



Update On October Activities

November 17, 2014

Top News

HathiTrust Member Meeting

HathiTrust held its first annual Member Meeting on October 10, 2014. Meeting Notes, presentations, and other documentation from the meeting are [posted online](#), as is a [new blog post](#) containing reflections on the meeting by Executive Director Mike Furlough.

Research Center Request for Proposals

The HathiTrust Research Center released a [Request for Proposals](#) for Advanced Collaborative Support (ACS). ACS is a newly launched scholarly service that pairs individuals with expert staff at the HTRC over an extended period of time, to facilitate computational research on the HathiTrust corpus and use of HTRC tools. Details are provided at the link above. Interest parties are invited to submit proposals by 5:00 pm on January 8th, 2015.

Ingest

Locally-digitized Content

HathiTrust advised Texas A&M University, Columbia University, Emory University, Yale University, and the University of Washington on issues of validating content, and provided information about content submission to the University of British Columbia.

Google-digitized Content

HathiTrust ingested more than 530,000 new public domain volumes from Harvard University, and more than 200,000 volumes that had previously be held in escrow by Google from Indiana University, Pennsylvania State University, and University of Illinois Urbana Champagne.

Internet Archive-digitized Content

HathiTrust communicated with the University of North Carolina, Chapel Hill about correcting problems with images and bibliographic data and about submission of new content.

Bibliographic Data Management

The California Digital Library loaded 773,823 new or updated bibliographic records into Zephyr.

Working Groups and Committees

November Forecast

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

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

Migrate to Solr 4.10 and re-index the collection.

There's an
elephant in
the library.™





Update On October Activities

Program Steering Committee

The Program Steering Committee (PSC) held its second in-person meeting in Washington, DC, on October 11th, the day following the first annual Members meeting. In addition to reviewing work under way in the [currently active working groups](#), the Committee received and began discussing a draft report and recommendations from the Government Documents Initiative Planning and Advisory Group. After further review, the PSC expects to forward the report to the Board in December, with recommendations for action. The remainder of the meeting focused on four broad areas that have been identified for further planning and activity in the coming year: Non-Text Formats; Quality Assurance and Validation; Services for Users who have Print Disabilities; and Metadata Strategies and Policies (view the [planning briefs](#) in these areas for more information). Through the remainder of the fall the PSC will use its biweekly calls to take up each of these areas in turn, and develop action plans for programmatic activities.

Projects

Copyright Review

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

	October		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	525	896	167,338	317,403
CRMS-World	6,996	11,690	83,564	158,893
Total	7,521	12,586	250,902	476,296

Government Documents Registry

Project staff continued to refine a relationship detection algorithm for US government documents, and hope to have an initial algorithm finalized by mid-November. Staff also continued to identify improperly cataloged records for US government documents in HathiTrust, and to seek to determine the comprehensiveness of selected government documents titles. An update of project activities for the [past six months](#) is now available from the [Registry web page](#).

HathiTrust Research Center

On October 23rd, J. Stephen Downie, Jacob Jett, Peter Organisciak and Loretta Auvil of the University of Illinois and Pip Wilcox of Oxford University presented an overview of the HTRC at the 2014 Chicago Colloquium on Digital Humanities and Computer Science. Panel members presented on the following topics:

Papers & Presentations

Mike Furlough, "Linking Print and Digital Strategies", Harvard University, October 1, 2014.

Mike Furlough, "Sharing Collections through Shared Stewardship: A HathiTrust Progress Report", Northwestern University, October 21, 2014.

Mike Furlough, "Why Digitize? or The Limits of Preservation", TEI, DHCS, October 23, 2014.

Jeremy York, "Big Collections in an Era of Big Copyright: Proactical Strategies for Making the Most of Digitized Heritage", DLF Fall Forum, October 28, 2014.

Presentations from the HathiTrust Member Meeting are linked to from the [Meeting notes](#).





Update On October Activities

- Introduction to HTRC (Downie)
- WCSA/Collection Building (Jett & Wilcox)
- Feature Extraction (Organisciak)
- HTRC Bookworm (Auvil)

More information on HTRC's panel at DHCS 2014 can be found on the conference website.

CLIR Fellows Sayan Bhattacharyya from the University of Illinois and Matt Davis from North Carolina State University were awarded a CLIR micro-grant to research and develop use cases for new tools to conduct large-scale algorithmic analysis of text corpora. The use cases are intended to support the development of tutorials for such tools, including tools to be used in the HathiTrust Research Center.

