Perm.pub pilot project

2022-08-09

E. Castedo Ellerman ⓘ (castedo@castedo.com)

Abstract

This document describes a pilot project hosted at https://perm.pub/ and its motivation. This pilot project involves two technologies of interest: one of them JATS, and the other a new technology leveraging the Software Heritage archive. The formation of working groups may be appropriate to share opinions and suggestions. Feedback is encouraged.

Objectives

The perm.pub pilot project aims to demonstrate how the new technology of Digital Succession Identifiers (DSI) can enable an alternative to current preprint servers with the following combined improvements:

1. the benefits of articles in modern web page format,

2. freedom for readers to choose different websites to access those documents, and

3. facilities for readers to ascertain the distinct versions of a document (manuscript, preprint, article) as they have become available at distinct points in time.

Notable benefits of a modern web page format are:

1. improved discovery of documents via Internet search engines,

2. an improved reader experience for readers using popular electronic devices of the 21st century, such as computers and mobile phones, rather than physical paper, and

3. the opportunity for innovative web site experiences that encourage researcher communication and collaboration, such as Manubot [1].

Digital Succession Identifiers

A digital succession contains multiple digital objects. In this application to perm.pub, the digital objects are directories with JATS XML files. Although a digital succession expands over time, each digital object within the succession does not change. A technical specification of Digital Succession Identifiers (DSIs) can be found in the Digital Succession Identifier Specification.

The capabilities of JATS XML, DSIs and underlying technologies discussed in the Digital Succession Identifier Specification, enable a trisection of the role of a preprint server.

Preprint servers trisected

The Software Heritage archive [2] opens dramatic new possibilities in the preservation of written outputs by researchers. A possibility the perm.pub pilot project aims to demonstrate is a decentralized trisection of a preprint server into three separate entities:

1. the archiver (e.g. Software Heritage),

2. the eprinter, and

3. the locator.

Today, the decentralized archiver is Software Heritage. Due to the use of intrinsic identifiers [3], this archiver role can be performed by multiple independent parties.

Eprinter role

The eprinter is a website which generates webpages and potentially alternative PDFs based on JATS XML stored in an archive. It is mostly up to the eprinter, and the community it serves, to decide which documents are eprinted and how they are presented. The lifetime of an eprinter is potentially short. By rendering webpages and implementing novel technological enhancements for a certain audience, an eprinter might undermine it's ability to sustainable exist long-term. This is not of great concern to the extent that the research community does not depend on an eprinter for long-term preservation.

Locator role

https://perm.pub/ aims to demonstrate a locator which serves a similar role to a DOI registrar or the ID system managed by a preprint server. The mandatory minimal long-term mission of the perm.pub locator is to serve static pages which identify which JATS XML and PDF files in the Software Heritage archive correspond to a given identifier of those documents. An appropriate subset of DSI located by perm.pub can also exist under a DOI registrar namespace.

References

1. Himmelstein DS, Rubinetti V, Slochower DR, et al (2019) Open collaborative writing with Manubot. PLOS Computational Biology 15:e1007128–. https://doi.org/10.1371/journal.pcbi.1007128
2. Cosmo RD, Gruenpeter M, Zacchiroli S (2020) Referencing Source Code Artifacts: A Separate Concern in Software Citation. Computing in Science & Engineering 22:33–43. https://doi.org/10.1109/MCSE.2019.2963148

3.   Di Cosmo R, Gruenpeter M, Zacchiroli S (2022) 204.4 Identifiers for Digital Objects: The case of software source code preservation. https://doi.org/10.17605/OSF.IO/KDE56