
A collaborative approach to preserving at-risk open access journals

Dom Mitchell, DOAJ

**In association with CLOCKSS, Internet Archive, Keepers
Registry and PKP**

Summary

Overview

Who is involved

Problems to solve

Project objectives

Scope

Phase 1

Method

Permanent article
identifiers

Ingesting the content

Deliverables

Estimated Timeline

Overview

- “Long-tail” journals are at-risk of disappearing.
- Project group of five leading organisations best placed to make this work
- A central hub where from preservation agencies can harvest consistent metadata and full-text
- Offer free and low-cost options for preservation and access.
- Phase 1 targets “diamond” OA journals (those with no article processing fees)

Who is involved



Problems to solve

1 Journals vanish from the Internet taking with them the research they published, leading to “reference rot” and broken URLs.^[1]

2 Some archiving and preservation solutions that exist today come with financial and technical barriers that prevent some journals from taking part

3 There is an initial investment in time and tech required by each service to set up a new journal before content is ingested. This can be expensive.

4 Communications with thousands of small publishers and a lack of standard practices makes archiving difficult.

5 Awareness among journal owners, editors and publishers about why archiving and preservation are important is low.

¹. *Open is not forever: a study of vanished open access journals* Laakso et al. <https://arxiv.org/abs/2008.11933>

Project objectives

- Increase the number of journals in DOAJ whose content is actively archived and preserved.
- Provide an important piece of infrastructure for the open access publishing community.
- Provide a solution that overcomes the financial and technical barriers that hinder unfunded journals from taking part in archiving solutions.
- Foster best practice and standardization among the long tail of DOAJ-indexed journals.
- Raise awareness among editors and journal owners on why archiving journal content is so important.

Scope



Phase 1

There are 7500 DOAJ-indexed, “diamond” open access journals not currently archived.

Only journals in DOAJ are accepted:

- ensures that they have an ISSN
- greatly reduces the possibility that they are predatory.

Over 50% of the journals use PKP's OJS journal hosting platform.



Method

An initial survey will be sent to 7500 journals to gather information on how ready the journal is to take part in one of three solutions. Responses will divide the journals into 3 groups:

- 01 *Those using an OJS version compatible with the PKP PN (3.2+)*** will be directed to resources that tell them how to take part.
- 02 *Those that can produce and export metadata and full text*** will be encouraged to participate in CLOCKSS
- 03 *Those that cannot produce and export metadata*** will be encouraged to allow Internet Archive to crawl their sites



Permanent article identifiers facilitate discovery and archiving

Using a DOI facilitates discovery and archiving.

Due to technical and financial barriers, not all journals use DOIs *and* ensure that the DOIs resolve.

DOAJ is hoping to collaborate with Crossref and have permission to assign DOIs on behalf of those journals that do not use any form of permanent article identifier.

Ingesting the content

01



Internet Archive and DOAJ will maintain an FTP server

Approved publishers will send metadata and full text to DOAJ. DOAJ will place this in the FTP server, hosted by Internet Archive, for CLOCKSS to collect daily.

03



Information on preservation status in Keepers Registry

The preservation data (status, which service, range of content and whether its active or not) is aggregated in the Keepers Registry.

Using the PKP PN

Journals will find out if they are on the right version of OJS and given help on how to add their content to the PKP PN. Journals that are not on a compatible version of OJS will be given advice on how to upgrade, or participate in Options 2 or 3.

02



Internet Archive will crawl websites

The IA crawlers will ingest content. This is done in an automated way with only moderate “human QA”, although there are tools for the journal to report/improve holdings completeness.

04



Deliverables

A sustainable solution that facilitates archiving for unfunded open access journals.

Aggregated archiving data available in Keepers Registry.

Goal is at least a 50% reduction in the number of “unarchived” journals in DOAJ.

Better documentation and training tools for raising awareness on why archiving is good for journals.



Estimated timeline

Discovery phase

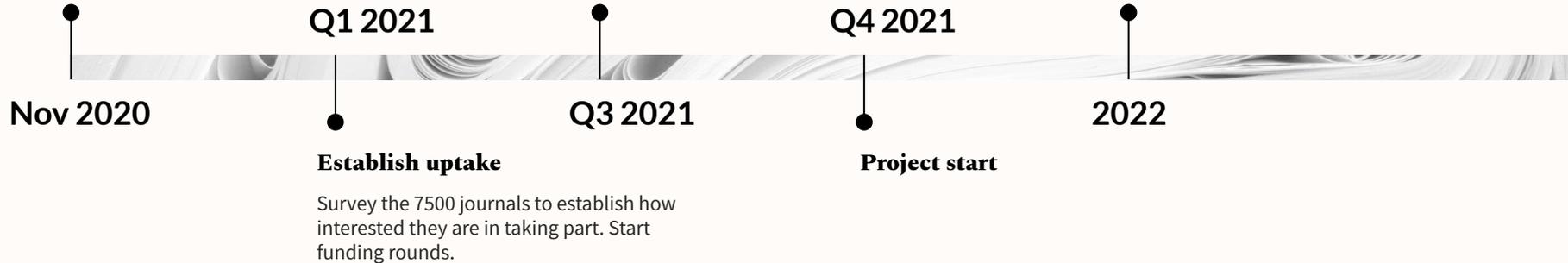
Planning, scoping. Kick-off. Seek funding. Attract publicity and support from the scholarly publishing communities.

Project Pilot

A selected group of journals covering each group will be taken through the process. Experiences from the pilot will help form clearcut processes for the rest of the project

Project review

The bulk of journals that indicated interest are in the process. No finite end date as experience shows that uptake can be slow and take time. Project review to ascertain whether a Phase Two is viable or not.



Thank you.

For more information, contact:

CLOCKSS Craig Van Dyck, cvandyck@clockss.org

DOAJ Dom Mitchell, dom@doaj.org

Internet Archive Jefferson Bailey, jefferson@archive.org

Keepers Registry Gaëlle Béquet, gaelle.bequet@issn.org

PKP James MacGregor, jbm9@sfu.ca