

NASIGuide: Digital Preservation 101

Prepared by the Digital Preservation Committee

Updated 24 April 2024

[PDF](#)

What is digital preservation?

Conscious implementation

Digital preservation involves conscious planning and implementation to maintain digital materials over the long term. It refers to “the series of managed activities necessary to ensure continued access to digital materials for as long as necessary” (definition from the [Digital Preservation Coalition](#)). It is a suite of services and an ongoing process.

Activities in an overall digital preservation strategy include but are not limited to:

- metadata management (such as administrative, technical, provenance, and metadata preservation)
- fixity checking and auditing
- file format migration
- infrastructure development and maintenance

Why is digital preservation crucial?

Maintaining your digital legacy

As there is a growing amount of digital information being created and consumed, it is imperative that we ensure the content lasts through time and responds to changes in user communities, formats and software. Preservation of digital content is key to ensuring the valuable digital assets we create today remain accessible and useable for future generations.

If a publisher discontinues a title or removes content from their site, it could become unavailable to future scholars. This breaks the citation chain so others cannot verify cited works, thereby putting much of the scholarly system at risk.

Responsibility

Libraries have long preserved scholarly output and continue to do so today. As Librarians we are responsible for the digital cultural heritage. Both libraries and publishers have a mutual obligation to ensure long-term preservation of the scholarly record – leveraging library expertise in preservation, publisher expertise in content production, and the author relationships and financial means of both parties.

Agreements between the parties are good places to document these mutual obligations. These might be subscription agreements, open access publishing agreements, or agreements that blend the two.

Backups and Snapshots

While the following practices and services are important aspects of, or related to, digital preservation, any of **these approaches alone does not represent a sufficient digital preservation strategy**. These are not considered digital preservation because they are more temporary in nature, or not over several locations and most importantly there is a more established method. See [Digital Preservation-Services and Initiatives](#).

Digitization

Digitization is a key process for providing access to analog collections, including text, photographic, and other audiovisual materials. When planning for digitization, it is important to also make plans for handling and preserving the digitized masters to ensure the digitized masters continue to be authentic and available into the future.

Backups

Backups are snapshots or full copies of a work. These copies are ideally refreshed periodically. However, they do not involve archival or curation activities to ensure ongoing integrity, access, usability, or understandability of materials. It is essential to follow the best standard practices for backing up your data. There should be at least three copies of the files, preferably on at least two different types of media, with at least one copy off-site (i.e. in different physical locations) or cloud-based storage. It is best to have a digital preservation strategy in addition to back-ups.

Commercial Server Backups

There are a multitude of new software being launched and hosted on commercial cloud providers for scholarly communication services. Most commercial server service providers have a version of backup for data from planned and unplanned events to the software. Please note that you must also check the GDPR and personal data compliance of these services as well. Services include but are not limited to:

- [Microsoft Azure Storage Redundancy](#)
- [Amazon Web Service WORM](#)

Internet Archive (IA)

The Internet Archive is an online repository for providing access to digital content, including archived versions of web pages and scholarly publications. [Project JASPER](#) To learn more about IA's digital preservation collaborations head to section, Project JASPER.

The Internet Archive is a useful platform for providing ongoing access to website snapshots. The Internet Archive's [Wayback Machine](#) captures web pages in situ for future use. If your site is down temporarily, the Wayback Machine is one of the first places a reader may look.

If you are adding your web page to the Wayback Machine, keep in mind that web pages in the Wayback Machine represent a snapshot of the page taken at a specific time, not a complete record of the page/site. Be sure to check that your page or site has been crawled with all the associated content (PDFs, videos etc.).

You can add a single crawl of a specific page to the IA by adding the URL to the "[Save Page Now](#)" box. Here is a blog post with [additional methods](#) to save individual pages.

A more robust method for archiving web pages and websites is the Internet Archive's [Archive-It](#) service. Archive-It provides more control over crawled webpages, including metadata creation,

access control, and file export for active preservation tasks. If the publication is connected with an institution that has a subscription to Archive-It, you could ask that the subscriber (probably the institution's library) to crawl your content.

Additional web archiving tools and software can be found on the International Internet Preservation Consortium's [Tools and Software page](#).

Digital Preservation - Services and Initiatives

Regional and National Archiving at your Service

As awareness about the need for digital preservation grows, the choice of archiving services has grown with it. Today, regional and national initiatives, some led by research library consortia, now operate alongside large third party archiving agency services with a global reach. Participating in these services or initiatives can be part of your preservation strategy. The following are more established methods of digital preservation.

LOCKSS (Lots of Copies, Keep Stuff Safe) Community of Services

LOCKSS is a general-purpose peer-to-peer open source digital preservation technology. The software underpins a variety of different digital preservation services.

Global LOCKSS Network (GLN)

[The GLN](#) preserves content that is generally available online, including materials in both open access and subscription-only journals and books. Libraries collaborate through a distributed network to preserve content as it appears on the publisher's site, with regular integrity checks of the data. The preserved content can be used for failsafe access and post-cancellation access.

Private LOCKSS Networks (PLN)

The PLN typically preserves content either from one type of software or a specific geographic area. PLNs may also preserve any digital content, not only books and journals. LOCKSS has a list of [additional PLNs](#).

