



Update On September/October Activities

November 18, 2015

Top News

2015 Member Meeting Registration and Call for Presentations

The 2015 HathiTrust Member Meeting will be held on December 9 at the Big Ten Center near O'Hare Airport in Chicago. This is a members-only meeting, and all member institutions may send up to two persons to attend this meeting, including the official member representative. If the representative cannot attend, please designate an attendee who can, if needed, vote on behalf of the institution.

Following the model of the 2011 Constitutional Convention, library directors from state university systems or consortia that are HathiTrust members may also attend, although only the member representative for the consortia may vote. The final agenda for the meeting will be announced no later than the first week of December.

Please register by November 30 using the form at <http://goo.gl/forms/wByh4aaVyj>

With the additional registration space available, we are also inviting attendees to give very brief presentations describing what your institution is doing with HathiTrust on your campus. Presentations may focus on the following topics:

- how you are using HathiTrust for collection development decisions or in the area of preservation
- projects focused on the ingest of locally-digitized materials into HathiTrust
- descriptions of outreach efforts to your patrons
- how you've implemented HathiTrust in your catalog, discovery systems, or library websites
- providing access to users with print disabilities
- unique cases of how your faculty and staff are using HathiTrust
- how your faculty using HTRC services or extracted feature sets for use in research.

Other topics are also welcomed. The presentations will be in lightning talk format, and each presenter will have somewhere between three and eight minutes to speak (final duration will depend on the number of responses). Timing will be strictly enforced, so practice ahead of time. One or two slides may be used by presenters and will need to be submitted to member organizers ahead of time.

Please submit your topic by Monday, November 23, 2015 using the form below. Speakers are encouraged to register for the meeting at the time of submission using this form: <http://goo.gl/forms/HJX5YqTfrA>

HathiTrust on the Road

HathiTrust staff will be attending the following events in November and December. Please contact us if you wish to meet us at any of these events:

CNI Fall 2015 Membership Meeting, Washington, D.C., December 14-15 - Mike Furlough

There's an
elephant in
the library.™





Update on September/October Activities

November 18, 2015

Kenney, Tabb, and McNeil Elected to Board of Governors

The members of HathiTrust have elected three individuals to serve on its Board of Governors beginning on January 1, 2016. [Anne R. Kenney](#), [Carl A. Kroch Librarian, Cornell University](#), and [Winston Tabb, Sheridan Dean of University Libraries, Johns Hopkins University](#), will serve three-year terms that end in December 2018. [Beth McNeil, Dean of Library Services at Iowa State University](#), will serve a one-year term. “Each of these three accomplished librarians will bring vision, wisdom and know-how to the Board of Governors, helping to ensure that the HathiTrust Digital Library will continue to innovate and thrive,” said Sarah Michalak, past chair of the Board and chair of the 2015 Nominating Committee.

https://www.hathitrust.org/election_results_announced

ECL/Copyright Comments

In June, the United States Copyright Office (USCO) released “[Orphan Works and Mass Digitization](#)” a detailed report proposing new orphan works legislation and proposing a pilot extended collective license for in-copyright, published works that have been digitized. In response to the USCO’s request for feedback on the collective licensing proposal, over 80 individuals and organizations submitted [comments](#). HathiTrust is opposed to this pilot project, and its response, filed by Executive Director Mike Furlough, can be read [HERE](#).

2016 Budget and Fees

HathiTrust is finalizing its budget and the 2016 member fees. These will be presented to the membership for review within the next several weeks.

Collections Survey Completed

Thank you to all of the member institutions that participated in the Survey on Collection Development Priorities. Preliminary results will be presented at the Member Meeting on December 9th.

MUSAG Update

The Metadata Use and Sharing Advisory Group (MUSAG) launched this month. The group held its first meeting and has begun work to establish its framework and goals. Look for an update and more information in the next newsletter.

ZAG Update

The Zephir Advisory Group continues work on drafting policies and procedures to address requests for Zephir data and functionality.

