HathiTrust Digital Library

SPECIAL EDITION - 2014 Mid-Year Review

July 7, 2014

The first half of 2014 included several significant milestones for HathiTrust. In February, the HathiTrust Board selected Mike Furlough to be the new Executive Director of HathiTrust, marking a major transition and beginning of a new phase for the partnership. In February also, partners surpassed 11 million volumes in the digital repository collection. In June, the U.S. 2nd Circuit Court released its ruling on the lawsuit brought by the Authors Guild and others against HathiTrust. The ruling re-affirmed the lawful work HathiTrust has undertaken to expand access to library collections. Throughout the year, partners have been increasing their involvement through working groups and committees tasked with moving forward our broad initiatives in collections, shared print monograph archiving, U.S. federal government documents, and rights and access. We enter the summer and fall with great momentum and an ever-increasing knowledge and appreciation for the tremendous amount we can accomplish, for our institutions and the world, when working together.

Highlighted Achievements and Activities

Details on each item can be found in the monthly updates from 2014, available at http://www.hathitrust.org/updates.

New Executive Director

HathiTrust announced the appointment of Mike Furlough as the Executive Director of HathiTrust. Mike began on May 19.

New Partners

4 new partners joined HathiTrust in the first half of 2014:

- Montana State University
- Mount Holyoke College
- University of Maine
- University of Texas System

New Content

HathiTrust partners contributed 384,536 volumes to the repository. 306,317 of these are in the public domain or have been made open access with permission. In addition to content from partners, HathiTrust ingested more than 80,000 volumes from Keio University, 326 volumes from the Sterling and Francine Clark Art Institute Library, and a set of 19 open access volumes made available through Knowledge Unlatched. Contributions of new content are shown in the table at the end of the update.

HathiTrust released a full-volume validation and packaging service for locally-digitized materials (see http://www.hathitrust.org/ingest_tools). If you are interested in receiving updates related to these tools, please subscribe to the HathiTrust Ingest Google Group.

Ruling in Authors Guild Lawsuit Appeal

The U.S. Second Circuit Court released its decision in the appeal of the Authors Guild lawsuit against HathiTrust. View HathiTrust's statement on the ruling.

"Heartbleed bug"

HathiTrust released a statement describing the scope of the impact of the "Heartbleed bug" on HathiTrust infrastructure and services.

CRMS Milestone

Staff from several partner institutions completed the review of in-copyright works in HathiTrust published in the United States from 1923 to 1963. This marked a major milestone in the work the Copyright Review Management System was established to carry out. Review of works in the CRMS-World project (reviewing works published outside the US) is ongoing. CRMS-US and CRMS-World are projects gener-

ously funded by the Institute of Museum and Library Services.

11 Million Volumes

HathiTrust surpassed 11 million volumes in the digital repository. A history of HathiTrust's road to the first 10 million volumes is available on the HathiTrust blog.

Orphan Works Roundtable

Executive Committee chair Sarah Michalak and Mike Furlough participated in a Roundtable discussion organized by the U.S. Copyright Office on March 10 and 11 on Orphan Works and Mass Digitization. Melissa Levine, Lead Copyright Officer at the University of Michigan Library also participated. View HathiTrust's written comments on the Roundtable.

Government Documents Call for Records

More than 40 institutions, including HathiTrust partners and non-partners, submitted records in response to HathiTrust's call for US federal government document records, issued in November 2013. The records were requested for analysis purposes as part of HathiTrust's US government documents initiative.

Governance and Working Groups

Board of Governors

The HathiTrust Board of Governors met on May 9, 2014 in Columbus, OH for one of two in-person meetings held each year (two additional meetings are held by phone). A summary of the meeting and outcomes can be found at http://www.hathitrust.org/updates_may2014#Board.

The Board appointed 2 new members to the Program Steering Committee to serve 2-year terms, beginning in June. The new members are Robert McDonald, Associate Dean, Library Technologies, Indiana University, and Chris Freeland, Associate University Librarian, Washington University in St. Louis.

