



Update On May Activities

June 8, 2012

Top News

Report on Board of Governors Meeting

The new Board of Governors met in Chicago in conjunction with the ARL membership meeting in May. The group spent some time before the meeting identifying priorities, focusing primarily on the organizational work of the Board. The Board quickly formed an Executive Committee, as stipulated in the [Constitutional Convention ballot proposal](#). The new Executive Committee members include Paul Courant, Carol Diedrichs, Laine Farley, Sarah Michalak and Bob Wolven. Another group chaired by Pat Steele was charged with initiating the process to assemble by-laws. This group will also attend to issues such as the duration of the appointment of the Executive Committee, and expects to conclude its work by the end of November. A third group will be formed to focus on the development of a Charter.

The Board of Governors will meet by teleconference for the next several months, targeting one meeting per month, as the process of developing by-laws moves forward. In these meetings the Board plans to review HathiTrust's past work, which will include a review of the HathiTrust budget as well as HathiTrust's committees and working groups. Although it was not able to discuss HathiTrust's existing committees and working groups in detail in Chicago, the Board expressed a deep appreciation for the work the Strategic Advisory Board, the Collections Committee, and the current operational working groups and committees have done. The Board asked that the existing groups continue their work (with the Board's enthusiastic support) until and while a review of committees can take place.

New Resources and Guides

We are pleased to announce the new [HathiTrust Resources and Guides](#) page, where we bring together overviews, instructional materials, and guides created by HathiTrust partner libraries, the Communications Working Group, and beyond. Materials posted on the page include reusable handouts, a detailed guide to using HathiTrust, lively blogs and dynamic videos. Please use, repurpose, and enjoy!

Have you created HathiTrust user guides or instructional materials? We encourage you to submit them to feedback@issues.hathitrust.org.

Embed Your Favorite Work

It is now possible to embed HathiTrust volumes in web pages. Code snippets to do this can be found at <http://www.hathitrust.org/embed>.

Ingest

Local Digitization

Staff at the University of Michigan completed development of the first iteration of tools to help depositors create and validate content packages prior to submission to HathiTrust. The tools will be made available in early June to several partner in-

What's In Your Collection?

Read our [new blog post](#) on building HathiTrust collections.

June Forecast

Rebuild the Large Scale Search Solr/Lucene index with CJK (Chinese, Japanese, Korean) indexing improvements

Distribute first iteration of tools to aid in preparing content for ingest into HathiTrust

There's an
elephant in
the library.™





Update On May Activities

stitutions that are working on ingest of locally-digitized materials.

Google

HathiTrust ingested approximately 150,000 additional public domain volumes from Harvard University Library.

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

As noted in Top News, the Communications Working Group released a [new Web page](#) featuring HathiTrust instructional materials from across the partnership, including guides developed for public services use. In addition, the group submitted a briefing to the new Board of Governors with recommendations for carrying out communications activities in the future. The working group also launched a [Pinterest account](#) for HathiTrust.

User Experience Advisory Group

The User Experience Advisory Group focused its attention on a project being undertaken by University of Michigan staff to redesign the HathiTrust home page (www.hathitrust.org). The group will begin consulting regularly on this project in June.

User Support Working Group

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

Projects

Bibliographic Data Management

Staff at California Digital Library (CDL) refined the code for loading bibliographic records into Zephir (the new bibliographic management system) and reloaded all HathiTrust records in the test environment. Work continued to code a process to sync rights information in Zephir with the HathiTrust rights database. The CDL team is developing documentation and guidelines for submitting bibliographic re-

User Support Issues

	May	April
Content	168	231
Quality	159	222
Non-partner Digital Deposit	0	1
Collections	3	4
Cataloging	51	33
Access and Use	129	112
Copyright	64	76
Permissions	12	7
Takedown	2	2
Print on Demand	1	0
Inter-library loan	0	2
Full-PDF or e-copy requests	22	10
Datasets	4	4
Data Availability and APIs	4	1
Reuse of content	2	1
Web applications	12	14
Functionality problems	3	4
Problems with login specifically	0	1
General questions about login	0	4
Partners setting up login	3	3
Usability issues	0	0
Feature requests	1	0
Partner Ingest	2	5
General	81	129
Partnership	6	5
Infrastructure	1	0
Miscellaneous	74	124
Total	443	519

*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 May Activities

cords to Zephir, and documentation of reports to be provided to institutions when records are loaded.

mPach (previously jPach)

University of Michigan staff continued work on modifications to the HathiTrust PageTurner to display JATS XML. Staff began development of wireframes for the Dashboard module and are close to the completion of a specification for mapping JATS metadata elements to MARC fields to create analytic records for journal articles.

HathiTrust Research Center (HTRC)

SAVE THE DATE!

HathiTrust Research Center UnCamp:

A 1.5 Day Event

Sept 10-11, 2012 Indiana University, Bloomington, IN.

Mark your calendars. HTRC is hosting its first annual HTRC UnCamp in September 2012 at Indiana University in Bloomington. The UnCamp is different: it is part hands-on coding and demonstration, part inspirational use-cases, part community building, and part informational, all structured in the dynamic setting of an un-conference programming format. It has visionary speakers mixed with boot-camp activities and hands-on sessions with HTRC infrastructure and tools. Through the HTRC Data API, attendees will be able to browse and run applications (yours or ours) against the full 2.8M volumes of the public domain corpus of HathiTrust. Bloomington is lovely in September, and the IU campus is noted as one of the most beautiful public university campuses in the nation.

Who should attend? The HTRC UnCamp is targeted to the digital humanities tool developers, researchers and librarians of HathiTrust member institutions, and graduate students. Attendance will be capped at 60 participants, so plan to register early!

