HathiTrust Digital Library

Update On September Activities

Top News

Executive Director Search

The Executive Director Search Committee has been actively working on building the pool of prospective candidates for the Executive Director position. We are very pleased to report that, as of October 1, 26 individuals have expressed interest in the position, individuals who represent the broad spectrum of the information landscape. We are very gratified by the response and expect to begin reviewing applications shortly.

Program Steering Committee Update

The Program Steering Committee held a full-day meeting on September 13th, and developed plans to initiate action in several programmatic areas. New subgroups will be appointed to advance the establishment of a distributed print archive on monographic holdings corresponding to the digital content in HathiTrust, and to expand and enhance access to U.S. federal government publications. These actions address ballot initiatives approved at the HathiTrust Constitutional Convention. Additional subgroups will be appointed to continue the work of the Collections Steering Committee setting priorities for expanding collections, and to define potential initiatives to expand access by clearing rights and incorporating open access publications. In future meetings the Committee will review a proposal for a distributed program to certify the quality of volumes within HathiTrust, and review recommendations for policies regarding metadata use. The Committee has also begun work to develop principles for evaluating and prioritizing proposals for programmatic initiatives.

HathiTrust Research Center UnCamp

HTRC held its 2nd Annual UnCamp at the University of Illinois at Urbana-Champaign Sep 8-9. 134 attendees from 19 US States, Canada, and Costa Rica participated in this 1.5-day event. The participants came from diverse backgrounds: administrators, developers, researchers, students, and librarians, making it a well-blended event for the community. This year's UnCamp featured hands-on access to the HTRC Data API, a newly designed portal for creating and building work sets, and an informational session on metadata enhancement accomplished through mini-grants funded by the Andrew W. Mellon Foundation. Topics highlighted in breakout sessions included computational access to the copyrighted corpus, and ways in which metadata enhancements could be coordinated between HTRC, the HathiTrust repository, and Zephir, a new bibliographic management system for HathiTrust under development by the California Digital Library. Attendees welcomed the volume-level word counts and gender information now available in the HTRC, and requested word counts at the page level, which HTRC is working on. After the UnCamp, a participant wrote and contributed code that facilitates authorization to HTRC services. Ted Underwood of UIUC, as UnCamp came to an end, tweeted a promising high note:
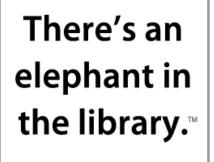
October Forecast

Continue to explore improvements to relevancy ranking.

Work on adding support for indexing of JATS articles.

Complete the development of ePub and PDF generation from JATS.

Papers & Presentations

Please see HTRC UnCamp 2013 for presentations given at the second annual HathiTrust Research Center UnCamp.

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Update On September Activities

"I thought that was a pretty great #HTRC13; it feels like a user community is coalescing; many thanks to @modernmuchness & everyone at IL/IN"

Information about accessing the HTRC production portal and HTRC sandbox can be found on the HTRC Getting Started FAQ.

Ingest

General

HathiTrust corresponded with several institutions about content formats, specifications, ingest package composition, and other issues related to ingest of locally-digitized materials. HathiTrust received bibliographic records for, and answered questions about, deposit of materials digitized by the Internet Archive.

Working Groups and Committees

User Support Working Group

The User Support Working Group created a new Bibliographic Corrections Subgroup and modified workflows for receiving, investigating, and routing inquiries related to bibliographic metadata corrections. This was done in preparation for the introduction of the new bibliographic management system, Zephir, and corresponding changes to the processes for correcting bibliographic records in HathiTrust.
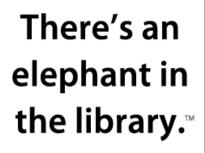
Projects

Bibliographic Data Management

California Digital Library (CDL) and University of Michigan staff achieved parity between Zephir and the current system operated by the University of Michigan; both systems included and were able to output the same bibliographic records. The systems entered a parallel phase, expected to last several weeks, where parity will be confirmed on a daily basis to confirm Zephir's readiness to operate as the production system. Since August, institutions depositing content have been submitting bibliographic records to both the University of Michigan and CDL, and are asked to continue to do so until Zephir goes into production. Please see http://www.hathitrust.org/ingest_checklist for information about submitting records to HathiTrust. Any questions about Zephir or content ingest should be directed to feedback@issues.hathitrust.org.