Program Steering Committee

The Program Steering Committee finalized the charges and membership of 4 new groups to carry forward HathiTrust activities (see http://www.hathitrust.org/working_groups). The groups are:

- Collections Committee
- Government Documents Planning and Advisory Group
- Print Monographs Archive Planning Task Force
- Rights and Access Working Group

Updates on the activities of these groups will be reported in future newsletters.

The PSC also began to review HathiTrust's use of automated quality metrics provided by Google to reduce the number of poorer quality volumes that are ingested. The PSC will be appointing a task force to assess the issues and make recommendations.

User Support Working Group

A summary of the issues received by the User Support Working Group is shown in a table at the end of the review.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities from the first half of 2014 is given below. See CRMS-US and CRMS-World for further information.

SPECIAL EDITION - 2014 Mid-Year Review

	Jan-Jun 2014		Overall	
	Public Domain	All	Public Domain	All
CRMS-US	7,564	9,361	165,725	314,873
CRMS-World	19,759	39,476	63,987	124,513
Total	27,323	48,837	229,712	439,386

Government Documents Registry

Work of the Government Documents Registry project team focused on the development of functional objectives for the Registry, and the development of strategies and processes to 1) identify duplicate records and understand relationships between different record sets and 2) identify gaps in government documents holdings, with an eye toward being able to determine the comprehensiveness of certain sets of materials in the HathiTrust repository.

HathiTrust is seeking an applications developer to design, implement, and populate a Registry. See the University of Michigan Jobs site for the full description and application details.

HathiTrust Research Center (HTRC)

Activities of the HTRC included the following:

- Release of a new dataset of features extracted from a subset of HathiTrust volumes.
- Continuation of monthly HTRC user group meetings. Notes from the meetings are posted on the HTRC Wiki.
- The awarding of mini-grants to 4 projects as part of the Workset Creation For Scholarly Analysis: Prototyping Project, funded by the Institute for Museum and Library Services.
- Preparations to include in-copyright material from the HathiTrust corpus.

Numerous presentations and workshops (see http://www.hathitrust.org/papers).

mPach

Michigan staff continued to develop and make improvements to mPach workflow modules designed to normalize and prepare born-digital publications for ingest into HathiTrust. Staff also focused on user interface issues, with specific attention to accessibility. A revised timeline for mPach implementation is posted at http://www.hathitrust.org/mpach.

Repository Updates

Activities in the first half of 2014 included the following:

Bibliographic Data Management

- Loading of more than 650,000 new or updated bibliographic records for volumes from 27 sources into Zephir, HathiTrust's bibliographic metadata management system.

New Functionality / Application Changes

Authentication and Authorization

- Development of a new application to improve management of staff who have special access to restricted materials (e.g., for copyright review or as a proxy for users who have print disabilities). Deployment is expected to occur in June.

Full-text search

- Extensive work to improve relevance ranking of search results, including in-depth testing of new indexing strategies (see blog posts from May and June).
- Integration and testing of a spelling suggestion feature developed by the California Digital Library.
- Work to install new high-performance storage for full-text search, which has been delayed due to issues encountered with hardware from the supplier.

SPECIAL EDITION - 2014 Mid-Year Review

Google Analytics

- Configuration of HathiTrust's Google Analytics to track the usage of HathiTrust Collections in addition to individual items.

ImageServer

- Release of a new version of HathiTrust's imgsrv application. The new version more effectively supports the generation of derivative versions of HathiTrust content for delivery to users and other HathiTrust applications.
- Release of an update to generate EPUB versions of content, delivered only through the mobile interface, using HTML coordinate OCR when HTML OCR is available.

PageTurner

- Addition of an "Embed this Book" feature and improvements and bug fixes to the "search in this text" functionality.

Repository and Infrastructure Changes

Server Replacement

- Completion of the replacement cycle for production web servers at the Michigan and Indiana repository instances.
- Ordering of replacement servers for HathiTrust full-text search infrastructure.

Storage Replacement

- Completion of installation of new and replacement storage for 2014.

