HathiTrust Digital Library

Update On December Activities

Top News

Executive Director Search

The Search Committee for the HathiTrust Executive Director position completed telephone interviews with nearly a dozen top candidates. In January, the Committee will conduct in-person "airport" interviews with a smaller number of semi-finalist candidates.

Approval of HathiTrust 2014 Budget and Fees

Partner institutions approved HathiTrust's 2014 budget and fees with no dissenting votes. Invoices will be sent to partners in January.

HathiTrust and Knowledge Unlatched

We are pleased to announce that HathiTrust will be preserving and providing access to works made available through Knowledge Unlatched, an organization that is "helping stakeholders to work together for a sustainable open future for specialist scholarly books". More information about Knowledge Unlatched can be found at http://www.knowledgeunlatched.org/.

Government Documents Initiative

The deadline for institutions and organizations to submit bibliographic records for US federal government documents to HathiTrust is January 31, 2014. We will be using the records to gain a greater understanding of the total corpus of US federal government documents and the proportion of the documents that have been digitized. This analytical work will support HathiTrust's broader objective to expand and enhance access to US federal government documents. More information about the initiative is available at http://www.hathitrust.org/usgovdocs. Please email feedback@issues.hathitrust.org with any questions. We are very grateful to the institutions that have submitted records so far, and hope to garner as broad participation as possible in this important undertaking.

Nominations for User Support Working Group

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. To submit nominations and for further information about the working group, please visit http://tinyurl.com/m9qlyyg.

Ingest

Validation service for locally-digitized materials

HathiTrust received feedback from partner institutions on a new single-page validation tool, developed to aid in the ingest of locally-digitized materials, and plans

January Forecast

Complete addition of quick links to the PageTurner to embed HathiTrust volumes.

Continue work to add support for the indexing of JATS articles and indexing volumes in chunks.

Continue development of ePub and PDF generation from JATS.

Continue exploring relevance ranking solutions.

HathiTrust Digital Library

Update On December Activities

to release a full-volume validation tool in January. If you are interested in receiving updates related to these tools, please subscribe to the HathiTrust Ingest Google Group.

General

HathiTrust ingested a second batch of locally-digitized volumes from Texas A&M University, and worked with Universidad Complutense de Madrid and the University of Illinois on ingest of new sets of locally-digitized materials. The University of Massachusetts, Amherst prepared to submit several thousand volumes digitized by the Internet Archive (IA), and Boston College completed steps for HathiTrust to begin ingest of several hundred IA-digitized volumes, expected to occur in January.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in December is given below. See CRMS-US and CRMS-World, projects funded by IMLS, for further information.

	December		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	2,375	5,436	158,167	305,593
CRMS-World	1,991	4,912	43,872	84,524
Total	4,366	10,348	202,039	390,117

Government Documents Registry

The project team continued to develop functional specifications for the registry and analyze metadata records to develop methods to detect duplicate records and match records of related items. The list of known federal agencies mentioned in the last monthly update is available for public viewing. Please note that it is a work in progress, and feedback (additional agencies; name/date corrections, etc.) is welcomed! Please contact valglenn@umich.edu with information or suggestions.

HathiTrust Research Center

The HTRC released version 2.0 of the Research Center software. Details can be found at http://wiki.htrc.illinois.edu/display/COM/HTRC+Release+2.0. The HTRC also made preparations to receive in-copyright content from the HathiTrust repository, considering possible design, architectural, and security implications. The HTRC continued to pursue grant opportunities, and is preparing a business plan for consideration by the HathiTrust Board. Stephen Downie will be leading a seminar on the HathiTrust Research Center at Oxford University on January 22,

Papers & Presentations

J. Stephen Downie, Tim Cole, Beth Plale, Katrina Fenlon, Karen Wickett, Megan Senseney. "The Workset Creation for Scholarly Analysis (WCSA) Prototyping Project: Background and Goals", 2013 Chicago Colloquium on Digital Humanities, December 6, 2013.

Beth Plale, "Opportunities and Challenges of Text Mining HathiTrust Digital Library", British Library, December 12, 2013.

There's an elephant in the library.™

www.hathitrust.org

Update On December Activities

2014.

mPach

University of Michigan staff continued to develop strategies for processing bibliographic metadata for mPach articles, and to scope services to be offered by Michigan Publishing using the mPach platform. See http://www.hathitrust.org/mpach for more information.

