HathiTrust Digital Library

Update On November/December Activities — February 26, 2016

Top News

March Webcast: Copyright Reviews and Access in HathiTrust

HathiTrust's collections include over 5.4 million volumes that are open for reading. A substantial number of these volumes are open because of cooperative copyright investigations conducted by staff as part of the Copyright Review Management System project (CRMS). Join Mike Furlough and Kristina Eden on March 16 at 3:30pm ET to learn more about the factors that govern access to volumes in HathiTrust and how the copyright review program works. All staff from member libraries are encouraged to attend. Register here. Once registered, attendees will be provided with information to access the webcast.

2015 Member Meeting

HathiTrust held its second annual Member Meeting on December 9th at the Big Ten Center in Chicago. Ninety-four attendees represented eighty-four member institutions at the all-day event. Slides of all presentations, as well as community-drafted notes, can be found here.

Executive Director Mike Furlough reported on 2015 progress, noting that HathiTrust added its 5 millionth open book and now contains 110 members. We expect to add several new members in early 2016 and to reach 14 million total volumes during the year. Bob Wolven, chair of the Program Steering Committee (PSC), explained that the committee had spent the year focused on metadata strategy and policy, quality assessment, and discussed the viability collecting of non-text formats. Working groups had been put in place on the first two of these items, while the Collections Committee surveyed the membership on non-text formats in the fall.

In addition to continuing work on those matters, Wolven reported that in 2016 PSC will work with HathiTrust staff processes to give the membership more direct opportunities to advise on the development agenda for HathiTrust. Furlough highlighted several areas of focus for the coming year, including the expansion of the federal documents initiative, startup of the shared print monograph archive program, and a focus on improving services for users with print disabilities and ingest of materials. The Collections Committee expects to have completed its analysis of the survey results in early 2016 and will publish the results.

HathiTrust Research Center Co-Director Stephen Downie reported on the development of the Research Center's services, including the release of a large-scale dataset of "extracted features," the beta release of the Bookwork+HathiTrust service, and a competitive program to award staff/support time to specific research projects.

Executive Director Mike Furlough

PSC Chair Bob Wolven

Update on November/December Activities
February 26, 2016

A session titled "HathiTrust at Your Library" featured representatives from ten institutions giving lighting talks to share how they using HathiTrust to improve or expand services for their users.

- "HathiTrust at Georgia Tech" by Jeffrey Carrico, Georgia Tech
- "Speaking the Same Language: Why NYPL Copied Rights Codes from HathiTrust" by Greg Cram, New York Public Library
- "Updating HathiTrust Links in WorldCat" by Joseph Hafner, McGill University
- "Scholar's Commons and HathiTrust Tools and Services" by Robert McDonald, Indiana University
- "Bringing Buried Treasure to Light: Creating article level discovery metadata for HathiTrusts resources" by Michelle Paolillo, Cornell University
- "Enhanced Access to HathiTrust for Patrons with Disabilities: Experiences of the State University Libraries of Florida" by Ben Walker, University of Florida
- "Book Inventory Management System for HTSPMP" by John Wang, University of Notre Dame

Board member Anne Kenney led a discussion about the shared print monograph archive program, soliciting feedback from members on the value of the program to their institutions, the potential costs and benefits of the program, and the services that could be implemented to insure the success of the program. Some audience members compared experiences in developing other shared print programs. Others highlighted the potential cost as a concern, and wondered what infrastructure and labor would be required to run the program, and how well that could be distributed. Several representatives voiced their hope that the program, once implemented, would prompt renewed national-level discussions to plan and coordinate shared print programs at the regional level.

The membership also received a report on the 2016 budget and fees, and unanimously voted to change the bylaws governing selection of the Program Steering Committee Chair. The board may now select the chair from the PSC membership or newly appoint a chair, rather than assigning a Board member to the committee.

The date for the 2016 Member Meeting will be announced in the spring.

2015 Fall Board of Governors Meeting
The 2015 Fall Board Meeting was held December 10th, 2015 at the Big Ten Center in Chicago. The Board reviewed the following matters and took actions as noted.

