



Update On January Activities

February 10, 2011

Top News

Washington University Libraries Join HathiTrust

We are very pleased to welcome Washington University to the partnership. The full press release is available from the [Washington University website](#).

Board of Governors

The process of electing and appointing members to the new HathiTrust Board of Governors is proceeding on schedule. According to the [Governance ballot proposal](#) accepted by partners at the Constitutional Convention, 6 members of the Board will be appointed by the founding partner institutions and 6 will be elected by the partnership. The [full process for the elections](#), including schedule, as well as the Board of Governors [charge](#), are available on the HathiTrust website. As reported in the January Executive Committee [meeting minutes](#), members appointed to the new Board by the founding institutions include:

- Committee on Institutional Cooperation: Carol Diedrichs (Ohio State) and Wendy Lougee (Minnesota)
- Indiana University: Brad Wheeler
- University of California: Laine Farley and Brian Schottlaender
- University of Michigan: Paul Courant

Advanced Full-text Search

University of Michigan staff completed and released the first phase of advanced search functionality for full-text search. New features support a variety of operations for searching bibliographic metadata in combination with full-text. Results can be limited to specific publication years, languages, and original formats. The next iteration of work will begin in February and introduce options for building queries with greater Boolean complexity.

California Digital Library staff continued work on the spelling suggester feature, focusing on automatically building a dictionary (including unigrams with language information and frequencies, and bigrams with frequencies) from a test index of public domain materials.

Changes to Tab-delimited files

The changes HathiTrust intended to make to the [tab-delimited files](#) (“hathifiles”) beginning February 1 resulted in some unexpected problems, which staff at Michigan are in the process of resolving. We currently plan to roll back the changes so that the files are in their pre-February state and pursue a March 1 date to add a total of 5 new fields to the files. Notification of 3 new fields was included in the [Update on December 2011 Activities](#). Two additional fields will be added, so that the tab-delimited files will include new fields for publication date, publication location, language, bibliographic format, and whether or not a volume has been identified as a U.S. federal government document. Updates on the status of the files will be send

February Forecast

Continue to work with partners on ingest of locally-digitized materials

Continue working on improvements to advanced full-text search

Resume work on Data API security

You can follow HathiTrust on Twitter at <http://www.twitter.com/hathitrust>

There's an
elephant in
the library.™





Update On January Activities

via [HathiTrust's account](#) on Twitter, and posted on the [tab-delimited files download page](#).

Year in Review

HathiTrust released a [Year in Review](#) of its 2011 activities, highlighting achievements in its repository services, partnership, and position in the library community.

Ingest

Local Digitization and Internet Archive

HathiTrust discussed deposit of an additional set of locally-digitized volumes with Yale University, and worked with Columbia University on packaging locally-digitized materials to HathiTrust specifications. Penn State University began preparations to deposit Internet Archive-digitized content into HathiTrust, and Getty Research Institute continued discussions with HathiTrust regarding bibliographic data for its locally- and Internet Archive-digitized materials.

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

Communications

The Communications Working group announced HathiTrust's major milestone of reaching 10 million volumes in January, and continued its work to develop a public services informational package. The group also engaged in looking for opportunities to highlight HathiTrust within the media and conference landscape.

User Experience Advisory Group

The User Experience Advisory Group discussed user interface issues related to possible changes to the Pageturner default view, and potential interface improvements to the list of user-created collections.

Total Volumes Added

	January	Overall
Columbia University	0	64,176
Cornell University	645	384,605
Duke University	0	4,522
Harvard University	1	53,441
Indiana University	38	186,950
Library of Congress	0	89,411
North Carolina State University	0	3,196
University of North Carolina - Chapel Hill	0	8,087
Northwestern University	407	6,056
New York Public Library	13	259,466
Penn State University	29	42,946
Princeton University	0	249,679
Purdue University	0	887
University of California	4509	3,292,163
University of Chicago	1091	11,699
University of Illinois	0	14,503
Universidad Complutense	15	108,683
University of Michigan	8503	4,512,644
University of Minnesota	342	90,581
University of Wisconsin	1244	528,578
University of Virginia	0	47,396
Utah State University	0	46
Yale University	0	23,674
Total	16,837	9,983,509*

Public Domain (~27% of total)

Total**	18,803	2,731,429
---------	--------	-----------

* Does not include archival and image materials in the Minnesota Digital Library project

**Includes volumes opened through copyright review or rights holder permissions.





Update On January Activities

User Support Working Group

In addition to regular activity responding to user inquiries, the User Support Working Group has spent the last several months evaluating its processes, workflows, and performance since it began in March 2011. This was done to prepare recommendations on a future structure and processes for responding to user feedback, which is part of the group's [charge](#). A number of ideas to improve efficiency in responding to inquiries and communicating within the group surfaced and have been implemented. The group completed a draft report on recommendations that it expects to submit to the Executive Committee in February.

