Mapping the Current Landscape of Research Library Engagement with Emerging Technologies in Research and Learning:

Stewarding the Scholarly and Cultural Record

By Sarah Lippincott

Edited by Mary Lee Kennedy, Clifford Lynch, and Scout Calvert

May 27, 2020

ASSOCIATION OF RESEARCH LIBRARIES

born-digital RESEARCH + CONSULTING

cni Coalition for Networked Information

EDUCAUSE

Table of Contents

This is the fourth installment of a forthcoming report, *Mapping the Current Landscape of Research Library Engagement with Emerging Technologies in Research and Learning*, that will be published in its entirety by late spring 2020.

The following installments are being published as they become available at https://doi.org/10.29242/report.emergingtech2020.landscape:

- Executive Summary [published March 26, 2020]
- Introduction, Methodology, and Cross-Cutting Opportunities [published April 2, 2020]
- Facilitating Information Discovery and Use [published April 14, 2020]
- Stewarding the Scholarly and Cultural Record [published May 27, 2020]
- Advancing Digital Scholarship
- Furthering Learning and Student Success
- Building and Managing Learning and Collaboration Spaces

Landscape Overview

Libraries bear responsibility not only for providing immediate access to broad and deep research collections, but for the long-term preservation of the scholarly record and the documentary evidence that comprises society's digital cultural heritage. The practices of information stewardship are being challenged by an expanded scholarly and cultural record that is "mutable and dynamic,"[1] unwieldy in its size and complexity, inextricably networked (that is, dependent on other components for context and interpretation), and ephemeral.[2] Many digital outputs are created within closed systems using proprietary technologies that further complicate content harvesting and preservation. Digital formats also pose new challenges for libraries in ensuring authenticity of digital content. Memory institutions are built on trust: the trust that materials under their stewardship are authentic, immutable, and preserved in perpetuity or deaccessioned through a transparent and well-understood process.

The complexity of digital stewardship, and the inversion of value brought about by the networked environment, make preservation of local collections all the more critical. Unique holdings, rather than mass-distributed scholarly resources, are becoming the research library's most valuable assets; libraries have a key role in stewarding this "hyperlocal digital memory."[3]

Stewarding the digital record requires new approaches to managing, "in a transparent and authentic way, support and context for the massively increasing volume of digital content at levels of rapid upward scalability."[4] All of these characteristics of the digital record—its diversity, scale, ephemerality, disaggregation of scholarly communications, and restrictive licensing of digital content—complicate this challenge. They require that memory institutions engage in proactive, upstream, capture processes, rather than the retroactive collecting that has characterized archival and collection development work for centuries.[5]

Yet, while funding and cooperation around mass digitization of physical artifacts has been robust over the last two decades, a similar approach has yet to crystallize for born-digital materials. A proactive approach to the preservation of the born-digital record requires technical, social, and legal solutions. Several of the experts interviewed for this report indicated a pressing need for coordinated, cross-institutional collaboration in order to adequately preserve the digital scholarly and cultural record.

The following sections explore several of the emerging technologies that pose new challenges and offer new solutions to managing digital content throughout its life cycle. These sections address the library's role in advancing open research and publishing practices, reinforcing integrity and trust in the scholarly and cultural record, and preserving the evolving scholarly and cultural record.

Strategic Opportunities

Advance open research and publishing practices

Long-term preservation is in some ways contingent on, or at least the beneficiary of, advances in open scholarship. By supporting open research practices—including the adoption of open metadata standards, creation of machine-readable publications, and depositing outputs (including underlying data and code) in open repositories—libraries make research more discoverable, reusable, reproducible, and durable. Libraries themselves have established open access publishing programs, leveraging new and existing technology infrastructure to develop, host, and distribute scholarly and creative works.[6] Libraries also play a critical role in achieving FAIR (findable, accessible, interoperable, and reusable) research data through their curation, education, and preservation activities.[7] Realizing the vision of FAIR scholarship will be a central challenge for the research community over the next decade.[8] Supporting and engaging in open research and publishing practices improves both the quality of scholarship itself and the quality and manageability of the scholarly record.