Zephir

California Digital Library (CDL) loaded 143,552 new or updated records into HathiTrust via Zephir, HathiTrust's new bibliographic metadata managment system. CDL suspended loading of partner records during holiday closures and pending the release of a bug fix in the loading program. Records submitted by partners were received during this time, and normal loading resumed in January. Partners interested in submitting bibliographic metadata to Zephir can visit http://www.hathitrust.org/bib_data_submission for an overview of the process.

Development Updates

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

Full-text Search

Staff began coding to index JATS XML content and to support indexing of volumes into a configurable number of "chunks" to improve relevance ranking of large documents.

PageTurner

California Digital Library staff began working on a variety of improvements to HathiTrust applications. The first is to add quick links to embed HathiTrust volumes in other Web pages to the PageTurner interface. CDL staff created mock-ups and began development.

Server Replacement Cycle

Staff completed the upgrade of servers in HathiTrust's development environment, combined with a move to a new Linux distribution to better support Ruby-based applications. Development staff tested all HathiTrust applications in the upgraded development environment, and shortly afterward, configured and put into service two new production web servers at the Indiana site as part of the periodic replacement cycle. In January, production web servers at the Michigan instance, which are not yet due for retirement, will be rebuilt to match the slervers in Indiana, and all old servers will be retired.

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Update On December Activities

Storage Replacement Cycle

Staff obtained pricing and submitted orders for new and replacement storage hardware as part of its regular purchase and replacement cycle. The cycle has been moved earlier by a few months to better coincide with the HathiTrust fiscal cycle.

Outages

HathiTrust users may have experienced slow or incomplete page viewing on Wednesday, December 11 from 4:10-4:18pm due to a problem at the Michigan site related to restart mechanisms built into the software release process. Service monitoring was enhanced to detect and repair the problem shortly after this outage.

Zephir bibliographic record loading was suspended from December 19, 2013 until Monday January 6, 2014 for the holiday break. The suspension was five days longer than originally announced, as CDL staff waited to deploy a new version of the loader software containing a bug fix until the Zephir Team was fully staffed on January 6, 2014.

Total Volumes Added	December	Overall
Boston College	0	2,363
Columbia University	1	65,036
Cornell University	16	437,491
Duke University	0	4,525
Harvard University	0	237,435
Indiana University	1	195,580
Library of Congress	0	89,724
North Carolina State University	0	3,196
Northwestern University	38	37,502
New York Public Library	0	288,370
Penn State	1,713	68,204
Princeton University	1	251,710
Purdue University	0	44,695
Texas A&M	1,175	1,201
Universidad Complutense	1	112,014
University of California	2,292	3,448,170
University of Chicago	3,067	38,635
University of Florida	0	9,763
University of Illinois	10	112,975
University of Michigan	515	4,666,032
University of Minnesota	3,076	115,935
UNC - Chapel Hill	0	17,025
University of Wisconsin	3	555,924
University of Virginia	0	50,821
Utah State University	0	117
Yale University	0	23,678
Total	11,909	10,878,121
Public Domain (~32% of total)		
Total*	16,595	3,542,155

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

HathiTrust Digital Library

Update On December Activities

User Support Issues	December	November
Content	**188**	**253**
Quality	179	244
Collections	8	8
Cataloging	**151**	**134**
Access and Use	**130**	**105**
Copyright	88	49
Permissions	7	6
Takedown	0	0
Print on Demand	0	1
Inter-library loan	0	0
Full-PDF or e-copy requests	11	15
Datasets	2	1
Data Availability and APIs	1	2
Reuse of content	2	7
Web applications	**21**	**27**
Functionality problems	7	4
Problems with login specifically	0	2
General questions about login	3	0
Partners setting up login	0	5
Usability issues	0	0
Feature requests	1	2
Partner Ingest	**4**	**5**
General	**77**	**66**
Partnership	2	5
Infrastructure	0	0
Miscellaneous	75	61
Total	**571**	**590**

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

Most-accessed volumes

Farm Implements, V. 13 (1899)

Patrons are People: How to be a Model Librarian, by Sarah Leslie Wallace.

The Sunlight Book of Knitting and Crocheting, by Adelaide Gray.

The Five Laws of Library Science, by S. R. Ranganathan

The Human Figure, by John H. Vanderpoel

Godey's magazine. v.40-41 1850

Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.

Goin' Fishin': the Story of the Deep-Sea Fishermen of New England, by Wesley George Pierce.

Analytic Geometry, by Frederick S. Nowlan.

The Book of a Hundred Hands, by George Brant Bridgman.