

ORIGINAL ARTICLE OPEN ACCESS

Securing the Future of Books: A Collaborative Guide to Preservation

Graham Bell¹ | David Cirella² | Gali Halevi³  | Claire Holloway⁴ | Neil Jefferies⁵ | Roxanne Missingham⁶ | Sylvia Pegg⁷ | James Phillpotts⁵ | Alison Pope⁸ | Alicia Wise³

¹EDItEUR, United House, London, UK | ²Yale University, Preservation, New Haven, Connecticut, USA | ³CLOCKSS, Stanford, California, USA | ⁴OCLC, Operations, Dublin, Ohio, USA | ⁵University of Oxford, Publishing & Content Operations, Oxford, UK | ⁶Australian National University, Acton, Australia | ⁷Royal Society of Chemistry, Thomas Graham House, Cambridge, UK | ⁸IOP Publishing, Content Management and Discovery No. 2 The Distillery, Bristol, UK

Correspondence: Gali Halevi (ghalevi@clockss.org)

Received: 17 March 2025 | **Revised:** 26 January 2026 | **Accepted:** 12 February 2026

ABSTRACT

This paper examines critical challenges in preservation, such as infrastructure limitations, technological barriers and resource constraints, and explores potential solutions through collaborative initiatives, standardised reporting and targeted investments. Drawing on existing literature and previous efforts to preserve academic content, including journal preservation and digital content archiving, the paper outlines actionable steps to safeguard the future of scholarly books. A working group was formed to create a comprehensive guide for preservation, providing a framework that will support publishers, libraries and researchers in securing the ongoing availability of digital monographs.

1 | Introduction

Preservation of books is a complex and multifaceted challenge that requires coordinated action from publishers, libraries and preservation agencies. In the past 5 years, the production of digital scholarly monographs, including the digitisation of back catalogues, has grown significantly. This surge in digital monograph content has been driven by various factors, including the increasing maturity of scholars in the humanities and social sciences in adopting online content. Additionally, external challenges such as the COVID-19 pandemic accelerated the need for remote access to scholarly materials, reshaping how researchers and learners engage with content. Today, digital monographs dominate university library collections. For instance, data from the Council of Australian University Librarians (CAUL) highlights that for over a decade, less than 10% of monograph expenditure in Australian university libraries has been on print materials. This shift to digital formats is also reflected in usage patterns; by 2020, 96.75% of monograph usage in Australian and New Zealand university libraries was digital.¹

These trends underscore a pivotal transformation in both the production and consumption of scholarly monographs, emphasising the urgent need for robust preservation strategies to ensure the longevity and accessibility of these critical academic resources.

Addressing issues of infrastructure, technological barriers, unclear roles and resource constraints will be critical to ensuring the long-term availability of these materials. While existing literature highlights significant gaps in preservation efforts, it also offers valuable insights and potential pathways for improvement. Collaborative initiatives, standardised reporting mechanisms and targeted investments in preservation infrastructure will be essential to overcoming these challenges and safeguarding the future of books.

Recognising the urgency and importance of addressing these issues, a working group was convened to create a comprehensive guide for the preservation of books. This effort builds on a tradition of collaborative initiatives aimed at safeguarding

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Learned Publishing* published by John Wiley & Sons Ltd on behalf of ALPSP.

Key Points

1. *Increased risk of digital book loss:* As digital formats become the primary medium for scholarly books, the risk of permanent loss has grown significantly due to factors like the closure of small publishers, website updates that remove historic works, changes to the technical standards for digital book formats and cybersecurity threats.
2. *Fragility of the digital ecosystem:* Unlike print materials, which were backed up in multiple copies by libraries, the digital book ecosystem lacks robust and consistent backup practices, making it more vulnerable to loss and degradation.
3. *Urgency of preservation strategies:* The immediate and extensive risks to digital content highlight the critical need for reliable preservation strategies to ensure continued access to scholarly knowledge for future generations.
4. *Importance of clarifying responsibilities and copyright navigation:* Efforts to clarify roles, improve copyright management and develop comprehensive preservation guidelines are essential for safeguarding the long-term availability of digital books.
5. *Collaboration and standardisation for Long-term success:* Continued collaboration, transparency and the development of standardised preservation systems, such as the creation of an EPUB preservation standard, will be crucial for building a sustainable and adaptable global preservation framework.

scholarly outputs, drawing inspiration from projects such as the development of best practices for journal preservation and the creation of community-driven guidelines for digital content archiving. By tackling the unique challenges posed by books, the working group aimed to provide actionable solutions that will benefit publishers, libraries and researchers alike, ensuring that these valuable resources remain accessible for generations to come.

2 | Literature Review

The dissemination, usability and sustainability of books in the preservation arena extend well beyond the OA ecosystem. Many of the challenges and opportunities discussed apply equally to all book formats, including commercially published, hybrid and digitally restricted works. The review situates OA books within the broader scholarly communication landscape. This review, therefore, highlights issues that are systemic rather than format-specific, offering insights relevant to the preservation and future of scholarly books as a whole.

The preservation of books remains an underdeveloped yet increasingly critical area within the digital preservation landscape, and challenges related to their long-term availability persist. These challenges include issues of infrastructure, technological barriers, unclear stakeholder roles and resource constraints.

One of the most pervasive challenges in preserving books is the lack of consistent infrastructure for long-term archiving. Barnes et al. (2022) report that the current preservation ecosystem for monographs is fragmented, with few organisations systematically preserving these materials. The scoping report highlights that, unlike journals, which benefit from established preservation initiatives, such as the KEEPERS Registry,² books lack comparable systems for tracking and managing preservation efforts.