METAARCHIVE

The [MetaArchive](#) is an example of a PLN service. It is expanding services beyond the PLN.

PUBLIC KNOWLEDGE PROJECT (PKP) PLN

[PKP provides a PLN service](#) with the Open Journal Systems (OJS) 23.4 and above.

CLOCKSS

[CLOCKSS](#) is a community-governed dark archive of scholarly content and is another Private LOCKSS Network. Copies of all of the content are held at twelve leading [libraries](#) around the world, running the LOCKSS software to ensure that the content is preserved in perpetuity. If content that is held within the CLOCKSS archive disappears from the publisher's site, or is about to disappear, CLOCKSS will trigger that content and make it available to everyone open access. The CLOCKSS [Board](#) is comprised by twelve publishers and twelve libraries. Both publishers and libraries [support](#) CLOCKSS. The cost to publishers is based on annual journal or ebook revenue. Information on [joining as a publisher](#). Information on [joining as a library](#). Note that a library that publishes may choose to join both as a library and as a publisher.

CLOCKSS has published [a guide to help book publishers get started with digital preservation](#).

Portico

[Portico](#) is a community supported dark archive committed to ensuring that scholarly content published in electronic form remains accessible for the long term. Portico operates a proprietary preservation service that has two mirror copies of the content. Portico's primary access scenario is a "trigger event." When content is no longer available online from the publisher or any other source, Portico makes it available for use. Access is only for those who purchased the book or journal, unless the publication was open access. The cost to publishers is based on annual journal or ebook revenue. Information on [joining as a publisher](#). Information on [joining as a library](#). Note that a library that publishes may choose to join both as a library and as a publisher.

Consortial Trusted Digital Repositories

Some libraries have partnered to create a [Trusted Digital Repository \(TDR\)](#). A TDR meets specific international standards (ISO 16363). Currently there are six that have been [certified](#) (including CLOCKSS and Portico). The four others are:

- [Canadiana.org](#)
- [Chronopolis](#)
- [Hathitrust](#)
- [Scholars Portal](#) (a consortia of 21 university libraries in Ontario, Canada)

Library of Congress

The Library of Congress is as committed to the acquisition and preservation of digital content, as it is to analog content. They are currently dedicated to a robust digital collecting strategy and serve as one of the Keepers in the Keepers Registry. More information about [LC's digital preservation](#).

Project JASPER

[Project JASPER](#) (JournAIS are Preserved forever) is a collaboration between CLOCKSS, DOAJ (Directory of Open Access Journals), Internet Archive, the Keepers Registry and PKP to preserve diamond OA journals. Journals are preserved by either the PKP PLN, CLOCKSS, or on a best-efforts basis by the Internet Archive.

Over to You: How to get started

A policy to help your organization

The NASIG Digital Preservation Committee has put together a model policy for you and your team to use. Not only is this policy a framework of organizational decision making, it also helps with the language and method of how to communicate the importance of preserving your content for good and for all. Having a policy will help you on your journey to measure, grow, and publicize your organization's commitment to preserving its scholarship.

Download the [NASIG Digital Preservation Model Policy](#).

Examples of policies that were developed using the model policy:

- [NYU Libraries Digital Preservation Policy](#)

Other Tips and Sources

Not sure what journals are at higher risk? Check out the Keepers Registry [The Keepers Registry](#) is an index of journals that have been preserved by one or more [archiving agencies](#) committed to ensuring long-term access to the scholarly and cultural record. Librarians use the service to check that important titles in their collection development priorities have been preserved. As a publisher, you should check that your titles are correctly listed and that the preservation coverage of your titles is as you expect. If your title's bibliographic information is wrong, we suggest you follow up with the Keepers Registry. If your preservation coverage is not as complete as you expect, we suggest you ask the archiving organization to better understand the situation. Ideally, your journal—and every issue—is held by three different keepers. For additional information, see our [Guide to the Keeper's Registry](#).

Library of Congress' Recommended Formats Statement

The Library of Congress' [Recommended Formats Statement](#) (updated annually) provides good guidance for publications in all formats to make sure they will last through time.

Digital Preservation Handbook

The Digital Preservation Coalition has created a [handbook](#) which “provides an internationally authoritative and practical guide to the subject of managing digital resources over time and the issues in sustaining access to them. It will be of interest to all those involved in the creation and management of digital materials.”

Other Organizations

- [ALA ALCTS Digital Preservation Interest Group](#) regularly has sessions at ALA, as well as other activities hosted by [ALCTS Preservation and Reformatting Section](#)
- [Digital Preservation Coalition](#) is a membership organization helping to support digital preservation practitioners around the world.
- [National Digital Stewardship Alliance](#), hosted by Digital Library Federation
- Open Access Scholarly Publishers Association, includes archiving in their [“Principles of Transparency and Best Practice in Scholarly Publishing”](#)
- Center for Research Libraries has several [“Digital Preservation Metrics”](#)
- [Northeast Document Conservation Center](#)
- [Preserving digital Objects With Restricted Resources \(Digital POWRR\)](#)

Further Information:

- [Digital Preservation Questions and Talking Points](#)
- [Guide to the Keeper's Registry](#)
- [NASIG Model Digital Preservation Policy](#)
- [Center for Research Libraries, LibLicense Digital Preservation](#)



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).