PRESERVING BOOKS FOR FUTURE GENERATIONS

CLOCKSS

AUTHORS

This guide was produced in working groups convened by CLOCKSS in 2023 and 2024. Many thanks to the following people for their expertise and helpful contributions:

Graham Bell
(EDItEUR)
David Cirella and Euan Cochrane
(Yale University Library)
Gareth Cole
(COPIM and University of Loughborough)
Rebecca Cook
(Wiley)
Sarah Fricker
(IOPP)
Thib Guicherd-Callin
(LOCKSS Program, Stanford University Library)
Clare Hodder
(rights^2 Consultants)
Claire Holloway
(OCLC)
Neil Jefferies
(Bodleian Libraries, University of Oxford)

Ruth Jones
(RJ Digital Consulting)
Roxanne Missingham
(Australian National University)
Lynette Owen
Sylvia Pegg and **Becky Roberts**
(Royal Society of Chemistry)
James Phillpotts
(Oxford University Press)
Wendy Reid
(Rakuten)
Denis Shannon
(University of Wyoming)
Tzivya Siegman
Alicia Wise
(CLOCKSS)

For further information please contact info@clockss.org

AUDIENCE

This guide is intended for senior leaders in book publishing organizations. The target audience is board and executive leadership, with scope across editorial, legal, marketing, metadata, production, rights, and related functions. The guide is focused on the foundational importance of digital preservation for responsible publishing organizations and the steps that senior leaders can take to facilitate this. No assumptions are made with regard to knowledge or a direct role in the legal or technical mechanics of meeting preservation requirements.

SCOPE

The scope of this guide is concerned with the preservation of the digital formats of books created and distributed by publishing organizations. This scope explicitly excludes some related areas which might be of interest to the intended audience:

- Publishing organizations may also be concerned with the appropriate archiving of the physical formats of books. However, this requires a specialist skill-set covering the physical conservation of perishable materials, and is not in the scope of this guide.

- Self-publishing authors may rightly be concerned with the preservation of their work. However, this guide specifically addresses the considerations of organizations undertaking the ongoing publication of books, which entails particular responsibilities for ensuring a stable and repeatable infrastructure, including for digital preservation. Self-publishing authors may want to read this guide for a broad understanding of these considerations, but should be aware that they are not its intended audience.

- Many book publishing organizations will also undertake the publication of other works, such as journals, magazines, or datasets. These organizations should also be concerned with the appropriate archiving of such materials. However, the differences in publishing models, formats, and processes between book and non-book content are substantial and there are different established norms to be aware of with regard to digital preservation. As such, the scope of this guide is solely limited to book publishing.

WHAT IS DIGITAL PRESERVATION?

The word 'archive' can be used in different ways. Sometimes it just means a collection of content you've digitized or published (for example, you may use it as part of a product name) and at other times it means an actively curated collection that is not meant for short-term access but rather for long-term preservation. It's the latter sense of the word that we focus on in this guide. For clarity, archiving in this sense is commonly referred to as 'digital preservation'. This differs from any secure storage arrangements you might already have in place, in that it explicitly accounts for the preservation of published materials agnostic of - and potentially outlasting - any specific publishing organization.

WHY PRESERVE?

Long-term digital preservation is a shared interest for book publishing stakeholders. There are many reasons a responsible book publishing organization should be concerned with archiving for long-term preservation, including that:

- **Your customers require it.**
  This is especially true if you publish books used in higher education institutions. Preservation is a standard requirement in agreements between publishers and university libraries/ consortia.

- **Your authors expect it.**
  Authors expect their contributions to be part of our cultural and intellectual heritage; and if used for research or teaching, they will also expect them to form part of the scholarly record. This means that this content needs to be preserved in perpetuity to be available to the readers and researchers of the future, outlasting your current organization.

- **You need it.**
  Long-term preservation should be part of your disaster recovery planning and strategy. This provides your organization with insurance for your valuable content, and a safety net to meet an array of commercial obligations. Publishing organizations should also be aware of any legal or regulatory requirements in specific jurisdictions, although it is important to note that these should not be considered an alternative to digital preservation.

DEVELOPING A PRESERVATION POLICY

A preservation policy can help your organization articulate the scope, activities, and principles that support the digital preservation of materials in your care. The size, type, needs, and resources of book publishing organizations will vary considerably. However, all publishers have a responsibility to consider the preservation requirements for the materials they publish.

Each publishing organization is also responsible for communicating its preservation policies and practices to customers, authors, readers, and users. Having an explicit digital preservation policy helps to ensure that preservation activities are repeatable, visible, and transparent to stakeholders. Critically, this includes transparency regarding what your organization is, and is not, preserving.

You are not alone in this. This guide seeks to clearly explain the importance of digital preservation for book publishing organizations, but there are also resources and services available to provide assistance with developing your own preservation policy. For example, the NASIG Digital Preservation Committee has developed a Model Digital Preservation Policy: **https://nasig.org/NASIG-model-digital-preservation-policy** to help you get started. This tool is designed to help you measure, grow, and publicize your organization's commitment to preserving the scholarship entrusted to it.