The ease of publishing digital content has engendered a shift away from a federated scholarly record produced by established journal and monograph publishers and distributed through libraries. Decentralization of the scholarly record into an assortment of institutional repositories, disciplinary repositories, social sharing sites, small web-only publications, personal blogs, and other channels, creates the need for a more resource-centric approach to dissemination, discovery, evaluation, and preservation of scholarship. Resource-centric scholarly communications relies on making research outputs "discretely exposed, portable, networked, and pluggable in a common way, presenting a rich content layer that serves as the foundation for the development of value added services, like peer-review, social networking, recommender systems, usage measures, and so on."[9] In an environment of "network-enabled literature," content filtering, currently enabled through peer review of individual papers for particular journals, will be superseded by "powerful, online filters" that "distil communities' impact judgements algorithmically, replacing the peer-review and journal systems."[10] The application of machine learning (ML) in scholarly communications processes could accelerate this trend, potentially replacing traditional publishing processes with "a set of decentralized, interoperable services that are built on a core infrastructure of open data and evolving standards."[11]

Many of the experts interviewed for this report cited research libraries' contributions to advocating for and facilitating the use of unique, persistent identifiers as key to enabling this new model of open scholarship and scholarly communications, calling such identifiers "crucial" and "paramount" at every stage of the research life cycle. Unique, persistent identifiers for research outputs can help define provenance, enable discovery, and ensure researchers receive appropriate credit for their work, among other important uses.[12] They are also imperative to support a shift toward a more "researcher-centric model of scholarly communication," in which individual scholars themselves become a key organizing principle.[13] This shift, which is evident in new tools that facilitate discovery, collaboration,

impact assessment, and other scholarly communications activities, depends on the ability of individual researchers to assert and define their unique digital identity and associate it with their intellectual outputs, their collaborators, affiliations, credentials, and other information. A 2018 survey of scientific researchers found that many are "actively engaged in defining their online identity to assert links to their work and communicate their research beyond conventional channels."[14] The authors cited ORCID[15] as the most widely adopted researcher identifier. Research libraries can contribute to addressing ongoing challenges related to the adoption and utility of persistent identifiers. For example, identifier registries remain siloed and limited in their scope: major services such as ORCID, CrossRef, and DataCite focus on one segment of the identifier landscape (researchers, articles, and data sets, respectively) and do not adequately cover the entities that comprise the scholarly communications network.[16] With their expertise in standards and discovery systems, and their relationships with the research community, research librarians are well-positioned to collaborate with identity registries to promote interoperability, encourage common practices, and move towards a more networked scholarly communications system.

Highlighted initiatives

Next Generation Repositories
Confederation of Open Access Repositories (COAR)
https://ngr.coar-repositories.org/
The COAR Next Generation Repositories Working Group aims to achieve interoperability between research repositories by "making the resource, rather than the repository, the focus of services and infrastructure." The group's technical vision centers on encouraging and enabling widespread adoption of unique identifiers to support dissemination and discovery of scholarship, and enable collaboration at scale.

TOME (Toward an Open Monograph Ecosystem)
Association of American Universities (AAU), Association of Research Libraries (ARL), and Association of University Presses (AUPresses)
https://www.openmonographs.org/
A joint initiative of AAU, ARL, and AUPresses, the TOME project is coordinating the production of open access digital monographs in support of a robust and sustainable scholarly publishing ecosystem. The project distributes its outputs through multiple open repositories.

Reinforce integrity and trust in the scholarly and cultural record

Memory institutions are built on trust: the trust that materials under their stewardship are authentic, immutable, and preserved in perpetuity or deaccessioned through a transparent and well-understood process. Emerging technologies pose new challenges to fulfilling the role of trusted steward. The assurance of authenticity, for example, is threatened by the ease of manipulating and altering digital media, and the complexities of determining provenance of digital materials. Deep fakes—counterfeit video, audio, still images, and textual content created using ML—pose a particular challenge. Research libraries have a range of digital forensics tools at their disposal to authenticate digital artifacts and collections at the time of accession and throughout their life cycle. They are also identifying secure pathways—possibly involving distributed ledger technologies (such as blockchain) and public key infrastructure (PKI)—to acquire copies of digital objects from sources they trust, documenting and proving the chain of custody, and any changes that have been made to it along the way.[17] After accessioning, fixity checking continually proves objects and collections do not change over time, due to degradation of the content, or to intentional or accidental manipulation. Underlying all of these processes is the need to maintain security and integrity of computing and storage operations in the face of cyberattacks and natural disasters. In their roles as educators, librarians can also help their constituents develop the skills needed to assess and critically engage with the integrity and reliability of information.

