



Update On June Activities

July 11, 2014

Top News

Save the Date: HathiTrust Member Meeting

The HathiTrust bylaws passed in 2013 call for “an Annual Meeting of the Members...for the transaction of such business as may come before the meeting.” We are pleased to announce that our first Annual Meeting will be held in Washington, DC on Friday October 10, 2014.

We expect the meeting to include progress reports on ballot initiatives, official business, and opportunities to discuss future strategy. More details on location, schedule and agenda will be forthcoming in the next few weeks. For now we ask all official [Member Representatives](#) to plan to attend this meeting. If a representative cannot attend, a designate may attend in his or her place.

Ingest

Locally-digitized content

HathiTrust ingested a second batch of locally-digitized content from the University of Illinois and prepared to ingest materials from Boston College. HathiTrust also began conversations about ingest with Penn State University and Yale University, and continued communications about ingest with Emory University, University of Illinois at Urbana Champagne, and University of Washington.

Internet Archive-digitized content

HathiTrust began ingest of content from McGill University (see <http://bit.ly/1xSm5Aq>) and corresponded with University of Massachusetts, Amherst about ingest of new materials.

Google-digitized content

Many volumes scanned from partner institutions by Google in the last year were not ingested due to a change in a quality metric provided by Google that HathiTrust uses to create thresholds for content that enters the repository. In June, HathiTrust updated its use of the metric to restore the quality threshold for Google-digitized content to its previous level. The update will eventually bring more than 200,000 new volumes into the repository.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in June is given below. See [CRMS-US](#) and [CRMS-World](#) for further information.

July Forecast

Correct a bug in navigation of large scale search results.

Continue work on new Image Server capabilities for continuous text content.

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

Improve processes for building and indexing collections, and improve sorting of serial publications in the Collection Builder application.

There's an
elephant in
the library.™





Update On June Activities

	June		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	320	510	165,725	314,873
CRMS-World	4,321	7,146	63,987	124,513
Total	4,641	7,656	229,712	439,386

Government Documents Registry

Project staff continued to develop strategies to identify and make relationships between publications based on bibliographic information. This included work on rules to normalize descriptive terms and enumeration and chronology information, and rules to merge records. Staff continued to investigate methods to identify gaps in metadata, and began to think more concretely about how to engage the community in efforts to identify gaps and duplicate volumes.

Development Updates

HathiTrust institutions performed the following work related to applications and Web interfaces:

Authentication and Authorization

Staff deployed a new system for managing users who have special access to restricted materials (e.g., for copyright or quality review). The system includes functions to register new users for specific time frames, renew access with appropriate authorization, and automatically expire access, as well as back-end scripts for individual and batch renewal or expiration.

Full-text Search

The software update that is expected to resolve performance and stability problems with the high-performance storage system for full-text search was delayed, and staff continued regular communications with the storage supplier on its availability. In the meantime, staff made improvements to the new daily index update process that is currently running in a test mode on the new storage system to more smoothly handle the large data updates that occur when the search index is fully rebuilt.

Staff investigated the suitability of the INEX 2007-2010 test collections to inform choices about relevance ranking algorithms for HathiTrust full-text search.

Tom Burton-West wrote the second in a series of blog posts: "[Practical Relevance Ranking for 11 Million Books, Part 2: Document Length and Relevance Ranking](#)".

Papers & Presentations

J. Stephen Downie, "Unlocking the Secrets of 3 Billion Pages: Introducing the HathiTrust Research Center." Taipei, Taiwan, June 3, 2014.

Bob Wolven, "HathiTrust Past, Present, and Future: A Brief Introduction", Metropolitan New York Library Council, June 5, 2014.

Thomas H. Teper, "How Can Digital Collections Support Shared Print Initiatives", ALA Annual, June 27, 2014.

Partner Presentations

Melissa Levine, "HathiTrust's Copyright Review Management System: From Theory to Practice", Metropolitan New York Library Council, June 5, 2014.





Update On June Activities

PageTurner and Image Server

Staff prototyped new imgsrv capabilities for continuous text (e.g., JATS encoded articles without page breaks) in PageTurner, demonstrating in-article search.

Server replacement cycle

Staff began installation of new full-text search servers. The servers are tentatively planned to be put into service in July.

Availability

Repository

Cumulative 12-month availability of repository access: 99.867%*

No outages were reported in June.

*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.

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

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

There's an
elephant in
the library.™





Update On June Activities

User Support Issues	June	May
Content	168	131
Quality	157	124
Collections	10	7
Cataloging	163	285
Access and Use	188	142
Copyright	125	88
Permissions	6	6
Takedown	0	0
Print on Demand	0	0
Inter-library loan	2	0
Full-PDF or e-copy requests	2	17
Datasets	0	3
Data Availability and APIs	3	3
Reuse of content	3	4
Web applications	18	18
Functionality problems	7	8
Problems with login specifically	2	1
General questions about login	1	0
Partners setting up login	1	0
Usability issues	0	0
Feature requests	2	1
Partner Ingest	4	7
General	86	93
Partnership	7	7
Infrastructure	0	0
Miscellaneous	79	86
Total	627	676

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

Most-accessed volumes

Advanced accounts; a manual of advanced book-keeping, by R.N. Carter.
The Human Figure, by John H. Vanderpoel.
Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.
Hortus gallicus pro Gallis in Gallia scriptus..., Symphoriano Ca[m]pegio ... auctore; [Analogia medicinarum indaru[m] et gallicaru[m]
Investigation of the Ukrainian Famine, 1932-1933: report to Congress, by the Commission on the Ukraine Famine.
Liberty bell, a collection of original poems by Abraham Lewis
The Book of a Hundred Hands, by George Brant Bridgman.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.2.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.3.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.5.

There's an elephant in the library.™





Update On June Activities

Total Volumes Added	June	Overall
Boston College	0	3,197
Columbia University	128	65,165
Cornell University	33,857	487,762
Duke University	0	7,774
Harvard University	630	238,065
Indiana University	416	196,082
Keio University	1,124	90,080
Knowledge Unlatched	5	24
Library of Congress	1	108,883
McGill University	893	893
New York Public Library	4	291,794
North Carolina State University	0	3,196
Northwestern University	18,754	56,398
Ohio State University	3,007	26,859
Penn State	285	81,492
Princeton University	212	251,925
Purdue University	0	44,698
Sterling & Francine Clark Art Institute	32	358
Texas A&M	0	1,201
Universidad Complutense	2	112,153
University of California	20,514	3,520,634
University of Chicago	12,459	51,630
University of Florida	101	9,866
University of Illinois	6,600	142,899
University of Massachusetts, Amherst	0	11,115
University of Michigan	16,823	4,689,072
University of Minnesota	303	120,180
UNC - Chapel Hill	0	17,025
University of Virginia	377	51,202
University of Wisconsin	1,151	557,252
Utah State University	0	117
Yale University	0	23,678
Total	117,586	11,262,697

Public Domain (~34% of total)

Total*	92,066	3,848,472
--------	--------	-----------

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

**There's an
elephant in
the library.™**

