

# A Model for Integrating the Publication and Preservation of Journal Articles

Kevin S. HAWKINS  
University of North Texas (USA)

@KevinSHawkins

# The deficit in journal preservation

While libraries once acquired and preserved paper copies of journal issues, today journals are usually hosted online by the publisher. How can libraries ensure preservation?

- Legal deposit requirement in national copyright law
- Acquire the right to keep a copy of issues from the period of subscription even after canceling a subscription
- LOCKSS and CLOCKSS: web crawling
- KB/Elsevier agreement and Portico: publisher delivers all content to the archive

What if libraries instead operated the very system used to publish journals?

# Library-based publishing

In North America, 43 out of 144 academic libraries responding to a survey in 2010 were offering publishing services to their university, usually on one of the following platforms:

- Open Journal Systems (OJS): an open-source all-in-one solution
- DSpace: open-source software for institutional repositories
- Digital Commons: a hosted platform from bepress

A Library Publishing Coalition survey in 2013 found that OJS and Digital Commons were the most commonly used platforms.

*With these platforms libraries cannot control software used to create files (usually PDFs) and therefore can only do bitwise preservation.*

# Beyond bitstream preservation

In institutional repositories and OJS:

- Bitstream preservation leaves content at risk of format obsolescence.
- Data can be migrated to newer formats, but what if files are not well-formed in the first place?
- Even if you have valid files in a preservation-quality format (e.g., PDF/A), you can't reformat for a non-paginated display or easily use with a screen reader or for data mining.

# HathiTrust

- a partnership of research libraries around the world
- a shared digital repository certified to be preservation-quality
- over 11 million digitized volumes (501 terabytes of data)



HATHI  
TRUST

# Why use HathiTrust?

- HathiTrust already archives and provides access to digitized library collections, so this shared digital repository is a natural place to archive and provide access to journal literature to ensure its long-term preservation and discoverability.
- We can build a system for publishing journals in which archiving happens as a byproduct of publication rather than after the fact.

# Preservation-quality formats for born-digital journal articles in HathiTrust

- **JATS** (formerly the NLM DTDs) is an application of ANSI/NISO Z39.96–2012 for encoding journal articles in XML.
- **TIFF or JPEG 2000** for images, checked for validity before publication. (Only bitwise preservation for multimedia objects and supplementary material at this time.)
- **METS** for structural and preservation metadata.

*The University of Michigan Library is building a system for publishing journals and archiving them in HathiTrust using these formats.*

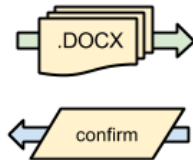
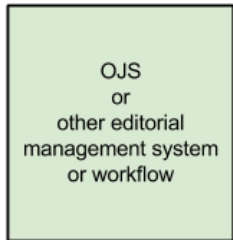


# mPach: what are we creating?

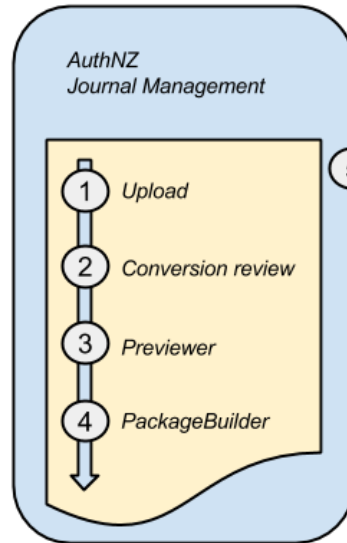
- modular platform
- tightly coupled with the HathiTrust repository
- for open-access journals
- all you need to publish and preserve an OA journal
- will integrate with Open Journal Systems (OJS)

# mPach

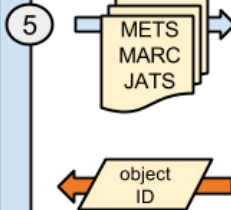
## Peer Review and Editorial Workflow



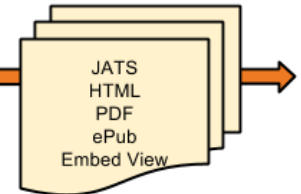
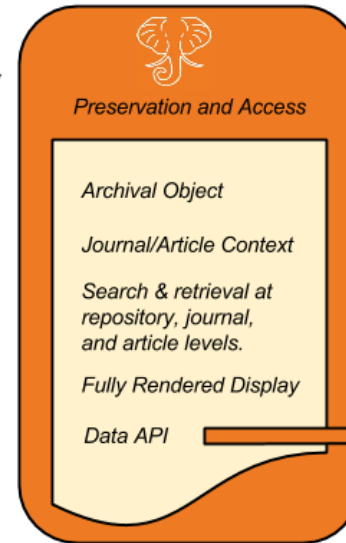
## mPach Prepper



Submitter operated by  
Michigan Publishing



## HathiTrust



# When can I use it?

- Will be open source, but ...
  - We've still got quite a bit of development to do, so it's not ready to use yet!
  - Once U-M Library works out policy for sharing sourcecode, we'll do so in case others want to contribute even while we're still actively developing.
  - Requirement to authenticate ingest mean that Michigan Publishing will need to mediate ingest by hosting the Submitter component.
- We plan to offer it as a hosted service (shared infrastructure for scholarly publishing!) rather than just providing open-source software.
- We need to determine how membership in HathiTrust will affect your rights to deposit content.

[www.lib.umich.edu/mpach](http://www.lib.umich.edu/mpach)

[kevin.s.hawkins@ultraslavonic.info](mailto:kevin.s.hawkins@ultraslavonic.info)  
[@KevinSHawkins](#)