PREPARING TO PRESERVE

Before addressing potential barriers to implementation, there are a few core considerations which will help to shape your organization's approach to preservation:

- **Consider what to preserve.**
  Any part of your portfolio (up to and including all of it) which can be identified as having lasting cultural or intellectual value should be preserved. For example, this should include different editions of books. It is important that only final, published content is preserved.

- **Consider content and functionality.**
  Archive at least the content you are aware would be needed to recreate your portfolio, if this were to be resupplied back to you and you had to start from scratch. Ideally, archive any additional information that a future organization without your institutional

knowledge would need to prepare this content for use.

- **Consider which formats to preserve.**
  In order for the archived content to be most readily accessible and usable in the future, general guidance is that more open formats will be preferable. For example, the Library of Congress has a guide to recommended formats: h**ttps://www.loc.gov/preservation/resources/rfs?loclr=blogsig** that sets out the most preservable approach to different types of content (including text, image, video, audio, database, geospatial, software, and more). There are also resources available to assist with overcoming the preservation challenges presented by highly interactive web-based scholarly resources (such as the Guidelines for Preserving New Forms of Scholarship: **https://doi.org/10.33682/221c-b2xj**).

- **Consider where to preserve.**
  Ideally books should be preserved in a 'dark archive' while the publisher can still provide access to the electronic files, to ensure continuity of availability. These dark archives do not allow any access to preserved material and only open content for access following very specific and catastrophic events (known as 'trigger events' or 'post-disappearance events'), as defined with participating publishers. There are a number of third-party digital preservation services which you can consider partnering with. These provide the necessary infrastructure and technical expertise to address long-term preservation in dark archives and the release of content from these as and if appropriate. These services operate under several different models, including some which are not-for-profit and/or are community-governed. Different services may maintain different types of content and metadata, as well as having different processes for collecting, ingesting and preserving content. As such, it is important to consider whether their policies and procedures align with yours, as well as the potential costs involved in preserving your content with them.

OVERCOMING POTENTIAL BARRIERS

FUNDING

"Isn't this going to be expensive?"

Different digital preservation services have different models to cover the costs associated with archiving. When considering where to preserve, book publishing organizations should be aware of these models and factor anticipated costs into their approach and partnering decisions. For example, depending on the service, these costs may scale on factors such as your organization's revenue or volumes of material being archived.

BUSINESS MODELS

"Doesn't having our content available elsewhere undermine our sales?"

Digital preservation services for publishers, such as **https://clockss.org/** and **https://www.portico.org/**, operate as 'dark archives' and will only permit access if a catastrophic event occurs and the publisher is not able to provide access to customers. Not preserving may be a barrier to some library customers making a purchase and therefore will be more harmful to sales.

"Doesn't the availability of older content undermine our business model?"

When considering what to preserve, book publishing organizations should pay particular attention to the preservation value of different editions or versions of books, and the most suitable definition of the specific and catastrophic events ('trigger events', 'post-disappearance events') that would result in permitted access to these. With consideration of these factors, book publishing organizations can make the most appropriate decisions regarding preservation in light of their specific business models. Preservation policies should be explicit regarding what materials will or will not be preserved.

RIGHTS ISSUES

"Do we actually have the necessary rights to do this?"

- Copyright is an important issue when thinking about digital preservation and authors need to be engaged early and often. It can be challenging to determine the specific rights held in books, for example, with regard to image clearances. As such, entering into a preservation agreement might be thought to present the risk of inadvertently violating the rights of an author or other creator, agent, or third-party publisher. Some of the rights considerations for book publishing organizations include:

- Author agreements are often entirely silent about digital preservation.

- Author agents can have a lot of influence for some books, which may result in the publisher being granted quite a limited and/or short license.

- Where author agreements address digital preservation, limitations by geography and variations by territory might still impact practical arrangements. For example, a publisher might hold the European rights to authorize digital preservation but not those for the US.

- It is common practice for rights to revert to authors. However, it is not always possible for book publishing organizations to track which rights have reverted and when.

- Third-parties may have relevant rights that the book publishing organization is unaware of because these have been granted by the author or author's agent.

- The chapters in edited volumes can each have a different rights position. For example, one chapter may be a work for hire and another might involve a copyright transfer with royalties.

- Print on demand rights may mean that books remain available via this route after their primary sales window. Availability of print-on-demand versions should be confirmed prior to making books accessible from a digital archive as, although they may not be available digitally, they may still not be appropriate to trigger.

- Books can become 'orphaned works', which are known to be subject to copyright, but with rightsholder(s) who are either unknown or

cannot be found. This leaves it unclear where or what rights are held and hence how to determine whether the necessary rights are in place to permit digital preservation.

In order to help overcome these rights challenges, it is recommended that book publishing organizations talk to their authors (and/or their agents) about digital preservation and include digital preservation language in author contracts. Effective rights management is also critically important. CLOCKSS have created a map: **https://clockss.org/bookwill/** that illustrates the ways that relevant rights to carry out the digital preservation of books can flow to digital preservation services.

COMPLEXITY

"Isn't this complex and resource intensive?"