Fraudulent or altered content could enter the historical record and be presented as reality either because its inauthenticity was not detected at the time of accessioning, or because bad actors were able to introduce it by hacking into a records management system.[18] Even regular curatorial practices introduce opportunities for content alteration. For example, the practice of offering access copies of digital archival materials in non-original formats (for example, providing an MP4 video file in lieu of an original in an obsolete or proprietary format) significantly improves the usability of digital archival content, but also creates an opening for nefarious or incidental changes during the format conversion process.[19] If these changes go undetected and undocumented, they could have serious implications for research integrity.

ML makes such manipulation of content by bad actors attainable at scale. Individuals, corporations, and governments can engage in ever more sophisticated forms of information control, taking advantage of the curation algorithms that serve digital content, thereby "recursively intermediating our realities according to evolving internal logics that we cannot see."[20] Bad actors may also be motivated and increasingly able to "tamper dynamically with the historical record."[21]

As the gravity and imminence of threats to the integrity of the historical record become increasingly apparent, librarians, archivists, and their collaborators are exploring new methods to ensure and reinforce trust in cultural heritage institutions as stewards. Ideally, workflows and technological protocols document an immutable chain of custody, providing an assurance of authenticity throughout a digital object's life cycle. Emerging technologies can also be applied to authenticating digital records, that is, tracking their provenance and chain of custody (for example, using distributed ledger systems) and comparing suspected fakes with a library of authenticated content to identify common elements that may have been co-opted.[22]

Projects including ARCHANGEL, from the National Archives of the UK and the InterPARES TrustChain project, for example, are exploring

the application of distributed ledger technologies (such as blockchain) as a tool for ensuring the integrity of digital archival records. The ARCHANGEL project aims to use "blockchain to record checksums (cryptographic hashes) and other metadata derived from either scanned physical records or born-digital records to allow verification of their integrity over decade- or century-long time spans" and to preserve those hashes in a distributed peer-to-peer network.[23] The project is also experimenting with the use of deep neural networks to refine the process of ensuring the integrity of records while allowing broader access. For example, the project is using research from the University of Surrey Computer Vision Centre "to create a hash which is invariant to changes in format, but changes more drastically if the file is manipulated in other ways."[24] This means that a video file converted to a more accessible format could be verified as an authentic version of the original, while one with frames removed would be flagged as altered. The commercial service ARTiFACTS[25] provides blockchain-based registration of a scholar's intellectual outputs, allowing them to manage their intellectual property prior to publication and validate the origins of outputs attributed to them. ORCID, which provides unique, persistent identifiers for scholars, recently announced integration with ARTiFACTS, making it easier for scholars to link their scholarly identity to their research outputs.

A number of scholars have problematized the use of distributed ledgers for ensuring archival integrity and have pointed out the discrepancies between blockchain's theoretical advantages and the reality of implementation. The promise of blockchain as a comprehensive digital preservation solution may be exaggerated. At this time, blockchain technology has only demonstrated success in addressing one component of digital preservation: ensuring the integrity of metadata records.[26] The digital objects themselves are not integrated into blockchain's distributed network and must undergo a separate preservation process. Other authors have argued that blockchain comes up short even in accomplishing its core goal of ensuring authenticity. The premise that blockchain's distributed peer-to-peer networks

ensure their neutrality, has received scrutiny, given that their operation generally depends on a core group of developers.[27] Longstanding methods of ensuring authenticity—entangling hashes and the protocol underlying the LOCKSS system—exceed blockchain's capabilities to provide "tamper-resistant storage against a powerful adversary."[28]

While blockchain may not "fundamentally alter archival practices," it may have a place as one element of the digital archivist's technology toolkit.[29] Blockchain will not replace other methods of ensuring provenance; will not obviate the need for migration, emulation, and other core approaches to content preservation; and will not eliminate the possibility of accidental or malicious corruption of digital records. However, it may become a useful underlying technology in records management systems and one method among many for ensuring the integrity of the scholarly and cultural record.

In addition to technologies that securely document provenance, collections stewards also need tools to detect altered or manipulated content in order to make strategic curatorial decisions: either refusing to accession the object or ensuring it is appropriately described. ML-powered tools can help effectively identify subtle indicators of faked media. For example, researchers have successfully used video analysis algorithms to analyze eye-blinking and detect heartbeats in order to identify fake videos.[30] These techniques are precarious, as the creators of deep fakes continuously enhance their processes to elude detection.