Travel funds and Registration. HTRC anticipates funding a small number of travel grants that can be used by an attendee to bring along a graduate student or for a HathiTrust member librarian/technologist to bring along a researcher from their organization who is interested in engaging with our research center. The Uncamp will have a minimal registration fee so as to make the Uncamp as affordable as possible for you to attend.

IMLS Quality Grant

All of the data collection for English language volumes was completed in May, including double-review of subsets of volumes for quality assurance. Review of volumes in the grant's final 1,000-volume sample, which includes volumes from 6 major non-Roman languages (Chinese, Japanese, Korean, Arabic, Cyrillic and Hebrew), is still in progress. At the end of May, staff had reviewed 77,115 of the total 95,086 pages sampled for review. Data collection is expected to be complete

Papers & Presentations

Jeremy York: [HathiTrust Overview](#), Michigan Chapter of the Association of Law Libraries Spring Meeting, May 18, 2012

Jeremy York: [We're Preserving the Past, What About the Present?](#), NISO Webinar, May 23, 2012

Partner-specific:

Brian Vetruba: [HathiTrust--A Govdocs Repository?](#), Regional Government Documents Conference, May 4, 2012

Heather Christenson: [How HathiTrust Serves the UC Community](#), May 21, 2012

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

There's an
elephant in
the library.™





Update On May Activities

in mid-June.

Work in June will focus on analysis of the collected data, as well as research, development, and data collection for use case studies, which will comprise the final portion of the grant. Staff will also undertake a specialized study of errors in digitized illustrations to try to more accurately describe the types of errors that are observed and their impact on use.

Current findings of the project will be presented at the ALA Annual Meeting in June 2012. The [project website](#) is being updated with a new graphic design; further [initial findings](#) will be forthcoming. Please see the website for details on the volumes samples, error models, and other grant activities.

Development Updates

Data API

Staff at the University of Michigan made minor changes to the Data API and the Data API's key service and Web client to better manage user privileges. A Data API security monitoring and reporting script was also deployed that runs on a daily basis.

Full-text Search

Michigan staff undertook work to improve indexing and searching of CJK languages (a discussion of the issues is available on the [large-scale search blog](#)). All 10+ million volumes are being re-indexed using the new [CJKBigramFilter](#) available in Solr 3.6, and a custom filter that will create a separate unigram index of Han characters (to support queries consisting of a single Han character). Staff revised the Solr indexing schema to eliminate unused fields and filters and to take advantage of upgraded Solr 3.6 filters. Staff also made changes in development to the full-text search and "search within a book" Web applications in preparation for the improved CJK indexing. Testing and production release of the application enhancements and newly-created index are anticipated in early June.

Staff at Michigan downloaded and began to index the [INEX Book Track](#) "Prove It" task corpus to use as a testbed to investigate various relevance ranking issues in HathiTrust full-text search.

Total Volumes Added	May	Overall
Columbia University	0	64,184
Cornell University	3,334	399,871
Duke University	0	4,523
Harvard University	146,064	199,739
Indiana University	26	187,664
Library of Congress	0	89,416
North Carolina State University	0	3,196
University of North Carolina - Chapel Hill	0	8,088
Northwestern University	0	7,203
New York Public Library	2	259,559
Penn State University	14	43,322
Princeton University	10	250,849
Purdue University	1	24,781
University of California	6,811	3,336,782
University of Chicago	364	20,821
University of Illinois	5	96,151
Universidad Complutense	0	111,827
University of Michigan	4,379	4,539,368
University of Minnesota	4,320	99,470
University of Wisconsin	4,337	539,208
University of Virginia	0	48,922
Utah State University	0	90
Yale University	0	23,678
Total	169,667	10,358,712

Public Domain (~29% of total)

Total*	153,218	3,033,255
--------	---------	-----------

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

There's an elephant in the library.™





Update On May Activities

Staff at California Digital Library (CDL) completed development of fast lookup data structures in the language-sensitive dictionary that will support a spelling-suggestion feature in full-text search (last reported on in the [Update on February Activities](#)). Staff used probabilistic techniques to fit the massive dictionary into RAM, allowing very fast lookup of bigram and unigram data. Staff also ported code from the CDL-developed XTF system that ranks spelling suggestions to the new structure, though the code is not yet fully functional. Next steps include modifying the ranking algorithm to take advantage of data from the language-sensitive dictionary, and evaluating and revising the algorithm to produce quality suggestions.

PageTurner

Staff at Michigan deployed fixes to the code that allows users to embed PageTurner views in Web pages using an iframe. Staff also added improved wording and an explanatory link to the PageTurner interface, recommended by the UX Advisory group in April, to clarify when full-PDF download of volumes in HathiTrust is or is not available.

Storage Hardware Replacement Cycle

Michigan staff completed the steps necessary to retire all storage that was scheduled for replacement in 2012. Staff had completed the installation of replacement and additional storage at the Michigan and Indiana sites in March.

Web Hosting Infrastructure Changes

HathiTrust's VuFind-based bibliographic catalog was successfully moved from University of Michigan Library Web hosting infrastructure to HathiTrust's Web hosting infrastructure. This completes a migration project that also involved HathiTrust's Drupal-based informational website and will greatly simplify future Web development.

Outages

Full-text search in HathiTrust was unavailable on Wednesday, May 9 from 6:00-8:30am EDT due to a problem with an index server. Shibboleth authentication to HathiTrust was unavailable on Monday, May 21 from 9:23-9:28am EDT due to a problem with a helper service required by Shibboleth.

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.

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

**There's an
elephant in
the library.™**

