



Update On February Activities

March 8, 2013

Top News

HathiTrust Bylaws

HathiTrust institutions voted unanimously to accept [bylaws](#) put forward by the Board of Governors. The Board will have its first meeting following the acceptance of the bylaws in April.

HathiTrust Research Center Software Release

The HathiTrust Research Center (HTRC) is planning a release of its software infrastructure on March 31, 2013. The release is of two separate cyberinfrastructure stacks, the “HTRC sandbox” and “HTRC production stack”, each loaded with the latest and greatest service components and tools. The “HTRC sandbox” is an open test bed for community experimentation. It contains a subset of public domain volumes. The “HTRC production stack” hosts the full HathiTrust public domain corpus. Both the sandbox and production stack offer the suite of services and tools, with bug fixes and other improvements, that debuted during the September 2012 UnCamp. Looking ahead, a June 2013 release of the HTRC production stack will include the HTRC-Sloan-Cloud in early support of “non-consumptive research”. Non-consumptive research is research in which the content of digital works is read or operated on (“consumed”) in an automated way only; no portions of works are displayed or available to be read by researchers themselves during computation or in computational results.

We welcome anyone to join HTRC email lists for [general announcements](#), [technical discussion](#) about the HTRC, or announcements [specific to the 2012 UnCamp](#) (see <http://www.hathitrust.org/htrc> for details on each list).

Ingest

Local Digitization

HathiTrust provided support to several institutions on use of the HathiTrust [ingest tools](#) and issues related to image conversion and metadata.

Internet Archive Digitization

HathiTrust ingested new content from Columbia University, the Library of Congress, Penn State, the University of California, the University of Illinois, and the University of North Carolina-Chapel Hill. HathiTrust also loaded bibliographic records from Boston College in preparation for content deposit.

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.

March Forecast

Continue work to improve relevance ranking for full-text search results.

Continue to develop deployment process for full-text-search spelling suggestion feature.

Continue work on unified Web application framework and website redesign.

Presentations

Jeremy York, “Your Library, Now Online! Putting HathiTrust in the Context of New (and Traditional) Library Services”, MCLS Webinar, February 6, 2013.

Tom Burton-West, “Practical Relevance Ranking for 10 Million books”, Code4Lib, February 12, 2013.

Jeremy York, “Putting It All Together: HathiTrust Vision, Practice, and Implementation”, SENYLRC Webinar, February 20, 2013.

There’s an
elephant in
the library.™





Update On February Activities

Operational

User Experience Advisory Group

The User Experience Advisory Group continued to review elements of HathiTrust Web applications identified through user feedback and other means as being in need of improvement.

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

Staff from the California Digital Library (CDL) and the University of Michigan continued to plan for the transition to Zephir for HathiTrust bibliographic data management. CDL staff began loading records into the Zephir production database.

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in February is given below.

Copyright Determinations

	February		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	3,438	7,169	125,185	231,165
CRMS-World	2,067	4,048	19,087	35,756
Total	5,505	11,217	144,272	266,921

mPach

Staff at the University of Michigan made architectural and procedural decisions related to the creation, storage, and revision of “aboutware” (information about editorial boards, submission policies, etc.) for journals deposited via mPach. Staff finalized journal article ingest procedures, and continued discussions about optimal handling of non-textual content that is embedded in articles or submitted as supplementary material. Staff also reviewed and updated the [project timeline](#).

You can follow HathiTrust on [Twitter](#) or [Facebook](#)
Subscribe to email updates (via [Google Groups](#))

There's an elephant in the library.™





Update On February Activities

Development Updates

Full-text Search

HathiTrust received a revised estimated shipping date for high-performance storage that will be used to improve full-text search services. The systems are expected to arrive in March for installation and testing.

Michigan staff made changes to optimize synchronization of the full-text search index from the HathiTrust repository instance in Michigan to its mirror site in Indiana. At the same time, staff performed a full update of the index to be sure that all volumes in the repository were indexed appropriately. Staff developed a process to automatically update the full-text index when print holdings information submitted by partner institutions is updated or added. Staff also continued work to improve relevance ranking of search results.

PageTurner

HathiTrust updated the way URL parameters are sent to Google Analytics in order to improve usage reporting. The change will especially affect full-text searches performed within individual volumes.

Print on Demand

HathiTrust finalized and deployed processes for producing PDFs that are optimized for printing on Espresso Books Machines.

Storage Hardware Replacement Cycle

HathiTrust began processes to replace production web servers at the Indiana data center, and development index and web servers at the Michigan data center. The server upgrades are planned to begin in late March or April.

Website Redesign

Staff developing HathiTrust Web applications continued to implement a unified framework for application code. This work is being done as part of a larger effort to redesign the look and feel of HathiTrust's Web interfaces. The redesign is sched-

Total Volumes Added	February	Overall
Boston College	0	1,842
Columbia University	643	65,033
Cornell University	305	416,962
Duke University	0	4,523
Harvard University	39	236,027
Indiana University	46	195,202
Library of Congress	1	89,723
North Carolina State University	0	3,196
Northwestern University	0	12,948
New York Public Library	94	259,672
Penn State	653	45,415
Princeton University	46	251,700
Purdue University	45	44,682
Universidad Complutense	54	111,982
University of California	1,099	3,386,296
University of Chicago	7	28,540
University of Florida	0	2,068
University of Illinois	4,201	109,306
University of Michigan	11,126	4,629,824
University of Minnesota	5	104,250
University of North Carolina - Chapel Hill	755	16,588
University of Wisconsin	3,924	554,955
University of Virginia	6	50,805
Utah State University	0	117
Yale University	0	23,678
Total	22,930	10,645,334

Public Domain (~31% of total)

Total*	11,723	3,308,664
--------	--------	-----------

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

There's an elephant in the library.™



Update On February Activities

User Support Issues

	February	January
Content	430	428
Quality	421	414
Non-partner Digital Deposit	1	1
Collections	6	10
Cataloging	82	99
Access and Use	96	148
Copyright	51	85
Permissions	5	17
Takedown	1	0
Print on Demand	0	1
Inter-library loan	0	0
Full-PDF or e-copy requests	16	23
Datasets	7	10
Data Availability and APIs	3	0
Reuse of content	3	4
Web applications	15	27
Functionality problems	3	2
Problems with login specifically	0	0
General questions about login	3	4
Partners setting up login	2	4
Usability issues	0	1
Feature requests	4	3
Partner Ingest	1	4
General	74	55
Partnership	12	20
Infrastructure	0	0
Miscellaneous	62	35
Total	698	761

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

Most-accessed volumes

Title	#Visits
Quicksand, by Nella Larsen.	3,075
Investigation of Korean-American relations. U.S. House of Representatives, October 31, 1978.	3,063
Noblesa catalana: cavallers y burgesos honrats de Rossello y Cerdanya, v.1., by Philippe Laz-erme.	2,199
The Conjure-Man Dies: a Mystery Tale of Dark Harlem, by Rudolph Fisher.	2,045
A genealogical record of the descendants of Quartermaster George Colton, by John Milton Colton.	1,762
Bradshaw's handbook for tourists in Great Britain & Ireland, sec.1 1866.	989
United States government organization manual, 1970/71.	974
The blacker the berry: a novel of negro life, by Thurman, Wallace.	942
The Magistrates of the Roman Republic, Vol. 1, by T. Robert S. Broughton.	803
Deutsche Rundschau. v. 1 (1874).	760

uled for release in April.

Outages

HathiTrust was unavailable for some users from Tuesday, February 26 at 11:45pm to Monday, February 27 at 5:45am due to a network problem affecting the entire University of Michigan campus.

