



Update On November Activities

December 12, 2014

Top News

HathiTrust Budget

HathiTrust submitted the 2015 budget to members for approval. Fee invoices are expected to be sent to members in January.

New Staff Member

We are pleased to announce the hiring of a new Applications Developer for HathiTrust, Josh Steverman. Josh began work December 1st and will be the primary developer for the HathiTrust Government Documents Registry.

New Full-text Search Blog Post

Tom Burton-West authored [the third](#) in a series of blog posts on relevance ranking in HathiTrust, this one on document length normalization.

Ingest

Locally-digitized Content

HathiTrust ingested new locally-digitized volumes from the Getty Research Institute and the University of Illinois, and continued working with Texas A&M University and Emory University on new deposit. Utah State University and the University of Missouri are also preparing content for ingest.

Google-digitized Content

HathiTrust continued to ingest content from Harvard University and also volumes that had been previously held by Google in escrow, adding a large number of volumes from Penn State in particular.

Internet Archive-digitized Content

HathiTrust began working with the University of Pennsylvania on content submission, and began ingesting content from the Getty Research Institute (both Internet Archive- and locally-digitized), which it is submitting now on a weekly basis.

Bibliographic Data Management

The California Digital Library (CDL) loaded 58,128 new or updated bibliographic records into Zephir.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in

December Forecast

Reassess accessibility features of PageTurner with particular attention to supporting new content types.

Continue working on migration to Solr 4.10 and re-index the collection.

HathiTrust on the Road

HathiTrust administrative staff will be attending the following meetings in January 2015. Please get in touch if you would like to meet with us there.

Jeremy York, Assistant Director, HathiTrust: Modern Language Association 2015 Convention, Vancouver, BC. January 8-11.

Mike Furlough, Executive Director, HathiTrust: ALA Midwinter 2015, Chicago, IL. January 29-February 2.

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

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

There's an
elephant in
the library.™





Update On November Activities

November is given below. See [CRMS-US](#) and [CRMS-World](#) for further information.

	November		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	327	533	167,577	317,784
CRMS-World	3,504	6,692	86,117	163,731
Total	3,831	7,225	253,694	481,515

Government Documents Registry

Project staff continued to test an initial algorithm to detect relationships between government documents, including when documents are duplicates, and experimented with ways to automate the addition of SuDoc stems to records that lack them based on agency author. Project staff also contacted HathiTrust members to investigate making corrections to records of more than 6,000 government documents volumes in HathiTrust that are believed to be improperly cataloged.

HathiTrust Research Center

A paper by Sayan Bhattacharyya, Peter Organisciak and J. Stephen Downie, was accepted for publication in a special issue of the peer-reviewed journal *Interdisciplinary Science Reviews*, covering “The Future of Reading”. The paper focuses on feature extraction from a digital humanities/digital culture standpoint and was supported by the HTRC.

On November 17th, Sayan Bhattacharyya and Harriett Green conducted a workshop on the HTRC Portal at the Scholarly Commons at the University of Illinois Library. The workshop covered how to create and modify worksets, how to run algorithms on worksheets, and how to interpret the results obtained when running selected algorithms (see the [event description](#) for further details).

Beth Plale and Robert McDonald represented HTRC at the recent Supercomputing 14 conference, November 17 to 20th. Their exhibit of HTRC featured a sphere visualization, i.e. viewing HTRC-related data on a globe. The visualization included texts published per country, HTRC UnCamp 2013 participants’ geolocations, and HathiTrust Google analytics. Follow [this link](#) to view the slides from the presentation.

Development Updates

Development updates and activities by HathiTrust institutions included the following:

Analytics:

- Modified the configuration for Google Analytics to track uses of volumes

Papers & Presentations

Mike Furlough and Jeremy York, “[Collective Stewardship Through HathiTrust Digital Library](#)”, Workshop on African Studies in the Digital Age, University of Michigan, November 4, 2014.

Sarah Michalak, “[HathiTrust: An Above Campus Solution](#)”, Research Libraries UK 2014 Conference, Birmingham, England, November 14, 2014.

Harriett Green and Sayan Bhattacharyya, “[Introduction to the Hathi Trust Research Center Portal for Text Mining Research](#)”, Workshop presented at the University of Illinois Library, November 17, 2014.

Beth Plale and Robert McDonald, “[The HathiTrust Research Center: Big Data Analytics in a Secure Data Framework](#)”, Supercomputing 2014, New Orleans, LA, November 17, 18, 19, 2014.