The adjacent table contains a summary of the issues received by the User Support Working Group in January.

Strategic

Collections Committee

The Collections Committee made good progress on a process for responding to requests and offers to include additional materials in HathiTrust, among other pending items on its work agenda.

Projects

Bibliographic Data Management System

The California Digital Library team continued to load and test records in Zephir, the new management system. The team finished a proposal for a minimum record submission standard, and completed work on a refined migration timeline -- both to be reviewed by University of Michigan in early February. CDL also performed a successful test to sync data from the HathiTrust rights database with records in Zephir.

HathiTrust Publishing (HTPub)

MPublishing staff at the University of Michigan Library created a timeline for work through early 2013. Work continued on a process to convert styled Word documents into JATS XML, focusing on extraction of metadata, and on adaptation of the HathiTrust PageTurner application to display JATS XML.

User Support Issues	January	December
Content	144	81
Quality	117	71
Non-partner Digital Deposit	0	2
Collections	10	6
Cataloging	38	30
Access and Use	79	107
Copyright	33	59
Permissions	20	4
Takedown	0	2
Print on Demand	0	1
Inter-library loan	0	2
Full-PDF or e-copy requests	15	28
Datasets	0	2
Data Availability and APIs	0	0
Reuse of content	1	1
Web applications	24	18
Functionality problems	7	9
Problems with login specifically	1	1
General questions about login	3	0
Partners setting up login	1	1
Usability issues	5	2
Feature requests	4	1
Partner Ingest	4	5
General	127	50
Partnership	7	7
Infrastructure	1	0
Miscellaneous	119	43

*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.™





Update On January Activities

IMLS Quality Grant

The primary focus of project staff in January was to complete page-level review of volumes in the third production run, performed on a sample of 1,000 Internet Archive-digitized volumes published pre-1923. As of January 31st, review of more than 97% (over 97,000 digital pages) of the volumes was complete. This included double-review of 10% of the volumes as a check on inter-coder reliability.

Physical review of the volumes sampled in the first production run continued in January. By the end of the month, volunteers from the University of Michigan School of Information had reviewed 848 of the 1,000 volumes.

Project staff at the University of Michigan began testing a beta version of the newly developed quality review interface, targeted specifically for review of volume-level errors such as missing, duplicate, and out-of-order pages. A test sample of known problematic volumes was developed to test the strength of the error model and application. Official data coding of whole-volume errors is expected to begin by the end of February. Please visit the [project website](#) for updates.

Development Updates

Logging Usage of In-Copyright Materials

HathiTrust implemented processes to track accesses to in-copyright works, in cases where access is permitted. The new processes will provide a means for HathiTrust to detect problematic activity such as bulk downloading operations, which may, for example, indicate a compromised user account.

New Web Servers and Web Load Balancers

Michigan staff transitioned two new web servers at the Michigan repository instance into service, replacing two older ones. During the same cutover, all Web service was moved to new Web load balancers which, as compared to the previous load balancing mechanism, provide a better distribution of traffic across all servers at both sites, as well as a faster response when individual servers or sites fail. Michigan staff routinely use these load-balancing systems to mask maintenance or upgrade processes that require individual servers or an entire site to be taken offline.

Storage Hardware Replacement Cycle

University of Michigan staff received final 2012 volume projections from partners and requested a price quote from Isilon for the purchase of new storage capacity and the annual storage hardware replacement cycle, which since last year have been combined into a single large acquisition. The new capacity is expected to be online in the first quarter of 2012.

There's an
elephant in
the library.™





Update On January Activities

Outages

The HathiTrust web site, including the bibliographic catalog and full-text search (but excluding page viewing and persistent URL resolution), was down on Friday, January 27 from 8:30-9:00pm EST due to a Drupal software upgrade.

Full-text search web pages may have generated incorrectly from Friday, January 27 at 7:30pm to Saturday, January 28 at 3:10pm due to an accidental, premature release of modifications to the full-text search software related to internationalization support.

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.

Papers and Presentations

- Jeremy York, [Panel Presentation](#). Session 9. Large Digital Libraries: Beyond Google Books. Modern Language Association Annual Meeting.
- Jeremy York, Panel Presentation ([remarks only](#)) Session 129. What's Still Missing? What Now? What Next? Digital Archives in American Literature. Modern Language Association Annual Meeting.
- John Wilkin, [Digital Preservation: A Matter of Trust](#). Session 444. Preservation Is (Not) Just Another Word for Nothing Left to Lose. Modern Language Association Annual Meeting.
- Jeremy York "[HathiTrust: The Elephant in the Library](#)". Library Issues Vol. 32 No. 3, January 2012.
- Sarah Pritchard "[HathiTrust Libraries Map a Shared Path: A Turning Point in Information Access](#)". Libraries and the Academy Vol. 12 No. 1, January 2012.

There's an
elephant in
the library.™