Government Documents Registry

Project team members held five open focus groups in late September/early October to gather feedback on proposed metadata elements and functionality for the HathiTrust government documents registry. The focus groups were well attended,

with representation from a wide variety of interested groups. The comments received and the discussions that took place were very useful for the project team. A summary of the feedback received in all of the sessions is available at http://bit.ly/16QeM3i. Thank you to everyone who participated!

Copyright Review

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

	September		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	2,386	6,400	148,411	283,343
CRMS-World	2,258	4,445	37,149	69,496
Total	4,644	10,845	185,560	352,839

mPach

Staff at the University of Michigan defined the preservation levels to be associated with content submitted through mPach. Integrated and supplemental materials that meet existing HathiTrust specifications will be preserved at the bit level with format migration. Materials that do not meet specifications will be preserved at the bit level only. Staff continued development of Norm, adding support to convert OpenDocument ("ODT") files to JATS XML and improving support for Unicode.

Development Updates

HathiTrust institutions performed the following work related to applications and Web interfaces:

Data API

Staff sent e-mail notices to registered Data API users about the availability of version 2 of the Data API and associated documentation. Version 1 of the Data API will be taken out of service on November 1, 2013.

Full-text Search

Staff created a production size, page-level index of volumes in HathiTrust to assist with tests in performance and relevance ranking of full-text search results (previously only a volume-level index existed). The page-level index was created in particular to investigate scalability issues related to Solr's grouping functionality. Staff also discussed and designed methods of "chunking" book OCR at indexing-time (indexing chunks other than at the page-level), as a part of experiments to improve relevance ranking.

HathiTrust Digital Library

Update On September Activities

Special Access to In-copyright Materials

Staff discussed ways of improving the efficiency of workflows for managing special access to in-copyright materials. This kind of access is granted on a limited basis for purposes such as copyright review. Staff continued work to supplement existing security measures surrounding in-copyright materials through the identification and tracking of potentially unauthorized accesses (e.g., that might result from a compromised user account).

Outages

No outages were reported in September.

Total Volumes Added	September	Overall
Boston College	0	2,363
Columbia University	2	65,035
Cornell University	848	430,600
Duke University	1	4,524
Harvard University	3	236,072
Indiana University	0	195,349
Library of Congress	0	89,724
North Carolina State University	0	3,196
Northwestern University	768	37,188
New York Public Library	7	288,364
Penn State	12	64,786
Princeton University	4	251,709
Purdue University	0	44,692
Universidad Complutense	14	111,998
University of California	12,008	3,419,334
University of Chicago	3	33,545
University of Florida	2,001	9,587
University of Illinois	1,017	112,151
University of Michigan	3,378	4,657,201
University of Minnesota	611	110,338
UNC -Chapel Hill	3	17,025
University of Wisconsin	56	555,871
University of Virginia	0	50,817
Utah State University	0	117
Yale University	0	23,678
Total	20,736	10,815,264

Public Domain (~32% of total)

Total*	10,641	3,462,764

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

HathiTrust Digital Library

Update On September Activities

User Support Issues	September	August
Content	**243**	**344**
Quality	225	335
Collections	17	6
Cataloging	**169**	**111**
Access and Use	**107**	**183**
Copyright	57	120
Permissions	5	4
Takedown	0	1
Print on Demand	0	0
Inter-library loan	2	0
Full-PDF or e-copy requests	17	21
Datasets	3	4
Data Availability and APIs	1	1
Reuse of content	4	4
Web applications	**22**	**26**
Functionality problems	9	9
Problems with login specifically	3	0
General questions about login	3	3
Partners setting up login	4	0
Usability issues	0	1
Feature requests	2	1
Partner Ingest	**9**	**8**
General	**90**	**64**
Partnership	8	7
Infrastructure	0	0
Miscellaneous	82	57
Total	**640**	**736**

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

Most-accessed volumes

Title
Elementary Catechism on the Constitution of the United States, by Arthur J. Stansbury.
Quicksand, by Nella Larsen.
Health in Africa: a Medical Handbook for European Travellers and Residents, by David Kerr Cross.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.1.
The five laws of library science, by S. R. Ranganathan.
The Rise of the Chinese Empire, Vol. 1, by Chun-shu Chang.
Railroad Gazette, v.7, 1875.
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.2.
Rogers' Inorganic pharmaceutical chemistry, by Charles Herbert Rogers.

There's an elephant in the library.™

www.hathitrust.org