

A perpetual motion machine: The preserved digital scholarly record

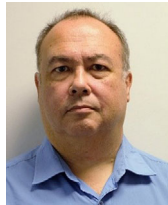
Tom Cramer ¹, Chip German ², Neil Jefferies ³, and Alicia Wise ^{4*}



Tom Cramer



Chip German



Neil Jefferies



Alicia Wise

¹Stanford University Library, Stanford, California, USA

²University of Virginia Library, Charlottesville, Virginia, USA

³Bodleian Libraries, University of Oxford, Oxfordshire, UK

⁴CLOCKSS, London, UK

ORCID:

T. Cramer: [0000-0002-8972-1941](https://orcid.org/0000-0002-8972-1941)

C. German: [0000-0003-3574-7940](https://orcid.org/0000-0003-3574-7940)

N. Jefferies: [0000-0003-3311-3741](https://orcid.org/0000-0003-3311-3741)

A. Wise: [0000-0002-7898-3428](https://orcid.org/0000-0002-7898-3428)

*Corresponding author: Alicia Wise, CLOCKSS, London, UK.

E-mail: awise@clockss.org

Key points

- Digital preservation will never be a solved problem: it needs constant reinvention, and is going to become harder over time.
- Scholarship is changing and this is affecting what needs to be preserved and what preservation means to the future of knowledge discovery.
- The diversification of outputs means that knowledge exists in a network of contextual metadata, data, software, standards and publications—requiring multilateral management of this complex knowledge graph.
- Preservation demands new skills, technologies and resources from librarians, publishers, funders and institutions—and more joined-up thinking about archiving.

Keywords: archive, digital preservation, preservation

INTRODUCTION

Digital objects, unlike a treasured book or vinyl record, have no meaning on their own. They are not static and reified, and they do not remind us of their value by sitting on a shelf within reach with their titles in view on their spines or on their covers. Worse yet, as much as the objects may exist out of sight, so do the threats that could make those objects irretrievable when we need them or when our children or grandchildren will. Clearly, a comprehensive means of preserving, safeguarding and making accessible digital objects for the future, especially those objects that comprise the record of advances in human knowledge, is an

essential foundation for human progress. That is what digital preservation is all about. It gets more complicated from there.

In these early years, we have learned that digital preservation can never be a solved problem. It is work that does not finish, and it becomes harder over time as formats, software and hardware fade into memory, and the creators and publishers move on to new challenges.

Scholars communicate ever more of their discoveries digitally, so the work of choosing what and how to preserve their scholarship requires much more effort. That is just the beginning of the challenges. Digital preservation is also becoming more complex as artificial intelligence, dynamic databases, interactive

resources, knowledge graphs and more expand the scholarly record. How do you preserve an entire online ecosystem in which scholars collaborate, discover and share new knowledge?

The disappointing answer is, of course, we cannot. The practice of preservation lags behind the development of ways of knowing and new ways of communicating knowledge. Therefore, while we work to advance preservation practices in order to catch up, we do what we can. Now, for example, we can take snapshots and preserve meaningful points in a scholar's journey through this ecosystem, and it is essential that we do just that. We also need to capture and preserve the persistent identifiers that link related components of the scholarly record together. They are represented by an alphabet soup of acronyms: DOIs, Funder IDs, ORCID, RORs. We mention them again a bit later in the article.

Digital preservation of human knowledge starts from a simple goal: ensure that the rapidly changing scholarly record exists into the future so students and researchers can learn from it and build on it. That, as all preservation practitioners know, is much more easily written in a sentence than done. It requires first an understanding of the social, organizational and intellectual environments where metadata and content exist, as well as an understanding of the content and technology itself. Acceleration of change in all these spheres is going to drive changes in the way we organize to do the job of digital preservation. It is no wonder we are worried. The challenges on the uncomfortably near horizon are even bigger than the ones we confront today.

SO WHERE ARE WE?

Long-term preservation is at the core of mission for memory organizations—libraries, museums, and archives. Long-term *digital* preservation has been recognized as important since the 1990s (e.g., Joint Funding Council, 1993 which led to Beagrie 2004) as early storage formats and hardware became obsolete, and compelling information was lost despite the existence of backup copies and remote storage.