To maintain their status as trusted sources of information, memory institutions will need to deeply engage with current societal debates on the nature of trust and trusted systems. Decentralization and distribution, such as blockchain's distributed ledger, have emerged as new and explicitly anti-institutional methods of establishing provenance and authenticity. Blockchain's original use case as a cryptocurrency system, for example, was developed out of a distrust of traditional banks and financial institutions for financial transactions. Memory institutions have long relied on more traditional notions of institutional trust, a form of trust that is rapidly eroding along with

trust in governments and many other institutions. Memory institutions face a formidable challenge moving forward: maintaining their current status as authoritative keepers of the historical record, while also embracing emerging technologies that distribute, decentralize, and open up digital trust relationships.

Highlighted Initiative

ARCHANGEL
UK National Archives
ARCHANGEL, from the National Archives of the UK and the InterPARES TrustChain project, is exploring the application of distributed ledger technologies (such as blockchain) as a tool for ensuring the integrity of digital archival records. The ARCHANGEL project aims to use "blockchain to record checksums (cryptographic hashes) and other metadata derived from either scanned physical records or born-digital records to allow verification of their integrity over decade- or century-long time spans" and to preserve those hashes in a distributed peer-to-peer network.[31]

Preserve the evolving scholarly and cultural record

A complex and expanding digital record has amplified the technical, social, and legal barriers to achieving digital preservation at scale. Over the last several decades, research libraries and their collaborators have made impressive headway in core digital preservation methodologies such as normalization, refreshing, migration, and emulation.[32] Yet, longstanding challenges have persisted even as new ones emerge. On the technical front, software, 3D data, dynamic web content, and massive data sets, among other media, push the limits of established digital preservation practices. The sheer volume of digital information produced each year means only a fraction can be reasonably preserved.

On the social and legal fronts, the increasingly distributed and licensed nature of scholarly content presents legal and administrative barriers. In addition to copyright challenges posed by digital materials, much "substantive digital content" resides within "proprietary social media

systems and news platforms, potentially requiring agreements with a complex array of private entities to acquire or rescue content for preservation."[33] Content that resides within proprietary platforms is also particularly at risk of being irrevocably lost, as evidenced by numerous examples of abrupt service shutdowns that allowed little time for users or other entities to migrate data.[34] As libraries and archives contend with ever-growing quantities of digital information, the financial and human resources required to perform digital preservation at scale present a growing challenge. Even as digital storage costs continue to decline dramatically, they remain prohibitive for institutions preserving petabytes of data. Making archived data instantaneously retrievable, a core goal of many digital preservation efforts in academic libraries, exacerbates these costs. Increasingly, the environmental impact of storing digital information is coming under scrutiny.[35]

Even as emerging technologies have destabilized the digital preservation environment, they also offer new solutions and opportunities. Libraries and their collaborators are following developments in containerization, distributed ledger technologies (blockchain), new storage media, and automation of digital preservation practices through ML to help ensure that the expanded scholarly record remains accessible well into the future.

The expanding range of file types and formats that require preservation—from software and code, to three-dimensional data, to dynamic websites—presents a daunting challenge. As contemporary academic research moves away from static, immutable, end products, and towards dynamic and diverse networks of outputs, the assets that require digital preservation grow exponentially. Libraries are not only preserving a journal article, for example, but (multiple versions of) data and code that informed its results; comments, annotations, and reactions from the scholarly community; articles that reproduced, validated, or built upon the original scholarship; and more. Data in particular is rapidly moving from a static research product to a continuous flow of information. Libraries lack sufficient

tools and protocols to manage and preserve these streams of networked information.

In order to ensure that the diversity of digital content remains usable over the long term, software preservation is an essential component of any digital preservation program. At the British Library, for example, the digital library program aims to preserve any and all software needed to access the digital objects in its collection, including "software required to open the file directly on current institutional computing technology; the migration and rendering software for such a preservation strategy; and emulators, base operating systems, and any other dependencies necessary to render the digital objects in question."[36]

Several collaborative initiatives aim to make software preservation attainable for libraries of all sizes, including the Emulation-as-a-Service project,[37] led by Yale University Library and supported by the Software Preservation Network, and ReproZip,[38] a software that allows users to capture digital content along with the environment in which it was produced, creating a preservation-ready package. By "capturing computing environments in which research takes place, [ReproZip] could be used to preserve software down to the operating system on which it runs."[39]