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

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

There’s an
elephant in
the library.™





Ingest

Locally-digitized Content

A number of different institutions have submitted content for ingest: University of Washington and Utah State University have submitted some content and work is ongoing to bring these volumes in, while Universidad Complutense de Madrid, University of Missouri, University of Illinois at Urbana-Champaign, University of Washington, Columbia University (through Internet Archive) have successfully ingested content. We anticipate receiving locally digitized materials from Emory University soon, via the Internet Archive.

Bibliographic Data Management

In September and October 2015, Zephir loaded 86,563 new and 179,289 updated records from HathiTrust content contributors and has established metadata submission processes for several new content streams. Technical staff on the Zephir Operations team have been engaged in an infrastructure migration which will be completed by the end of 2015. In the coming months, the team will address how Zephir data is integrated into the HathiTrust User Support Working Group bibliographic corrections workflow.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in September/October is given below. See [CRMS-US](#) and [CRMS-World](#) for further information. The CRMS projects are funded by the Institute of Museum and Library Services.

	October		Overall	
	Public Domain Determinations	All Determinations	Public Domain Determinations	All Determinations
CRMS-US	508	711	175,794	329,520
CRMS-World	2,704	6,273	130,374	244,469
Total	3,212	6,984	306,168	573,989

Volumes Added

Updated ingest numbers and Collection statistics are updated daily and can be found on our website here: https://www.hathitrust.org/visualizations_deposited_volumes_current

There's an elephant in the library.™





Update On September/October Activities

November 18, 2015

US Federal Documents Registry

As of November 1, 2015 there are 653,821 US federal government documents in HathiTrust.

Work has continued to improve the [US Federal Government Documents Registry](#). Project staff have held discussions about the basic infrastructure necessary for a beta release, which will include one Registry record with a unique identifier instead of the current model of clustering duplicate and related records. Additional work has been focused on refining the duplicate detection algorithm, incorporating stricter matching on identifiers such as OCLC number and SuDoc call number, and testing and modifying algorithms for searching the title and SuDoc fields. Staff continue to identify records for out-of-scope materials - to-date, roughly 5,000 individual records and over 30 categories of records based on individual publishers have been identified for removal.

A more detailed overview of US Federal Government Documents Registry [activities from May-October 2015](#) can be found on the [project's web page](#).

Development Updates

Recent development updates and activities by HathiTrust institutions have included the following:

Full-text Search

Work continued on a testing framework for relevance ranking. Click logging (i.e., logging what the user clicks on in the browser) using "Balanced Interleaving" which allows comparing different relevance ranking algorithms and settings has been put into production. The framework has been configured for A/A testing (testing without any changes to ranking to make sure that the testing framework works correctly and to estimate the number of clicks needed to have confidence in test results). Initial work on log analysis software has begun.

Core Services Update

Core Services staff have been regularly meeting with the HathiTrust Executive Director to plan the storage replacement and expansion strategy for 2016 and beyond.

To handle load due to increased usage of HathiTrust, Core Services deployed an additional web server for HathiTrust at the Indiana site. An additional web server for the Michigan site is on order.

Forecast

Continue work on a test framework for relevance ranking, including interleaving of search results for the comparison of ranking algorithms

Work to remove all dependencies on Cosign in favor of complete reliance on Shibboleth.

Research ways to support alternative text formats.

There's an
elephant in
the library.™





Papers and Presentations

Workshops

Bhattacharyya, Sayan and Eleanor Dickson. "Text Analysis with the HathiTrust Research Center: Tools and Datasets for Research and Teaching." University of Chicago Digital Humanities Forum. Chicago. October 2, 2015. [Abstract](#); [Slides](#), (.pptx version) ; [Slides](#), (.pdf version)

Bhattacharyya, Sayan and Eleanor Dickson. "Getting Started with the HathiTrust Research Center's Tools for Text Analysis." Digital Libraries Federation Forum (DLF Forum), Vancouver, October 26-28, 2015.