Laakso et al. (2022) further emphasise the inadequacy of preservation coverage for books. Their study reveals that while some preservation initiatives exist, they are limited in scope, often failing to capture the full breadth of Open Access (OA) book publishing. Similarly, Derrot et al. (2014) note that many digital preservation efforts focus on digitised versions of print books rather than born-digital books, leaving significant gaps in coverage.

Preserving books is inherently complex due to the diversity of file formats, platforms and metadata standards. Kirchhoff and Morrissey (2014) highlight the challenges posed by the rapid evolution of digital publishing technologies, which create compatibility issues and demand continuous updates to preservation systems. Formats such as EPUB and PDF, while widely used, often lack consistent metadata, making them difficult to archive systematically (Romano 2003).

Rosenthal et al. (2005) underline the importance of robust preservation systems capable of adapting to technological changes, but they also caution that building such systems requires substantial investment and collaboration among stakeholders.

Derrot et al. (2014) further stress that preserving born-digital content involves unique challenges, such as platform dependencies and the risk of format obsolescence.

Another major challenge is the ambiguity surrounding who should take responsibility for preserving books. Wise (2022) observes that preservation efforts often fall into a grey area between publishers and libraries, with neither party clearly accountable for ensuring long-term access. Laakso (2023) similarly notes that small and medium-sized publishers, which contribute significantly to OA book production, often lack the resources or expertise to implement preservation practices.

Ferwerda et al. (2017) argue that the lack of a clear division of labour among stakeholders exacerbates preservation challenges. Their study of eight European countries found that while national policies often emphasise the importance of OA book preservation, they rarely provide actionable guidance on how to allocate responsibilities among publishers, libraries and preservation agencies.

Preservation efforts are also hindered by resource limitations, particularly for smaller publishers and independent authors. Barnes et al. (2022) highlight the financial and logistical barriers faced by smaller actors in the ecosystem, who often lack the capacity to invest in preservation infrastructure. Kirchhoff (2011) echoes this concern, noting that even

well-funded institutions struggle to balance preservation priorities with other pressing demands.

Laakso et al. (2022) emphasise that existing preservation services, such as CLOCKSS and Portico, are limited in their capacity to archive the growing volume of books. This limitation is compounded by the high costs associated with maintaining digital preservation systems, including technological upgrades, staff training and ongoing maintenance.

Laakso et al. (2022) advocate for the development of a registry or similar tool to monitor book preservation status, akin to existing systems for journals. Without such mechanisms, stakeholders are left without the data needed to identify gaps and allocate resources effectively. Barnes et al. (2022) identify this as a critical gap, arguing that the absence of standardised reporting makes it difficult to assess the scope and effectiveness of current preservation initiatives.

The challenges faced in preserving books mirror those encountered in journal preservation, suggesting potential lessons and strategies. Ferwerda et al. (2017) highlight the importance of collaborative initiatives, such as the KEEPERS Registry and CLOCKSS, which have successfully addressed some of the challenges in journal preservation. Applying similar models to books could help build more robust and scalable preservation systems.

Romano (2003) and Wise (2022) also emphasise the need for cross-sector collaboration, involving publishers, libraries and preservation agencies. By pooling resources and expertise, stakeholders can overcome many of the technological and financial barriers to preservation.

If the preservation of books is a critical yet often overlooked aspect of ensuring long-term access to scholarly knowledge, the preservation of open access (OA) books deserves special attention. Recent research by Mikael Laakso (Laakso 2023), sponsored by CLOCKSS, DOAB (Directory of Open Access Books) and OAPEN (Open Access Publishing in European Networks), revealed a concerning gap: fewer than 50% of OA books are preserved in digital form within long-term archives. A specific challenge for books lies in the false assumption that their openness inherently ensures preservation. Many believe that because these books are freely accessible, someone will eventually identify them and include them in a preservation system. Unfortunately, this is not a reliable assumption. Open access books require deliberate action to ensure their long-term survival.

This gap in preservation raises pressing questions about roles and responsibilities. Who is ultimately accountable for safeguarding OA books? How can the smallest players in the publishing ecosystem, often working with limited budgets and resources, be supported to ensure their content is not lost to future generations? These questions remain largely unresolved, leaving a landscape filled with uncertainty and inconsistencies.

3 | Aims of the Working Group

The working group (see Appendix A for full list and roles of members), convened regularly throughout 2023 and 2024 with

the overarching goal of addressing the pressing challenges of preserving books. This collective effort was driven by four key aims, each designed to address critical gaps in the preservation landscape and to lay the groundwork for sustainable solutions.

3.1 | Developing a Comprehensive Preservation Guide

A primary aim of the group was to create a practical, accessible guide for stakeholders involved in book preservation. This guide is intended to serve as a roadmap for publishers, libraries and preservation agencies, clarifying roles, responsibilities and best practices for ensuring long-term access to books. Given the lack of systematic resources available for preserving digital monographs, this guide helps unify fragmented efforts and provide actionable recommendations for organisations of all sizes.

3.2 | Exploring the Copyright Arena

Understanding and navigating the complex copyright landscape for books was another critical focus. Copyright restrictions, licensing agreements and usage terms often present significant barriers to preservation efforts. The working group aimed to identify common challenges and propose solutions to facilitate the legal and ethical preservation of books. This exploration was especially important for addressing the needs of smaller publishers and independent authors, who may lack the expertise or resources to manage copyright-related complexities.

3.3 | Delivering a Proof-of-Concept Project to Make Discoverable What Books Are Preserved Where

To demonstrate the feasibility of preservation status discovery, the group undertook a proof-of-concept project. This initiative explored how the current supply chain infrastructure could be leveraged to identify which books are being preserved, where and by whom. By mapping and testing the data flows indicating preservation activities, the preservation status of selected titles, the project provided valuable insights into existing gaps and highlighted opportunities to improve transparency and collaboration across the preservation ecosystem. The proof-of-concept also emphasised the potential for scalable solutions that could benefit the broader community.

