



Update On July Activities

August 9, 2014

Top News

HathiTrust Research Center Award and Job Announcement

The HathiTrust Research Center (HTRC) was awarded a grant from the National Endowment of the Humanities for its project, "Exploring the Billions and Billions of Words in the HathiTrust Corpus: HathiTrust+Bookworm". View the [full announcement](#).

The HTRC is also seeking a Manager of Operations and Lead R&D Architect. Please see the [job posting](#) for more information. Applications are being accepted until August 14, 2014, or until the position is filled.

HathiTrust Member Meeting

As announced in the [Update on June Activities](#), HathiTrust's first Annual Meeting will be held in Washington, D.C. on Friday, October 10, 2014. We ask all official Member Representatives to plan to attend. Following the model of the 2011 Constitutional Convention, library directors from consortia that are HathiTrust members may also attend. Details on the location, schedule and agenda will be distributed soon.

Ingest

Locally-digitized content

HathiTrust corresponded with the University of Washington, University of Iowa and Princeton University about ingest of locally-digitized content.

Internet Archive-digitized content

HathiTrust began ingest of content from the University of Connecticut and corresponded with Washington University, the University of Massachusetts, Amherst and Columbia University about ingest of new content.

Bibliographic Data Management

The California Digital Library loaded 98,850 new or updated bibliographic records into Zephir.

Projects

Copyright Review

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

August Forecast

Make improvements to the interface for navigating full-text 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.

Migrate to Solr 4.9 and reindex the collection.

There's an
elephant in
the library.™





Update On July Activities

	August		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	538	772	166,193	315,546
CRMS-World	3,640	6,285	67,366	130,405
Total	4,178	7,057	233,559	445,951

Government Documents Registry

Project staff documented possible methods for identifying items as U.S. federal government documents based on their bibliographic metadata, and continued work on an algorithm to detect relationships between items. These methods will be tested and refined in the coming weeks.

HathiTrust Research Center

Tim Cole and Peter Organisciak recently presented HTRC posters on [HathiTrust metadata evaluation](#) and [large-scale text analysis](#) at Digital Humanities 2014 in Lausanne Switzerland, July 7-12, 2014. The following week, J. Stephen Downie and Megan Senseney conducted instructional sessions about HTRC tools and services across multiple workshops at the [Digital Humanities Oxford Summer School](#), July 14-18, 2014.

Development Updates

Development activities by HathiTrust institutions included the following:

Authentication and Authorization:

- Enhancements to the workflow for updating access privileges for staff who have special access to restricted materials.

Collection Builder

- Improved performance of the Collection Builder application when sorting lists of items in large personal collections, and improved accuracy in the sorting of multi-part monograph and serial volumes when date information is available.

Full-text Search

- A determination that the INEX 2007-2010 Book Track test collections would not be suitable for use in testing HathiTrust full-text search relevance ranking algorithms due to several issues, including missing relevance judgments and underspecified queries. Staff are in the process of analyzing the issues to design criteria for creating a suitable test collection.

Papers & Presentations

Mike Furlough, "HathiTrust: Sharing, Access, and Stewardship", Association of European Research Libraries (LIBER) 2014 Annual Conference, Riga, Latvia, July 2, 2014.

K. Fenlon, T. Cole, M.J. Han, C. Willis, C. Fallaw, "Rethinking HathiTrust Metadata to Support Workset Creation for Scholarly Analysis", DH2014, Switzerland, July 10, 2014.

P. Organisciak, S. Battacharyya, L. Auvil, B. Plale, J. S. Downie, "Large-scale Text Analysis Through the HathiTrust Research Center", DH2014, Switzerland, July 10, 2014.

M. Senseney, J. Stephen Downie, Instructional sessions presented at the Digital Humanities at Oxford Summer School 2014, July 14-18, 2014.