Digital preservation can sound complex, but the good news is that the infrastructure to realize this has already been established and, by its nature, exists outside of your organization. After thinking through the core considerations set out above, simply get the appropriate content to your selected digital preservation services and once they have received this, they will handle the technical challenges. These organizations will also be able to provide you with help and advice on the practicalities of working with them to ensure ongoing preservation. For example, you might already be sending some of your content to other organizations that could help - with advice from your digital preservation service provider, it may be possible and appropriate to work through intermediaries you already partner with.

HOW TO PRESERVE

When working with a third-party digital preservation service, they will provide guidance to the relevant team in your organization on the details of working with them. For illustrative purposes, below is a high-level overview, based on the processes for CLOCKSS.

1. Provide samples of your full-text content and metadata to your chosen digital preservation partners for evaluation. They will likely provide guidance or feedback on the content, folder structure, and packaging of your content. For example, guidance on how to ensure that the relationships between content and metadata

files are clear and consistent.

2. Establish a mechanism to provide content for preservation in a structured way, for example, by SFTP. Some services may be able to harvest content from your website, with the right permissions.

3. Confirm with your preservation partner an appropriate frequency for deliveries and how to ensure that all deliveries can be uniquely identified. If a file needs to be subsequently revised, unique identification (for example, by file name) helps to ensure this. Services may be able to automatically detect replaced files and establish appropriate versioning of previously archived copies.

In terms of general guidance on the content and formats that digital preservation services would expect to receive, this is likely to include stipulations for:

- Clear and consistent file and format identifiers.
- Use of non-proprietary formats, such as PDF, HTML, and EPUB.
- Standard schemas for metadata, such as BITS or ONIX.
- Clear license information.

Creating and maintaining a book manifest page with listings of books published each year will allow automated discovery of new content by preservation services collecting content from the publisher website.

IN SUMMARY

Long-term digital preservation is a shared concern for book publishing stakeholders. Responsible book publishing organizations should be aware that this will be a key consideration for your customers and authors, as well as a critical need for your organization. A specific preservation policy can help your organization articulate the relevant scope, activities, and principles that support the digital preservation of materials in your care. There are resources and services available to provide assistance with the creation of such a policy and its practical implementation.

Factors which will shape your organization's approach to preservation include considering what to preserve, the content, functionality, and formats related to this, and where to preserve materials. There may also be several perceived barriers to preserving books, including access concerns, the complexity of doing so and the funding necessary, potential impacts on business models, and rights issues. While these factors require due consideration, there are mitigations to each. In terms of concerns regarding practical implementation, when working with a third-party digital preservation service, they will be able to provide direct guidance on the relevant details.

We hope that this guide provides a useful overview for senior leaders in book publishing organizations of the foundational importance of digital preservation and the practical and immediate steps that they can take to facilitate this.

FURTHER READING

Although far from an exhaustive list, provided below are links to a number of resources for further reading on this topic.

- A comprehensive resource on digital preservation is the Digital Preservation Coalition's "Digital Preservation Handbook": **https://www.dpconline.org/handbook/digital-preservation** This includes sections on organizational activities and technical solutions and tools, as well as further information on the context of why digital preservation matters.

- As an example of national library requirements, where these apply, the British Library provides guidance on legal deposit: **https://www.bl.uk/help/how-to-deposit-your-digital-publications**

- To help with the creation of your own preservation policy, NASIG have created a Model Digital Preservation Policy containing all the relevant sections which can be tailored to the circumstances and needs of your organization: **https://nasig.org/NASIG-model-digital-preservation-policy**

- The Library of Congress provides an annually revised "Recommended Formats Statement" which can help you identify the best format to use for different types of digital content: **https://www.loc.gov/preservation/resources/rfs/**

- Guidelines relating to the preservation of complex publications making use of new technologies were covered by Greenberg, Hanson, and Verhoff in their 2021 paper "Guidelines for Preserving New Forms of Scholarship": **https://doi.org/10.33682/221c-b2xj**

- The relevant rights to carry out the digital preservation of books can flow to digital preservation services. CLOCKSS have produced a map that illustrates the ways this can happen and encourage stakeholders to discuss their book's last will and testament: **https://clockss.org/bookwill/**

- Third-party digital preservation services will provide guidance on the details of working with them. As an example of this, the full guidelines for CLOCKSS digital preservation processes are available: **https://lockss.github.io/clockss-file-transfer-guidelines.html**

- Preservation provider metadata can be included in ONIX records for books. See page 4 of this EDItEUR application note: **https://www.editeur.org/files/ONIX%203/APPNOTE%20Did%20you%20also%20know.pdf**

- This page on the Royal Society of Chemistry website is an example of a book manifest, which allows automated discovery of new content by preservation services: **https://books.rsc.org/books/browse-by-year**

- In their 2023 paper "'Good, Better, Best': Practices in Archiving & Preserving Open Access Monographs", Barnes et al. explore potential preservation solutions for small and scholar-led publishers of open access monographs: **https://doi.org/10.5281/zenodo.7876048** This work was produced as an output of the Community-led Open Publishing Infrastructures for Monographs (COPIM) project.

For further information please contact info@clockss.org