



Update On April Activities

May 14, 2015

Late Breaking News

Research Center Releases Important Dataset

The HathiTrust Research Center has released the Extracted Features Dataset, derived from 4.8 million public domain volumes in the HathiTrust collection. The release will support analysis of large worksets of volumes in the HathiTrust public domain collection, at scales previously intractable for most individual researchers. For example, page-level token (word) counts, can be used to help build topic models, classifications and perform other text analytics. <http://www.hathitrust.org/htrc-releases-massive-dataset>

Top News

Spring Board of Governors Meeting

The Board of Governors held its spring 2015 meeting on May 1 in Berkeley, CA. In addition to regular updates, the Board discussed and took action on the following matters:

Print Monograph Archive Planning: The Board reviewed the recommendations of the [Print Monograph Archive Planning Task Force](#) and the cover report and recommendations of the Program Steering Committee. The Board commended the Task Force on their excellent report and discussed how the initiative would be implemented. A summary of the reports recommendations is being prepared for comment by the broader library community, and is expected to be available in advance of ALA.

Government Documents: To clarify the scope of the initiative, the Board has officially retitled it the US Federal Documents Initiative. The Board gave final budget approval to hire a program officer to oversee this initiative.

Staffing: The Board gave final budget approval for a staff position in the Executive Director's office to support user and member services, documentation, and project management.

Budget and Strategic Planning: In preparation for the 2016 budget process, the Board discussed principles for long-range financial planning and management.

Membership strategy: The Board discussed several membership inquiries and the current criteria for membership in HathiTrust. Mike Furlough was tasked to consult with a small group of directors to be named on membership strategies, and with drafting updated criteria for consideration by the Board and eventually the membership.





Update On April Activities

May 14, 2015

New Blog Post Highlights Efforts to Improve Quality

Jeremy York and Kat Hagedorn have written a [blog post](#) explaining how HathiTrust addresses reported problems with digitized volumes in HathiTrust.

Nominating Committee Named

Appointees to the 2015 HathiTrust Nominating Committee have been named:

- Alberta Comer, Dean of the J. Willard Marriott Library, and University Librarian, University of Utah
- Robert Gerrity, University Librarian, University of Queensland
- Lorraine Haricombe, Vice Provost and Director of Libraries, University of Texas Austin
- Karen Williams, Dean of University Libraries, University of Arizona

The 2015 Nominating Committee will be chaired by past chair of the Board Sarah Michalak, Associate Provost and University Librarian, University of North Carolina Chapel Hill.

The HathiTrust Nominating Committee has responsibility for soliciting nominees for the Board of Governors and candidates for the Program Steering Committee. In fall 2015 HathiTrust will hold its first regular election for new Board members since initiating the current governance model in 2012.

Program Steering Committee

In April, the Program Steering Committee focused primarily on review and analysis of the report and recommendations of the [HathiTrust Monograph Archive Planning Task Force](#). PSC forwarded its own cover report and recommendations to the Board of Governors for consideration at their May 1 meeting. PSC has also continued work on several major issues identified last fall: 1) developing a framework to plan for new collection formats 2) improving quality validation and assessment 3) improving metadata quality and policy development 4) creating a framework for development proposals.

User Support Working Group Nominations

The User Support Working Group is seeking nominations for up to 2 new members. We are seeking staff who have expertise in providing general user support and those who have expertise in cataloging in particular. The nomination period has been extended to May 22, 2015. To submit nominations and for further information about the working group, please visit <http://tinyurl.com/m9qlyyg>.

HathiTrust Member Update Webcasts

HathiTrust will host two webcasts this summer to provide an update on current activities and member services. All staff from member libraries, especially libraries that have recently joined HathiTrust, are encouraged to attend. Registration details and dates will be announced by early June.