Mike Furlough, “[Sharing Collections through Shared Stewardship](#)”, Association of Southeastern Research Libraries Fall 2014 Membership Meeting, Atlanta, GA, November 19, 2014.





Update On November Activities

(and searches within books) at the volume-level only rather than the page- and volume-level. This better reflects the way the Google Analytics data is being used, and aligns with Analytics' normal processing of heavily parameterized URLs.

Full-text Search

- A software release for full-text search high-performance storage that addresses performance and stability problems and is suitable for production deployment is expected to be received from the storage vendor for testing in December.

Storage Replacement Cycle

- Obtained pricing and submitted orders for storage hardware as part of HathiTrust's regular storage purchase and replacement cycle. This purchase follows a smaller, out-of-cycle purchase and installation of storage earlier in the fall, which was done to accommodate substantial repository growth that exceeded earlier projections. Installation is planned to start in January.

Availability

Repository

Cumulative 12-month availability of repository access: 99.949%* (+0.000%).

No outages were reported in November.

Zephir: Bibliographic metadata exports from Zephir were unavailable on November 4th due to a database network connection outage.

* Repository access refers to page viewing and full-text search functionality, i.e., user-facing applications. It does not refer to preservation or storage infrastructure, which is under continual operation.

User Support Issues	November	October
Content	129	153
Quality	118	142
Collections	11	10
Cataloging	151	198
Access and Use	120	229
Copyright	55	156
Permissions	6	9
Takedown	1	0
Print on Demand	0	0
Inter-library loan	6	2
Full-PDF or e-copy requests	14	19
Datasets	1	2
Data Availability and APIs	1	0
Reuse of content	1	3
Web applications	24	24
Functionality problems	13	6
Problems with login specifically	0	2
General questions about login	2	1
Partners setting up login	0	0
Usability issues	0	0
Feature requests	1	1
Partner Ingest	23	13
General	92	128
Partnership	7	4
Miscellaneous	85	124
Total	539	745

*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.™





Update On November Activities

Total Volumes Added

	November	Overall
Boston College	53	3,263
Columbia University	8,227	73,393
Cornell University	1,573	505,647
Duke University	26	7,801
Getty Research Institute	2,141	18,263
Harvard University	66,760	838,100
Indiana University	3,466	528,644
Keio University	0	90,094
Knowledge Unlatched	0	28
Library of Congress	0	108,892
McGill University	0	893
New York Public Library	1	294,825
North Carolina State University	0	3,196
Northwestern University	4	56,663
Ohio State University	1,821	54,299
Penn State	237,986	386,578
Princeton University	5	252,807
Purdue University	0	47,488
Sterling & Francine Clark Art Institute	0	358
Texas A&M	12	1,213
Universidad Complutense	1,784	117,229
University of Alberta	3	76,106
University of California	12,995	3,602,849
University of Chicago	7	51,966
University of Connecticut	8	4,637
University of Delaware	0	38
University of Florida	0	9,866
University of Illinois	9,850	316,633
University of Massachusetts, Amherst	13	11,128
University of Michigan	2,087	4,708,881
University of Minnesota	10	138,607
UNC - Chapel Hill	0	17,025
University of Virginia	1	51,207
University of Wisconsin	52	560,672
Utah State University	0	117
Yale University	0	23,678
Total	348,885	12,963,084

Public Domain (~37% of total)

Total*	128,174	4,843,992
--------	---------	-----------

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

Most-accessed volumes

The Lion Monument at Amhipolis, by Oscar Broneer.
Masterpieces of Furniture Design: A Collection of Measured Drawings, v.1-2 plates 1-50, by Verna Cook Salomonsky.
The Human Figure, by John H. Vanderpoel.
Quicksand, by Nella Larsen.
The Five Laws of Library Science, by S. R. Ranganathan.
Pennsylvania German Pioneers: A Publication of the Original Lists of Arrivals in the Port of Philadelphia from 1727 to 1808, Vol. 42, by Ralph Beaver Strasburger.
The Book of a Hundred Hands, by George Brant Bridgman.
Highway Safety, Design, and Operations: Freeway Signing and Related Geometrics. Hearings, Ninetieth Congress, second session.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.4
Roster of the Confederate soldiers of Georgia, 1861-1865, v.3.

There's an elephant in the library.™

