



Update On December Activities

January 11, 2013

Top News

Access to Out-of-Print and Brittle or Missing Items

One of the lawful uses of in-copyright works HathiTrust has been pursuing is to provide access on an institutional basis to works that fall under United States Copyright Law Section 108 conditions: works in HathiTrust that are not available on the market at a fair price, and for which print copies owned by HathiTrust member institutions are damaged, deteriorating, lost or stolen. As a part of becoming a member, institutions are required to submit information about their print holdings for fee calculation purposes. We have also been requesting information about the holdings status and condition of works, to facilitate uses of works where permissible by law (specifications for HathiTrust holdings data are available at http://www.hathitrust.org/print_holdings).

As of December 2012, we are using the holdings status and condition information submitted by United States member institutions, in combination with information about the market availability of works stored in the HathiTrust rights database, to determine whether or not access to applicable in-copyright works in HathiTrust is allowed. The specific terms of access are as follows:

- Access is only available to users affiliated with HathiTrust member institutions in the United States, and only from U.S. soil.
- In order to gain access, users from member institutions must be authenticated into HathiTrust via Shibboleth using their institutional login.
- Print copies of the works in HathiTrust must be owned currently or have been owned previously by the institution's library system.
- The number of users who can access a given digital copy at a time is determined by the number of print copies held (or previously held) in the library system. If a library system only has one print copy, only one user at a time will be able to access the digital copy.

A general scenario for how out of print determinations are made and communicated to HathiTrust is available in the HathiTrust rights database documentation: http://www.hathitrust.org/rights_database#op. Additional information on the service is available at <http://www.hathitrust.org/out-of-print-brittle>.

HathiTrust Bylaws

The Board of Governors completed a draft of HathiTrust bylaws, which was distributed to partner institutions in early December for comment. The Board is working on a final version with consideration for partner comments. The final version will be put forward to partners for voting in January.

HathiTrust Research Center Video

The Research Center released an informational video, following on the UnCamp that was held earlier in the fall of 2012. The video can be accessed at <http://www.hathitrust.org/htrc>.

Papers & Presentations

Jeremy York, "More, Better, Together: HathiTrust Aspirations and Accomplishments", The European Library/Europeana Libraries Joint Conference.

John Unsworth, Beth Sandore Namachchivaya, "Digital Humanities At Scale: HathiTrust Research Center", CNI Fall 2012.

John Wilkin, "HathiTrust: Strategies and Challenges in Consolidating the Published Record", National Diet Library, Japan.

Stephen Downie, "Introduction to the HathiTrust Research Center: A Briefing", University of Western Ontario.

See <http://www.hathitrust.org/papers> for a list of all papers, presentations, and reports.

There's an
elephant in
the library.™





Update On December Activities

Most-accessed Volumes in HathiTrust

This month we are including a new metric in our newsletter: the most accessed works in HathiTrust by pageview count. A table of volumes is included at the end of the update.

Ingest

Local Digitization

Staff at the University of Michigan met to discuss the next steps for HathiTrust's [ingest tools](#), created to aid institutions in validating and packaging locally-digitized content prior to deposit in HathiTrust. A conference call is planned in January, which will include members of several partner institutions that have been working with the existing tools, to discuss possibilities and options for the future. HathiTrust continued discussions about deposit of locally-digitized materials with the University of Illinois, and responded to questions from McGill University.

Internet Archive Digitization

HathiTrust ingested new content from Penn State University and loaded records for content from the University of Florida and University of North Carolina-Chapel Hill. Ingest of volumes from Florida and UNC, and additional volumes from Penn State, is expected to occur in January.

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 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 on preliminary testing of data exports from Zephir, the new HathiTrust bibliographic management system under development by CDL. CDL and Michigan staff continued to plan for the upcoming period when Zephir and the bibliographic management system at Michigan will be run in parallel, prior to the full transition to Zephir.

January Forecast

Hold meeting on next steps for HathiTrust ingest tools.

Begin testing features of new Solr relevance-ranking algorithms.

Complete purchase of storage for full-text search.

Continue work to consolidate CSS framework for Web applications.

There's an
elephant in
the library.™





Update On December Activities

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in December is given below. The numbers this month reflect a different methodology for aggregating statistics. In previous months, the number of Reviews was given, and the number of volumes reviewed that were Opened. In the majority of cases, volumes are reviewed more than once (by more than one person). This meant that the number of Reviews reported was larger than the number of actual volumes reviewed. Similarly, the number of volumes Opened represented volumes that may have been determined in more than one review to be in the public domain. The table below provides a more accurate representation of the number of volumes where a determination was made, and what the determination was. We will use this representation going forward.