Development Updates

Development updates and activities by HathiTrust institutions included the following:

Authentication, Authorization and Access:

- Continued to add support for "access profiles" (see the [Update on September Activities](#)), including modifications to mechanisms that display relevant rights information in OAI records, and watermarks in the HathiTrust PageTurner.

Full-text Search

- Fixed a bug affecting indexing and full-text searching of an estimated 50% or more of Chinese and Japanese volumes. Searching of these materials is now significantly improved.
- Performed benchmarking tests on the new high-performance storage system after installing new pre-release software. The system now performs as expected, and will be put into service when a software release suitable for production deployment is obtained from the provider.
- Made further enhancements to the search index update and release process that will be used with the new storage system.

Server Replacement Cycle

- Completed installation of new full-text search servers at the Indiana repository instance, and transitioned those and the new servers installed at Michigan in September into service.

Storage Replacement Cycle

- Purchased and completed an early installation of approximately half of the new storage for the 2015 cycle. The storage was purchased to accommodate substantial repository growth this fall, which exceed earlier projections.

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

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

There's an
elephant in
the library.™





Update On October Activities

Availability

Repository

Cumulative 12-month availability of repository access: 99.949%* (+0.105%).

Permanent links to HathiTrust volumes, including links from the HathiTrust catalog, were not working on Thursday, October 9 from approximately 4:30-5:10pm due to an outage with the CNRI Handle Service.

A bug in Zephir resulted in a failure to export full catalog metadata on October 31. The problem was corrected on November 4. As a result of the problem, the aggregate “[hathifile](#)” generally produced on the first of each month was not available until November 4.

* 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.

There's an
elephant in
the library.™





Update On October Activities

User Support Issues

	October	September
Content	153	172
Quality	142	161
Collections	10	10
Cataloging	198	223
Access and Use	229	110
Copyright	156	61
Permissions	9	8
Takedown	0	1
Print on Demand	0	1
Inter-library loan	2	2
Full-PDF or e-copy requests	19	16
Datasets	2	4
Data Availability and APIs	0	1
Reuse of content	3	5
Web applications	24	22
Functionality problems	6	10
Problems with login specifically	2	1
General questions about login	1	2
Partners setting up login	0	2
Usability issues	0	0
Feature requests	1	0
Partner Ingest	13	12
General	128	101
Partnership	4	14
Miscellaneous	124	87
Total	745	640

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

Most-accessed volumes

[The Lion Monument at Amhipolis, by Oscar Broneer.](#)

[Quicksand, by Nella Larsen.](#)

[Mitchell's Modern Atlas: A Series of Forty-Four Copperplate Maps.](#)

[The Human Figure, by John H. Vanderpoel.](#)

[Now and Then and Long Ago in Rockland County, New York, compiled by Cornelia F. Bedell.](#)

[Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.](#)

[Perfume and Flavor Materials of Natural Origin, by Steffen Arctander.](#)

[Highway Safety, Design, and Operations: Freeway Signing and Related Geometrics. Hearings, Ninetieth Congress, second session.](#)

[Godey's Magazine, v.40-41, 1850.](#)

[Coffee Processing Technology, v. 1, by Michael Sivetz and H. Elliott Foote.](#)

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





Update On October Activities

Total Volumes Added	October	Overall
Boston College	0	3,210
Columbia University	0	65,166
Cornell University	1,607	504,074
Duke University	0	7,775
Getty Research Institute	1	16,122
Harvard University	533,275	771,340
Indiana University	132,916	525,178
Keio University	14	90,094
Knowledge Unlatched	1	28
Library of Congress	9	108,892
McGill University	0	893
New York Public Library	6	294,824
North Carolina State University	0	3,196
Northwestern University	17	56,659
Ohio State University	1,909	52,478
Penn State	57,065	148,592
Princeton University	2	252,802
Purdue University	575	47,488
Sterling & Francine Clark Art Institute	0	358
Texas A&M	0	1,201
Universidad Complutense	2,067	115,445
University of Alberta	129	76,103
University of California	8,536	3,589,854
University of Chicago	56	51,959
University of Connecticut	0	4,629
University of Delaware	1	38
University of Florida	0	9,866
University of Illinois	11,747	306,783
University of Massachusetts, Amherst	0	11,115
University of Michigan	3,161	4,706,794
University of Minnesota	17	138,597
UNC - Chapel Hill	0	17,025
University of Virginia	0	51,206
University of Wisconsin	1,308	560,620
Utah State University	0	117
Yale University	0	23,678
Total	754,419	12,614,199

Public Domain (~37% of total)

Total*	704,260	4,715,818
--------	---------	-----------

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

There's an
elephant in
the library.™

