



## Update On November Activities

December 14, 2012

### Top News

#### Bibliographic Corrections

HathiTrust has received a number of inquiries recently about corrections to bibliographic data. HathiTrust's general policy on bibliographic data correction is available at [http://www.hathitrust.org/bib\\_metadata\\_correction](http://www.hathitrust.org/bib_metadata_correction). We consider the definitive records for volumes in HathiTrust (which are generally volumes digitized from print originals) to be those held by the depositing institutions. When institutions submit corrections to print records in HathiTrust, these corrections are not automatically propagated to WorldCat. Institutions must update the print records in WorldCat separately.

OCLC creates records in WorldCat for electronic versions of works as they become available in HathiTrust (OCLC uses the [hathifiles](#) to identify when new volumes enter the repository and then derives digital master records from the print records identified by the OCLC numbers in the hathifiles). These electronic versions are solely OCLC's responsibility and under its control. Institutions do not need to, and should not try to update records for electronic versions. We are working with OCLC to refine the process by which records for e-versions are updated to stay in sync with HathiTrust records, and records for print versions that institutions update. We will be providing more information on this in future updates. For the present, if you notice a problem with a record in WorldCat for a HathiTrust volume, please notify us at [feedback@issues.hathitrust.org](mailto:feedback@issues.hathitrust.org).

#### Infrastructure Changes for Out of Print and Brittle

HathiTrust completed changes that will incorporate the "in print" status of volumes (whether or not a volume is in print), as well as holding status and condition information provided by partners in their [print holdings data](#), in volume access determinations.

### Ingest

#### Local Digitization

Staff from Texas A&M University contacted HathiTrust to discuss deposit of locally-digitized volumes related to Texas agricultural history. HathiTrust provided ingest support to the University of Iowa, University of Illinois, and University of Utah, including elaboration of content specifications, help in running image validation tools, and assistance in diagnosing errors. The [information page](#) about the tools HathiTrust provides for packaging and validating locally-digitized materials has been revised and includes a link to an updated [HathiTrust Deposit Form](#), which in turn includes guidelines and specifications for deposit.

#### Internet Archive Digitization

The University of North Carolina submitted a sample of bibliographic metadata in

#### December Forecast

Continue work to consolidate CSS framework for Web applications.

Continue work on indexing of CJK languages and relevance ranking for full-text search.

Complete the separation of administrative data from code in HathiTrust Web applications.

#### Papers & Presentations

Kevin Hawkins and Jeremy Morse, "mPach: Publishing directly in HathiTrust", Digital Library Federation Fall Forum 2012.

John Weise, Chris Powell, Kat Hagedorn, "HathiTrust: Sharing the Care and Feeding of the Elephant", Digital Library Federation Fall Forum 2012.

Stacy Kowalczyk, "Digital Humanities at Scale: HathiTrust Research Center", Digital Library Federation Fall Forum 2012.

There's an  
elephant in  
the library.™





## Update On November Activities

anticipation of an upcoming deposit. The University of Florida began deposit of Internet Archive-digitized volumes and anticipates depositing 26,250 items over the next several months. HathiTrust ingested two additional batches of content (totaling nearly 400 volumes) from Penn State, with two more batches to be ingested in December. The University of Illinois deposited more than 800 volumes as part of an ongoing project.

You can follow HathiTrust on [Twitter](#) or [Facebook](#)  
Subscribe to email updates (via [Google Groups](#))

### Working Groups and Committees

Working groups and committees in HathiTrust may have an operational or strategic focus. See [http://www.hathitrust.org/working\\_groups](http://www.hathitrust.org/working_groups) for more information.

#### Operational

##### User Experience Advisory Group

The User Experience Advisory Group was pleased to welcome a new member, Nadaleen Tempelman-Kluit, to the group. Nadaleen is an Instructional Design Librarian at New York University.

##### 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 processes for exporting bibliographic data from Zephir for use in HathiTrust services. CDL improved the speed at which data could be exported from Zephir. CDL and Michigan continued to plan for the time when the current bibliographic management system at Michigan and the new system (Zephir) will run in parallel. This will occur prior to HathiTrust moving to Zephir as the bibliographic management system for HathiTrust.

#### Copyright Review

A summary of copyright review activities in November is given below.

	November		Overall	
	Opened	Reviewed	Opened	Reviewed
CRMS-US	7,102	14,406	185,974	352,869
CRMS-World	8,306	15,097	23,487	46,062
Total	15,408	29,503	209,461	398,931

There's an elephant in the library.™





## Update On November Activities

### IMLS Quality Grant

The project team continued to plan user studies to evaluate and contextualize findings of the grant project. Grant principal investigator Paul Conway traveled to the University of Minnesota to launch the first user study, which will investigate thresholds for error tolerance in digitized volumes among library collections managers. Focus group meetings and other activities for this study will continue through the first quarter of 2013. The team submitted its second narrative report to IMLS, summarizing activities in the past year. The report will be posted soon on the [project website](#).

### mPach

Staff at the University of Michigan continued work on a mockup of changes needed to the PageTurner interface to support navigation of XML-based articles. Staff began to develop functionality to render JATS articles in PDF (for download purposes). Staff also engaged in discussions about the mPach article ingest workflow and proposed modifications to HathiTrust's Collections feature to facilitate navigation among journal articles.