3.4 | Developing a Preservation Standard for EPUB Format Books

Recognising the widespread use of the EPUB format for books, the working group prioritised the development of a preservation standard tailored to this format. EPUB is versatile and widely adopted, but its lack of uniformity in metadata, versioning and file structure poses significant challenges for long-term preservation. The standard developed outlined essential technical specifications and preservation requirements, offering a foundation for consistent and reliable archiving of EPUB-based books. This effort aimed to ensure that EPUB files remain accessible,

interoperable and usable over time, regardless of technological changes.

By addressing these aims, the working group sought to make meaningful progress towards a more robust and sustainable system for preserving books, ensuring that these valuable resources remain accessible for future generations.

4 | Aim 1: Developing a Guide for Preserving Books

Digital books face similar preservation risks to any other materials produced in a digital format. However, the preservation landscape is complicated by the presence of a ‘long tail’ of smaller publishers. These smaller actors are often difficult to reach and may lack the resources or knowledge to engage with preservation initiatives effectively.

To address this, book publishers must actively ensure that their works are deposited in national libraries or with trusted preservation partners such as CLOCKSS and Portico. Preservation requires more than just archiving; it involves ongoing stewardship, including regular copying and management, to ensure content remains accessible over time. Without these proactive measures, even open access books, despite their open availability, risk being lost to future readers and researchers.

4.1 | Purpose and Scope of the Guide

The guide provided a framework for preserving books to ensure their long-term accessibility and usability.³ It highlights the unique challenges of preserving books and offers practical solutions for stakeholders across the publishing ecosystem, including publishers, libraries and preservation agencies.

4.1.1 | Key Themes and Topics Covered in the Guide

1. Why Preservation Matters
 - Books are vulnerable to loss due to the lack of systematic preservation practices, especially for smaller publishers and institutions.
 - Preservation safeguards scholarly knowledge and ensures accessibility for future generations.
2. Challenges in Book Preservation
 - Limited awareness and resources among publishers, particularly small and medium-sized ones.
 - Complexity in coordinating roles across publishers, libraries and preservation agencies.
 - Lack of standardised practices for reporting preservation status and coverage.
3. Roles and Responsibilities
 - *Publishers*: Ensure deposit of books in trusted digital repositories and collaborate with preservation services.
 - *Libraries*: Advocate for preservation and coordinate with publishers and preservation agencies.
 - *Preservation agencies*: Provide infrastructure and tools for long-term digital preservation.

4. Preservation Standards and Best Practices
 - Recommends EPUB as the preferred file format for digital books due to its widespread use and compatibility.
 - Encourages use of metadata standards like ONIX to enhance discoverability and interoperability.
5. Strategies for a Sustainable Preservation Ecosystem
 - Collaborative efforts between publishers, libraries and archives.
 - Greater transparency in reporting preservation status to improve coverage and accountability.
 - Funding models to support small publishers and ensure equitable participation.
6. Key Actions for Stakeholders
 - *Publishers*: Deposit books with a minimum of three recognised preservation services, for example CLOCKSS, Portico and national libraries through digital legal deposits.
 - *Libraries*: Partner with publishers to promote awareness and adoption of preservation practices.
 - *Preservation networks*: Build and share tools that simplify preservation workflows for stakeholders.

The guide underscores the critical importance of collaborative and systematic approaches to preserving books. By aligning efforts across the publishing ecosystem and leveraging shared tools and practices, stakeholders can ensure the long-term availability of these valuable resources. This includes service providers who create most of the publishers’ digital files. This dynamic is already evident in areas such as accessible EPUB production for print-disabled users, where specialised vendors typically have far more advanced technical knowledge than most publishing teams. As a result, preservation readiness is shaped not only by publisher policies but also by the practices, standards and capabilities of the service providers who generate, convert and manage the underlying digital assets.

5 | Aim 2: Exploring the Copyright Arena

To address copyright challenges in digital preservation, a dedicated sub-group was established within the working group (see Appendix B for list of members and roles). This team delved into the legal and practical issues surrounding the long-term care of books, recognising that copyright plays a pivotal role in ensuring preservation success.

Authors often appreciate the idea of their work representing their intellectual and cultural legacy being accessible centuries into the future. Many libraries also require long-term preservation clauses in their purchasing agreements. Yet, explicit discussions about what happens to books when the original caretakers are no longer present often remain overlooked. This gap creates uncertainty and risks for preservation efforts.

Preservation is a shared responsibility among stakeholders in the publishing ecosystem. Much like parents specifying guardians for their children in a will, publishers, authors and preservation partners need to establish clear agreements about the future of a book. These discussions ensure that works are safeguarded even if the original stakeholders can no longer provide direct care.

For commissioned and licensed content, copyright agreements can explicitly include preservation clauses. Such agreements with trusted archives, like CLOCKSS, enable books to be deposited and protected securely for future generations. Legal deposit legislation in many countries also provides an important safety net, mandating the collection and preservation of books by national libraries. Together, these efforts create a network of complementary services that strengthen preservation.

To ensure a book's survival, it must be preserved across multiple trusted services. The guiding principle of digital preservation is straightforward: the more copies, the better; the less correlation between copies, the safer; the higher the dependability of each copy, the more secure; and the faster failures are detected and repaired, the stronger the preservation outcome.

