipe Freitas, and José Borbinha, "Designing Digital Preservation Solutions: A Risk Management-Based Approach," *International Journal of Digital Curation* 5, no. 1 (2010), https://doi.org/10.2218/ijdc.v5i1.140. Illustrating a different opinion, one interviewee felt that digital preservation has always been about risk management; however, individuals who are less familiar with the best practices may not be cognizant of this principle. Agreeing with this perspective, another individual noted, "there seems to be a lack of awareness of the degree to which preservation has often been about deciding what not to save."

necessitate that libraries establish closer partnerships with companies that are creating or providing access to digital content to consider both organizational and technical risks. Another interviewee expressed concerns about the lack of accounts of failure and loss in digital preservation discourse: "More candid discussions around loss and failure will promote openness and transparency in our community and help us with risk management."[20]

"It is time to shift our focus to risk assessment beyond failure and understand how to document loss and the associated implications for interactivity"

Lack of assessment metrics. One of the interviewees recalled Clifford Lynch's statement about digital preservation being "a metric that has defied measurement."[21] There is broad recognition about the importance of assessment and outcome measures. However, there do not seem to be sufficient collaborative approaches to explore what constitutes success and how we identify and measure outcomes associated with digital preservation.

State of Digital Preservation: Potential Research Areas

The advancements in the digital preservation realm since the call by the Task Force on Archiving of Digital Information in 1996 is quite remarkable.[22] The experts interviewed highlighted several important developments during the last two decades that have started scaffolding a technical and organizational infrastructure. Nevertheless, there is continuing anxiety and a feeling of being left behind as the concept of what constitutes digital information and what preservation means in the realm of digital culture continues to change. Although archival and preservation methods and tools have evolved to respond to these changes, some assumptions and expectations are still influenced by the traditional practices. It is increasingly difficult to isolate the user experience without taking into consideration various contextual issues such as the software required to make

---

[20] The interviewee who made this point cited this report for further information: Carly Dearborn and Sam Meister, "Failure as Process: Interrogating Disaster, Loss, and Recovery in Digital Preservation," *Alexandria: The Journal of National and International Library and Information Issues* 27, no. 2 (2017): 83-93, http://journals.sagepub.com/doi/pdf/10.1177/0955749017722076.

[21] Joint Conference on Digital Libraries, 2006.

[22] Donald Waters and John Garrett, *Preserving Digital Information, Report of the Task Force on Archiving of Digital Information,* 1996, https://www.clir.org/pubs/reports/pub63/.

sense of preserved data or the application needed to experience new media artwork. The key to digital preservation is sustaining interactivity and variability to support future uses in addition to considering the core archival principles such as authenticity, fixity, and integrity.

We would like to conclude by proposing three overarching issues we believe may be fruitful to explore:

1. Although there has been significant progress on many fronts, what seems to be missing is a cohesive and compelling roadmap to guide the international community in knitting together the advances made and addressing the gaps based on the characteristics of the new digital realm. Creating such a roadmap would require first the development of a shared understanding of what digital preservation comprises, articulating the key technical and sociocultural problems, identifying impediments and barriers including resource limitations, and developing a stepwise action agenda accordingly.

- What are the conditions that would lead to an effective collaboration to explore the critical issues holistically in a steady and transparent manner?

- How can we build on the successes to date to inform a realistic vision for continuing the incremental advances?

2. At the crux of the challenges faced are the notions of ownership and control. The scholarly record and the primary sources necessary to inform tomorrow's scholarship are both increasingly more distributed, entailing a range of discovery, access, and use arrangements.

- What are the distributed, distinctive, and complementary roles and responsibilities of the key stakeholders such as research libraries, archives, preservation service providers, publishers, scholarly societies, federal agencies, funders, and other content creators and distributors in ensuring long-term access for future use?

- How can the digital preservation community coordinate the stewardship of valuable and vulnerable materials in a systematic and transparent way *at scale* during a time when the community itself is expanding in scope and variety?

3. Digital preservation programs are expensive and complicated to build and sustain, especially at an institutional level, relative to the demonstrated value they provide.[23] We need to make a case -- or perhaps a series of many cases -- for archiving our cultural heritage based on a strong set of value propositions (both from the public good and economic viability perspectives) and articulate the risks involved in potential loss.

- What future problems and risks are we trying to solve or mitigate?[24]

- What are the measurable and demonstrable benefits of digital preservation that can be promoted to present preservation as a communal responsibility that deserves funding?

- Are there ways to define a cohesive scope by mapping out the various content types in the digital realm that are unlikely to survive without the community's coordinated intervention?[25]

The purpose of conducting and summarizing these interviews was to examine the perspectives of a group of experts on the state of digital preservation with an emphasis on identifying areas for further research. This was not based on a balanced sampling of individuals involved in different stages of digital preservation, holding distinctive roles (e.g., senior library leaders, director of preservation services, preservation specialist/technician, IT specialist, archivist, etc.). Therefore, we are interested in opening up these preliminary findings for a community review to further expand and refine them. Do we capture the successes and needs in this arena as you see them? What is missing? Our objective in conducting this preliminary study is to offer a starting point in order to gather further input towards the development of a potential research and policy agenda.

---

[23] The DPC's *Make the Case for Digital Preservation in Your Organisation* webpage provides some helpful ideas and examples for promoting the importance of digital preservation within an organization. Available at https://www.dpconline.org/knowledge-base/make-the-case

[24] This comment from a reviewer of a draft version of this paper captures a critical issues, "The main risk is that one assumes that 'somebody else' will take care of the digital information."

[25] There are two related suggestions from the interviewees relevant to this point. The DPC's "Bit List" of Digitally Endangered Species is a crowd-sourced strategy to identify digital materials that are most at risk, available at https://www.dpconline.org/our-work/bit-list. Several communities, such as web archiving, AV, software, email, have articulated their challenges and concerns; however, such information remains scattered.

Acknowledgements

I am grateful for the valuable comments and suggestions from my Ithaka S+R colleagues Roger Schonfeld, Kimberly Lutz, and Deanna Marcum. I greatly benefited from the thoughtful feedback to a draft version of this paper provided by Svein Arne Brygfjeld (National Library of Norway), Kathleen Fitzpatrick (Michigan State University), Mike Furlough (HathiTrust), Chip German (APTrust), Steve Knight (National Library of New Zealand), Carol Mandel (NYU Libraries), Michelle Paolillo (Cornell University Library), Bernie Reilly (Center for Research Libraries), and William Kilbride (Digital Preservation Coalition). I am humbled by the digital preservation community's deep expertise, generosity in sharing their insights, and commitment to advancing the field.