HathiTrust Digital Library

Update On July/August Activities

September 2, 2015

Top News

HathiTrust is Hiring

HathiTrust has opened searches for three positions to oversee core operations and advance our programs to expand access to US federal documents and to develop the shared print monographs archiving program.   Full details and requirements of each position are posted at: https://www.hathitrust.org/jobs

HathiTrust Member Meeting to be held December 9 in Chicago

The 2015 HathiTrust Member Meeting will be held on December 9 at the Big Ten Center near O'Hare Airport in Chicago. We ask all official member representatives to plan to attend or to designate an attendee who can, if needed, vote on behalf of their institution. Following the model of the 2011 Constitutional Convention, library directors from consortia that are HathiTrust members may also attend, although only the member representative for the consortia may vote.

We plan to begin at 9am and conclude by 3pm, with a continental breakfast available at 8am. A detailed agenda will be sent to registered attendees in advance of the meeting. We expect the meeting to include an update on strategic initiatives, proposed by-laws changes, and an in-depth discussion of the planned Shared Print Monograph Archiving program, as well as finances, the legal landscape, and current priorities.

There is no charge to attend the meeting, but we will ask attendees to RSVP by November 25, 2015. Registration information will be sent to member representatives.   A block of rooms are available for attendees at the Aloft Chicago O'Hare, adjacent to the Big Ten Center, and other hotels are located within walking distance of the Big Ten Center. Please contact Melissa Stewart (mmstewa@ hathitrust.org) if you have questions.

Board of Governors Update

The HathiTrust Board of Governors held their summer meeting by telephone on July 29. Executive Director Mike Furlough provided the Board with a mid-year budget report, an update on planning for the 2016 budget, and a progress report on the Federal Documents and Shared Print Monograph Archiving Initiatives. Bob Wolven reviewed recent work of the Program Steering Committee, which he chairs. The Board took action to approve plans for the 2015 Member Meeting, the 2015 Board of Governors election process and schedule, and proposals to fill a position to support the Shared Print Initiative, and also appointed two new members to the Program Steering Committee. In addition, the Board discussed factors to be considered in developing formal membership criteria.

1

Board of Governors Elections to Begin Soon

The HathiTrust membership will elect three new members to the HathiTrust Board of Governors in an election that will begin on September 28 and conclude on October 19. The Nominating Period closed on August 24 and the final slate of candidates will be announced immediately before the election. For more information about the 2015 Elections Process see: https://www.hathitrust.org/elections2015

New Program Steering Committee Members Appointed

The Board of Governors is pleased to announce the appointment of two new members to the Program Steering Committee, serving two year terms that conclude in June 2017.

- Greg Raschke, Associate Director for Collections and Scholarly Communication at North Carolina State University
- Oya Rieger, Associate University Librarian for Scholarly Resources and Preservation Services at Cornell University and Program Director for arXiv

The Board received an exceptional group of nominations for the PSC. A new round of appointments will be made in spring 2016.

Metadata Policy, Strategy, Use and Sharing Advisory Group Appointed

The Program Steering Committee has appointed the Metadata Policy, Strategy, Use and Sharing Advisory Group (MUSAG), co-chaired by Todd Grappone, UCLA and Martin Kurth, Yale. MUSAG has been charged to help formulate policies for use, distribution, and quality assurance of HathiTrust's metadata assets, and will also advise on the development of strategy for management and development of these assets. The full membership of the Advisory Group includes:

- Tim Cole, University of Illinois, Urbana-Champagne
- Kristina Eden, University of Michigan
- Steven Folsom, Cornell University
- Valerie Glenn, HathiTrust
- Todd Grappone, University of California, Los Angeles, co-chair
- Martin Kurth, Yale University, co-chair
- Patricia Martin, California Digital Library
- Shana McDonald, Georgetown University
- Angelina Zaytsev, HathiTrust

The full charge for the group can be read at:
https://www.hathitrust.org/wg_musag_charge

HathiTrust on the Road (Continued)

IPRH Critical Digital Humanities @ Illinois workshop, Champaign, IL, October 7, 2015 - Harriett Green, Elanor Dickson

Digital Library Federation 2015 Forum, Vancouver, BC, October 26-28, 2015 - Mike Furlough, Angelina Zaytsev, Elanor Dickson, Sayan Bhattacharyya

Educause Annual Conference, Indianapolis, IN, October 28, 2015 - Robert McDonald, Beth Plale, Beth Namachchivaya, Dirk Herr-Hoyman

Update On July/July Activities

September 2, 2015

Ingest

Locally-digitized Content

Cornell University and the Bentley Library (University of Michigan) began ingest of locally-digitized materials. Emory University, University of Maryland, Northwestern University, Columbia University prepared for submission of locally digitized materials. McGill University, University of Delaware, University of Missouri, and Yale University submitted additional materials for ingest.