Collections Survey: The board reviewed preliminary results from the Collection Priorities Survey issued by the Collections Committee in Fall 2015 and requested a more formal analysis to be delivered in early 2016.

Greg Cram, New York Public Library

Joseph Hafner, McGill University

There's an elephant in the library.™

www.hathitrust.org

Program Steering Committee Chair: The Board discussed possible candidates to replace Bob Wolven as chair of the Program Steering Committee. Wolven and Executive Director Mike Furlough were asked to work on the matter further so that a new chair would be in place by Spring 2016.

2016 Board treasurer and chair-elect: The Board elected Wendy Lougee, Univerisity of Minnesota, to the position of treasurer and chair-elect for 2016.

Finances and budget: The Board reviewed five-year budget projections and discussed the impact of 2016 fee increases on members. Wendy Lougee and Mike Furlough will define the process for assessing HathiTrust's financial model and formula during 2016.

The meeting also included an update on the HathiTrust Research Center's programs and services by Stephen Downie and Robert McDonald, Indiana University. Mike Furlough updated the Board on several topics, including recruitment, the federal documents and shared print initiatives, services for users with print disabilities, and copyright review.

New Board of Governors Membership

The HathiTrust Board of Governors includes several new members this year. As previously announced, Beth McNeil, University of Iowa, Winston Tabb, Johns Hopkins University, and Anne Kenney, Cornell University were all elected to terms beginning January 1, 2016.

John Culshaw, University Librarian at the University of Iowa, was appointed to the Board by the member schools of the Committee on Institutional Cooperation. Culshaw takes the seat left vacant by Carol Pitts Diedreich, who recently retired as Vice Provost and Director of University Libraries at The Ohio State University.

The full roster of the 2016 HathiTrust Board of Governors is as follows:
- Ivy Anderson, California Digital Library (interim, appointed)
- Richard Clement, University of New Mexico, past chair 2016 (term ends December 2016)
- John Culshaw, University of Iowa (appointed)
- James Hilton, University of Michigan (appointed)
- Anne Kenney, Cornell University (term ends December 2018)
- Wendy Lougee, University of Minnesota, chair-elect and treasurer 2016 (appointed)
- Beth McNeil, Iowa State University (term ends December 2016)
- Brian Schottlaender, University of California, San Diego (appointed)
- Winston Tabb, Johns Hopkins Universitiy (term ends December 2018)
- Carolyn Walters, Indiana University (appointed)

- Lizbeth (Betsy) Wilson, University of Washington, chair 2016 (term ends December 2017)
- Bob Wolven, Columbia University (term ends December 2017)

The 2016 Executive Committee members are:
- Lizbeth (Betsy) Wilson, University of Washington, chair 2016
- Wendy Lougee, Universitiy of Minnesota, chair-elect and treasurer 2016
- Richard Clement, University of New Mexico, past-chair 2016
- Robert Wolven, Columbia University, Program Steering Committee Chair

2016 Budget and Fees

The Membership approved the 2016 HathiTrust budget via electronic ballot on December 21. The 2016 budget includes a 6.5% increase in total amount of member fees collected. Factors contributing to the increase include the addition of new staff positions to solidify operations, the addition of payments to member libraries that manage HathiTrust operations, and increased costs for storage backups. In 2016 each member pays $10,855 to support the preservation of public domain and open access items in HathiTrust. Members pay a variable amount to support preservation of in-copyright materials in the collection, which is based on the member's collections and their overlap with HathiTrust.

Betsy Wilson, Chair-Elect and Treasurer 2015

The HathiTrust budget includes expenses in two major categories: operations and programs. Operations expenses relate to the administration and core preservation and access services of HathiTrust, such as the cost of data storage and backup, data centers, servers, contracted services, travel, office expenses, and staff. Programmatic expenses support activities and initiatives that allow us to pursue new or programs or short-term projects that extend the value of the collection and provide added benefit to the members. These include, for example, the initiative to expand and enhance the US federal government documents, the initiative to establish a distributed network of print monograph archives, support for the HathiTrust Research Center, and copyright review.