## Development Updates

### Full-text Search

This past **June**, staff at Michigan discovered a **bug** in the Solr edismax processor that rendered search precision improvements for CJK (Chinese, Japanese, and Korean) materials smaller than expected. In November, conversations between staff at Michigan and Stanford about issues with CJK support led Michigan to contact Solr/Lucene developer and committer Robert Muir for advice. Muir (unaffiliated with Michigan or Stanford), an expert on multilingual issues, wrote and committed a code patch that fixed the bug. Staff at Michigan implemented the code patch and have seen orders of magnitude improvements (as an example, the query [東京スカイツリー] (Tokyo Sky Tree) produced about 450,000 hits without the patch and 16 hits after the patch). HathiTrust is very grateful for this assistance. Michigan staff made further improvements to indexing, which will be used in a full re-indexing of the full-text index in December. Staff also produced a sample bigram index, which will be used in ongoing work at California Digital Library on a spelling suggestion feature.

Staff at Michigan reviewed proposals received in response to an RFP issued in October for high-performance storage for full-text search, and are in the process of selecting the final systems to negotiate pricing. Installation and testing of the high-performance storage is tentatively scheduled for January.

### Web Applications

HathiTrust made a number of updates to Web applications, including:

There's an  
elephant in  
the library.™





## Update On November Activities

1. Initiation of work to remove sensitive information from application code for increased security.
2. Modification of the PageTurner to retrieve bibliographic data from the HathiTrust catalog's VuFind Solr index, rather than Michigan's bibliographic database (in preparation for the move to Zephir for metadata management).
3. Correction of a problem in PageTurner that caused execution to fail when DNS servers were unavailable.
4. Migration of mapping information between HathiTrust namespaces and depositing institutions to a database table for easier maintenance.
5. Migration of an access control list for special uses of in-copyright materials (e.g., for copyright or quality review purposes) to a database table to streamline maintenance.

### Website Redesign

Programmers for HathiTrust Web applications convened to develop a strategy for implementing a single Cascading Style Sheet (CSS) framework across all applications. A single framework will increase interface consistency and simplify future development, including a planned redesign of the HathiTrust home page and common portions of application interfaces.

### Outages

On Saturday, November 3, search within a volume was unavailable to some users from 3:00-8:30am and full-text search was unavailable to some users from 6:00-8:00am due to a temporary disk space shortage on a search server at one HathiTrust site.

HathiTrust sends notice upon discovery and resolution of unscheduled outages and in advance of scheduled outages and maintenance work that may result in an outage. We welcome and encourage additional recipients for these notices. If your institution is not receiving outage notifications and would like to, please contact [feedback@issues.hathitrust.org](mailto:feedback@issues.hathitrust.org).

### Total Volumes Added

	November	Overall
Boston College	0	1,816
Columbia University	204	64,390
Cornell University	3,209	415,363
Duke University	0	4,523
Harvard University	0	235,985
Indiana University	156	194,896
Library of Congress	0	89,722
North Carolina State University	0	3,196
Northwestern University	144	12,707
New York Public Library	0	259,574
Penn State University	390	44,525
Princeton University	0	251,650
Purdue University	70	44,525
Universidad Complutense	0	111,901
University of California	3,665	3,382,059
University of Chicago	7	26,663
University of Florida	1,034	1,034
University of Illinois	3,033	104,044
University of Michigan	5,608	4,602,578
University of Minnesota	304	103,839
University of North Carolina - Chapel Hill	0	8,088
University of Wisconsin	3,472	550,274
University of Virginia	0	50,799
Utah State University	0	117
Yale University	0	23,678
<b>Total</b>	<b>21,296</b>	<b>10,587,946</b>

### Public Domain (~31% of total)

Total*	17,122	3,269,229
--------	--------	-----------

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





## Update On November Activities

User Support Issues	November	October
<b>Content</b>	<b>304</b>	<b>310</b>
Quality	298	297
Non-partner Digital Deposit	0	1
Collections	4	6
<b>Cataloging</b>	<b>86</b>	<b>111</b>
<b>Access and Use</b>	<b>95</b>	<b>112</b>
Copyright	43	58
Permissions	4	11
Takedown	0	1
Print on Demand	0	1
Inter-library loan	0	4
Full-PDF or e-copy requests	15	13
Datasets	4	2
Data Availability and APIs	1	0
Reuse of content	2	0
<b>Web applications</b>	<b>13</b>	<b>21</b>
Functionality problems	4	8
Problems with login specifically	0	0
General questions about login	2	0
Partners setting up login	0	0
Usability issues	0	1
Feature requests	3	1
<b>Partner Ingest</b>	<b>3</b>	<b>9</b>
<b>General</b>	<b>141</b>	<b>61</b>
Partnership	18	14
Infrastructure	0	0
Miscellaneous	123	47
<b>Total</b>	<b>642</b>	<b>624</b>

\*See [User Support Working Group Issue Types](#) for a description of the types of issues included in each category.

There's an elephant in the library.™