Increasing diversity of the scholarly record

Early digital preservation initiatives in the scholarly communication field focused on print objects, such as PDF scans of book or journal pages or digitally published journal articles in their native PDF print-optimized formats. For libraries and publishers, this was a logical extension of their approach to the physical artefacts of scholarly discourse that they traditionally handled. Even so, scholars have always communicated using techniques beyond the formally published article, monograph or book, but such additional materials were generally considered outside the purview of publishers and libraries, and of secondary importance to the scholarly record.

The impact of digital technology on scholarly processes has resulted in a broadening of the range of possible scholarly communication outputs and created the potential for multiple

versions of an output to coexist. The wider range of outputs poses challenges to the bibliographic and archival assumptions that have historically served libraries and publishers well. Components such as preprints, data, software, workflows and methodologies all contribute to this more expansive conceptualization of the scholarly record. Scholars are re-evaluating what precisely needs to be preserved and, more crucially, where technical and fiscal responsibility lies.

Libraries and publishers have tracked these changes, needing to employ a new digital focus using a related array of digital skills. Those capabilities are quickly becoming essential competencies for both types of organizations.

New hybrid forms such as data and software 'papers' and methodological pre-registrations have emerged to anchor some of these new outputs in existing scholarly workflows, but those new forms require their own preservation strategies, including how to preserve their connections to the scholarly objects that are related to them.

Librarians have had to develop new descriptive skills and standards, and they will need to supplement those new foundational necessities with museum and archive practices created to address similar needs. Library catalogues and indices have historically been weak on context (references and citations are embedded in documents themselves rather than their metadata), whereas museums devote significant effort to better understand and record the context in which their objects existed. Context is critical to understanding the increasingly diversified scholarly record.

The use of digital methods varies widely across academic disciplines, not just at the aggregated levels of the humanities and social sciences or of science, technology and medicine. Now the methodologies vary at a far more granular level where individual disciplines develop and evolve their own standards and tools. This is a level of complexity and change that librarians have no hope of managing on their own with the resources that they have available. Instead, librarians must increasingly become advocates for digital preservation 'at birth'. Librarians must engage with content creators early in their endeavours to ensure that their creations can be preserved and accessed in the future, and that the relevant metadata can be collected at source. Funders can promote this approach as well, motivating researchers themselves to become the first leg of the ongoing relay race of digital preservation. In a similar manner, curation and selection also become collaborative tasks to be shared by the full range of stakeholders as the volume of digital content grows exponentially.

The importance of the knowledge graph

A direct result of this diversification of outputs is that knowledge is no longer effectively encapsulated in individual artefacts such as papers, but in the network of contextual metadata, data, software, standards and other publications linked to the paper. However, the need to construct, maintain and preserve this collection of connected components, or knowledge graph, is also driven by several additional factors:

- The increasing use of technology and digital methods has led to a crisis of reproducibility (Fidler & Wilcox, 2021) in many disciplines, where papers alone are not sufficient for findings to be reproduced and validated.
- Funders, institutions and other stakeholders are increasingly looking for metrics around the effectiveness of the research process, especially with respect to speed of access to and re-use of resources. This requires that instruments, grants, facilities and a host of other resources are documented in the graph.
- Contributors and creators are concerned that digital objects are correctly attributed, and that broader contributory roles are recognized. This attribution should persist through name and/or gender changes, further complicating the task of identity management in the scholarly record.
- Technology enables more frequent, wider and more diverse collaboration on research activities than has been the norm. Outputs are increasingly the responsibility of multiple organizations.

The maintenance of the graph therefore becomes a multilateral activity that has policy, legal and resource implications for the organizations involved, with libraries, and occasionally publishers, acting as brokers and facilitators. Library digital collections are usually curated and preserved locally, but libraries also licence content from publishers instead of purchasing it outright, so collaborative organizations such as CLOCKSS (<https://clockss.org/>) and Portico (www.portico.org/) have emerged to address the long-term digital preservation of licensed scholarly outputs. Legal deposit libraries around the world have evolved to collect and preserve digital publications and harvest websites in their countries.