Updated Volume Identifiers

- HathiTrust made a one-time, batch change to a set of approximately 320,000 volume identifiers. A full list of the updated identifiers is available at http://www.hathitrust.org/hathifiles. Any institu-

tions or individuals that save links to HathiTrust volumes locally should update these identifiers to ensure working links.

Availability

Cumulative 12-month availability of repository access (as of June 1, 2014): 99.867%.

Papers and Presentations

All papers and presentations are listed at http://www.hathitrust.org/papers.

HathiTrust Digital Library

SPECIAL EDITION - 2014 Mid-Year Review

Volumes Added	Jan-June	Overall
Boston College	834	3,197
Columbia University	129	65,165
Cornell University	50,271	487,762
Duke University	3,249	7,774
Harvard University	630	238,065
Indiana University	502	196,082
Keio University	90,080	90,080
Knowledge Unlatched	24	24
Library of Congress	19,159	108,883
McGill University	893	893
New York Public Library	3,424	291,794
North Carolina State University	0	3,196
Northwestern University	18,896	56,398
Ohio State University	26,859	26,859
Penn State University	13,288	81,492
Princeton University	215	251,925
Purdue University	3	44,698
Sterling & Francine Clark Art Institute	358	358
Texas A&M University	0	1,201
Universidad Complutense	139	112,153
University of California	72,464	3,520,634
University of Chicago	12,995	51,630
University of Delaware	28	28
University of Florida	103	9,866
University of Illinois	29,924	142,899
University of Massachusetts	11,115	11,115
University of Michigan	23,040	4,689,072
University of Minnesota	4,245	120,180
University of North Carolina - Chapel Hill	0	17,025
University of Virginia	381	51,202
University of Wisconsin	1,328	557,252
Utah State University	0	117
Yale University	0	23,678
Total	384,576	11,262,697

Public Domain (~34% of total)

Total*	306,317	3,848,472

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

HathiTrust Digital Library

SPECIAL EDITION - 2013 Mid-Year Review

User Support Issues

User Support Issues	Jan-June 2014	Jan-June 2013
Content	**960**	**1,057**
Quality	881	997
Collections	75	60
Cataloging	**969**	**605**
Access and Use	**936**	**822**
Copyright	608	463
Permissions	58	68
Takedown	5	2
Print on Demand	1	2
Inter-library loan	8	10
Full-PDF or e-copy requests	113	119
Datasets	20	34
Data Availability and APIs	9	7
Reuse of content	20	23
Web applications	**124**	**137**
Functionality problems	55	41
Problems with login specifically	10	6
General questions about login	8	12
Partners setting up login	5	10
Usability issues	0	13
Feature requests	10	11
Partner Ingest	**50**	**35**
General	**510**	**380**
Partnership	55	65
Miscellaneous	455	313
Total	**3,549**	**2,976**

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

Most-accessed volumes

Jan-June 2014

Title
Quintus Curtius [History of Alexander], Vol. 1, with an English translation by John C. Rolfe.
The Human Figure, by John H. Vanderpoel
Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.
Quintus Curtius [History of Alexander], Vol. 2, with an English translation by John C. Rolfe.
Godey's magazine. v.40-41 1850
Quicksand, by Nella Larsen.
History of wages in the United States from Colonial times to 1928, United States Department of Labor.
A family tour from ocean to ocean : being an account of the first amateur motor car journey from the Pacific to the Atlantic, whereby J.M. Murdock and family, in their 1908 Packard "Thirty" touring car, incidentally broke the transcontinental record, by J.M. Murdock.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.2.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.1.

About HathiTrust

HathiTrust is an international partnership of academic and research institutions dedicated to ensuring the preservation and accessibility of the vast record of human knowledge. The partnership owns and operates a digital repository containing millions of public domain and in-copyright volumes, digitized from partnering institution libraries and other sources. The preserved volumes are made available in accordance with copyright law as a shared scholarly resource for students, faculty, and researchers at the partnering institutions and as a public good to the world community. For more information, visit HathiTrust.org. You can follow HathiTrust on Facebook and Twitter.