As researchers use 3D scanning and virtual reality (VR) tools to capture archaeological sites and artifacts, with the goal of preserving the world's cultural heritage, they "are doing little to conserve their own digital products."[40] The data underlying 3D and VR models is complex and varied, and often requires specific software for reuse and interpretation. Algorithmically generated 3D data (such as data produced through 3D scanning) is particularly difficult to "decouple from the technology used to create" it.[41] A lack of tools, standards, and other resources for 3D and VR data curation means that "in many cases scholars are still reduced to creating screenshots or video documentation of their VR/AR experiences, at least for archival purposes."[42] In the past several years, this problem has received considerable attention. The editors of a 2019 Council on Library and Information Resources (CLIR) report on VR and

3D data in libraries urged the community to "consider 3D/VR as scholarly products in their own right, rather than as illustrations or supplemental material," and therefore worthy of attention to the full suite of data life cycle services, including long-term preservation.[43] A number of libraries are actively engaging in this work.

The University of Virginia (UVA) Library has developed an approach it describes as 3D cultural heritage informatics (LIB3DCHI), which encompasses the "full scope of 3D data curation through the collection, processing, archiving, and distribution of data and its derivatives to the scholarly community."[44] The UVA Library emphasizes access and use, and has implemented Web3D technologies to help conveniently distribute 3D content and data through web browser interfaces. Responding to an "absence of standards and best practices for producing, managing, and preserving 3D and VR content," the collaborative VR Preservation Project[45] will explore metadata standards, infrastructure, and other requirements for preservation to complement the library's active programs supporting VR content creation and use. ML techniques are also being explored as a way to deal with the complexity and breadth of VR data preservation. For example, the game company Electronic Arts (EA), is using ML and AI tools to automate the process of recording every possible interaction with its VR environments in order to capture a comprehensive archival version rather than a single representative experience.[46] As libraries build their own VR and 3D content and advise their communities on best practices, an emphasis on adopting web-native and open 3D formats, where possible, will facilitate use, sharing, and preservation.[47]

One of the most perplexing issues for digital preservationists, web archiving, has only grown more challenging as static websites are replaced by dynamic and personalized feeds of information. The cultural heritage community has not developed sufficient capacity to capture web content, including "contemporary source materials like news, blogs, and online discussion forums," that 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."[48] As

scholars increasingly seek to produce web-based outputs such as digital humanities projects and interactive visualizations in addition to or in lieu of more traditional publications, research libraries are experiencing high demand for digital preservation services. As long as these digital outputs remain fragile and ephemeral, they face significant obstacles to being considered equivalent to more durable forms of scholarship in tenure and promotion considerations.

Typical web crawlers, while able to operate at scale, lack the capacity to harvest dynamic information, instead gathering static snapshots.[49] Technologies like Webrecorder,[50] which allows users to capture live interactions with websites, offer one method of logging a representation of a website for long-term preservation, though these intensive methods break down at scale.[51] For scalable solutions to web archiving, researchers are exploring a number of options, including the potential application of a new web packaging standard introduced by Google in 2019[52] and the use of human-mediated web capture frameworks that can apply a set of heuristics defining elements to be captured to an entire class of web publications rather than individual websites.[53]

Finally, even as storage grows cheaper and more efficient, research libraries face an exponentially mounting volume of digital information; storage capacity remains a fundamental challenge for institutions aspiring to achieve large-scale digital preservation. The use of local and cloud servers for digital hosting and preservation seems likely to remain ubiquitous in the cultural heritage community as emerging storage options offer only marginal improvements over current technologies, or are far from ready for widespread adoption. Many emerging technologies are also unsuitable for providing instantaneous access, making them incompatible with the goals of many library digital preservation programs. Gains are being made as "engineers continue to eke out further performance and capacity gains from hard drives and flash storage—and researchers are developing next-generation technologies such as DNA storage, crystal etching techniques, and molecular storage that could hold massive amounts of data on a small object for hundreds of thousands of years or longer."[54]

Even as emerging technologies begin to provide solutions for automating the digital curation life cycle, it remains an expensive process that entails significant human intervention and judgment.[55] Curation needs depend on the nature of the collection: the format and characteristics of its contents, its intended uses and audiences, its sensitivity and cultural context. Collections that contain ethnographic materials or collections pertaining to or of marginalized communities require culturally appropriate curation methods that align with the values and interests of those communities.[56] Digital curation is therefore an active and collaborative process that requires interdisciplinary expertise and resists large-scale automation.