However, the decentralized nature of the knowledge graph and the ease with which digital objects can be updated add even more challenges:

1. Multiple jurisdictions—Copyright, privacy and information-security regulations are highly variable across the worldwide range of jurisdictions that digital distribution can cross in an instant. This can lead to interesting conditions where the legal availability of parts of a complex scholarly object may differ depending on the viewer, highlighting the need for increased harmonization and simplification of licensing and access protocols.
2. Validation—For journals, interested parties created the Keepers Registry (<https://keepers.issn.org/>), an international monitoring infrastructure and resource for discovering what has and has not been preserved for the long term. Unfortunately, there is nothing similar for books, data, software or the wide array of other scholarly outputs, although services such as CORE (<https://core.ac.uk/>) in the UK, which have the potential to address this issue, are beginning to emerge. Without such registries, it is hard to determine what is where and whether it is being preserved. Consequently, there may be

gaps in the record of important material, or, conversely, material may be preserved in more places and at more expense than is needed. The persistent identifier services described above provide potential surrogate capabilities to meet the needs, but success depends on their wide adoption.

3. Versioning—The scholarly record has become a more dynamic aggregation of linked components. It changes over time and has fuzzy boundaries. Ensuring that citations can accurately lead to the version(s) of objects current at the time that the citation was made requires repositories to keep and track previous versions of the material that they hold. To do so, they should support standards such as the Memento protocol (<https://mementoweb.org/about/>) to allow direct linking to such time-specific versions.
4. Permanent Identification and Linking—The relationships between digital objects, people and organizations provide essential context, and a host of permanent identifiers have emerged to provide connecting glue via unique identifiers (e.g., DOIs, ORCID, FunderID, ROR, RAIDs; see the National Transportation Museum reference National Transportation Library, 2022 for a more complete description). In addition, few of these resources, with the exception of CrossRef DOIs, are themselves preserved.
5. Retractions—Retractions are an important exception to the versioning requirements mentioned above, in that they need to be actively propagated to earlier versions of the retracted object and reflected in both metadata and directly retrieved files. All parties will need to agree to the additional workflows and policies needed.

Clearly, what we thought the scholarly record was in years past is not what it is today or what it will be. Even defining the term is not easy. In 2014, OCLC distributed a paper (Lavoie et al., 2014) that its research unit had developed on the evolving scholarly record. The authors presented a scholarly record framework, including something they called the ‘stakeholder ecosphere’, represented by Fig. 1:

For our purposes here, we see in the ecosphere two critical actions that are prerequisites for preservation as we understand it today: the actions are labelled ‘fix’ and ‘collect’. Our conclusion is that ‘fixing’, which we might conceive as a broadening of the traditional term ‘publishing’, of a modern digital scholarly record component is not a single event, but a sometimes-unending string of such events related to that component over time. Likewise, ‘collecting’ of that component (dependent as it is on the component being ‘fixed’ each time) can extend as long as that component is relevant to the dynamic scholarly record. If preservation is your calling, anticipate perpetual motion.

Redefining the target audience for scholarship

As we noted above, the use of scholarly outputs has evolved with digital dissemination and changing practices. They must now fill several roles: as dissemination to both scholarly peers and a wider public; as a record for stakeholders with interests in

