HathiTrust, MPublishing, and mPach

Digital Libraries and Digital Publishing at the University of Michigan

Kevin S. Hawkins

Slides are available at http://www.ultraslavonic.info/talks/20120619.pdf

Brief history of digital libraries at U-M

- Early 1990s: Began making full text of high-demand public-domain works like KJV, Quran, and Book of Mormon available through Gopher.
- 1996: Began hosting digital content for commercial publishers.
- 1999: Text Creation Partnership formed.
- 2004: Partnership with Google announced.

Brief history of publishing at U-M

- 1930: University of Michigan Press founded.
- 1996–2009: Various projects where Library would publish Press content online.
- Library increasingly intervening in scholarly communication
  - 2001: Scholarly Publishing Office created to provide sustainable electronic publishing services, support local control of assets, and explore opportunities to extend and disseminate library collections.
  - 2006: Deep Blue (institutional repository) launched.
  - 2006: Copyright Office formed.

A shared digital repository, a digital library, and a partnership of institutions

HATHITRUST

In light of the Google partnership, U-M Library wanted to …

- preserve and make use of its digital copies of works scanned by Google to the fullest extent of the law

- take this opportunity to build a new digital library architecture (for Google and non-Google content) rather than use its existing infrastructure, with its various flaws.

At first we built MBooks, but it became …

HathiTrust

- HathiTrust was formed in 2008 as a partnership of 13 institutions. Now has more than 60 members.

- It has developed an innovative cost model by which institutions share costs for sustaining the repository. Certain features available only to users at member institutions.

- Member institutions can contribute staff to development of infrastructure and to ongoing projects (like the Orphan Works Project).

- Held a "constitutional convention" in 2011 to establish a governance structure and set development priorities.

Stats and main features

- HathiTrust contains over 10 million digitized volumes, over 3 million of which are in the public domain in the US.

- Users can search catalog and full text of whole corpus but only see page images for works in the public domain in the user's country. Access also available through OCLC WorldCat and Serials Solutions' Summon.

- Users can build their own collections of digitized volumes for searching and browsing.

Taking advantage of copyright law

- US copyright law allows for reproductions of works protected by copyright for visually impaired users and for replacement of damaged originals by libraries. HathiTrust facilitates both.
- Copyright Review Management System (funded by IMLS) allows HathiTrust to make available full text of works that have risen into the public domain
  - e.g., certain US works published between 1923 and 1963.

Grant of access by rightsholders

Rightsholders can grant a license to HathiTrust to make a work available through open access.

Advanced uses of the data

- Bibliographic API and datasets for access to metadata

- Data API for access to content as allowed by copyright law

- HathiTrust Research Center for text mining of public domain and eventually works protected by copyright ("non-consumptive research")

A brand for services, a publisher, and a partner for scholars

MPUBLISHING

MPublishing

MPublishing is the primary academic publishing enterprise of the University of Michigan and part of the University Library.

Formed in 2010 as an umbrella for the combined publishing-related activities of the University of Michigan Library and Press.

You may ask

WHAT KINDS OF ACTIVITIES?

Monograph publishing

OPEN HUMANITIES PRESS *

EBOOK *

*print on demand of select titles

Print-on-demand programs

- Michigan Historical Reprint Series
- HathiTrust Reprints
- University of Michigan Faculty Reprints

Journal publishing: 20 active journals including …

The Bulletin of the American Society of Papyrologists

Fragments

Interdisciplinary Approaches to the Study of Ancient and Medieval Pasts

Institutional repository

Copyright Office

In addition to providing advice to campus community, staff are engaged in:

- Copyright Review Management System: identifying works that are not protected by copyright due to failure to register or renew

- Orphan Works Project: identifying works whose copyright holders cannot be identified or contacted

Text Creation Partnership

A project and a set of tools

MPACH

Here's the problem

Most of MPublishing's content is published in a system which:

- only MPublishing is actively developing

- is not modular

- was designed for digitized library content, not publishing of born-digital literature
  - can't share access with editors
  - items can't relate to each other

And here's the opportunity

- There's strong demand from editors and institutions for supporting open-access journal publication.
- Libraries aren't entirely happy with the tools available:
  - Open Journal Systems (OJS) has a difficult interface, an inflexible workflow, and no features to ensure preservation of content.
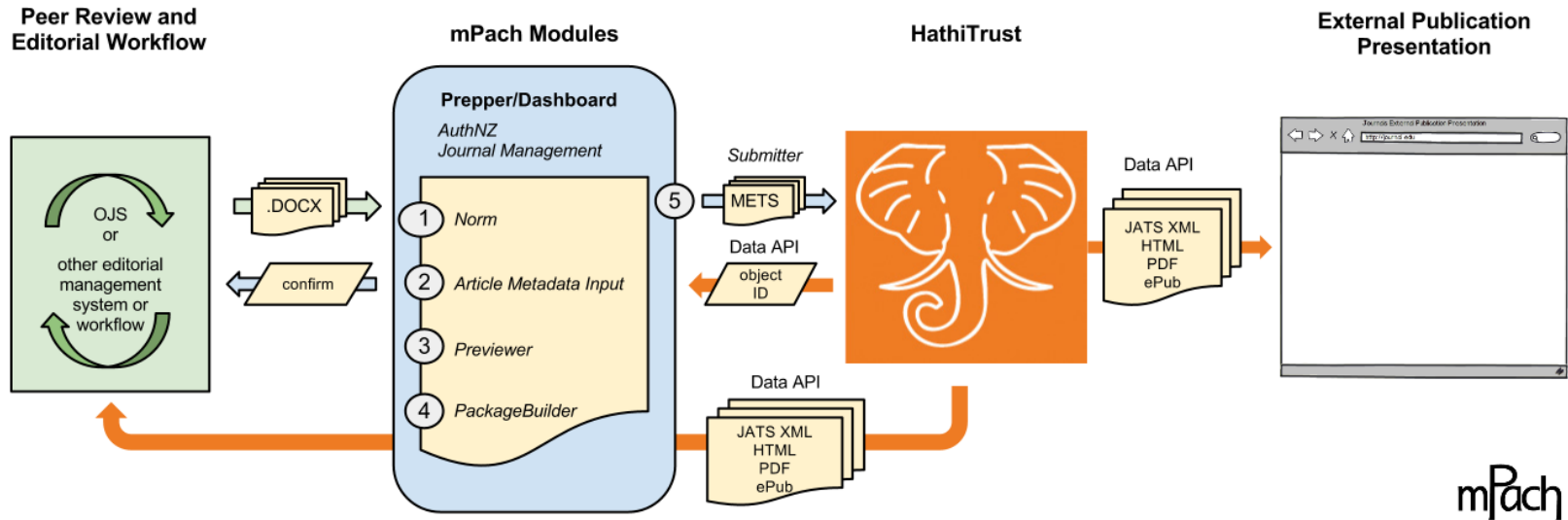 - Hosted solutions are expensive and lack preservation-quality content standards.

The University of Michigan Library is funding …

- development of a modular platform for publishing journals into the HathiTrust repository

- tools to meet the needs of MPublishing

- open-source code available for use by HathiTrust member institutions

More on modularity

- We want to build a system that lets us integrate available tools (instead of reinventing the wheel) and swap out components as needed.

- The default option will be to use of OJS as a submission and editorial management system and as the primary interface for readers.

mPach system architecture

Questions?

Slides are available at http://www.ultraslavonic.info/talks/20120619.pdf