A trusted archive demonstrates its reliability through several key attributes:

- *Certification*: Peer-reviewed certification (e.g., the CRL TRAC audit, ISO:16363 or the Core Trust Seal) signals a commitment to and expertise in long-term preservation.
- *Mandate and funding*: A clear mission and sustainable funding ensure the archive's longevity.
- *Track record*: Evidence of successfully preserving content over time builds trust.
- *Transparency*: Clear agreements, workflows and documentation of processes ensure accountability and accessibility.
- *Succession planning*: Plans that outline what happens to content if the archive ceases operation provide additional security.

Reflecting on these challenges and solutions, the group developed key recommendations for authors, publishers and preservation advocates:

1. *Communicate your wishes*: Discuss what should happen to your book when you're no longer actively involved in its care. Should it be forgotten or remembered? Should it remain privately preserved or publicly accessible?
2. *Collaborate across stakeholders*: Practical copyright issues often hinder preservation efforts. Engage with preservation partners early to explicitly determine where your book will be preserved, how it will be made accessible and how its preservation will be documented.
3. *Document rights clearly*: Publishers and authors must clarify the rights they hold and how these rights transfer to preservation agreements. Varying author agreements, territorial limitations, rights reversion clauses and third-party rights can complicate preservation efforts. Writing down these terms and agreements is essential to avoid future conflicts or roadblocks.

To help stakeholders navigate this complex landscape, the group created an infographic (see Figure 1) summarising best practices for ensuring books become part of the shared cultural and intellectual heritage for generations to come. By fostering collaboration, clarifying rights and making preservation a deliberate

and documented process, we can ensure that the books of today remain accessible far into the future.

5.1 | Legal Rights: Drivers and Challenges in Preserving Books

Preservation is deeply tied to the legacy of authors and the needs of the research community. Many authors, regardless of genre, are motivated by the idea that their work, often seen as a legacy, will remain accessible for hundreds of years. Among publishers, academic publishers are particularly attuned to the importance of preservation due to strong demand from research libraries, where long-term access is a standard expectation in purchasing agreements. In these cases, preservation is not just an aspiration but a requirement woven into the publisher-library relationship.

Despite the clear value of preservation, many publishers encounter significant hurdles. For some, particularly those driven only by legal deposit requirements, digital preservation feels like an obligation rather than a voluntary commitment. Even for publishers who are motivated, the complexities of rights management present substantial barriers:

1. *Rights uncertainty*: Publishers often struggle to determine the full extent of the rights they hold. This makes entering preservation agreements charged with the risk of inadvertently violating third-party or author rights.
2. *Territorial restrictions*: Publisher rights are sometimes limited by geography, complicating global preservation. For instance, a publisher might hold European rights but lack US rights, creating uncertainty about digital preservation.
3. *Reverted rights*: Authors frequently regain rights to their works, and publishers may lack systems to track when and how this happens.
4. *Edited volumes*: Books with multiple contributors often involve a mix of rights arrangements, making preservation agreements even more challenging.
5. *Legacy systems*: Older publishers may have decades or even centuries of agreements in their archives. Many are poorly documented or in formats no longer easily accessible, leaving a grey area around copyright status.

Several strategies can help publishers overcome these challenges and strengthen preservation efforts. One promising approach could be for organisations like CLOCKSS to offer indemnities to publishers, providing confidence to proceed with preservation agreements even when rights clarity is lacking. Another possibility is collective licensing, such as extended collective licensing for orphan works, to streamline the preservation process.

Clearer guidelines and increased confidence regarding CLOCKSS trigger events where content becomes open access when a publisher ceases to exist and the content is no longer available could help reduce hesitancy and improve access. Publishers need assurance that their content won't be triggered without proper checks on rights and availability. Alicia Wise,