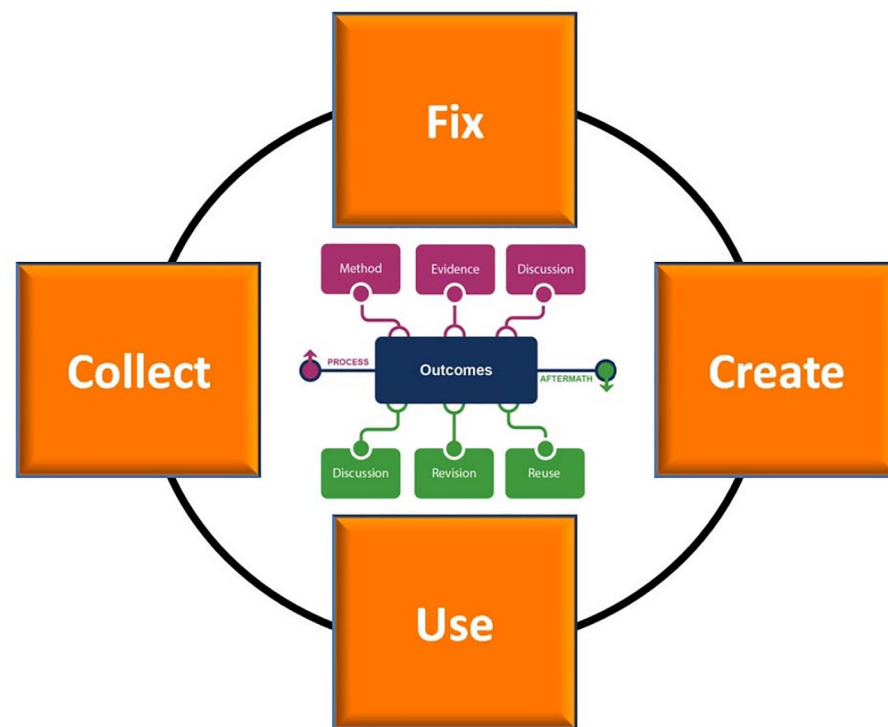


FIGURE 1 Roles in the Scholarly Record's Stakeholder Ecosystem.

Source: Credit: Lavoie et al. (2014).

research outcomes, value-for-money and re-use; and as a mechanism for societal and commercial engagement. This broader audience is also reflected in a more expansive approach to research metrics and reporting, beyond citation counts to an evolving, somewhat less concrete concept of 'impact' that may vary with the stakeholder concerned.

Similarly, it is difficult to make the economic case for long-term preservation without considering access as the essential outcome. Indeed, conceptual preservation frameworks such as OAIS explicitly emphasize the notion of 'target audiences'. However, when and how access is provided to preserved materials can be challenging in a dynamic, decentralized environment for the reasons we outlined in the previous section.

Most creators want their scholarly outputs to become part of the permanent scholarly record and assume that publishers and librarians are already taking care of this task. Realistically, that assumption only applies to those outputs that were already within publishers' and librarians' purview—papers, monographs and books along with their digital analogues—and even then, only partially. As the number of scholarly products in need of digital preservation continues to grow, more organizations and services become involved in the process.

Availability and access alone are not sufficient to reach these expanded audiences effectively. Digital outputs must also be discoverable in a sense that goes beyond the traditional library and publisher search tools. That discovery pathway, using the terms of today, must engage the wider Web. While open-access texts can be indexed by services such as Google and Google Scholar,

good metadata can greatly improve the accuracy of search results. More restricted access and non-textual materials are almost entirely dependent on good quality and open metadata for discovery. For libraries, publishers and other content holders, this means moving on from their domain-specific cataloguing and taxonomic skill sets to adopt Web-based standards and approaches.

It is important to note that making content open access does not ensure that it is preservable or preserved. Laakso et al. (2021) identified 174 open-access journals that completely disappeared from the web between 2000 and 2019, and the authors demonstrated that only 40% of the journals indexed in the Directory of Open Access Journals (DOAJ, <https://doaj.org>) in 2019 were preserved in a digital archive. Bosman et al. (2021) surveyed diamond open-access journal publishers in March 2021 and concluded that 68% of these journals did not have an archiving policy. The JASPER Project (<https://doaj.org/preservation/>) was launched in 2021 to address this preservation challenge. Collaborating project partners have developed infrastructure to enable open access journals that are indexed by DOAJ to be ingested by archives. As we write, a pilot is underway with a group of producers of such journals to understand what information or support they might need to preserve their work.

Shifts in digital preservation technology

Having considered how changing scholarly practice affects those tasked with the digital preservation of scholarly

materials, we also need to examine how digital preservation itself is changing. One key trend, characteristic of maturing markets, is consolidation both in the general technology space but also in digital preservation specifically. We can see it in three key areas:

