HathiTrust Digital Library

Update On October Activities

Top News

HathiTrust Board Officers

The HathiTrust Board of Governors has identified officers for the Executive Committee as follows:

Chair: Brian Schottlaender
Chair-elect/Treasurer: Sarah Michalak
Past Chair: Paul Courant
Chair of the Program Steering Committee: Bob Wolven
Executive Director (ex officio): John Wilkin

More information about the Board of Governors, including the charge and full membership is available at http://www.hathitrust.org/board_of_governors.

Funded PhD Opportunities with HathiTrust Research Center

The Graduate School of Library and Information Science (GSLIS) and the Illinois Informatics Institute (I3) at the University of Illinois are actively recruiting outstanding doctoral candidates interested in research assistantships with the HathiTrust Research Center (HTRC) to develop the HTRC infrastructure, create mechanisms for outreach and engagement with scholarly communities, and cross-pollinate ideas among HTRC stakeholders. View the full announcement for more information.

Infrastructure Change for Out of Print and Brittle

HathiTrust altered the semantics of the "out-of-print and brittle" ("opb") designation in the HathiTrust Rights Database to "out-of-print" ("op") only, as outlined in last month's update. Volumes with the "op" designation began appearing in the tab-delimited Hathifiles on November 2. All "op" volumes will be updated in the Hathifiles on November 12. Rights Database documentation, including a sample scenario, has been updated to reflect the change.

Ingest

Local Digitization

HathiTrust answered questions from staff at the University of Missouri, University of Utah, and University of Washington about ingest of locally-digitized content, including questions about the new ingest tools for packaging content prior to submission to HathiTrust.

Internet Archive Digitization

Penn State and Columbia University provided bibliographic records for new sets of Internet Archive-digitized volumes to be ingested. Content from Columbia University is from its Medical Heritage Library. The University of North Carolina contact-

HathiTrust Digital Library

Update On October Activities

ed HathiTrust staff to begin deposit of a second batch of Internet Archive-digitized volumes. The Getty Research Institute resumed discussions regarding deposit of its IA-digitized materials.

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

User Experience Advisory Group

The User Experience Advisory Group provided feedback on a new landing page for Limited (search-only) volumes in HathiTrust and a prototype of a new PageTurner design created by University of Michigan staff.

User Support Working Group

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

Projects

Bibliographic Data Management

California Digital Library (CDL) staff worked with staff at the University of Michigan to test data exports from Zephir that will be used in HathiTrust services such as bibliographic and full-text search. The testing examined issues of performance in data transfer, as well as the structure of the exports.
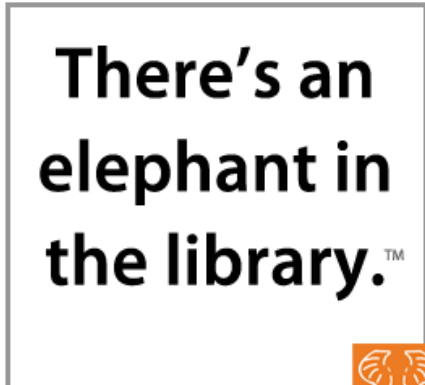
CDL staff completed testing of the Zephir bibliographic record submission process with the majority of institutions that are contributing records to HathiTrust on an ongoing basis. CDL and HathiTrust staff met to discuss the process for communicating with institutions about submission of bibliographic data and content once the cutover to Zephir occurs.

Copyright Review

A summary of copyright review activities in October is given below.

	October		Overall	
	Opened	Reviewed	Opened	Reviewed
CRMS-US	4,177	8,404	178,872	338,463
CRMS-World	4,933	8,699	15,181	30,965
Total	9,110	17,103	194,053	369,428

HathiTrust Digital Library

Update On October Activities

IMLS Quality Grant

Members of the project team continued preparations to launch the first of two user studies related to content quality. The first study will use image review exercises and focus groups to examine thresholds of error tolerance in digital volumes for library collection managers. Staff from the University of Michigan and University of Minnesota will participate in the study.

The project team analyzed outcomes of its meeting with imaging scientist Don Williams, which took place in September, and enhanced its catalog of commonly identified illustration errors based on information from the meeting.

The team worked to finalize a data curation profile and produce final datasets of the data collected during the grant project. More information on the project is available on the project website.

mPach

Staff at the University of Michigan completed a prototype of the Prepper module (see a list of all modules), as well as enhancements to PageTurner to display journal articles encoded in JATS XML, in time for a presentation and demo at the 2012 Digital Library Federation Forum.

