



Update On August Activities

September 14, 2012

Top News

HathiTrust Training and Information Sessions Survey

In the [Update on July Activities](#) we distributed a short survey to receive feedback on our next series of HathiTrust information and training sessions. We have received many responses. The deadline for completing the survey is September 21. If you have not already, please take a moment to provide input on the kinds of sessions you would like to attend or lead, and the form you would prefer these sessions to take (e.g., a webinar series, in-person meeting, or a combination of the two). The survey is available at <http://tinyurl.com/8n3k9nr>.

Data API Changes In Effect October 1

Beginning October 1, all requests to the Data API will need to be signed with an access key provided by HathiTrust. Access keys for programmatic uses of the Data API can be obtained at <http://babel.hathitrust.org/cgi/kgs/request>. HathiTrust has also created a [Web client](#) that employs a user's login credentials as a proxy for an access key to facilitate non-programmatic uses. Complete documentation of the security enhancements, methods of obtaining keys, how to sign requests, and how to access the Web client is available at http://www.hathitrust.org/data_api.

Also effective October 1, the host "services.hathitrust.org" will no longer exist for the Data API. The new host will be "babel.hathitrust.org", the same host as the PageTurner and other HathiTrust services. Calls to the Data API will therefore need to use URLs such as the following (note the additional "cgi" in the path):

<http://babel.hathitrust.org/cgi/htd/meta/mdp.39015019203879>

rather than

<http://services.hathitrust.org/htd/meta/mdp.39015019203879>

Shibboleth "library walk-in" Attribute

Later this year, HathiTrust will begin accepting the "library-walk-in" Shibboleth attribute from partner institutions to provide certain member privileges to guest users who do not have an institutional login. For instance, "Library-walk-in" users will have the ability to download full-PDFs of all public domain materials in HathiTrust. Partners who wish to use HathiTrust library-walk-in functionality must confirm in writing that they are asserting the library-walk-in affiliation only for users physically present in a library building at the time of session initiation. Please see [Shibboleth Login](#) for more information about Shibboleth in HathiTrust.

Ingest

Internet Archive Digitization

HathiTrust ingested nearly all of a set of approximately 2,000 volumes from Boston College, and loaded bibliographic records for additional volumes that will be

September Forecast

Work on spelling suggestion feature for full-text search.

Continue work on full-text search, including relevance ranking and support for CJK languages.

There's an
elephant in
the library.™





Update On August Activities

deposited by the University of Illinois. The University of Florida submitted sample bibliographic records to be analyzed in preparation for content ingest.

Working Groups and Committees

Working groups and committees in HathiTrust may have an operational or strategic focus. See http://www.hathitrust.org/working_groups for more information.

Operational

Communications Working Group

The Communications Working Group did not meet in August, taking its first break since the group's formation in May 2010. As the group awaits the solidification of the new HathiTrust governance, group members plan to address the results of their [survey on training](#), and look ahead to fall activities and meetings.

User Experience Advisory Group

The User Experience Advisory Group continued discussions about a new home page design and provided feedback on mockups created by the University of Michigan.

User Support Working Group

A summary of the issues received by the User Support Working Group is provided at the end of the update.

Projects

Bibliographic Data Management

California Digital Library (CDL) and University of Michigan staff agreed on a data workflow for updating rights information in the HathiTrust rights database when CDL takes responsibility for managing HathiTrust bibliographic data. The CDL team is refining and improving the performance of bibliographic data exports needed to support HathiTrust operations. Analysis continued to address issues with a small percentage of poor quality records.

Michigan staff successfully tested the bibliographic record submission process for Zephyr (the new management system) and commented on corresponding submission guidelines. In the coming month CDL will be contacting institutions that are currently, or were in the past, contributors of content to HathiTrust to test the new process for submitting records. The test will be aimed primarily at current content contributors, but all contributors will be invited. Please contact feedback@issues.hathitrust.org if your institution is not contributing content currently but you would like to test.

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

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

There's an
elephant in
the library.™





Update On August Activities

Copyright Review

A summary of copyright review activities in August is given below. For further information on these activities please see [CRMS-US](#) and [CRMS-World](#).

	August		Overall	
	Opened	Reviewed	Opened	Reviewed
CRMS-US	5,773	11,793	169,995	320,883
CRMS-World	2,423	5,615	6,592	15,075
Total	8,196	17,408	176,587	335,958

HathiTrust Research Center

The HTRC made preparations for its first “UnCamp”, held in Bloomington, Indiana on September 10-11. A full report on the gathering will be forthcoming.

IMLS Quality Grant

Project staff continued work to finalize the quality review datasets. This included reviewing datasets for completeness, accuracy, and missing data, and performing reliability and validation testing on data for volumes that were double-coded for quality assurance purposes.