- *Cloud providers* have emerged and then undergone a consolidation phase leaving a few major players (such as Amazon Web Services, Microsoft and Google) with business models and revenues based primarily on charging for computation and processing services and network utilization. For these, archival and backup storage is essentially an additional supporting service for their primary business, generating a relatively low return. A few specialist storage providers remain, generally with a focus on backup and disaster recovery.
 - *Managed digital preservation and archiving service providers* have emerged as the practice has matured. Some are non-profit collaborations, while others are for-profit businesses. Some resell cloud-based storage, supplemented with their own add-on services, while others maintain their own networks of data centres/storage nodes. These services are targeted at records and documents where practice and standards are reasonably well established. In most cases, the provider integrates a mix of proprietary and open tools into a curation and preservation workflow. CLOCKSS and Portico are examples of this category of providers, but the missions and practices of other entities in it range widely.
 - *Storage technology*, according to industry sources such as Spectra Logic (2022), has consolidated with just three hard-drive manufacturers (Seagate, Western Digital and Toshiba) and two tape-media manufacturers (Fuji and Sony) remaining. These traditional sources for direct purchase of storage technology are receding in visibility as cloud entities have gained prominence in backup, bulk storage and solid-state storage for high-performance access. While cloud providers do use the technologies from storage manufacturers, they typically achieve much greater utilization rates compared with on-premises storage because of resource sharing. Consequently, overall consumption of storage relative to data volumes has become more efficient.
- Consolidation and market pressure in the disk and tape industries has a more direct effect, risking the emergence of a single remaining supplier, which in turn can lead to price increases driven by scarcity and lack of competition. The IBM Magstar tape system is already a single-vendor product. Perhaps the best risk mitigation for this possibility is a feature of the general technical environment: the continuing development of new storage technologies. The explosion of demand for storage of digital materials of all kinds makes such a mitigation effect likely, but not without its own set of future complications for digital preservation (see below).

Preservation is forever; preservation solutions have limited lives

Digital preservation involves looking after content over time periods longer than the existence of any of the underlying technologies and software products, and longer certainly than the existence of many suppliers. Consequently, the ability for preservation organizations to repeatedly develop and execute exit strategies is a basic operational requirement.

Large organizations, although often exhibiting rapid growth, also experience shorter lifespans (Viguerie et al., 2021) and sometimes choose to pivot their business models rapidly to survive. For those adopting cloud storage either directly or indirectly through an added-value service provider, this should give pause for thought. It is essential to understand the risks of such approaches and what steps are needed to mitigate those risks in order to form a view about whether the providers under consideration are reliable partners in the short-to-medium term.

One of the temptations of outsourcing is to reduce or eliminate in-house specialist teams with expert skills in metadata, content-stewardship and applicable technologies to realize greater cost savings. That choice is usually ill-advised. Those skills are vitally necessary when the inevitable exit of a preservation service provider occurs. Migration to new suppliers or technology stacks, especially when a high level of proprietary integration previously took place, is the time that preserved information is most at risk.

The environmental cost of digital preservation

Although adequate redundancy and resilience are core concepts in effective digital preservation, so is environmentally-responsible behaviour (Pendergrass et al., 2019). Most storage forms other than tape require power continuously, and maintenance activities such as periodically checking a digital file for data corruption consumes further energy through additional network activity and processing. Increasingly, risk mitigation strategies need to consider environmental costs. For example, while data corruption is relatively easy to check for, modern storage technologies already have several layers of error checking and recovery built in, so the usefulness of carrying out additional checks is debatable. Similarly, checking for malicious tampering might be more efficiently achieved by logging storage access and performing audits rather than exhaustively rechecking every file. In particular, the energy costs of moving data to and from the cloud in any quantity can be significant—which can be seen in the ingress and egress charges of most providers.

There are existing low-power alternatives to tape such as optical disks and PIQL (film), but they suffer from relatively low density, limited bandwidth and high latency to retrieve data. At present, the alternatives also hold relatively weak positions in the marketplace, with consequent cost and long-term viability concerns. While there are potential low-power technologies on the horizon without these drawbacks—for example, DNA

(www.twistbioscience.com/products/storage) and 5D fused silica (www.microsoft.com/en-us/research/project/project-silica/)—it is not clear that long-term preservation is a large-enough market for them to be developed commercially. Memory organizations and those with aligned interests need to co-operate to make this case and be prepared to invest. In any eventuality, digital preservation activities must fully understand and disclose their environmental impacts and pursue meaningful actions to minimize them.

The digital preservation community

Partly in response to the challenges that we have highlighted but also as a sign of a maturing profession, the digital preservation community has become more organized. Bodies such as the Digital Preservation Coalition (DPC, www.dpconline.org/) and the National Digital Stewardship Alliance (NDSA <https://nds.org/>) are focused primarily on digital preservation, supported by international conferences such as iPres (<https://ipres-conference.org/>) and PASIG (www.preservationandarchivingsig.org/). However, to avoid siloization, the field is also represented in other, broader groups concerned with digital content including the Digital Curation Centre (DCC, www.dcc.ac.uk/) and the Research Data Alliance (RDA, www.rd-alliance.org/).

