



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

Kevin S. HAWKINS

University of Michigan, Ann Arbor  
(USA)

# Публикация и хранение журнальных статей. Модель объединения

Кевин С. Хокинс

Мичиганский университет  
г. Энн-Арбор  
(США)

@KevinSHawkins



# The deficit in journal preservation

Publishers increasingly host journals online instead of selling copies to libraries. 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

*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, OJS, and most other systems currently in use:

- 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 10 million digitized volumes (nearly 500 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

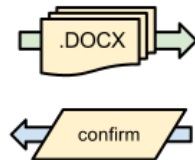
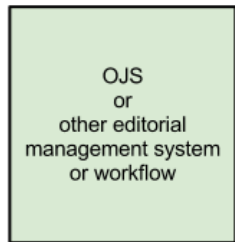
- **JATS** (formerly the NLM DTDs) is an application of ANSI/NISO Z39.96–2012 for encoding journal articles in XML.
- **TIFF or JPEG 2000** for images, checked for validity before publication. (Only bitwise preservation for multimedia objects and supplementary material at this time.)
- **METS** for structural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

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

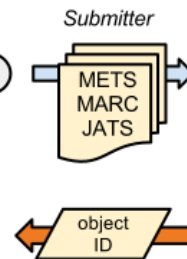
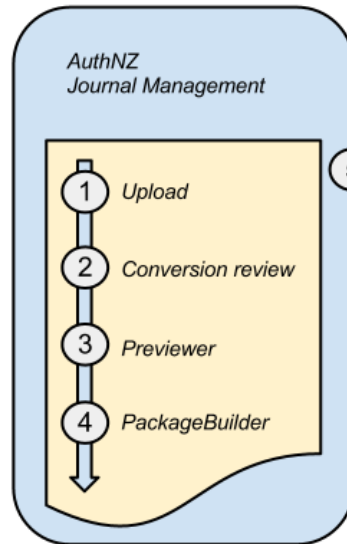


# mPach

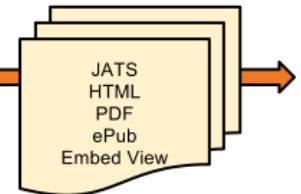
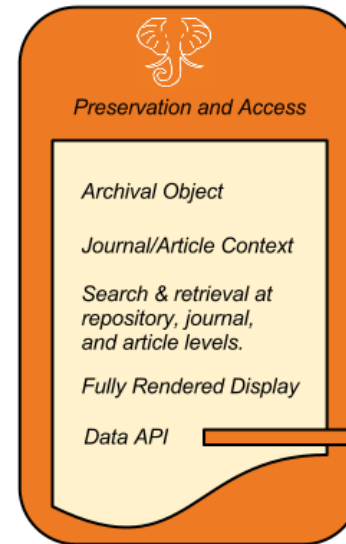
## Peer Review and Editorial Workflow



## mPach Prepper



## HathiTrust





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

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