



Update On October Activities

November 8, 2013

Top News

Announcing Zephir

HathiTrust released a new bibliographic management system, Zephir, developed by the California Digital Library. Zephir is custom-made to support the particular needs of bibliographic management in HathiTrust. A full announcement is available at http://www.hathitrust.org/zephir_announcement. See <http://www.hathitrust.org/zephir> for background on the project and system documentation. From this time, institutions submitting bibliographic metadata to HathiTrust need only submit metadata to Zephir (see http://www.hathitrust.org/bib_data_submission for details).

HathiTrust and DPN

HathiTrust announced its intention to become a “replicating node” in the [Digital Preservation Network](#) (DPN). The formal announcement can be read at http://www.hathitrust.org/hathitrust_dpn_announcement.

Call for US Government Documents Records

HathiTrust is issuing a broad call for bibliographic records for US federal government publications from HathiTrust partner and non-partner institutions alike, in support of its initiative to expand and enhance access to US federal government documents. Further information about the initiative and details about the call for records are available at <http://www.hathitrust.org/usgovdocs>. Records requirements, an information sheet to accompany bibliographic record submissions, and instructions on submission, are available at <http://tinyurl.com/kyw26fo>.

Board of Governors

The HathiTrust Board of Governors held an in-person meeting in Arlington, VA on October 11. The Board discussed a full agenda, including the [ballot initiatives passed at the 2011 constitutional convention](#), proposals to expand the availability of open access materials in HathiTrust and to broaden access to users who have print disabilities, the HathiTrust Research Center, the Executive Director search, and member issues arising from the bylaws, including an approval process for new members and an annual meeting. Actions on these items will be reported on in this and upcoming monthly updates.

Executive Director Search

The Executive Director search committee worked to narrow a rich pool of candidates and will be conducting phone interviews with selected individuals in November.

November Forecast

Continue to work on support for indexing of JATS articles.

Continue development to generate ePub and PDF from JATS XML.

Continue to explore relevance ranking solutions.

Papers & Presentations

Kevin S. Hawkins, “A Model for Integrating the Publishing and Preservation of Journal Articles ([slides](#)) ([paper](#))”, RCDL, October 15, 2013.

Seth Johnson, Bryan Smith, Kevin S. Hawkins, “[mPach: Integrated Publishing and Archiving of Journals in HathiTrust](#)”, Impromptu JATS Users Group Meeting, October 23, 2013

There's an
elephant in
the library.™





Update On October Activities

Ingest

Validation service for locally-digitized materials

Completion of a web-based service to validate single image files and a cloud storage-based service to validate entire volumes was delayed in October. These services are now planned for release in November, and December or January, respectively. The creation of these services is the result of conversations held with partners who plan to deposit locally-digitized content in HathiTrust, and a general survey about the usefulness of these services in facilitating local packaging of materials prior to submission to HathiTrust.

General

HathiTrust answered questions about content ingest from the University of Delaware and Vanderbilt University, reviewed sample content from Texas A&M university, and prepared for ingest of files from the University Press of Florida.

Working Groups and Committees

Program Steering Committee

The Program Steering Committee is developing charges for a Government Documents Initiative Planning and Advisory Group and a Rights and Access Working Group, and expects to recruit members for these groups in the coming weeks. Work continues on charges for groups to advance the establishment of a distributed print archive on monographic holdings corresponding to the digital content in HathiTrust and to continue the work of the Collections Steering Committee setting priorities for expanding collections. The Committee is also reviewing a proposal for a distributed program to certify the quality of volumes within HathiTrust.

Projects

Government Documents Registry

The project team reviewed feedback received during focus groups on the government documents registry conducted in late September/early October, and began to develop functional requirements for the registry. The team also began identifying potential strategies for detection of duplicate records. A project timeline has been added to the project [web page](#).

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in October is given below. See [CRMS-US](#) and [CRMS-World](#) for further information.

You can follow HathiTrust on [Twitter](#) or [Facebook](#)

[Subscribe to email updates](#) (via [Google Groups](#))

There's an
elephant in
the library.™





Update On October Activities

	October		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	3,849	7,825	152,266	291,275
CRMS-World	2,774	5,084	39,889	74,544
Total	6,623	12,909	192,155	365,819

HathiTrust Research Center

Members of the HTRC Executive Team attended an in-person meeting of the HathiTrust Board of Governors and will be crafting a business plan for HTRC operations going forward for the Board to review. The HTRC is also making plans to expand the texts included in its research environment to all works in HathiTrust, including those that are in-copyright.