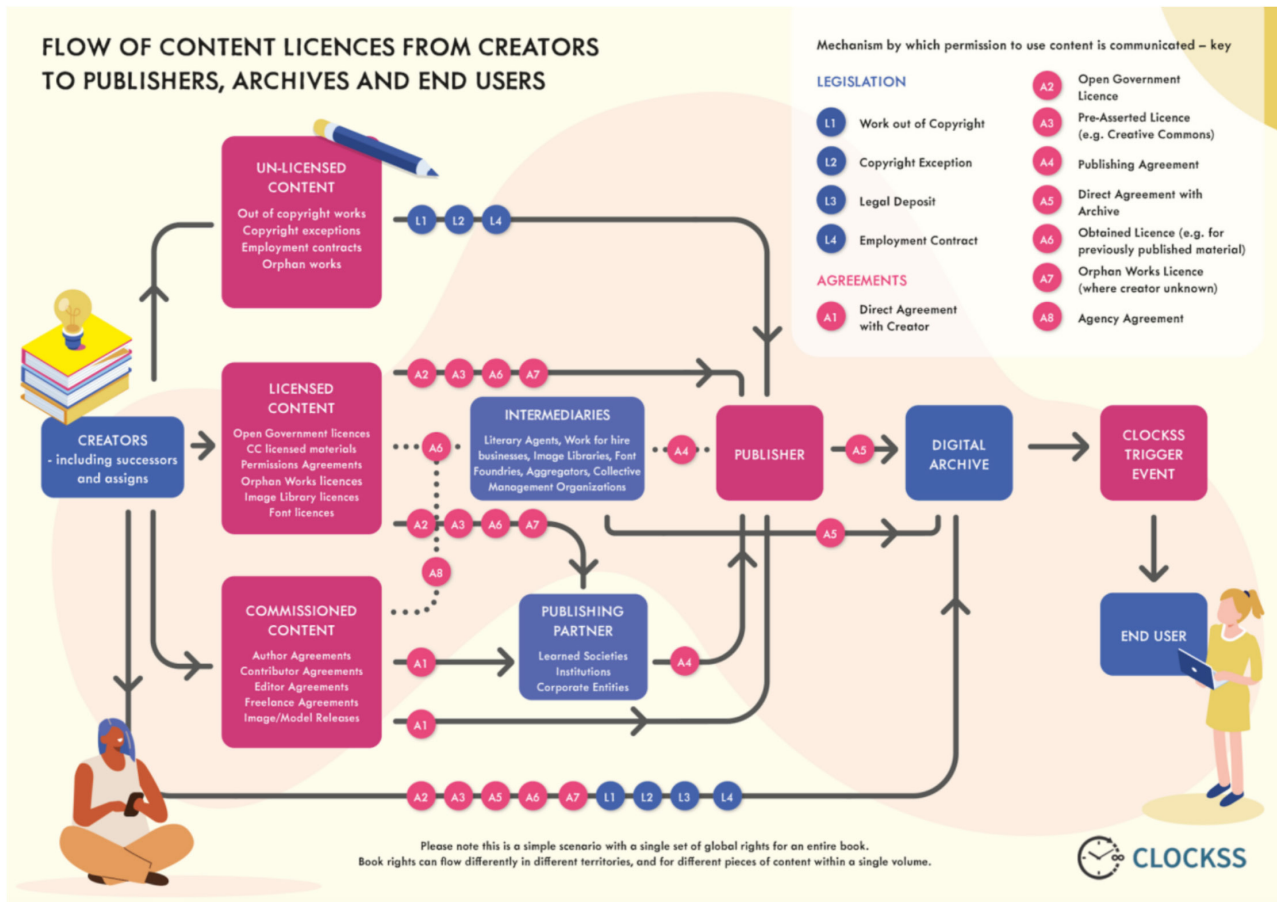


FIGURE 1 | Infographic describing the flow of rights from creator to archives via publishers and other intermediaries.

for example clarified that CLOCKSS has never triggered a book and would not do so without ensuring rights compliance.⁴⁵

Another solution lies in proactive communication. Publishers can start discussions with authors and agents about digital preservation during contract negotiations, ensuring that preservation rights are explicitly included. Even if rights revert to the author, the book could remain preserved in a trusted archive. Finally, publishers must invest in modernising rights management systems, moving from fragmented paper records and outdated systems to more transparent, organised processes.

Aggregators play an indirect, yet complex, role in preservation. While libraries often purchase e-books through aggregators, publishers typically view these platforms as access channels rather than preservation partners. Aggregators do not receive preservation rights from publishers and the royalty structures for content accessed via aggregators differ significantly from those for other sales channels.

This dynamic can create confusion among librarians, who sometimes conflate post-cancellation access rights with long-term preservation. For books, as with journals, changes in content availability through aggregators can spark concern about access and preservation. Publishers and libraries must clarify these distinctions and ensure proper preservation practices are in place independently of aggregators.

5.2 | Mapping the Rights Chain for Preservation

Preservation efforts require a clear understanding of the rights chain from creator to archive. In straightforward cases, creators assign global rights to publishers, enabling them to preserve works through agreements with trusted archives. However, complications arise with third-party content, multiple creators or complex legacy agreements. Collective licences might offer a practical solution for managing rights in such scenarios.

The rights chain also involves the transition of works to archives and ultimately to end-users. Content may reach archives via agreements, legal deposit legislation or pre-asserted licences like Creative Commons. For end-users, access is determined by archive agreements, copyright expiration or other legal frameworks.

5.2.1 | Illustrations: A Unique Challenge

Illustrations add another layer of complexity to preservation. Some cases are straightforward, with a single illustrator and clear permissions. Others involve multiple creators and permissions that may be incomplete or unclear. Publishers, wary of past lawsuits, are often cautious about reuse beyond original agreements. Ensuring that illustrations can be preserved alongside books requires careful rights management and clear documentation, whether handled by publishers, authors or editors.

5.2.2 | Third-Party Materials Under Open Access Licences

When it comes to digital preservation, the use of third-party materials under OA licences is generally straightforward. Open Access licences, particularly those under Creative Commons (CC), are designed to facilitate reuse while remaining tied to copyright terms. For example, materials licensed under CC-BY or CC-BY-NC are typically compatible with digital preservation efforts, as these licences explicitly permit redistribution and reuse with appropriate attribution.

Open access licences typically grant users the freedom to access, share and reuse content, but there are situations where open access materials contain content that is not fully open. For instance, some open access books may include chapters or sections with 'All Rights Reserved' content, often published under fair use or fair dealing exceptions, or material that has been copyright-cleared for a specific purpose. In these cases, the licensing needs in the metadata become crucial to clarify the permissions and restrictions associated with the content. The metadata should clearly indicate which parts of the work are under open access and which may have more limited usage due to copyright considerations. This helps users understand what they can legally do with different sections of the work and ensures that copyright laws are respected. Properly detailing these licensing distinctions in the metadata prevents confusion and ensures that access to the work aligns with the rights granted.

Even materials under the more restrictive CC-BY-NC-ND licence are often suitable for preservation. While this licence prohibits derivative works, copyright law clarifies that routine actions like format migration or reformatting for preservation purposes do not qualify as creating a derivative work. Thus, these activities remain within the bounds of what is permissible.

5.2.3 | The Complexity of Multi-Contributor Works

Works involving multiple contributors such as co-authors, illustrators or translators present unique challenges for preservation. Each contributor's rights need to be individually understood, documented and respected. Without a clear and comprehensive record of these rights, the preservation process becomes more difficult and risks potential conflicts with the contributors or their estates.

5.2.4 | Copyright Challenges After an Author's Death

The period between an author's death and the end of copyright protection can be especially challenging for publishers and archivists. During this time, copyright often reverts to the author's heirs, who may not be fully informed or engaged in managing these rights. An example of this complexity comes from older UK titles governed by the 1911 Copyright Act,⁶ where specific reversionary provisions apply. For contracts made before 1 July 1957, even if the contract assigned copyright to the publisher, the rights automatically revert to the author's estate 25 years after

the author's death. This process requires no formal paperwork to amend the original contract.

