HathiTrust Digital Library

Update On January Activities

Late Breaking News

HathiTrust Bylaws Accepted

HathiTrust institutions approved bylaws put forward by the Board of Governors for voting in January. The bylaws are available at http://www.hathitrust.org/documents/hathitrust-bylaws-201302.pdf.

Top News

HathiTrust Resources

HathiTrust hosts a page of resources including handouts and informational sheets created by the Communications Working Group, and links to information about HathiTrust that institutions have posted for their own constituencies. If you have posted resources about HathiTrust (including videos, library guides, etc.) that are not listed, please let us know (feedback@issues.hathitrust.org), and feel free to use and share those that are available.

Ingest

Local Digitization

HathiTrust hosted a conference call with members of several partner institutions that have been working with HathiTrust's ingest tools, to discuss development options for the next iteration of the tools. A summary of the meeting, including next steps, is posted in a Google Group forum on HathiTrust Ingest. Individuals interested in issues surrounding ingest of locally-digitized content into HathiTrust are welcome to join. HathiTrust continued to respond to inquiries about the ingest tools and ingest of locally-digitized materials.

Internet Archive Digitization

HathiTrust began ingest of new batches of volumes from the University of Florida, the University of Illinois, and the University of North Carolina at Chapel Hill. HathiTrust also loaded bibliographic records for new volumes from Columbia University and Penn State.

Working Groups and Committees

Working groups and committees in HathiTrust may have an operational or strategic focus. See http://www.hathitrust.org/working_groups for more information.

Operational

User Experience Advisory Group

The User Experience Advisory Group began to review elements of HathiTrust Web

---

February Forecast

Continue to work on authorization to support access for users who have print disabilities

Continue relevance ranking tests in full-text search

Begin steps to deploy spelling suggestion feature

Continue work on common Web application framework

There's an elephant in the library.™

www.hathitrust.org

applications that were previously identified as in need of improvement, and review feature requests that have been submitted to the User Support Working Group.

User Support Working Group

A summary of issues received by the User Support Working Group is given in the table at the end of the update.

Projects

Bibliographic Data Management

California Digital Library (CDL) continued to work with staff at the University of Michigan to test data exports from Zephir, and to plan for the upcoming transition to Zephir for HathiTrust bibliographic metadata management. CDL staff made modifications to support rights determinations on bibliographic records that conform to the Resource Description and Access (RDA) standard. CDL is in the process of implementing a new backup strategy for Zephir to ensure the service will be accessible in the event of an outage. An updated timeline for the project is posted at http://www.hathitrust.org/htmms.

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in January is given below.

Copyright Determinations

	January		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	3,725	7,847	122,167	224,678
CRMS-World	2,803	6,871	17,005	31,581
Total	6,528	14,718	139,172	256,259

mPach

Staff at the University of Michigan met to discuss possible changes to the HathiTrust PageTurner, item-level search, and item-level search results to accommodate materials submitted via mPach. Staff began working with sample analytic MARC records and METS objects to determine the scope of anticipated changes. Staff also reviewed the mPach ingest process and discussed workflows for processing issue- and journal-level metadata.

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Update On January Activities

Development Updates

Full-text Search

Staff at Michigan continued to work on relevance ranking for full-text search results. Experiments performed in January found that some very short documents were being ranked too highly by Solr's default ranking algorithm. Preliminary tests indicated that the issue may be resolved by using Solr's new BM25, DFR and Information Based ranking settings. Experiments will continue in February.

Michigan staff received official relevance judgments for full-text search data submitted to the 2012 INEX Prove-IT Book Track (see also the paper "Practical Relevance Ranking for 10 Million Books"). After evaluating the results, staff submitted new data using updated relevance judgments.

Staff at CDL made significant progress on a spelling suggestion feature for full-text search. Staff improved the relevance of results through the creation of a new dictionary of suggestions that uses a bigram index for an entire shard's worth of documents (about 800,000 volumes). The HathiTrust full-text index is broken into "shards" in Solr's search architecture. Staff also tuned the algorithm that provides suggestions, and made significant changes to suggestion scoring parameters, significantly increasing the quality of results. CDL will now look toward deployment of the new feature.