HathiTrust on the Road

Staff from HathiTrust will be attending the following meetings and conferences in June. Please be in touch if you would like to meet with us during our travels:

- American Libraries Association
2015 Annual Meeting, San Francisco, CA, June 26-30, 2015
- Mike Furlough, Executive Director
 - Jeremy York, Assistant Director

There's an
elephant in
the library.™





Update On April Activities

May 14, 2015

HathiTrust Participates in Planning Grant for Services to Students with Disabilities

Mike Furlough, Executive Director, and J. Stephen Downie, Co-director of the HathiTrust Research Center, will serve on the steering committee of an IMLS planning grant awarded to Tufts University. Titled “Repository Services for Accessible Course Content,” the project will be led by Larua Wood, Director of Tisch Library, Tufts University, and John Unsworth, University Librarian and CIO, Brandeis University. Over the course of one year, this planning project will bring together experts from disability/accessibility services with librarians, IT professionals, advocates, and legal counsel, to develop shared infrastructure within which universities can support their students with disabilities. HathiTrust currently provides students who have a print disability and who are enrolled at a member institution with access to in-copyright works in the collection.

You can follow HathiTrust on [Twitter](#) or [Facebook](#)
[Subscribe to email updates](#)
(via Google Groups)

Ingest

Internet Archive

Staff worked with Duke University, Washington University, and Columbia University. Tufts University successfully submitted their first batch of content for ingest.

Locally-digitized Content

Staff worked with Boston College, Virginia Tech, Northwestern University, Cornell University, Texas A&M, and Princeton University to resolve questions. Content was ingested from University of Illinois, Urbana Champaign. University of Washington successfully ingested their first batch of content.

Bibliographic Data Management

The California Digital Library (CDL) loaded 78,778 new, and 43,357 update records.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in March is given below. See [CRMS-US](#) and [CRMS-World](#) for further information. The CRMS projects are funded by the Institute for Museum and Library Services.

	April		Overall	
	Public Domain Determinations	All Determinations	Public Domain Determinations	All Determinations
CRMS-US	792	1,154	171,047	323,129
CRMS-World	4,188	7,594	105,105	197,495
Total	4,980	8,748	276,152	520,624

There's an elephant in the library.™

www.hathitrust.org



Update On April Activities

May 15, 2015

US Federal Documents Registry

As of April 30th there are 621,188 federal publications in HathiTrust.

Over 7.8 million records are currently included in the Registry and testing continues on algorithms to identify related items. An alpha version of the US Federal Documents Registry will be available in June, and staff will be seeking feedback on initial functionality and refining potential use cases for the Registry.

Two University of Washington iSchool students have been working with project staff since January on approaches to the manual review of record pairs identified by the relationship detection process as being potentially related. In April they began reviewing record pairs and making decisions as to whether or not the records were for duplicate items.

A [report on activities from the past six months](#) is now available.

HathiTrust Research Center

The Research Center has released the Extracted Features Dataset (v.0.2). View the full update here: <http://www.hathitrust.org/htrc-releases-massive-dataset>

The Research Center held its monthly user group meeting on April 30, 2015. Sayan Bhattacharyya presented the HTRC + Bookworm project and demoed the online interactive system to the participants, also fielding questions and feedback from the HTRC user base.

HTRC UnCamp 2015 was featured and overviewed in a blog of DLF (Digital Library Federation), written by three staff at University of Michigan Library. <http://www.diglib.org/archives/8289/>

J. Stephen Downie traveled to Brown University, Bryn Mawr College, Haverford College and Swarthmore College to present generally on SHARC services and their role in instruction in the classroom.

Volumes Added

Ingest numbers for April can be found on our website here: http://www.hathitrust.org/statistics_deposited_volumes_monthly

Collection statistics are updated daily here: http://www.hathitrust.org/visualizations_deposited_volumes_current

There's an
elephant in
the library.™





Update On April Activities

May 14, 2015

Development Updates

Development updates and activities by HathiTrust institutions included the following:

Full-text Search

- Re-indexing using additional new hardware was begun in April and is expected to be available in production in May. When completed this will result in improved search performance.
- Staff prototyped a method of relevance ranking using “balanced interleaving,” which allows for comparison of different weights in metadata fields and OCR.

Infrastructure

- Upgraded load balancers to better support current HTTPS best practices
- Changed ingest reports and ingest logs to be listed by content provider and digitization source instead of by namespace

PageTurner

- Content provider information is now sent to Google Analytics during item access.

Papers and Presentations

Loretta Auvil. HathiTrust + Bookworm. Panel at DPLAFest 2015. Apr 18, 2015. Indianapolis, IN.