Mike Furlough, "Sharing Collections through Shared Stewardship: A HathiTrust Progress Report", TRLN 2014 Annual Meeting, Chapel Hill, NC, July 23, 2014.

There's an
elephant in
the library.™





Update On July Activities

- Continued communication with the supplier of the high-performance storage system for full-text search. Staff now await a software update that is expected to resolve performance and stability problems.

PageTurner

- The release of a new user interface “skin” for the Copyright Review Management System. This update brings the CRMS interface into closer alignment with the public-facing PageTurner interface, and will address presentation bugs and facilitate future changes.

Server replacement

- Continued installation of new full-text search servers. The servers are now expected to be put into service in August at the Michigan site and in September at the Indiana site.

Availability

Repository

Cumulative 12-month availability of repository access: 99.844%*

Service was unavailable on Friday, July 25 from 6:30-8:30am EDT and full-text search was additionally unavailable until 9:15am EDT, when blocking measures were implemented against abnormally heavy search activity and all services were restored.

Personal collections were unavailable on Monday, July 28 from 5:00-5:10pm EDT for a database optimization designed to increase performance.

* 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 July Activities

User Support Issues	July	June
Content	197	168
Quality	182	157
Collections	14	10
Cataloging	179	163
Access and Use	178	188
Copyright	126	125
Permissions	3	6
Takedown	0	0
Print on Demand	0	0
Inter-library loan	2	2
Full-PDF or e-copy requests	10	2
Datasets	3	0
Data Availability and APIs	3	3
Reuse of content	7	3
Web applications	22	18
Functionality problems	5	7
Problems with login specifically	4	2
General questions about login	1	1
Partners setting up login	1	1
Usability issues	1	0
Feature requests	2	2
Partner Ingest	9	4
General	122	86
Partnership	10	7
Miscellaneous	112	79
Total	707	627

*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.
Coffee processing technology, v. 1, by Michael Sivetz and H. Elliott Foote.
Hortus gallicus pro Gallia scriptus..., Symphoriano Ca[m]pegio ... auctore; [Analogia medicinarum indaru[m] et gallicaru[m]
Quintus Curtius [History of Alexander], Vol. 2, with an English translation by John C. Rolfe.
Pearson's magazine. v.5 no.4 (Apr. 1901).
The Human Figure, by John H. Vanderpoel.
Kinematics and dynamics of plane mechanisms, by Jeremy Hirschhorn.
Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.3.
Quintus Curtius [History of Alexander], Vol. 1, with an English translation by John C. Rolfe.

There's an
elephant in
the library.™





Update On July Activities

Total Volumes Added	July	Overall
Boston College	13	3,210
Columbia University	1	65,166
Cornell University	6,108	493,870
Duke University	1	7,775
Harvard University	0	238,065
Indiana University	16	196,098
Keio University	0	90,080
Knowledge Unlatched	0	24
Library of Congress	0	108,883
McGill University	0	893
New York Public Library	3,024	294,818
North Carolina State University	0	3,196
Northwestern University	1	56,399
Ohio State University	15,064	41,923
Penn State	9,996	91,488
Princeton University	850	252,775
Purdue University	2,214	46,912
Sterling & Francine Clark Art Institute	0	358
Texas A&M	0	1,201
Universidad Complutense	1,129	113,282
University of California	47,213	3,567,847
University of Chicago	34	51,664
University of Connecticut	4,629	4,629
University of Delaware	0	28
University of Florida	0	9,866
University of Illinois	10,283	153,182
University of Massachusetts, Amherst	0	11,115
University of Michigan	8,702	4,697,774
University of Minnesota	18,247	138,427
UNC - Chapel Hill	0	17,025
University of Virginia	4	51,206
University of Wisconsin	1,398	558,650
Utah State University	0	117
Yale University	0	23,678
Total	128,927	11,391,624

Public Domain (~34% of total)

Total*	120,097	3,968,569
--------	---------	-----------

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

There's an
elephant in
the library.™