Michigan staff made adjustments to the full-text search indexing process to better support experimental indexing runs, and to synchronize data between the full-text search index and information in the HathiTrust print holdings database.

HathiTrust completed the purchase of high-performance storage for full-text search. The new systems are expected to be received in late February for installation and testing.

Lawful Uses

HathiTrust began to implement application-level changes to support access to materials by designated representatives at partner institutions on behalf of users at those institutions who have print disabilities. Designated representatives will need to register, and to access HathiTrust using their Shibboleth login from a fixed IP address. Further details on the service will be forthcoming.

PageTurner

HathiTrust completed stylistic changes to messages in mobile PageTurner that appear when special access to materials is granted (e.g., access to volumes that fall under Section 108 conditions or to users who have print disabilities).

Storage Hardware Replacement Cycle

HathiTrust completed projections for new and replacement storage needed for HathiTrust in 2013.

HathiTrust Digital Library

Update On January Activities

Website Redesign

Staff at Michigan drafted implementation guidelines for a unified Web application framework for HathiTrust. The framework will simplify the execution of a redesign of the HathiTrust website, which is expected to be completed in April.

Outages

HathiTrust was unavailable from 5:00-5:40pm EST on Monday, January 21 due to an error in a software release.

Total Volumes Added	January	Overall
Boston College	0	1,842
Columbia University	0	64,390
Cornell University	1,222	416,657
Duke University	0	4,523
Harvard University	3	235,988
Indiana University	83	195,156
Library of Congress	0	89,722
North Carolina State University	0	3,196
Northwestern University	226	12,948
New York Public Library	4	259,578
Penn State	30	44,762
Princeton University	3	251,654
Purdue University	8	44,637
Universidad Complutense	27	111,928
University of California	1,942	3,385,197
University of Chicago	1,813	28,533
University of Florida	60	2,068
University of Illinois	218	105,105
University of Michigan	8,862	4,618,698
University of Minnesota	33	104,245
University of North Carolina - Chapel Hill	7,745	15,833
University of Wisconsin	651	551,031
University of Virginia	0	50,799
Utah State University	0	117
Yale University	0	23,678
Total	22,930	10,622,285

Public Domain (~31% of total)

	January	Overall
Total*	18,311	3,296,941

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

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Update On January Activities

User Support Issues	January	December
Content	**428**	**274**
Quality	414	268
Non-partner Digital Deposit	0	3
Collections	10	6
Cataloging	**99**	**52**
Access and Use	**148**	**95**
Copyright	85	59
Permissions	17	9
Takedown	0	0
Print on Demand	1	0
Inter-library loan	0	0
Full-PDF or e-copy requests	23	11
Datasets	10	5
Data Availability and APIs	0	0
Reuse of content	4	2
Web applications	**27**	**16**
Functionality problems	2	5
Problems with login specifically	0	2
General questions about login	4	1
Partners setting up login	4	3
Usability issues	1	1
Feature requests	3	0
Partner Ingest	**4**	**1**
General	**55**	**48**
Partnership	20	10
Infrastructure	0	0
Miscellaneous	35	38
Total	**761**	**486**

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

Most-accessed volumes

Title	Count
Investigation of Korean-American relations, 1978	6,049
Noblesa catalana: cavallers y burgesos honrats de Rossello y Cerdanya, v.1., by Philippe Laz-erme	5,159
Revelation interpreted, by Rev. G. A. Kratzer	3,456
All about tank cars, by Standard Car Construction Company	2,598
Godey's magazine, v.40-41 1850	2,332
A genealogical record of the descendants of Quartermaster George Colton, by John Milton Colton	1,693
Bradshaw's handbook for tourists in Great Britain & Ireland, sec.1 1866	1,609
The effects of nuclear weapons. Compiled and edited by Samuel Glasstone and Philip J. Dolan	1,454
The cornet of horse: a tale of Marlborough's wars, by G. A. Henty	1,366
Lloyd's list, 1823	1,122

There's an elephant in the library.™