To illustrate, consider a book published in 1955 under such a contract. If the author died in 1985, the copyright remains valid until the end of 2055. However, from 1 January 2011, the copyright ownership would have automatically reverted to the author's heirs. Understanding and navigating these provisions require careful attention to both historical legislation and individual agreements.

For those seeking additional guidance, appendix I of *Clark's Publishing Agreements, 11th Edition* (Gadsby 2023) provides detailed insights into these reversionary provisions and their implications for publishers and preservation efforts.

Ultimately, addressing these challenges requires a concerted effort across the publishing industry. Digital preservation is not yet a standard part of third-party rights agreements and publishers often lack the confidence to address long-term preservation explicitly. Greater collaboration, clearer rights alignment within publishing house and industry-wide recognition of preservation's importance are critical.

To advance preservation, stakeholders must articulate the value of long-term preservation, describe good practices and advocate for collaboration across the industry. Developing toolkits, improving transparency around trigger events and exploring collective licensing solutions can pave the way for more effective and confident preservation efforts. By addressing these challenges, publishers can ensure that today's books remain a lasting part of our cultural and intellectual heritage.

6 | Aim 3: Delivering a Proof-of-Concept Project to Track Book Preservation

Unlike journals, where preservation and discovery systems are relatively well established, determining if a book, regardless of its business model, is actively preserved remains a challenge. Publishers have questioned the need for including preservation data, viewing it as additional work. However, librarians are increasingly demanding this information to inform collection development decisions. Knowing that a title is preserved can elevate its value in acquisition decisions, particularly when libraries are deciding between titles for electronic licences. Preservation status thus becomes a competitive advantage for publishers.

To address this, a proof-of-concept project was initiated to explore ways to accurately identify which books are preserved and where. The goal was to provide a clear picture of the preservation status of books and ensure this information is seamlessly integrated into library and publisher workflows.

6.1 | Leveraging ONIX for Archival Messaging

The project centered on using ONIX metadata, a standard format widely used by publishers to share information about their titles with libraries and retailers to communicate preservation

details. Publishers are already accustomed to using ONIX to describe books, business models, pricing, accessibility and licensing information. By extending this framework to include archival information, publishers can easily update metadata during regular processes such as price reviews or backlist updates.

ONIX Website role codes 47, 48 and 51⁷ have been identified as the mechanism to indicate whether a title is digitally preserved and where it is archived. This solution streamlines integration into existing workflows, making it less burdensome for publishers. Collaborative efforts with publishers such as Channel View, University of Amsterdam Press, Cambridge University Press, Taylor & Francis and others have been instrumental in piloting this approach.

In addition to ONIX, MARC records commonly used by libraries are another avenue for including preservation data. These records can indicate preservation arrangements in field 857,⁸ offering libraries a more direct way to access this information.

One significant challenge in this initiative is ensuring the seamless transfer of data from ONIX to MARC. Although ONIX-to-MARC crosswalks exist, they do not currently support ONIX's archival data field (codes 47, 48 and 51). This gap means that preservation information might fail to make it into MARC records used by libraries. Efforts are underway to improve the crosswalks, but in the meantime, publishers can submit MARC records directly to ensure preservation details are captured.

6.2 | Building the Ecosystem

The proof-of-concept project has already seen promising progress. Metadata providers like NielsenIQ BookData (Nielsen Book 2016) and library service platforms such as OCLC have begun receiving preservation data. Although the process is still evolving, the inclusion of preservation information is gaining traction among publishers, librarians and metadata aggregation/distribution agencies.

A natural extension of this project would be the creation of a centralised registry for book preservation, akin to the KEEPERS Registry for journals. Such a system would require coordinated community support and a dedicated organisation to collect and maintain the data. It would function like a parcel delivery system tracking preservation from initial deposit to confirmed receipt, providing confidence to both publishers and librarians.

This initiative has already helped publishers validate and improve their own preservation processes, ensuring that titles marked as archived are indeed preserved. With support from key stakeholders, the project is set for success as more publishers adopt the process and populate the metadata supply chain with preservation information. The long-term vision is to make preservation data easily discoverable, empowering librarians and publishers alike to ensure the enduring availability of scholarly content.

6.3 | Developing a Preservation Standard for EPUB Format Books

There is an established ISO (International Organization for Standardization) standard for creating preservation-ready PDF files, called PDF/A (PDF for Archiving),⁹ which helps publishers prepare PDF-formatted documents for long-term preservation and supports global archives in managing these books. PDF/A refers to a specific subset of the PDF standard, designed for preservable documents which, conveniently, also improves accessibility. About 18 months ago, a global initiative sought to develop a similar archival standard for EPUB-formatted e-books, and this effort was taken up via NISO (National Information Standards Organization). The proposed new format, dubbed 'EPUB/A' is a variation of the EPUB standard intended to ensure the 'preservability' of an EPUB file by restricting a document's reliance on external resources.

The process involved understanding ISO's procedures and navigating various challenges. ISO is a global body for standards across various industries, operating similarly to a 'United Nations for standards'. It coordinates with national standards bodies, such as NISO in the United States, which reports to ANSI (American National Standards Institute). ISO encompasses numerous committees, each addressing different fields. Understanding this structure was a significant initial challenge.

Another challenge arose when it became clear that ISO, while responsible for ratifying the EPUB standard, does not lead the development of the EPUB format itself. This role belongs to the W3C,¹⁰ which is focused on enhancing e-book accessibility. Developing an archival version of EPUB thus required engaging with a second international standards organisation.

