

DRAFT OUTLINE: PKP Private LOCKSS Network

The body of this document was originally shared with the PLN community (via the PLN listserv) on June 3/14 as a google document. Modifications to the document, and the opportunity to leave comments on it, closed on August 6/14. Definitions and other details related to the operations of the PKP PLN will continue to evolve.

Background

The Global LOCKSS Network (GLN) preserves content from over 10,000 electronic journals from a wide variety of commercial and learned society publishers. However, fewer than 120 journals published using OJS are included in this list. Economies of scale in dealing with publishers is the main reason for this lack of OJS content in the GLN, which is administered centrally by LOCKSS staff at Stanford. In addition, the process by which titles enter the GLN is driven by demand from within the institutional members of the LOCKSS Alliance, libraries that within typically prioritize titles for which they pay large sums of money. OA journals are the minority in the GLN, a condition which marginalizes OJS titles even more.

A number of Private LOCKSS Networks (PLNs) exist that use the same technology as the GLN but are governed by library consortia or other groups of organizations. Examples include MetaArchive, the COPPUL (Council of Prairie and Pacific University Libraries) PLN, and the ADPNet (Alabama Digital Preservation Network) PLN. Some of these PLNs, most notably COPPUL's, harvest and preserve small numbers of OJS journals, but these titles are typically limited to those hosted by libraries that are also part of the PLN.

Given the absence of OJS journals in the Global LOCKSS Network, and the local emphasis of the OJS journals that are being preserved in existing PLNs such as COPPUL's, a Private LOCKSS Network that will host any OJS journal, regardless of where it is hosted or who its publisher is, would serve an important role. Such a PLN would also be a welcome addition to the LOCKSS community ecosystem. Vicky Reich, Executive Director of the LOCKSS Program, has been advocating for individual communities to take more direct control of their digital preservation activities, and is aware of how OJS titles have been disadvantaged by the Global LOCKSS Network (and is also aware of how many journal managers are interested in having their content preserved by LOCKSS). PKP has discussed the proposed PLN for OJS journals with LOCKSS staff and has secured their endorsement.

Overview

The proposed PKP PLN will:

- provide preservation services for any OJS journal that meets minimum requirements. These are yet to be determined but may include having published one year's worth of content and having an ISSN

- allow OJS journal managers to register their journals for inclusion in the PLN by simply enabling a plugin in their OJS instance and agreeing to a small number of terms in the Preservation Agreement
- on registration of a journal, automatically determine if the registered journal meets the minimum inclusion requirements
- preserve the raw journal content as generated by OJS's existing export tools
- automatically harvest new content from registered journals and add the content to the PLN, eliminating manual staff intervention at PLN member sites
- provide access to preserved content after a “trigger event” has been communicated to PKP by the journal manager

The proposed PKP PLN loosely follows the model established by CLOCKSS, in which publishers' content is harvested into a Private LOCKSS Network and exposed for public access after a “trigger event,” such as the journal publisher going out of business, the publisher dropping a journal, the publisher ending access to the back issues of a journal, or catastrophic and permanent failure of the publisher's technical infrastructure. The nature of trigger events within the PKP PLN is yet to be determined but may be limited to direct contact by the participating journal manager with PKP staff. Alternatively, automated processes may be established to “monitor” registered journals to determine that access to them has ceased.

Addition of journals to the PKP PLN

As stated above, a request to be included in the PKP PLN will be initiated simply by journal managers activating a plugin with their OJS instances. After the minimum requirements for inclusion have been met by the journal, its content will be automatically harvested into the PLN with no further intervention by the journal manager or authors.

One important consequence of allowing journal managers to initiate inclusion in the PKP PLN, and of the automated (instead of manually curated) nature of harvesting journal content, is that the tools developed to allow these processes must protect the PLN from inadvertently preserving malicious, illegal, or otherwise harmful content. For example, a journal manager may intentionally publish material used without permission of the copyright holder. Another potential threat is that someone spoofing a “journal manager” makes very large files available for preservation with the intent of using all of the PLN's disk storage capacity. These situations will likely be rare, but the PLN must take precautions to reduce the risk associated with automating the harvesting of content for preservation.

Preservation of content

Content from participating journals is preserved in the native OJS export/import format. Each issue is exported and stored as a zipped Bag. Also included in the Bag is information about the journal (journal title, ISSN, etc.) and copyright and licensing information about each article in the issue. As in other Private LOCKSS Networks, the content is automatically replicated at each of the PLN member sites and is constantly monitored to ensure it is intact and secure.