Copyright Determinations

	December		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	2,433	5,028	118,442	216,831
CRMS-World	2,198	3,689	14,202	24,710
Total	4,631	8,717	132,644	241,541

IMLS Quality Grant

The project team will present a research poster at ALA Midwinter in Seattle, during the Preservation Administrators Interest Group Meeting on Saturday, January 26. The poster will focus on digitization error related to material characteristics of a book. The project team continues to focus on more complex analyses of the data collected in the past year and also on presentation of the findings. Additional findings and results will be posted on the project website later this month: <http://hathitrust-quality.projects.si.umich.edu>.

mPach

Staff at the University of Michigan revised the [list of modules](#) for mPach, to reflect recent changes in the planned system architecture. An extensive conceptual workflow for ingest of an mPach Submission Information Package into HathiTrust has been devised and will be finalized soon. Michigan staff finalized plans for modifications to the HathiTrust Data API to support the retrieval via the API of JATS XML, derivative formats, and supplemental materials that may be associated with a JATS XML article.

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





Update On December Activities

Development Updates

Full-text Search

Staff at the University of Michigan released a bug fix for the Solr edismax query parser and a new index into production in late December (See the [Update on November Activities](#) for details.). These changes will significantly improve the precision of CJK (Chinese, Japanese, and Korean) search results.

Michigan staff began preliminary analysis of HathiTrust document length statistics. The results of the analysis will aid in designing tests of length normalization features for the [new relevance ranking algorithms available in Solr 4.0](#). Staff built a test index using the new relevance ranking algorithms available in Solr 4.0 (DFR, BM25. IB). Experiments using the test index will begin in January.

Staff at Michigan made a final selection of high-performance storage for full-text search and completed pricing negotiations (see the [Update on November Activities](#) for background). Purchase of the storage is expected to be complete in January, with installation and testing to follow soon after in late January or early February.

Web Applications

Michigan staff completed the removal of sensitive information from source-controlled HathiTrust application code to designated system-level locations. Staff also completed the separation of privileges for accessing application databases. Different classes of applications now connect as different database users with different privileges.

Michigan staff began to implement improvements to the display of special access messages (e.g., for works that are out of print and brittle) in the mobile version of PageTurner.

The PageTurner scroll view now advances by full pages when the navigation controls are used (e.g., next page button), rather than advancing by half of a page at a time.

The HathiTrust feedback form now detects content and metadata-related feedback submissions by CRMS (Copyright Review Management System) reviewers, pre-

Total Volumes Added	December	Overall
Boston College	26	1,842
Columbia University	0	64,390
Cornell University	72	415,435
Duke University	0	4,523
Harvard University	0	235,985
Indiana University	177	195,073
Library of Congress	0	89,722
North Carolina State University	0	3,196
Northwestern University	15	12,722
New York Public Library	0	259,574
Penn State University	207	44,732
Princeton University	1	251,651
Purdue University	104	44,629
Universidad Complutense	0	111,901
University of California	1,196	3,383,255
University of Chicago	57	26,720
University of Florida	974	2,008
University of Illinois	843	104,887
University of Michigan	7,258	4,609,836
University of Minnesota	373	104,212
University of North Carolina - Chapel Hill	0	8,088
University of Wisconsin	106	550,380
University of Virginia	0	50,799
Utah State University	0	117
Yale University	0	23,678
Total	11,409	10,599,355

Public Domain (~31% of total)

Total*	9,401	3,278,630
--------	-------	-----------

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

There's an elephant in the library.™





Update On December Activities

User Support Issues

	December	November
Content	274	304
Quality	268	298
Non-partner Digital Deposit	3	0
Collections	6	4
Cataloging	52	86
Access and Use	95	95
Copyright	59	43
Permissions	9	4
Takedown	0	0
Print on Demand	0	0
Inter-library loan	0	0
Full-PDF or e-copy requests	11	15
Datasets	5	4
Data Availability and APIs	0	1
Reuse of content	2	2
Web applications	16	13
Functionality problems	5	4
Problems with login specifically	2	0
General questions about login	1	2
Partners setting up login	3	0
Usability issues	1	0
Feature requests	0	3
Partner Ingest	1	3
General	48	141
Partnership	10	18
Infrastructure	0	0
Miscellaneous	38	123
Total	486	642

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

filling problem tickets with CRMS-specific information to simplify the management of support requests.

Outages

No outages were reported in December.

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. If your institution is not receiving outage notifications and would like to, please contact feedback@issues.hathitrust.org.

Most-accessed volumes

Title	Count
Investigation of Korean-American relations, 1978	41,887
Investigation of Korean-American relations, 1977	4,169
The Tosa diary, tr. from the Japanese by William N. Porter, 1912	3,842
Le Tribun du peuple; ou le Defenseur des droits de l'homme, 1966	1,760
Away from the work-a-day world. An honest effort to picture the charms of Frankfort--the loveliest spot on Lake Michigan, 1904	1,737
Railway age, v.70 1921 Jan-Jun	1,638
Woman's home companion, v.36 1909	1,498
Thomae Actii De Ivdo scacchorvm in legali methodo tractatvs; nunc primùm in lucem ed. cum summaris & indice, 1583	1,296
Forest leaves, v.1-3, no.1-32 1886-1892	1,277
Railway age, v.69, 1920	1,217

There's an elephant in the library.™