A detailed description of the pricing model is available at https://www.hathitrust.org/cost. Partners communicate the volumes held in their print collections through print holdings data submitted to HathiTrust (see https://www.hathitrust.org/print_holdings). While the 2016 budget reflects a 6.5% increase in total fees for 2016, the fees of individual members increased by differing percentages, due to the fact that the fee model is based on the profile of each member's collection.

The Board of Governors will spend time in 2016 assessing the current financial model and report back to the membership by end of year.

Inputs for 2016 Fees

Date Calculated: October 1, 2015
Total volumes: 13,737,592
Total Pub domain volumes: 5,361,303
Total per volume cost: $0.2207
Number of members: 110
Each member's PD fee: $10,855

There's an elephant in the library.™

www.hathitrust.org

Recruiting

HathiTrust has been conducting searches for three new positions: Director of Services and Operations, Program Officer for Federal Documents and Collections, and Program Officer for Shared Print Initiatives. Results of these searches will be announced before early spring.

HathiTrust Research Center

On December 8, 2015, Andy Patterson and Inna Kouper, of Indiana University, won the Indiana University School of Informatics and Computing 2015 Fall Projects and Research Symposium Award for Best Undergraduate Research Project for their work on "HTRC Visualization," a visualization of publication metadata from the HaithiTrust database of published works. Finding meaningful trends in a large corpus of big data.

Robert McDonald, HTRC Exec. Committee

Looking Ahead for HTRC

The entire HTRC team is committed to an ongoing program of constant improvement of our tools and the delivery of our services. Thus we will be putting substantial effort into large-scale improvements of our services this year, with focus on Workset Builder and Data Capsule, and the tools that our users can employ in their research on the HathiTrust corpus.

Additionally, HTRC plans to reinvigorate its translational research efforts by working closely with new HathiTrust staff, the Advisory Board and other HTRC stakeholders. We also plan to explore HTRC's new feature extraction, metadata creation, Bookworm visualizations and linked open data efforts could play in the enhancement of HathiTrust services beyond those intended for analytic researchers.

From the HathiTrust Bookworm project

Ingest

HathiTrust paused most ingest activity at the end of 2015 to focus on planning for a new storage upgrade to be completed in early 2016.
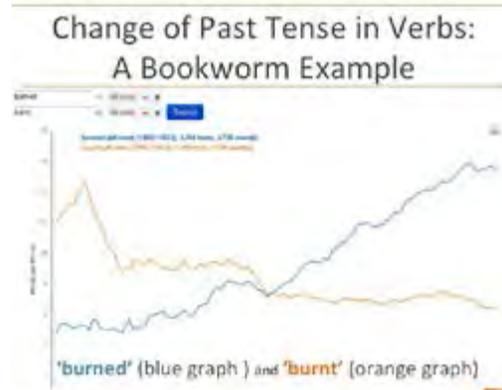
Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in September/October is given below. See CRMS-US and CRMS-World for further information. The CRMS projects are funded by the Institute of Museum and Library Services.

	December		Overall	
	Public Domain Determinations	All Determinations	Public Domain Determinations	All Determinations
CRMS-US	310	435	176,390	330,350
CRMS-World	2,704	6,273	130,374	244,469
Total	3,212	6,984	306,168	573,989

US Federal Documents Registry

The US Federal Documents Registry is currently available in alpha release at at https://www.hathitrust.org/usdocs_registry/. There are 6,258,658 records in the Registry, derived from over 25 million records contributed by more than 50 libraries. A complete list of contributors can be found on the About the Registry page.

Work continues to refine duplicate detection based on identifiers (OCLC number, ISSN, SuDoc number, etc). This has been hampered somewhat about data quality issues, most notably that different libraries and different integrated library systems store information in a variety of locations. The project team has begun working on the automated matching of records with similar titles but no common identifiers. This work will continue in spring 2016, and will continue to be refined.