End-user access to journals in the PKP PLN

The PKP PLN is a dark archive. End users will not have access to the preserved content until after a trigger event. After a trigger event, PKP staff will approve the importing of the preserved content into one or more OJS instances hosted by PKP member institutions. Once loaded into these host OJS instances, the content will be available to the public under a suitable Open Access license.

Which license depends on the copyright and licensing policies an OJS journal had in effect at the time the content was added to the PLN. The OJS plugin that packages content for preservation in the PLN will detect existing copyright and licensing configuration options and ask the journal manager to confirm them. Upon confirmation, the journal's content is harvested into the PLN. Copyright and licensing information that applies to the content is preserved with it in the PLN, so that it will be easily available at the time of a trigger event. The four general categories of copyright and licensing options that apply to the dissemination of the OJS content after a trigger event are:

- the journal's content is available under a Creative Commons license
- the journal's content is available under a non-CC Open Access license
- the journal or publisher owns the copyright to the content and will be required to assign PKP the right to make it available under a CC license after a trigger event
- any other copyright/licensing arrangement (e.g., the journal is a toll journal and the publisher is not in a legal position to allow the content to be released under a CC license)

In all but the last of these situations, the journal content will be made available to the public after a trigger event by PKP and its partners under the same terms that applied before the trigger event. In the fourth situation, PKP will need to decide what the most appropriate means are to make the content available, including negotiating with the journal managers if they are available at the time of the trigger event. In practice, the vast majority of journals using currently OJS fall into the first three types.

Organizational infrastructure of the PLN

A Private LOCKSS Network requires the participation of a minimum of seven nodes (LOCKSS servers), ideally distributed over a wide geographic area, to ensure sufficient redundancy. For the proposed LOCKSS PLN, PKP is seeking seven LOCKSS Alliance members who would be willing to participate as one of the seven hosting sites and also be comfortable with extending this support to the general OJS community. Many of the participants in the latter category are not likely to be LOCKSS Alliance members nor able to provide the technical and server support. PKP does not envision this will impose an undue load on the LOCKSS Alliance sites, especially with the addition of the automated support for adding content.

PKP PLN members

PKP members who have confirmed participation in the proposed PLN are:

- University of Alberta
- University of British Columbia
- Indiana University
- University of Pittsburgh
- University of Victoria, and
- Simon Fraser University

and one other site is giving active consideration:

- University of Illinois

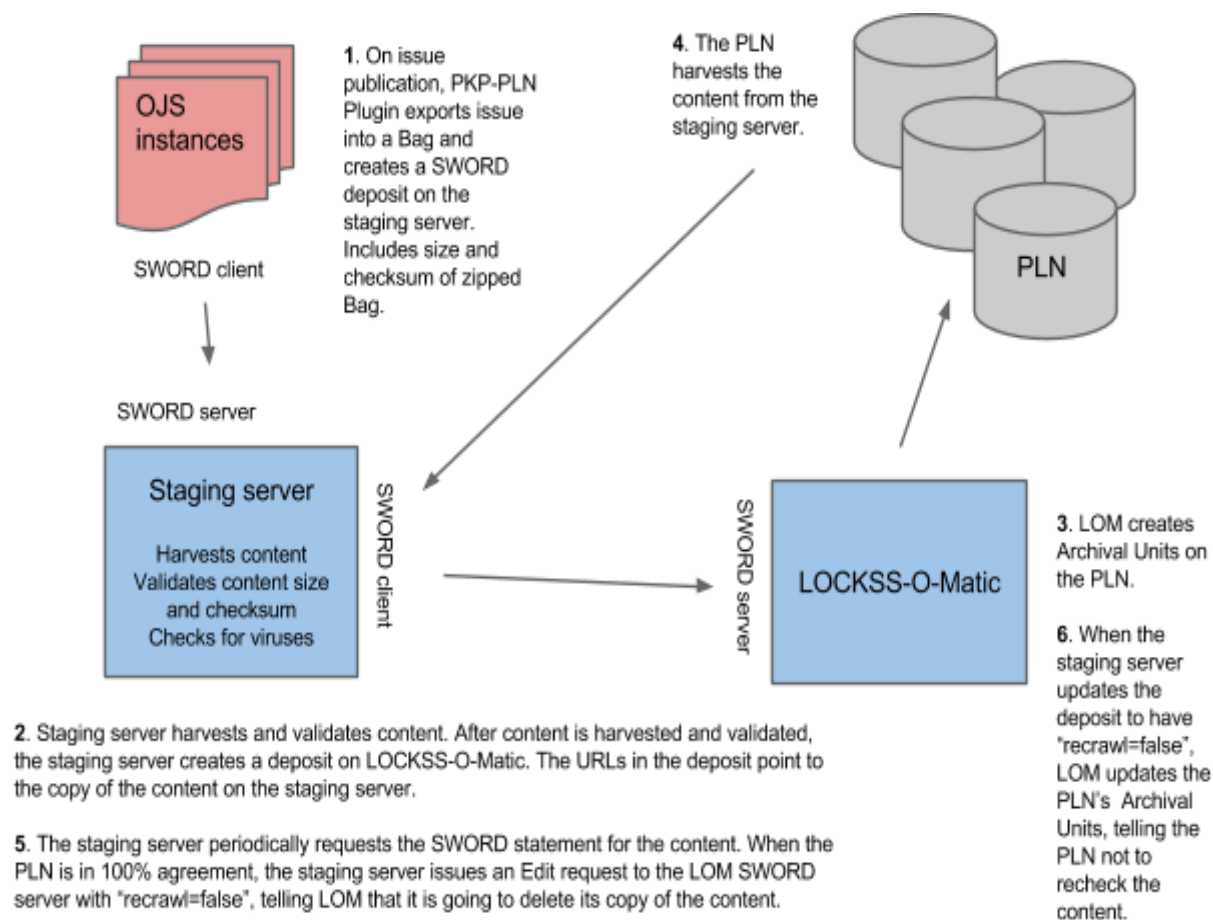
In addition, PKP is working with the SAFE PLN in Europe (<http://www.safepln.org/>) led by the Université libre de Bruxelles (ULB) to host a LOCKSS server there, and also with the Brazilian national PLN managed by the Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), to provide additional redundancy and geographical distribution within the PKP PLN.