Highlighted Initiatives

Emulation-as-a-Service Infrastructure (EaaSI)
Yale University Library
https://www.softwarepreservationnetwork.org/projects/emulation-as-a-service-infrastructure/
EaaSI is building a network of institutional partners to build capacity for emulation beyond what any individual institution can offer. The program aims to offer third-party emulation services for memory institutions that allow them to provide access to digital media in an interactive (and where appropriate, secure and restricted) format via a standard web browser.[57]

VR Preservation Project
University of Oklahoma Libraries
http://vrpreservation.oucreate.com/
The University of Oklahoma Libraries aim to develop a set of common standards and best practices for the archiving and preservation of VR-related data. Led by Zack Lischer-Katz, CLIR Postdoctoral Fellow in Virtual Reality Preservation and Archiving for the Sciences, the two-year project will focus on developing both the guidelines and technologies necessary for VR content and software preservation.

Key Takeaways

- **The growth of dynamic, networked, interactive information presents new challenges for digital preservation.** The scale, diversity, and complexity of digital artifacts complicates efforts to effectively steward the scholarly record and digital cultural heritage. The prevalence of dynamic digital formats such as VR, the curation of the web by inscrutable algorithms, and the siloing of digital content in proprietary formats and platforms, create obstacles for achieving large-scale digital preservation. Preservation of born-digital content depends not only on appropriate technologies to capture and curate it, but on the upstream practices that make content discoverable and harvestable. Both open standards and open licensing therefore are imperative to enabling collection and stewardship of scholarly information.[58]
- **The ability to easily manipulate digital archival materials threatens trust in memory institutions.** Malicious actors, including individuals, corporations, and governments, have more methods than ever before to attempt to rewrite history through the creation of deep fakes, exploiting file format changes, and hacking digital archives. Memory organizations rely on trust from communities they serve that the information they provide accurately reflects the historical record, and this trust cannot easily be regained after it has been lost.
- **Emerging technologies present both new solutions and challenges for long-term digital preservation.** Containerization technologies, advances in emulation, distributed ledgers, and ML tools all provide promising new approaches to long-term digital preservation. However, many digital preservation efforts are rooted in a historical, print-centric model of retroactive collecting and need to transition to coordinated and proactive upstream processes.
- **Digital preservation at scale requires collaboration.** Many individual institutions are engaging in innovative digital preservation initiatives. However, achieving trustworthy, representative digital preservation at scale requires that these

technologies become part of a coordinated, cross-institutional, or even national approach to digital preservation. This coordinated approach must leverage institutional strengths and capacity, and also requires that research libraries continue to advocate for the adoption of the open standards, technologies, and protocols that make digital content available for harvesting and curation. Combining open standards and technologies with collaborative governance will allow for a more comprehensive approach to preserving the digital historical record.

Endnotes

1. Brian Lavoie et al., *The Evolving Scholarly Record* (Dublin, OH: OCLC Research, 2014), https://doi.org/10.25333/C3763V.
2. Carol A. Mandel, *Can We Do More? An Examination of Potential Roles, Contributors, Incentives, and Frameworks to Sustain Large-Scale Digital Preservation* (Arlington, VA: Council on Library and Information Resources, September 2019), https://clir-dlf.app.box.com/s/31tc6nrua3cj8jjwoymee78gl3plzlo2.
3. Jason Griffey, interview by author, November 14, 2019.
4. NDSA Coordinating Committee and NDSA Working Group co-chairs, *2015 National Agenda for Digital Stewardship* (National Digital Stewardship Alliance, September 2014), http://hdl.loc.gov/loc.gdc/lcpub.2013655119.1.
5. Mandel, *Can We Do More?*
6. *Library Publishing Directory 2020* (Atlanta, GA: Library Publishing Coalition, 2020), https://librarypublishing.org/wp-content/uploads/2020/02/library_publishing_directory_2020.pdf.
7. Mark D. Wilkinson et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3 (2016), https://doi.org/10.1038/sdata.2016.18.

8. Directorate-General for Research and Innovation (European Commission) et al., *Future of Scholarly Publishing and Scholarly Communication: Report of the Expert Group to the European Commission*, Copyright, Fair Use, Scholarly Communication, etc. 97 (Luxembourg: Publications Office of the European Union, January 2019; Lincoln, NE: University of Nebraska–Lincoln Libraries, January 2019), https://digitalcommons.unl.edu/scholcom/97.

9. Confederation of Open Access Repositories, Eloy Rodrigues, and Kathleen Shearer, *Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group*, Copyright, Fair Use, Scholarly Communication, etc. 64 (Lincoln, NE: University of Nebraska–Lincoln Libraries, November 28, 2017), https://digitalcommons.unl.edu/scholcom/64.