Presentations

Jett, Jacob. "HTRC Workset Ontology." GSLIS CIRSS E-Research Round Table, 21 October 2015. Slides: <http://cirss.lis.illinois.edu/Events/eventDetails.php?id=259>

Underwood, Ted. "The Rhythms of Genre." 2nd NovelTM Workshop, Montréal, Quebec, Canada, October 23.

Miao Chen and Eleanor Dickson. "HathiTrust Research Center Tools and Services." (Webinar). Georgetown University Library, November 2, 2015.

Muhammad Saad Shamim and Sayan Bhattacharyya. "Culturomics: New Developments in Analyzing Digitized Texts." [Rice University Digital Humanities Group](#), November 9, 2015.

Sayan Bhattacharyya. "Big Textual Data in Undergraduate Student Writing for Literature Courses: Affordances of the HathiTrust Research Center's Extracted Features Dataset". Chicago Colloquium on Digital Humanities & Computer Science (DHCS 2015). University of Chicago, November 13-15, 2015. [Abstract](#).

Eleanor Dickson and Sayan Bhattacharyya. "Using the HathiTrust Research Center's Tools for Text Analysis." Chicago Colloquium on Digital Humanities & Computer Science (DHCS 2015), November 15, 2015.

Beth Plale, Robert McDonald. "Visualizing the HathiTrust Research Center (HTRC) data". Supercomputing 2015. Austin, Texas, November 16-19, 2015.

Robert McDonald. "The HathiTrust Research Center: Enabling New Knowledge Through Shared Infrastructure" in NISO Webinar: Text Mining: Digging Deep for Knowledge, November 18, 2015.

Repository Availability

Cumulative 12-month availability of repository access: 99.975% (-0.006%).

From Tuesday, August 18 at 07:00 through Wednesday, August 19 at 15:55 EDT, approximately 1 in 4 HathiTrust users were unable to authenticate if they attempted to log in, due to a failure in underlying authentication services on a web server.

On Wednesday, August 19, from 17:00-17:17 EDT the HathiTrust Collection Builder application was unavailable while the database was optimized to improve response times when viewing large collections.

On Wednesday, September 9, from 16:15-16:45 EDT HathiTrust users were unable to access resources via their handle.net persistent URL due to a software failure following an operating system upgrade.

There's an
elephant in
the library.™





Update On September/October Activities

November 13, 2015

User Support Issues	Sept-Oct	July-Aug
Content	281	286
Quality	248	264
Collections	27	21
Cataloging	233	265
Access and Use	231	206
Copyright	89	90
Permissions	14	17
Takedown	7	0
Print on Demand	1	2
Inter-library loan	6	2
Full-PDF or e-copy requests	65	54
Datasets	2	3
Data Availability and APIs	3	0
Reuse of content	9	8
Web applications	41	74
Functionality problems	11	27
Problems with login specifically	4	11
General questions about login	2	3
Partners setting up login	0	1
Usability issues	1	1
Feature requests	4	5
Partner Ingest	35	21
General	218	193
Partnership	16	16
Miscellaneous	202	177
Total	1039	1045

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

Most-accessed volumes Sept/Oct 2015

- [Quicksand, by Nella Larsen.](#)
- [Godey's Magazine, v. 40-41, 1850.](#)
- [Priscilla Wool Crochet Book.](#)
- [America is in the Heart: A Personal History, by Carolos Bulosan.](#)
- [Solid Mensuration, by Willis F. Kern and James R. Bland.](#)
- [The United States Practical Receipt Book, or, Complete Book of Reference: For the Manufacturer, Tradesman, Agriculturalist or Housekeeper: Containing Many Thousand Valuable Receipts, in all the Domestic Arts, by a practical chemist.](#)
- [Kinematics and Dynamics of Plane Mechanisms, by Jeremy Hirschhorn.](#)
- [Babbitt, by Sinclair Lewis.](#)
- [The Human Figure, by John H. Vanderpoel](#)
- [The Conjure-Man Dies: a Mystery Tale of Dark Harlem, by Rudolph Fisher.](#)

There's an elephant in the library.™