Sayan Bhattacharyya and Harriett Green. “[The HathiTrust+Bookworm tool for lexical trend discovery](#)”. Workshop, [Scholarly Commons](#), University of Illinois at Urbana-Champaign Library, April 29, 2015.

Boris Capitanu, Ted Underwood, Peter Organisciak, Sayan Bhattacharyya, Loretta Auvil, Colleen Fallaw and J. Stephen Downie. ‘[Extracting features from text for non-consumptive reading with the HathiTrust Research Center](#).’ Graduate School of Library and Information Science Research Showcase, University of Illinois at Urbana-Champaign. April 3, 2015.

Dirk Herr-Hoyman. [SHARC: Secure HathiTrust Analytics Research Commons](#). University of Wisconsin Digital Humanities Plus Art: Going Public, 2nd Annual Conference. Apr 17, 2015. Madison, WI.

May Development Forecast

Put Solr plug-in to reduce memory use into production and complete the process of full-text reindexing.

Continue work on a test framework for relevance ranking, including interleaving of search results for the comparison of ranking algorithms.

Add social sharing options to PageTurner and Collection Builder.

There’s an
elephant in
the library.™





Update On April Activities

May 14, 2015

Papers and Presentations Continued

Mike Furlough. “[HathiTrust Partner Update](#),” Indiana University, Bloomington, IN, April 16, 2015.

Mike Furlough (panel member). “Gaps in Distributed National Digital Capacity,” IMLSFocus Strategic Priorities 2015: The National Digital Platform, Washington, DC, April 28, 2015

Mike Furlough (panel member). “Shared Print Repositories: Partnerships and Scalable Solutions”, ARL Spring 2015 Membership Meeting, Berkeley, CA, April 29, 2015

Robert McDonald. The HathiTrust Research Center: An Overview of Advanced Computational Services. DPLAFest 2015. Apr 18, 2015. Indianapolis, IN. <http://www.slideshare.net/rhmcDonald/the-hathitrust-research-center-an-overview-of-advanced-computational-services>

Heather Christenson. “[HathiTrust: UC Collections & Services](#)”. Webinar, University of California Libraries Advisory Structure series. May 12, 2015. Oakland CA.

Jeremy York. “[HathiTrust Digital Repository: Under the Hood](#)”, Guest lecture in University of Michigan School of Information course on digital preservation, April 20, 2015.

J. Stephen Downie. “[HathiTrust Research Center: Your Analytic Gateway to the HathiTrust’s 4.5 billion pages](#)”, organized by Tri-Co Digital Humanities.

Availability

Repository: Cumulative 12-month availability of repository access*: 99.975% (+.004%)

* 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 April Activities

May 14, 2015

User Support Issues

	April	March
Content	156	148
Quality	140	144
Collections	16	13
Cataloging	158	154
Access and Use	133	130
Copyright	73	65
Permissions	14	9
Takedown	2	0
Print on Demand	0	0
Inter-library loan	4	2
Full-PDF or e-copy requests	29	29
Datasets	0	2
Data Availability and APIs	4	0
Reuse of content	4	3
Web applications	47	47
Functionality problems	15	27
Problems with login specifically	1	1
General questions about login	3	1
Partners setting up login	0	1
Usability issues	0	0
Feature requests	2	0
Partner Ingest	15	10
General	115	138
Partnership	17	6
Miscellaneous	98	132
Total	607	637

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

Most-accessed volumes

- [Quicksand, by Nella Larsen.](#)
- [Modern California Houses: Case Study Houses, 1945-1962, by Esther McCoy.](#)
- [Design of Equilibrium Stage Processes, By Buford D. Smith.](#)
- [Godey's Magazine, v. 40-41, 1850.](#)
- [Solid Mensuration, by Willis F. Kern and James R. Bland.](#)
- [The Lesson of Japanese Architecture, By Jiro Harada.](#)
- [War in New Guinea, Official War Photographs of the Battle for Australia.](#)
- [The Human Figure, by John H. Vanderpoel](#)
- [Quintus Curtius \[History of Alexander\], Vol. 1, with an English translation by John C. Rolfe.](#)
- [Roster of the Confederate soldiers of Georgia, 1861-1865, v.3.](#)