Google-digitized content

University of Texas has begun work to include their Google-digitized materials in HathiTrust. This would include 500,000 in-copyright materials and 6,000 public domain materials.   The schedule for ingest is still being determined.

Bibliographic Data Management

The California Digital Library (CDL) loaded 96,183 new, and 112,035 update records to Zephir.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in May 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.

	July		Overall	
	Public Domain Determinations	All Determinations	Public Domain Determinations	All Determinations
CRMS-US	1,423	1,486	174,071	327,123
CRMS-World	3,491	6,115	119,741	223,606
Total	4,914	7,601	293,812	550,729

Update On July/August Activities

US Federal Documents Registry

As of August 31, there are 651,200 open US federal documents in HathiTrust.

An alpha version of the US Federal Documents Registry was launched in June. Currently the Registry includes 5,479,188 records, contributed by 42 different libraries. Project staff are currently working on incorporating additional records into the Registry, improving duplicate detection by exploring ways to parse item description information (enumeration and chronology), and identifying records for out-of-scope materials such as state government documents. Feedback on Registry functionality and content is welcomed and encouraged. The Registry may be accessed at http://www.hathitrust.org/usdocs_registry/.

Near-term plans for Registry development include the incorporation of fuzzy text matching into the duplicate detection process, improving the Registry interface by adding an advanced search and displaying additional fields in records, incorporating a mechanism for removing records for out-of-scope materials, and developing better integration with the HathiTrust Digital Library.

Development Updates

Recent development updates and activities by HathiTrust institutions have included the following:

Full-text Search

- Work continued on a testing framework for relevance ranking. Staff consulted with experts in information retrieval evaluation and received valuable feedback. Initial improvements to logging were put into production. An initial prototype for click logging (i.e., logging what the user clicks on in the browser) using "Balanced Interleaving" which allows comparing different relevance ranking algorithms and settings has been completed. Plans are to test the interleaving and click logging code and put it into production in September.

Collection Builder

- A "Share this Collection" URL is now provided for ease of reuse. Additionally, a social toolbar has been added to share collections with popular services.
- Code to improve performance on large collections was put into production in August.

Fall Development Forecast

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

Research ways to support alternative text formats.

There's an elephant in the library.™

www.hathitrust.org

Update On July/August Activities

PageTurner
- A social toolbar has been added to share items with popular services.
- Full book PDF downloads are now logged directly to Google Analytics when the user downloads the final, built PDF.
- Logging improvements were added to PageTurner to enable more detailed analysis of use and to be used in conjunction with the click logging (i.e., logging what the user clicks on in the browser) framework for Full Text search.

Papers and Presentations

Workshops

McDonald, Robert, Jaimie Murdock and Jiaan Zeng. "Topic Exploration with the HTRC Data Capsule for Non-Consumptive Research." Workshop, JCDL15, Knoxville, TN. 21 June 2015.

Furlough, Mike. Panel member: The Shared Print Management and Planning Viewpoint: Present Needs and Potential Areas of Synergy. Preserving America's Print Resources II: a North American Summit, Berkeley, CA 25 June 2015

Bhattacharyya, Sayan. "The HathiTrust Research Center: Large-scale Computational Analysis with the World's First Massive Digital Library." Linguistic Society of America (LSA)'s Biennial Linguistic Institute, The University of Chicago, 13 July 2015. Slides.