Plans for early-mid 2016:

We anticipate moving the Registry to beta in first half of 2016 - this will include a single Registry record with a unique identifier; a more accessible user interface; and ongoing updates from the HathiTrust repository. We also plan to move forward with analysis, identifying needed metadata as well as those items which have records in the Registry but have not been digitized.

In late spring 2016, project staff will undertake an initial assessment of the Registry, based partially on current use cases, previously established success criteria, and user feedback.

About the Federal Documents Initiative

Ballot Initiative passed at the 2011 HT Constitutional Convention

Focus: expanded coverage & enhanced access to U.S. Federal documents.

Near term activities:
- Bibliographic and collections analysis: Developing a registry of US Federal Government Documents
- Digitze! Focus first on known and cataloged materials: Gap analysis driven; prioritize print, post-1976 materials; Identify collections for inclusion (and get them)
- Publicize the efforts: Within the library community and the general public

Development Updates

Full-text Search

Work was initiated on a unified logging and log analysis framework for HathiTrust applications. An assessment of the ability of the current logging programs to log data from which evidence of successful and unsuccessful searches, user tasks, and relevance can be derived was performed. A first round of modifications was made to the current logging and log analysis programs. A number of approaches to analyzing sequences of user actions were investigated.

There's an elephant in the library.™

www.hathitrust.org

Update On November/December Activities
February 26, 2016

PDF Downloads

Began research into improving the accessibility of downloaded PDFs. The initial outcome of this work has been deployed which improves the use of downloaded PDFs with audio readers (e.g. Adobe Read Aloud).

Storage

The storage replacement and expansion strategy for 2016-2019 was completed and approved by the HathiTrust Director. HathiTrust will now move to a four-year replacement cycle for data storage, and will replace all storage once every four years to gain more favorable pricing and reduce costs passed along to members. The purchase was executed, and the new equipment was received at both locations before the close of 2015. Staff will work in the first quarter of 2016 to bring the new equipment online and retire old equipment.

Papers and Presentations

Publications

Underwood, Ted. "How Scholars Can Support Digital Libraries" *Europeana Research.* November 16, 2015.

Presentations

Shamim, Muhammad Saad and Sayan Bhattacharyya. "Culturomics: New Developments in Analyzing Digitized Texts." Rice University Digital Humanities Group. November 9, 2015.

Bhattacharyya, Sayan. "The HathiTrust Research Center's Extracted Features Dataset: An Opportunity for 'Distant" Reading of Millions of Books from the World's Great Research Libraries." Part of "Big Data Case Studies" panel, *Big Data Summit 2015.* Research Park, University of Illinois at Urbana-Champaign. November 11, 2015. Slides.

Bhattacharyya, Sayan, Boris Capitanu, Peter Organisciak, Loretta Auvil, Colleen Fallaw and J. Stepen Downie. "Big Textual Data in Undergraduate Student Writing for Literature Courses: Affordances of the HathiTrust Research Center's Extracted Features Dataset." 2015 Chicago Colloquium on Digital Humanities & Computer Science (DHCS 2015). University of Chicago. November 13-15, 2015. Abstract

Dickson, Eleanor and Sayan Bhattacharyya. "Using the HathiTrust Research Center's Tools for Text Analysis." 2015 Chicago Colloquium on Digital Humanities & Computer Science (DHCS 2015). University of Chicago. November 15, 2015.

Repository Availability

Cumulative 12-month availability of repository access: 99.975% (-/+0.000%).

You can follow HathiTrust on Twitter or Facebook

Subscribe to email updates (via Google Groups)

Volumes Added

Ingest numbers and Collection statistics are updated daily and can be found on our website here: https://www.hathitrust.org/visualizations_deposited_volumes_current

Update On November/December Activities
February 26, 2016

Bhattacharyya, Sayan. Class session in Prof. Christi Merrill's class 'Comparative Literature 322: Writing World Literatures' on Nov 19, 2015, at the University of Michigan, Ann Arbor, showcasing the use of HT+Bookworm. Slides , Pre-session blog post, Post-session blog post

Bhattacharyya, Sayan. "Text Analysis with the HathiTrust Research Center." Workshop, part of the University of Michigan Digital Scholarship Series. University Library Instructional Center (ULIC), Shapiro Undergraduate Library, University of Michigan, Ann Arbor. Nov 20, 2015.