Development Updates

Authentication

HathiTrust fixed a bug that prevented authentication for users who had certain character entity references (e.g., "é") in their Shibboleth displayName attribute. HathiTrust also implemented functionality to map users from multiple authentication Identity Providers (IdPs) to a single partner institution. This functionality comes into play when multiple campuses or organizations are members under the aegis of a single institution.

Data API

HathiTrust completed final development work associated with supporting OAuth signatures on requests to the Data API. HathiTrust also began work on version 2 of the Data API, and tested new features that will support the delivery of PDFs for print-on-demand purposes, and include improved URI syntax to better support new formats such as JATS XML for mPach.

Full-text Search

Staff at the University of Michigan conducted a series of tests to gather technical requirements for an RFP for a new high-performance storage system to improve the response time of full-text search, increase the volume of searches the system can handle, and accommodate the extra load that new relevance ranking features would introduce. The tests resulted in specific numerical requirements that were

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Update On October Activities

incorporated as minimum specifications into the RFP, which was completed and released to ten suppliers in October, with proposals due back in early November. Evaluation and final pricing negotiation is expected to continue through November and December, with system installation to take place in early 2013.

Michigan staff made changes to full-text search, as well as the HathiTrust bibliographic catalog, to improve faceting on the Author field for works with multiple authors.

Staff continued research geared toward improving relevance ranking and indexing of works in Chinese, Japanese, and Korean.

Imgsrv

Imgsrv is the web application that serves derivatives of HathiTrust's master images to Web applications such as the PageTurner. HathiTrust made changes to the way Imgsrv constructs PDFs for download to optimize for size. When possible, the original JP2 and TIFF images stored in the repository are included in the PDF. If there is a risk that the final PDF will be over 2GB, a lower resolution derivative is extracted from JP2 images and compressed as a JP2; TIFF images are scaled down and compressed as JPEGs.

PageTurner

In conjunction with recommendations from the UX Advisory Group, the default view in HathiTrust was changed to "scroll" view. HathiTrust also improved processes for caching images and made modifications to the landing page for the Limited (search-only) works.

Website Redesign

Over the last several months, University of Michigan UX department staff have been working on new designs for the HathiTrust home page and application interfaces. In October, developers at Michigan began to explore options for a consolidated framework of Cascading Style Sheets (CSS) across HathiTrust applications.

Total Volumes Added	October	Overall
Boston College	0	1,816
Columbia University	2	64,186
Cornell University	3,317	412,154
Duke University	0	4,523
Harvard University	2	235,985
Indiana University	7,057	194,740
Library of Congress	0	89,722
North Carolina State University	0	3,196
Northwestern University	5,342	12,563
New York Public Library	3	259,574
Penn State University	4	44,135
Princeton University	6	251,650
Purdue University	3,989	44,455
Universidad Complutense	2	111,901
University of California	4,522	3,378,394
University of Chicago	1,739	26,656
University of Illinois	1	101,011
University of Michigan	14,426	4,596,970
University of Minnesota	919	103,535
University of North Carolina - Chapel Hill	0	8,088
University of Wisconsin	1,014	546,802
University of Virginia	9	50,799
Utah State University	0	117
Yale University	0	23,678
Total	42,354	10,566,650
Public Domain (~30% of total)		
Total*	40,347	3,252,107

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

HathiTrust Digital Library

Update On October Activities

Outages

No outages were reported in October.

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	October	September
Content	**310**	**248**
Quality	297	242
Non-partner Digital Deposit	1	0
Collections	6	2
Cataloging	**111**	**80**
Access and Use	**112**	**116**
Copyright	58	71
Permissions	11	5
Takedown	1	2
Print on Demand	1	0
Inter-library loan	4	4
Full-PDF or e-copy requests	13	11
Datasets	2	3
Data Availability and APIs	0	0
Reuse of content	0	1
Web applications	**21**	**12**
Functionality problems	8	4
Problems with login specifically	0	0
General questions about login	0	0
Partners setting up login	0	0
Usability issues	1	0
Feature requests	1	0
Partner Ingest	**9**	**3**
General	**61**	**55**
Partnership	14	10
Infrastructure	0	0
Miscellaneous	47	45
Total	**624**	**514**

*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.™

www.hathitrust.org