Bhattacharyya, Sayan and Eleanor Dickson. "Introduction to the HathiTrust Research Center (HTRC): Teaching and research using the power of data and metadata in large text corpora." Workshop, Humanities Intensive Learning and Teaching (HILT) 2015, 28 July 2015.
Slides (pptx format), Slides (pdf format)

Bhattacharyya, Sayan and Eleanor Dickson. "Advanced Topics in Text Analysis with the HathiTrust Research Center (HTRC)". Workshop, Humanities Intensive Learning and Teaching (HILT) 2015, 29 July 2015.
Slides (pptx format), Slides (pdf format)

Update On July/August Activities

Presentations

Hinze, Annika, Craig Taibe-Schock, David Bainbridge, Rangi Matamua, J. Stephen Downie. "Improving access to large-scale Digital libraries through Semantic-enhanced Search and Disambiguation." Full paper, JCDL 15, Knoxville, TN. 23 June 2015.

Nurmikko-Fuller, Terhi, Kevin Page, Pip Willcox, Jacob Jett, Chris Maden, Timothy Cole, Colleen Fallaw, Megan Senseney, J. Stephen Downie. "Building Complex Research Collections in Digital Libraries: A Survey of Ontology Implications." Short Paper, Joint Conference on Digital Libraries (JCDL)15, Knoxville, TN. 23 June 2015.

Bhattacharyya, Sayan and J. Stephen Downie. "Approaching textuality with the metaphor of the digitized workset." Short paprer, Digital Humanities 2015 (DH 2015) Conference, Sydney, Australia. 29 June - 3 July 2015. Abstract, Slides

Organisciak, Peter, Loretta Auvil, J.Stephen Downie. "Remembering books: A within-book topic mapping technique." Short paper, Digital Humanities (DH) 15 Conference, Sydney Australia. 29 June – 3 July 2015.

Downie, J. Stephen. "Managing Modern Data for Humanities Research, Case Study: HathiTrust Research Center." Digital Humanities at Oxford Summer School (DHOxSS) 2015, Oxford, UK. 24 July 2015.

Furlough, Mike and Zaytsev, Angelina. New Member Webinar, 30 July 2015. Slide Presentation.

Furlough, Mike and Zaytsev, Angelina. New Member Webinar, 5 August, 2015. Slide Presentation.

Repository Availability

Cumulative 12-month availability of repository access: 99.981% (+0.006%).

On Thursday, June 25, from 01:04-01:27 EDT users may have been unable to access HathiTrust due to a problem between one of the search servers and its underlying storage.

On Thursday, August 6, from 15:35-16:42 EDT some users may have had difficulty accessing HathiTrust services due to a network broadcast storm that severely crippled network traffic to services hosted at the Ann Arbor datacenter.

There's an elephant in the library.™

www.hathitrust.org

Update On July/August Activities

September 2, 2015

User Support Issues	July–Aug	May
Content	**286**	**154**
Quality	264	139
Collections	21	15
Cataloging	**265**	**161**
Access and Use	**206**	**140**
Copyright	90	65
Permissions	17	12
Takedown	0	0
Print on Demand	2	0
Inter-library loan	2	0
Full-PDF or e-copy requests	54	21
Datasets	3	4
Data Availability and APIs	0	1
Reuse of content	8	1
Web applications	**74**	**29**
Functionality problems	27	10
Problems with login specifically	11	0
General questions about login	3	0
Partners setting up login	1	0
Usability issues	1	0
Feature requests	5	2
Partner Ingest	**21**	**13**
General	**193**	**97**
Partnership	16	6
Miscellaneous	177	91
Total	**1045**	**594**

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

Most-accessed volumes

Roster of the Confederate soldiers of Georgia, 1861-1865, v.1.

Roster of the Confederate soldiers of Georgia, 1861-1865, v.2.

Roster of the Confederate soldiers of Georgia, 1861-1865, v.5.

Solid mensuration, by Willis F. Kern and James R. Bland.

Roster of the Confederate soldiers of Georgia, 1861-1865, v.3.

Roster of the Confederate soldiers of Georgia, 1861-1865, v.4

The Human Figure, by John H. Vanderpoel

Biographical souvenir of the states of Georgia and Florida : containing biographical sketches of the representative public, and many early settled families in these states.

The Power of Sexual Surrender, by Marie N. Robinson.

Modern California Houses: Case Study Houses, 1945-1962, by Esther McCoy.

There's an elephant in the library.™

www.hathitrust.org