10. Jason Priem, "Beyond the Paper: The Journal and Article Are Being Superseded by Algorithms That Filter, Rate and Disseminate Scholarship as It Happens," *Nature* 495, no. 7442 (March 28, 2013): 437–41, https://doi.org/10.1038/495437a.

11. Priem, "Beyond the Paper."

12. Angela Dappert et al., "Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research," *Data Science Journal* 16 (June 15, 2017): 28, https://doi.org/10.5334/dsj-2017-028.

13. Roger C. Schonfeld, "Who Is Competing to Own Researcher Identity?," *Scholarly Kitchen*, January 6, 2020, https://scholarlykitchen.sspnet.org/2020/01/06/competing-researcher-identity/.

14. Michela Bello and Fernando Galindo-Rueda, "Charting the Digital Transformation of Science: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2)," OECD Science, Technology and Industry Working Papers no. 2020/03 (Paris: OECD Publishing, April 9, 2020), https://doi.org/10.1787/1b06c47c-en.

15. ORCID website, accessed May 19, 2020, https://orcid.org/.
16. Dappert et al., "Connecting the Persistent Identifier Ecosystem."
17. Mark Bell et al., "Underscoring Archival Authenticity with Blockchain Technology," *Insights* 32, no. 1 (2019): 21, https://doi.org/10.1629/uksg.470.
18. Melanie Ehrenkranz, "How Archivists Could Stop Deepfakes from Rewriting History," News, *Gizmodo*, October 16, 2018, https://gizmodo.com/how-archivists-could-stop-deepfakes-from-rewriting-hist-1829666009.
19. Bell et al., "Underscoring Archival Authenticity with Blockchain Technology."
20. Glenn D. Tiffert, "Peering Down the Memory Hole: Censorship, Digitization, and the Fragility of Our Knowledge Base," *American Historical Review* 124, no. 2 (April 2019): 550–68, https://doi.org/10.1093/ahr/rhz286.
21. Tiffert, "Peering Down the Memory Hole."
22. Jacquelyn Burkell and Chandell Gosse, "Nothing New Here: Emphasizing the Social and Cultural Context of Deepfakes," *First Monday* 24, no. 12 (December 2, 2019), https://doi.org/10.5210/fm.v24i12.10287.
23. Bell et al., "Underscoring Archival Authenticity with Blockchain Technology."
24. Bell et al.
25. ARTiFACTS website, accessed May 19, 2020, https://artifacts.ai/.
26. Angela Woodall and Sharon Ringel, "Blockchain Archival Discourse: Trust and the Imaginaries of Digital Preservation," *New Media & Society*, November 22, 2019, https://doi.org/10.1177/1461444819888756.

27. Victoria L. Lemieux, "Blockchain Recordkeeping: A SWOT Analysis," *Information Management* 51, no. 6 (2017): 20–27.

28. David Rosenthal, "Blockchain Solves Preservation!," *DSHR's Blog*, September 13, 2018, https://blog.dshr.org/2018/09/blockchain-solves-preservation.html.

29. Victoria Louise Lemieux, "Trusting Records: Is Blockchain Technology the Answer?," *Records Management Journal* 26, no. 2 (2016): 110–39, https://doi.org/10.1108/RMJ-12-2015-0042.

30. Burkell and Gosse, "Nothing New Here."

31. Bell et al., "Underscoring Archival Authenticity with Blockchain Technology."

32. Oya Y. Rieger, *The State of Digital Preservation in 2018: A Snapshot of Challenges and Gaps*, Issue Brief (New York: Ithaka S+R, October 29, 2018), https://doi.org/10.18665/sr.310626.

33. Mandel, *Can We Do More?*

34. Richard Ovenden, "Libraries' Role in Preserving Digital Information," *Carnegie Reporter*, October 30, 2019, https://www.carnegie.org/news/articles/libraries-role-in-preserving-digital-information/.

35. Morgan David and Yuji Sekiguchi, "The Role of Optical Storage Technologies in Future Digital Archives," in *Sustainable Audiovisual Collections through Collaboration: Proceedings of the 2016 Joint Technical Symposium* (Bloomington, IN: Indiana University Press, 2017), 170–76.

36. Peter May, Maureen Pennock, and David A. Russo, "The Integrated Preservation Suite: Scaled and Automated Preservation Planning for Highly Diverse Digital Collections" (paper presented at iPRES 2019: 16th International Conference on Digital Preservation, Amsterdam, The Netherlands, 2019), 10, https://osf.io/smq5w/.

