A Collaborative Approach to Preserving At-Risk Open Access Journals: "Journals Preserved Forever"

iPRES October 2021, Beijing

Alicia Wise, Executive Director, CLOCKSS Archive

Abstract

Scholars rely upon ongoing access to the published research in their fields. For journals to disappear is to harm the ability of scholars to conduct their work. Recent research has indicated that journals published by the long tail of smaller publishers may be especially at   risk for disappearing from the Web. Five leading organizations are collaborating to significantly improve the number of Open Access journals that are being preserved for the long-term, in a project named "Journals Preserved Forever" (nicknamed "JASPER").

Introduction

The recent paper by Laakso et al. (1) showed that 174 Open Access journals have disappeared from the Web since 2000. This finding serves as a "call to arms" for organizations which are committed to the long-term preservation of the online scholarly literature. Five partners -- the Directory of Open Access Journals (DOAJ), the Keepers Registry, the CLOCKSS Archive, the Internet Archive, and the Public Knowledge Project (PKP) -- are working together on a project to provide preservation options that are easy   and affordable for small publishers to work with. The goal is to increase the number of journals being preserved by hundreds of titles in the near-term, and thousands over the next few years. Phase One focuses on diamond OA journals.

Why Do Journals Disappear?

Many journals run on very small budgets, and have little or no revenue. Often such journals are "labors of love" by an individual professor, who operates the journal from their   academic department as a side-project. Diamond OA journals are discussed in detail in Bosman et al. (2). These journals may not participate in the broader community of academic publishing, and may be unaware of many of the principles of academic publishing, such as long-term preservation. Such journals may lack technical expertise and infrastructure. According to Bosman et al., 75% of journals are unable to format their content either in XML or HTML, providing only PDF in most cases. The journals might not generate income through an Author Processing Charge (APC), and may have no other source of revenue, other than, in some cases, modest financial support from the Editor-in-Chief's academic department. If the journal does not gain traction in its community, it may have few manuscript submissions, and the Editor might choose to cease publishing the journal, and might take it down from its website. In this way, the journal's published content could disappear from the Web.

How Can Access to Disappearing Journals Be Preserved?

Digital Preservation systems such as the Public Knowledge Project's Preservation Network (PKP PN), the CLOCKSS Archive, and the Internet Archive, provide services to ensure that the published scholarly literature will continue to be available online, even if the publisher has ceased to make it available. These systems make copies of the journals and store them securely, using technical solutions to ensure the health of the bits, so that if they need to make the journals available, they know that they will be able to do so. Each of these systems is different, and offers different pluses and minuses.

How the Project Will Improve the Preservation Coverage of OA Journals

The DOAJ (Directory of Open Access Journals) serves as a kind of central hub within the fully OA journal ecosystem. Being included in DOAJ is a mark of credibility for an OA journal. Once listed in DOAJ, a journal can choose to send metadata of its published articles to DOAJ, along with other basic information. DOAJ is unique in that it is the one place that thousands of fully OA journals from all over the world interact. The "Journals Preserved Forever" project will build on this important piece of infrastructure. Centralizing the communications and interactions with so many small publishers will greatly strengthen the ability of preservation agencies to successfully work with this diverse community.

7000 of the journals that are listed in DOAJ do not ensure their content is made available to any kind of preservation service.

The project aims to provide three easy options for the preservation of these DOAJ-listed journals, and motivate their sponsors to join the scheme.

Three Preservation Options

OPTION 1. PKP PN: Forty percent of the DOAJ-listed journals that are not being preserved are using Open Journal Systems (OJS), which is an open source publishing tool from the Public Knowledge Project. PKP also offers its Preservation Network (PKP PN), which OJS users can choose to participate in via an easy opt-in workflow, provided their OJS version is compatible. PKP is strengthening its documentation, and conducting outreach, to encourage OJS users to participate in the PKP PN. There is no cost for an OJS-using publisher to preserve their journals in the PKP PN.

OPTION 2: CLOCKSS. DOAJ's central hub will begin to accept full-text files from OA publishers. This transmission will be integrated with the metadata submissions that publishers already make. Collecting metadata and full-text from many OA publishers into a centralized standardized interface will greatly simplify the process   for digital preservation service providers. The CLOCKSS Archive will harvest the full-text from this system, to preserve in its international distributed archive.

OPTION 3. INTERNET ARCHIVE: Publishers will also have the option for their material to be archived in the Internet Archive , especially publishers who lack the ability to provide metadata or are unable to upload full-text copies, but who can provide website URL(s) for the best-effort automated web harvesting of their articles. IA will provide different levels of no-cost service options including file upload, web crawling, and user tools.

And a Single Point of Information

PKP PN, CLOCKSS, and Internet Archive report their archiving operations to the Keepers Registry (keepers.issn.org) which is managed by the ISSN International Centre. The archival status of the OA journals preserved under this scheme will thus be made available for free on the web to fuel research and professional practices. As Keepers Registry aggregates archival metadata provided by 10+ archiving agencies worldwide and will bring in more, the plan is to involve them in the scheme at a second stage.

The Project Plan

Phase One focuses on "diamond" OA journals, i.e. those journals with no Author Processing Charge. These titles are viewed as being the most at-risk of disappearing, because they may have the least financial sustainability.

Future phases will open the system to other types of journals, and to more preservation agencies.

A pilot will be conducted in Q3 2021, with the Minimum Viable Product to be launched in Q4. The five project participant organizations are contributing funding to launch the project. The organizations are seeking grant funding to support the further building of the system, and the start-up period of operation. During this time, the sustainable ongoing funding will be defined.

The project will also encourage the use of Digital Object Identifiers (DOIs), which are an essential component of the joint scholarly publishing infrastructure. DOAJ looks to collaborate with Crossref, to support small OA publishers to adopt DOIs.

All five project participants are contributing to the development of the system, with technical work by DOAJ, IA, and CLOCKSS/LOCKSS related to a central repository and data sharing and distribution.

How the Digital Preservation Works

Preservation requires more than just making a back-up copy. There is a spectrum of preservation solutions. The National Digital Stewardship Alliance publishes Levels of Preservation, which shows different levels of preservation for five functional areas: storage, integrity, control, metadata, and content.

CLOCKSS and PKP PN both use the LOCKSS preservation software, with multiple copies continuously validating the accuracy of the preserved content by cross-checking among themselves. Each copy is in a different geographic location, and being managed under different organizational regimes; this ensures no single point of failure.

Internet Archive will store a minimum of four copies in a minimum of two geographic areas. IA owns and operates its own data centers, thus limiting the risk and costs of reliance upon commercial cloud storage or reliance upon member-contributed infrastructure controlled by external organizations.

These preservation systems ensure that the content will continue to be available, if the publisher ceases to host the content.

Conclusion

The ongoing availability of research resources is of paramount importance to scholars, who must be able to access and build upon all of the published research in their fields. As an industry, we are pledged to eliminate the possibility that these high-value resources can disappear. "Journals Preserved Forever" aims to close the gap in preservation coverage that currently exists among diamond Open Access journals.

Acknowledgments

References

1. M Laakso, L Matthias, N Jahn. Open is not forever: A study of vanished open access journals. *Journal of the Association for Information Science and Technology*, 21 February 2021. https://doi.org/10.1002/asi.24460
2. J Bosman, J E Frantsvag, B Kramer, P-C Langlais, V Proudman. OA Diamond Journals Study, Part 1: Findings. 9 March 2021. DOI: 10.5281/zenodo.4558704