Preservation agreement

Journal managers who want their journal included in the PKP PLN will need to accept the terms of the Preservation Agreement, via the OJS plugin. This agreement will describe what is being preserved, the nature of the PKP's commitment to preserve their journal, PKP's right to preserve the journal, the type of license the content will be made available under after a trigger event (as described above), and how they should contact the PKP to initiate a trigger event.

Technical infrastructure

The technical infrastructure necessary for such a network is modest. It is comprised of 1) a new OJS plugin, 2) a set of OJS-specific tools to enable automated registration of journals, testing minimum requirements for inclusion in the PLN, and exposing the journal content for harvesting, 3) LOCKSS-O-Matic, software to automate the harvesting of content into the PLN, and 4) the LOCKSS network itself. The following diagram illustrates how these pieces work together:



OJS plugin

The existing OJS plugin that generates a LOCKSS manifest and permission statement will not work within the proposed PKP PLN, so a separate OJS plugin will be included with OJS 2.4.5 (to be released in July 2014). Journal managers will opt into the PKP PLN by activating this plugin. As described above, the plugin also creates Bags containing OJS issues, and communicates with the staging server via the SWORD protocol to deposit the Bags into the LOCKSS network.

Staging server and validation tools

A small set of tools and services will test registered journals for minimum inclusion requirements, validate content on the staging server, and deposit content into the PLN. The PKP development team is currently undertaking the creation of these tools, which will be ready for testing when version 2.4.5 of OJS is released. The GitHub repository for this work is <https://github.com/mjordan/pkppln>.

LOCKSS-O-Matic

Simon Fraser University Library is developing a tool, LOCKSS-O-Matic, to automate the harvesting of content into a PLN. Most PLNs are governed by their communities but for the most part, technical administration is handled by LOCKSS staff at Stanford. This tool will eliminate this reliance on LOCKSS support staff and allow for a very high degree of automating the process of adding content to PLNs.

LOCKSS-O-Matic is not specific to the proposed PKP PLN; it will be able to automate the ingestion of a variety of content into a Private LOCKSS Network. The development roadmap for LOCKSS-O-matic is available at <https://github.com/mjordan/lockss-o-matic/wiki/Roadmap>.

LOCKSS Private Network

The servers that run the LOCKSS instances making up the PLN will be standard LOCKSS boxes with 2 TB of accessible storage.

Development timeline

The OJS plugin to enable participation in the PKP PLN will be released with OJS 2.4.5, probably in July 2014. All other development and organizational work will also be complete by that time.