HathiTrust on the Road

HathiTrust staff will be attending the following events in early 2016. Please contact us if you wish to meet us at any of these events:

CRL 2016 Global Resources Collections Forum, Chicago, IL, April 14-15 - Mike Furlough

DPLAfest 2016, Washington, D.C., April 14-15 - Kristina Eden and Angelina Zaytsev

Open Scholarship Initiative 2016, Fairfax, VA, April 19-22 - Mike Furlough

ARL Spring Membership Meeting, Vancouver, BC, Canada, April 26-28 - Mike Furlough

Beth Plale, HTRC co-director and chair, presenting "HTRC Visualization"

Tracey Berg-Fulton
@BergFulton

Beautiful paper marbling in this book "Les anciennes armées françaises" via HathiTrust
babel.hathitrust.org/cgi/pt?id=mdp….

View translation

Tweet of the Month: https://twitter.com/bergfulton/status/677603361261228032

JF Ptak
@ptak

From NC to CA on a train, w/o a ticket, #Hobo 1903 Full txt via @hathitrust ow.ly/Wqz7B #books

FROM NORTH CAROLINA TO SOUTHERN CALIFORNIA

WITHOUT A TICKET AND HOW I DID IT

GIVING MY EXCITING EXPERIENCES AS A "HOBO"

Tweet of the Month: https://twitter.com/ptak/status/681897450882334720

BY JOHN PEELE

There's an elephant in the library.™

www.hathitrust.org

Update On November/December Activities
February 26, 2016

Most-accessed volumes Nov/Dec 2015

Quicksand, by Nella Larsen.
Solid mensuration, by Willis F. Kern and James R. Bland
A short guide to New Zealand.
History of the descendants of John Hottel (immigrant from Switzerland to America) and an authentic genealogical family register of ten generations from the first of the name in America, 1732, to the present time, 1929, with mumerous brief biographical sketches, collected and compiled from many indisputable sources: court and church records, old and late family records and tombstones of the many states in the Union. Begun by Rev. W. D. Huddle, B.S., and completed by his wife, Lulu May Huddle.
Descendants of Governor William Bradford (through the first seven generations) / compiled by Ruth Gardiner (Mrs. Francis C.) Hall under auspices of Bradford Family Compact.
Locomotive cyclopedia of American practice, 1950-52 : definitions, drawings and illustrations of diesel, steam, electric and turbine locomotives for railroad, industrial and foreign service; their parts and equipment; descriptions and illustrations of locomotive shops and servicing facilities / compiled and edited for the Association of American Railroads, Mechanical Division ; editor, C.B. Peck.
Harper's bazaar v.40 1906.
The Saturday evening post. v.195 1922 Oct-Dec.
Economic Concentration. : hearings before the United States Senate Committee on the Judiciary, Subcommittee on Antitrust and Monopoly, Eighty-Eighth Congress, second session, Eighty-Ninth Congress, Ninetieth Congress, Ninety-First Congress.
Annual report on transport statistics in the United States, 1956.

User Support Issues	Nov-Dec	Sept-Oct
Content	**249**	**281**
Quality	229	248
Collections	19	27
Cataloging	**250**	**233**
Access and Use	**253**	**231**
Copyright	87	89
Permissions	28	14
Takedown	2	7
Print on Demand	0	1
Inter-library loan	6	2
Full-PDF or e-copy requests	64	65
Datasets	5	2
Data Availability and APIs	2	3
Reuse of content	8	9
Web applications	**51**	**41**
Functionality problems	17	11
Problems with login specifically	7	4
General questions about login	1	2
Partners setting up login	1	0
Usability issues	1	1
Feature requests	6	4
Partner Ingest	**37**	**35**
General	**190**	**218**
Partnership	19	16
Miscellaneous	171	202
Total	**1030**	**1039**

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