As we have noted, when preserving materials for decades and even centuries, material is likely to outlast the services preserving it at any one time. Few companies and fewer technologies will survive as long as digital scholarly outputs will be required, so a vibrant and resilient community provides the continuity that such technical constructions cannot. Every preservation activity must consider the next step in the 'relay' of technologies, organizations and systems, enabling the transition to the next phase to be accomplished safely and economically, and working with the community to ensure that there is *always somewhere to go*. Key community resps fall into the following areas.

Skills

Digital preservation is problematic in that it requires a high level of broad technical skills (e.g., <https://digicurv.gla.ac.uk/>) at various points in the archival and preservation lifecycle. In addition to current operational skills, awareness of and engagement with upcoming technologies are essential for planning for future migrations, and for ensuring that systems and standards can accommodate new digital forms and materials. Equally, experience of older or obsolete systems may be required when accessioning material from older sources and projects. Creating communal pools of expertise reduces the challenge of maintaining these less-frequently accessed skill sets that can be beyond the means of a single organization.

In facing the challenges of a changing world of scholarly products, academic libraries and preservation services must manage through disruptive trends that require their own new combinations of skills and practices. Most have found that they need the perspectives of cross-functional teams encompassing

competency areas such as collections, communication, curation, engagement, finance, library/archival science, copyright and licensing, metadata, publishing, software and IT/core systems. As digital preservation increasingly requires resources beyond the walls of the library (which will negotiate partnerships/collaborations for some needs and buy services for others), its leaders must retain in-house skills in several important areas. One is to ensure that preserved content is described well on its way to its destination (and that description is iteratively updated over time) so it can be recognized upon its return years (perhaps many years) later. Other areas of retained skills are digital-content risk assessment, procurement of preservation services and detailed planning for ingress and egress to change providers. Even core library skills are difficult to scale for the expanding volume of digital materials. Libraries are often surprised to discover that they are not well prepared to select from the growing waves of digital scholarly output the materials that are most important to preserve, especially considering their limited budgets.

Trust and certification

Those entrusting their materials to others for the long term need some assurance that they will be cared for appropriately. The community has developed, and continues to work on, certification schemes for these practitioners. They are of different types and value ranging from the DPC's rapid assessment model (www.dpconline.org/digipres/dpc-ram), which is lightweight and widely deployed, to the Trustworthy Repositories Audit & Certification (TRAC) scheme of the Center for Research Libraries (www.crl.edu/archiving-preservation/digital-archives/certification-assessment), which has been completed by only six services. There is an array of ISO standards relating to preservation that can be helpful for experts in assessing whether they will get what they need from a given service or a supplier. However, many preservation service providers, as well as memory institutions that operate their own preservation services, may not submit to relatively expensive formal certification processes although they will use the frameworks and standards for internal evaluation and documentation. Ultimately, preservation practitioners may judge providers less on documentation of trustworthy practices, and more on a record of successful delivery, and trust earned from respected colleagues. This may partly explain the relative popularity of CoreTrustSeal (www.coretrustseal.org/, formerly the Data Seal of Approval), which combines formal elements with a peer review component.

Advocacy and engagement

The economic and cultural justification for digital preservation is predicated almost entirely on future access and availability of the materials preserved. This places the activity in an unusual position: the majority of its beneficiaries cannot be readily engaged. They have not yet been born. Instead, digital preservation today depends on the interest and engagement of stakeholders for whom it is not a primary concern—funders,

publishers, researchers and institutions. The community has worked to produce robust risk and cost models (National Archives, 2022) to identify, justify and communicate the costs of doing the right things now, such as creating good metadata, using machine-friendlier formats such as EPub rather than just PDF and keeping multiple archival copies. Vitally, the community also needs to demonstrate that digital preservation and archiving is a sufficiently cohesive and viable market that future preservation-enabling technologies can be commercialized at reasonable cost. Not doing so risks jeopardizing the next step in the scholarly-record relay.