The HTRC held its second monthly HTRC Usergroup meeting, on educational materials related to the HTRC. Notes from the meeting are available on the [HTRC wiki](#). These meetings are open to all who are interested (see [this link](#) to sign up; you can also sign up to participate in the [HTRC wiki](#)). More information about the HTRC, including directions to sign up for a general HTRC announcements list and announcements related to the HTRC UnCamp are available at <http://www.hathitrust.org/htrc>.

mPach

University of Michigan staff presented on mPach at [RCDL'2013](#), at the [Impromptu JATS Users Group Meeting](#), and locally at the University of Michigan.

Development Updates

HathiTrust institutions performed the following work related to applications and Web interfaces:

Collection Builder

Staff corrected a problem that resulted in long collection names being truncated.

Data API

Version 1 of the Data API was taken out of service on November 1, 2013. [Version 2](#) is the current version.

Development Environment

Staff continued working on web server upgrades for the HathiTrust development environment, and selected developers began to test the servers.





Update On October Activities

Full-text Search

Staff continued work on issues related to relevance ranking and began testing of new algorithms to measure document homogeneity. Staff also corrected an indexing issue that resulted in some Full view works being represented as Limited (search-only) in the online catalog.

Outages

HathiTrust users may have experienced slow page loading or errors in page loading on Friday, October 18 from 12:30-9:45pm due to a software release earlier in the day that left page viewing at the Indiana site in a non-working state that was subtle enough to be undetectable to monitoring systems.

Total Volumes Added	October	Overall
Boston College	0	2,363
Columbia University	0	65,035
Cornell University	3,268	433,868
Duke University	0	4,524
Harvard University	1,358	237,430
Indiana University	71	195,420
Library of Congress	0	89,724
North Carolina State University	0	3,196
Northwestern University	100	37,288
New York Public Library	3	288,367
Penn State	526	65,312
Princeton University	0	251,709
Purdue University	3	44,695
Universidad Complutense	3	112,001
University of California	16,125	3,435,459
University of Chicago	1,842	35,387
University of Florida	0	9,587
University of Illinois	275	112,426
University of Michigan	5,551	4,662,752
University of Minnesota	1,831	112,169
UNC - Chapel Hill	0	17,025
University of Wisconsin	7	555,878
University of Virginia	4	50,821
Utah State University	0	117
Yale University	0	23,678
Total	30,967	10,846,231
Public Domain (~32% of total)		
Total*	31,549	3,494,313

*Includes works opened via copyright review and rights holder permissions.

There's an elephant in the library.™



Update On October Activities

User Support Issues

	October	September
Content	249	243
Quality	242	225
Collections	7	17
Cataloging	189	169
Access and Use	164	107
Copyright	90	57
Permissions	8	5
Takedown	2	0
Print on Demand	0	0
Inter-library loan	0	2
Full-PDF or e-copy requests	12	17
Datasets	4	3
Data Availability and APIs	2	1
Reuse of content	3	4
Web applications	36	22
Functionality problems	11	9
Problems with login specifically	1	3
General questions about login	1	3
Partners setting up login	0	4
Usability issues	0	0
Feature requests	2	2
Partner Ingest	3	9
General	105	90
Partnership	6	8
Infrastructure	0	0
Miscellaneous	99	82
Total	746	640

*See [User Support Working Group Issue Types](#) for a description of the types of issues included in each category.

Most-accessed volumes

Title
Primitive culture: researches into the development of mythology, philosophy, religion, art, and custom, Vol. 1, by Edward Tylor.
The coquette; or, The history of Eliza Wharton, by Hannah Webster Foster.
Godey's Magazine, v.40-41 1850.
Ammianus Marcellinus; with an English translation by John C. Rolfe
Quicksand, by Nella Larsen.
Building a nation and where to build ideal American homes, by Jere Johnson Jr.
Education in East Africa; a study of East, Central and South Africa by the Second African education commission
Fugitives; the story of Clyde Barrow and Bonnie Parker, by Emma Parker
Godey's magazine. v.66-67 1863.
Coffee processing technology, Vol. 1, by Michael Sivetz and H. Elliott Foote.

There's an elephant in the library.™