Fortunately, there was alignment between these two organisations on the idea that improving the archivability of EPUB-formatted e-books would also enhance accessibility. By adopting an EPUB archival standard, publishers could be encouraged to create e-books that are both more durable and more accessible, benefiting users with disabilities.

Once this strategic alignment was established, organisational politics became easier to navigate. The next step was to engage a wide range of stakeholders, including librarians, publishers, digital preservationists, e-book retailers and DRM (Digital Rights Management) system providers. It was essential to understand and incorporate the perspectives of international stakeholders to develop a comprehensive and effective standard.

EPUB is designed primarily for e-books and digital reading. While both formats serve a similar purpose in the publishing landscape, reflowable EPUB offers greater flexibility in adapting to different screen sizes and devices. Excessive use of page-oriented features in EPUB (fixed layout) not only hinders accessibility but goes against the original design goals for the format. Consequently, EPUB/A avoids such functionality.

A key decision in developing the EPUB/A was to avoid external links to content like fonts or multimedia. Since external links will become inaccessible over time, it is recommended that such

content be embedded within the EPUB file. This approach also benefits accessibility as external resources pose challenges for accessibility tools and offline access.

EPUB, like PDF, can support a range of advanced features, such as executable code; however, the vast majority of publications do not take advantage of them. Such features are excluded from PDF/A because they have the potential to pose significant accessibility, security and preservability issues in the long term. EPUB/A specifies an equivalent set of exclusions. Publishers can optionally include such features by maintaining and archiving regular EPUB files, but with the understanding that accessibility and preservation may be compromised.

Another issue discussed was whether DRM should be included in EPUB/A. DRM is associated with external decoding keys, which pose risks for long-term preservation, as they might become inaccessible if a publisher or DRM service vendor no longer exists to provide them. To mitigate this risk, contracts between publishers and archives can offer legal protections for DRM-free digital content.¹¹

Additional requirements, such as enhanced metadata for preservation, were also considered. However, it was determined that adding extra requirements would complicate the process and could present a barrier to adoption. EPUB/A will remain a strict subset of EPUB without imposing additional requirements.

The goal is for EPUB/A to be easy to produce from existing EPUB workflows. While some content might push the limits of the format, the hope is that the standard will be straightforward and accessible for most publishers.

The draft specification for EPUB/A has been submitted to ISO, and international experts will guide it through the ISO process. This process is expected to take from 18 to 24 months, with the new international standard anticipated by the end of 2026.

7 | Discussion

The preservation of books is a growing concern within the scholarly community, yet it remains under-addressed compared to the preservation of other digital content, such as journal articles. The preservation landscape for books is fragmented, with no central mechanism to track or standardise efforts, leaving both researchers and publishers uncertain about the long-term accessibility of these resources.

The question of responsibility for preserving books remains unresolved, with no clear consensus on whether publishers, libraries, or other stakeholders should bear the responsibility. Smaller publishers, in particular, face unique challenges in managing preservation efforts due to limited resources and expertise. For these smaller players, collaboration with larger institutions and access to accessible resources and guidelines could significantly alleviate the burden of ensuring their content is preserved for future generations. Additionally, navigating the complex copyright landscape poses another barrier to book preservation, particularly for independent authors and publishers who may struggle with the legal intricacies of digital archiving.

The working group's efforts, including the creation of a practical guide for book preservation, represent an essential step in addressing these challenges. By clarifying roles, responsibilities and best practices, the group provided much-needed guidance for all stakeholders involved in book preservation. However, these efforts also highlight the necessity for broader collaboration and transparency in the preservation ecosystem, particularly in tracking which books are being preserved, where and by whom. The proof-of-concept project demonstrated the importance of transparency in identifying existing gaps and fostering cooperation among preservation organisations.

Furthermore, the development of a preservation standard for EPUB-format books addresses one of the most pressing technical challenges in book preservation. While EPUB is a widely adopted format for digital books, its lack of uniformity and standardisation as well as its richness and variety, makes it difficult to ensure long-term preservation. By establishing a clear set of technical specifications for 'preservable' EPUB-based books, the working group's efforts will contribute to the creation of a more reliable and consistent framework for archiving these important resources.

8 | Conclusions

The preservation of books is critical to ensuring continued access to scholarly knowledge, yet it remains an area in need of significant attention and improvement. As digital formats have become the primary medium for scholarly books, the risk of permanent loss has grown significantly. Many titles have already been lost due to the closure of small publishers, updates to websites by associations and universities that remove historic works, and the damage caused by cybersecurity attacks on web servers. This has created a fragile ecosystem, as libraries no longer maintain multiple backup copies as they did with print materials. The immediacy and extent of this risk continue to grow, highlighting the need for reliable preservation strategies to safeguard digital content for future generations.

Efforts to clarify responsibilities, improve copyright navigation and develop comprehensive preservation guidelines will play a crucial role in safeguarding books for future generations. The working group's initiatives, including the creation of a preservation standard for EPUB books, represent meaningful steps forward in addressing the technical and logistical challenges of book preservation. However, long-term success will depend on continued collaboration, transparency and the establishment of standardised systems to ensure that books are preserved reliably and consistently across the globe. The progress made so far is promising, but ongoing efforts will be needed to build a sustainable, unified preservation framework for books that can adapt to the evolving landscape of digital publishing.

Author Contributions

Gali Halevi: writing – review and editing. **Graham Bell:** investigation, writing – review and editing. **David Cirella:** writing – review and editing. **Claire Holloway:** investigation, writing – review and editing. **Neil Jefferies:** investigation, writing – review and editing. **Roxanne Missingham:** writing – review and editing. **Sylvia Pegg:** writing

– review and editing. **James Phillpotts**: writing – review and editing. **Alison Pope**: writing – review and editing. **Alicia Wise**: conceptualisation, investigation, writing – review and editing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Endnotes

¹ <https://www.caul.edu.au/programs-projects/statistics-services#%3A~%3Atext%3DTheCAULStatisticsprovideddatafromuniversitylibrariesCdataforthe profilingofANZacademiclibraries>.

