NASIGuide: Digital Preservation 101

Prepared by the NASIG Digital Preservation Committee

Updated January 2020

Now that scholarly publishing has transitioned to the web, it is imperative that we ensure the content lasts through time. 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. Libraries have long preserved scholarly output and continue to do so today. However, there are steps publishers can and should take to make sure the content they publish remains available.

"Digital preservation 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.

Not Preservation

Backups

Make sure you are following standard best 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.

Internet Archive (IA)

If your site is down temporarily, the Internet Archive Wayback Machine is one of the first places a reader may look. Be sure to not only check that your site has been crawled but that all the content (PDFs, videos etc.) are accessible through the Internet Archive (IA). You can add a single crawl of a specific page to the IA by adding the URL to the "Save Page Now" box (blog post with additional methods to save individual pages). If the publication is connected with an institution that has a subscription to Archive-It, you could ask that the subscriber (probably the library) crawl your content.

Additional web archiving tools and solutions for Internet citations can be found in this guide from the European University Institute.

Preservation

As awareness in preservation grows, the number 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 global reach.

LOCKSS (Lots of Copies, Keep Stuff Safe)

Libraries collaborate through a distributed network to preserve content as it appears on the publisher's site, with regular integrity checks of the data. More information.

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. There is no cost for a publisher to join, but there is limited space. Information on joining as a publisher.

Private LOCKSS Network (PLN)

PLNs such as MetaArchive or the PKP PLN (which is for anyone using Open Journal Systems (OJS) software) typically preserve content either from one type of software or a specific geographic area. PLNs may also preserve any digital content, not only books and journals. If you use OJS, please opt in to the PKP PLN. MetaArchive is expanding services beyond PLN. LOCKSS has a list of additional PLNs listed by LOCKSS.

CLOCKSS

CLOCKSS is a community-governed dark archive of scholarly content, with copies of all of the content at twelve leading libraries around the world, running the LOCKSS software to ensure that the data remain valid. When CLOCKSS triggers journals that would otherwise disappear, the journals are always Open Access. As of June 2018 CLOCKSS has triggered 53 journals. The CLOCKSS Board is comprised by twelve publishers and twelve libraries. Both publishers and libraries support CLOCKSS.

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'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. 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)

Other Tips and Sources

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 Keepers 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
- 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