37. Emulation-as-a-Service website, accessed May 19, 2020, https://www.softwarepreservationnetwork.org/projects/emulation-as-a-service-infrastructure/.

38. ReproZip website, accessed May 19, 2020, https://www.reprozip.org/.

39. Vicky Steeves, Rémi Rampin, and Fernando Chirigati, "Using ReproZip for Reproducibility and Library Services," *IASSIST Quarterly* 42, no. 1 (2018): 14, https://doi.org/10.29173/iq18.

40. David Koller, Bernard Frischer, and Greg Humphreys, "Research Challenges for Digital Archives of 3D Cultural Heritage Models," *Journal on Computing and Cultural Heritage* 2, no. 3 (December 2009): 7:1–7:17, https://doi.org/10.1145/1658346.1658347.

41. Jennifer Moore and Hannah Scates Kettler, "Who Cares about 3D Preservation?," *IASSIST Quarterly* 42, no. 1 (2018): 15, https://doi.org/10.29173/iq20.

42. Victoria Szabo, "Collaborative and Lab-Based Approaches to 3D and VR/AR in the Humanities," in *3D/VR in the Academic Library: Emerging Practices and Trends*, ed. Jennifer Grayburn et al. (Arlington, VA: Council on Library and Information Resources, February 2019), https://www.clir.org/pubs/reports/pub176/.

43. Zack Lischer-Katz et al., "3D/VR Creation and Curation: An Emerging Field of Inquiry," in *3D/VR in the Academic Library: Emerging Practices and Trends*, ed. Jennifer Grayburn et al. (Arlington, VA: Council on Library and Information Resources, February 2019), https://www.clir.org/pubs/reports/pub176/.

44. Will Rourk, "3D Cultural Heritage Informatics: Applications to 3D Data Curation," in *3D/VR in the Academic Library: Emerging Practices and Trends*, ed. Jennifer Grayburn et al. (Arlington, VA: Council on Library and Information Resources, February 2019), https://www.clir.org/pubs/reports/pub176/.

45. "3D Collection Strategies," Virginia Tech University Libraries, accessed May 19, 2020, https://lib.vt.edu/research-teaching/lib3dvr.html; see also VR Preservation Project website, accessed May 19, 2020, http://vrpreservation.oucreate.com/.

46. John Tilbury, "Digital Preservation Futures: Looking ahead to 2030...," *Preservica Blog*, March 20, 2019, https://preservica.com/blog/digital-preservation-futures-looking-ahead-to-2030.

47. Szabo, "Collaborative and Lab-Based Approaches to 3D and VR/AR in the Humanities."

48. Rieger, *The State of Digital Preservation in 2018*.

49. Sawood Alam et al., "Supporting Web Archiving via Web Packaging," preprint, submitted June 17, 2019, http://arxiv.org/abs/1906.07104.

50. Webrecorder website, accessed May 19, 2020, https://webrecorder.io/.

51. Martin Klein et al., "The Memento Tracer Framework: Balancing Quality and Scalability for Web Archiving," in *Digital Libraries for Open Knowledge*, ed. Antoine Doucet et al., Lecture Notes in Computer Science, vol. 11799 (Cham, Switzerland: Springer, 2019), 163–76, https://doi.org/10.1007/978-3-030-30760-8_15.

52. Alam et al., "Supporting Web Archiving via Web Packaging."

53. Klein et al., "The Memento Tracer Framework."

54. Samuel Greengard, "The Future of Data Storage," *Communications of the ACM* 62, no. 4 (April 2019): 12, https://doi.org/10.1145/3311723.

55. Carole Palmer, interview by author, October 30, 2019.

56. Uttaran Dutta, "Digital Preservation of Indigenous Culture and Narratives from the Global South: In Search of an Approach," *Humanities* 8, no. 2 (2019): 68, https://doi.org/10.3390/h8020068.

57. Euan Cochrane, "Making Things EaaSIer: Overview from EaaSI's PI," Software Preservation Network, February 12, 2019, https://

www.softwarepreservationnetwork.org/blog/making-things-eaasier-overview-from-eaasis-pi/.

58. Cameron Neylon, "More Than Just Access: Delivering on a Network-Enabled Literature," *PLOS Biology* 10, no. 10 (October 23, 2012): e1001417, https://doi.org/10.1371/journal.pbio.1001417.