FINAL THOUGHTS

In this article, we have examined some aspects of the digital scholarly record and challenges of trying to preserve it. Much of our discussion has been about how things change constantly, but at the core are some key points that we think will not change for those doing this work. Here are a few:

1. Think of preserving the digital scholarly record not for 10 or 20 years, but for a hundred and more (much more). Plan accordingly.
2. Whatever your definition of the digital scholarly record, it is probably obsolete. Track the behaviour of your research community.
3. Every time you think you've chosen the right preservation technology for the long term, you're mistaken. See point 5.
4. Spending too much money and compute cycles verifying that a record has not changed is looking through the wrong end of the preservation future scope. The big risks lie elsewhere.
5. Changes in technologies, formats, standards, content-versions, ownership, business practices and more means that the digital scholarly record will never truly be at rest.
6. Automated, outsourced practices will never outpace the ever-growing need for skilled staff in dealing with number 5.

Preserving the digital scholarly record is difficult, and all of us must feel a sense of ownership in doing the work, regardless of our specific roles in the preservation ecosystem. We have described it here as a relay event, but it is not like the Olympics. There will not be a winner because there is no finish line, because the course changes radically with each lap, and because the baton in this run gains weight with every step. Our job—the job of all stakeholders in the digital scholarly record—is to do the very best we can, working in partnership wherever possible, then hand it off to those who will carry it forward after us.

AUTHOR CONTRIBUTIONS

Alicia Wise conceived the project; Tom Cramer, Chip German, Neil Jefferies, and Alicia Wise developed an initial topic list to be explored; Chip German, Neil Jefferies, and Alicia Wise wrote the article; and Tom Cramer, Chip German, Neil Jefferies, and Alicia Wise edited and reviewed the paper.

REFERENCES

- Beagrie, N. (2004). The continuing access and digital preservation strategy for the UK Joint Information Systems Committee (JISC). *D-Lib Magazine*, 10, 7–8. <https://doi.org/10.1045/july2004-beagrie>
- Bosman, J., Frantsvåg, J. E., Kramer, B., Langlais, P. C., & Proudman, V. (2021). *OA diamond journals study. Part 1: Findings*. Zenodo. <https://doi.org/10.5281/zenodo.4558704>
- Fidler, F., & Wilcox, J. (2021). Reproducibility of scientific results. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (summer 2021 edition). Metaphysics Research Lab, Stanford University. Retrieved from. <https://plato.stanford.edu/archives/sum2021/entries/scientific-reproducibility/>
- Joint Funding Council. (1993). *Libraries Review Group report (The Follett Report)*. Retrieved from www.ukoln.ac.uk/services/papers/follett/report/
- Laakso, M., Matthias, L., & Jahn, N. (2021). Open is not forever: A study of vanished open access journals. *Journal of the Association for Information Science and Technology*, 72, 1099–1112. <https://doi.org/10.1002/asi.24460>
- Lavoie, B., Childress, E., Erway, R., Faniel, I., Malpas, C., Schaffner, J., & van der Werf, T. (2014). *The evolving scholarly record*. OCLC Research. Retrieved from. www.oclc.org/research/publications/2014/oclcresearch-evolving-scholarly-record-2014-overview.html, <https://doi.org/10.25333/C3763V>
- National Archives. (2022) Safeguarding the nation's digital memory. Retrieved from www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/research-collaboration/safeguarding-the-nations-digital-memory/
- National Transportation Library. (2022). *Persistent identifiers (LibGuide)*. National transportation library. US Department of Transportation, Retrieved from. https://transportation.libguides.com/persistent_identifiers
- Pendergrass, K. L., Sampson, W., Walsh, T., & Alagna, L. (2019). Toward environmentally sustainable digital preservation. *The American Archivist*, 82(1), 165–206. <https://doi.org/10.17723/0360-9081-82.1.165>
- Spectra Logic Corp. (2022). *Data Storage Outlook Report* (updated annually). Retrieved from <https://spectrallogic.com/data-storage-outlook-report/>
- Viguerie, S. P., Calder, N., & Hindo, B. (2021). *2021 Corporate longevity forecast*. Innosight (Huron Consulting Group) Retrieved from. <https://www.innosight.com/insight/creative-destruction/>