The IMLS grant advisory board met for its second time in mid-August. The project team presented its findings to-date and advisory board members provided input on work to be completed in the final stages of the project, as well as on research directions in the future. Over the next several months the project team will focus on completing the design of user studies to further investigate quality in relation to the usefulness of digitized volumes, collecting data to support the user studies, and conducting the user studies themselves.

Efforts continue to develop a framework for certifying the quality of volumes in HathiTrust. This includes the development of a modified data collection Web interface based on the interfaces used in the grant thus far.

For more information on the project, please visit the [project website](#).

mPach

The mPach team at the University of Michigan updated the project timeline on the [HathiTrust project page](#). Work continued on modifications to the HathiTrust PageTurner to display JATS XML, and on refinements to the METS specification for mPach Submission Information Packages. Michigan staff made progress on enhancements to the Norm tool (part of content preparation), specifically enhancements to normalize bulleted lists, figures with captions, and tables. Wireframes are nearly complete for the Dashboard module (see the [list of mPach modules](#) for more information on mPach modules). Michigan staff will be presenting on mPach at the 2012 DLF Forum.





Update On August Activities

Development Updates

Accessibility

Staff at the University of Michigan continued work to improve general accessibility for HathiTrust Web applications.

Data API

Michigan staff extended functionality of the Data API to serve full PDFs of volumes for print-on-demand services on Espresso Book Machines (EBM) via the ExpressNet sales network. Staff also augmented Data API usage monitoring to explicitly track signed requests, and made enhancements that will enable the Data API to deliver dynamically-generated image derivatives (such as PNG images as opposed to TIFF or JP2 images).

First Full Repository Upgrade

Development and testing for the metadata upgrade reported in the [Update on July Activities](#) has been completed, and the upgrade will begin in October.

Full-text Search

Michigan staff continued to investigate the Solr edismax parser bug that is preventing CJK searching from working properly. Staff confirmed that the bug also affects Solr 4.0 and submitted sample documents and queries demonstrating the problem to the Solr JIRA issue tracking system: see <https://issues.apache.org/jira/browse/SOLR-3589>. Staff investigated possible workarounds for this issue, and conducted e-mail discussions with several Blacklight developers who are working on CJK issues.

Staff also made changes to the automated full-text search indexing process so that failures caused by server errors are automatically re-queued.

The INEX (Initiative for the Evaluation of XML Retrieval) Book Track accepted a paper by Michigan developer Tom Burton-West on full-text search relevance ranking in HathiTrust. The paper will be published in the INEX 2012 Pre-proceedings as part of the CLEF Labs Working Notes.

Total Volumes Added	August	Overall
Boston College	1,816	1,816
Columbia University	0	64,184
Cornell University	5,307	408,755
Duke University	0	4,523
Harvard University	1,637	235,983
Indiana University	14	187,683
Library of Congress	1	89,722
North Carolina State University	0	3,196
University of North Carolina - Chapel Hill	0	8,088
Northwestern University	6	7,214
New York Public Library	8	259,571
Penn State University	35	44,018
Princeton University	781	251,644
Purdue University	10,361	38,048
Universidad Complutense	71	111,899
University of California	26,493	3,373,076
University of Chicago	2,240	24,679
University of Illinois	823	101,001
University of Michigan	8,533	4,560,303
University of Minnesota	2,105	102,501
University of Wisconsin	3,559	542,795
University of Virginia	1,868	50,790
Utah State University	0	90
Yale University	0	23,678
Total	65,658	10,495,257
Public Domain (~30% of total)		
Total*	60,100	3,187,744

*Includes volumes opened through copyright review and rights holder permissions.





Update On August Activities

PageTurner

Michigan staff made changes that will make it easier to support new formats in the PageTurner interface. The mPach project will make use of the changes to add support for JATS XML.

Outages

HathiTrust was unavailable on Monday, August 13 from 7:30-8am EDT for a security-related database reorganization.

HathiTrust sends notice upon discovery and resolution of unscheduled outages and in advance of scheduled outages and maintenance work that may result in an outage. We welcome and encourage additional recipients for these notices. If your institution is not receiving outage notifications and would like to, please contact feedback@issues.hathitrust.org.

User Support Issues	August	July
Content	286	326
Quality	279	318
Non-partner Digital Deposit	1	0
Collections	3	4
Cataloging	142	113
Access and Use	119	112
Copyright	62	66
Permissions	15	16
Takedown	0	1
Print on Demand	1	4
Inter-library loan	8	6
Full-PDF or e-copy requests	21	16
Datasets	7	4
Data Availability and APIs	1	0
Reuse of content	4	3
Web applications	22	27
Functionality problems	8	3
Problems with login specifically	1	0
General questions about login	1	1
Partners setting up login	4	0
Usability issues	0	12
Feature requests	2	2
Partner Ingest	4	2
General	74	108
Partnership	9	7
Infrastructure	0	0
Miscellaneous	65	101
Total	647	688

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

There's an elephant in the library.™