² <https://keepers.issn.org/>.

³ https://clockss.org/wp-content/uploads/2024/04/Clockss_Guide_WEB.pdf.

⁴ <https://clockss.org/bookwill/>.

⁵ <https://scholarlykitchen.sspnet.org/2024/02/20/kitchen-essentials-alicia-wise-clockss/>.

⁶ <https://www.legislation.gov.uk/ukpga/Geo5/1-2/46/enacted>.

⁷ <https://ns.editeur.org/onix/en/73>.

⁸ <https://www.loc.gov/marc/bibliographic/bd84188x.html>.

⁹ <https://pdfa.org/resource/iso-19005-1-pdf-a-1/>.

¹⁰ <https://www.w3.org/publishing/epub3/>.

¹¹ <https://clockss.org/wp-content/uploads/2024/10/2025-CLOCKSS-Participating-Publisher-Agreement.pdf>.

References

Barnes, M., E. Bell, G. Cole, J. Fry, R. Gatti, and G. Stone. 2022. “WP7 Scoping Report on Archiving and Preserving OA Monographs.” https://repository.lboro.ac.uk/articles/report/WP7_scoping_report_on_archiving_and_preserving_OA_monographs/20222166.

Derrot, S., J.-P. Moreux, C. Oury, and S. Reecht. 2014. “Preservation of e-Books: From Digitized to Born-Digital.” In *11th International Conference on Digital Preservation (IPRES)*. IPRES. <https://bnf.hal.science/hal-01088755/>.

Ferwerda, E., F. Pinter, and N. Stern. 2017. “A Landscape Study on Open Access and Monographs: Policies, Funding and Publishing in Eight European Countries.” <https://apo.org.au/node/117361>.

Gadsby, O. 2023. “Clark’s Publishing Agreements: A Book of Precedents, Edited by Lynette Owen.” *Logos* 34, no. 1: 71–72.

Kirchhoff, A. 2011. “E-Books: The Preservation Challenge.” *Against the Grain* 23, no. 4: 10. <https://doi.org/10.7771/2380-176X.5935>.

Kirchhoff, A., and S. Morrissey. 2014. “Preserving e-Books: DPC Technology Watch Report.” *DPC Technology Watch Series* 14. <https://doi.org/10.7207/twr14-01>.

Laakso, M. 2023. “Open Access Books Through Open Data Sources: Assessing Prevalence, Providers, and Preservation.” *Journal of Documentation* 79, no. 7: 157–177. <https://doi.org/10.1108/JD-02-2023-0016>.

Laakso, M., A. Wise, and R. Snijder. 2022. “Peering Into the Jungle: Challenges in Determining Preservation Status of Open Access Books.” In *Proceedings IPRES 2022 Glasgow 12 16 September 2022*. IPRES. https://helda.helsinki.fi/bitstream/10227/531055/1/ipres2022_proceedings_laakso_et_al_2.pdf.

Nielsen Book. 2016. “Nielsen Book Launches Pubtrack Digital UK.” http://emarketing.nielsenbook.co.uk/files/amf_nielsen/project_82/NielsenBook_PubTrackDigitalLaunch_Feb2016.pdf.

Romano, F. 2003. “E-Books and the Challenge of Preservation.” *MFIR* 32, no. 1: 13–25. <https://doi.org/10.1515/MFIR.2003.13>.

Rosenthal, D. S., T. S. Robertson, T. Lipkis, V. Reich, and S. Morabito. 2005. “Requirements for Digital Preservation Systems: A Bottom-Up Approach.” <https://arxiv.org/abs/cs/0509018>.

Wise, A. 2022. “Short Paper: Open Access Books and Digital Preservation. IPRES 2022.” <https://www.digipres.org/publications/ipres/ipres-2022/papers/short-paper-open-access-books-and-digital-preservation.html>.

Appendix A

Working Group Members

- Graham Bell, (EDITEUR)
- David Cirella and Euan Cochrane, (Yale University Library)
- Gareth Cole, (COPIM and Loughborough University)
- Rebecca Cook, (Wiley)
- Sarah Fricker, (IOPP)
- Thib Guicherd-Callin, (LOCKSS Program, Stanford University Library)
- Clare Hodder, (rights2 Consultants)
- Claire Holloway, (OCLC)
- Neil Jefferies, (Bodleian Libraries, University of Oxford)
- Ruth Jones, (RJ Digital Consulting)
- Roxanne Missingham, (Australian National University)
- Lynette Owen, (Lynette Owen Consulting)
- Sylvia Pegg and Becky Roberts, (Royal Society of Chemistry)
- James Phillpotts, (Oxford University Press)
- Wendy Reid, (Rakuten)
- Denis Shannon, (University of Wyoming)
- Tzivya Siegman, (W3C)
- Alicia Wise, (CLOCKSS)

Appendix B

Sub-Working Group Members

A sub-group was established specifically to look at copyright issues related to digital preservation. Members included:

- Rebeca Cook, (Wiley)
- Sarah Fricker, (IOPP)
- Clare Hodder, (RightsZone)
- Roxanne Missingham, (Australian National University)
- Lynette Owen, (Lynette Owen Consulting)
- Alison Pope, (IOPP)
- Becky Roberts (Royal Society of Chemistry)
- Denis Shannon, (University of Wyoming)
- Diane Spivey (Publishing Rights and Contracts Consultant)
- Alicia Wise, (CLOCKSS)