



Update On September Activities

October 14, 2011

Late Breaking News

Constitutional Convention

On October 8-10, 2011, 130 representatives from 64 HathiTrust partner institutions, including library directors, chief information officers, and senior library administrators, gathered in Washington D.C. for an unprecedented “Constitutional Convention” to reflect on the accomplishments of HathiTrust since its launch in 2008, and determine directions and priorities for the partnership in its next phase. The business portion of the meeting consisted of deliberations and voting on 7 ballot initiatives presented by partner delegations prior to the convention. The final proposals and outcomes are available at http://www.hathitrust.org/constitutional_convention2011. A large portion of the Convention was also spent in general discussions on a variety of topics including the new pricing model for partner institutions, lawful uses of library-owned materials, and international cooperation. A more complete report on the Convention, its outcomes, and what they mean for the partnership, is forthcoming. The following presentations from the Convention are available on the HathiTrust website:

- Opening remarks (view [text](#) or [presentation](#)): John Wilkin, Executive Director, HathiTrust
- Report on HathiTrust 3-year review and Q&A (view [presentation](#)): Ed Van Gemert and Trisha Cruse, HathiTrust Strategic Advisory Board

University of Miami Joins HathiTrust

The University of Miami [announced membership](#) in HathiTrust in early October. We are very pleased to welcome Miami to the partnership.

HathiTrust Mobile

Following a soft release in August, HathiTrust is pleased to formally announce its new mobile interface (visit <http://m.hathitrust.org>). The interface offers mobile-friendly access to key functionality including searching the HathiTrust catalog and reading HathiTrust “Full view” texts. Users from HathiTrust partner institutions can download texts in PDF or ePub format. Since the mobile interface is web-based, it works on all platforms, and may be viewed either from mobile devices or from desktops and laptops. The interface has special functionality for tablets where there are two ways to read texts: either in the vertical scrolling format, or in a horizontal flip format. Please give the new mobile interface a try and don’t hesitate to [send your comments and feedback!](#)

Top News

Author’s Guild Lawsuit

On September 12, the Author’s Guild, the Australian Society of Authors, the Union

October Forecast

- Release advanced full-text search
- Re-index entire corpus to support advanced search and improve relevance ranking
- Continue work on the spelling suggestion feature

Presentations

Cory Snaveley “[Scaling Full-text Search](#)” Library of Congress, September 26, 2011.

Jeremy York “[Collaborating Globally, Planning Locally](#)” GWLA/UNM Webcast, September 26, 2011.

There’s an
elephant in
the library.™





Update On September Activities

Des Écrivaines et des Écrivains Québécois (UNEQ), and eight individual authors filed a lawsuit against HathiTrust, the University of Michigan, the University of California, the University of Wisconsin, Indiana University, and Cornell University for copyright infringement. The suit was updated on October 8. We believe this is a misguided and unnecessary lawsuit. A [full statement by HathiTrust](#) is available online, and links to statements by the University of Michigan and analysis from a variety of sources are available at http://www.hathitrust.org/authors_guild_lawsuit_information.

Requirements for New Partners

Beginning January 1, 2012, partners joining HathiTrust will need to provide information about their library holdings at the time of joining. The holdings data will be used for partner fee calculations and to offer access on a limited basis to in-copyright materials (see the Holdings Database update in the [July newsletter](#) for details). Partners must be configured with [Shibboleth](#) for their users to authenticate for partner services in HathiTrust.

Ingest

Local Digitization Ingest

University of Michigan staff continued work with several partner institutions on ingest of locally-digitized materials, including Northwestern University, Universidad Complutense de Madrid, the University of Florida, the University of Iowa, the University of North Carolina-Chapel Hill, the University of Pittsburgh, and the University of Utah.

Working Groups and Committees

User Experience Advisory Group

The UX Advisory Group compiled and discussed a list of possible interface features and improvements that have been requested by users and staff at partner institutions. Three improvements were identified as high priority and will be ongoing topics of discussion until solutions are reached which can be passed to the University

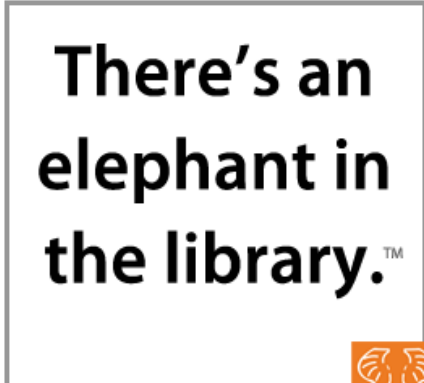
Total Volumes Added

	September	Overall
Columbia University	0	64,042
Cornell University	10,815	368,146
Harvard University	24	52,838
Indiana University	33	186,172
Library of Congress	0	71,418
North Carolina State University	240	3,194
Northwestern University	165	5,349
NYPL	115	259,158
Penn State University	1,438	40,807
Princeton University	3,132	248,914
University of California	102,280	3,141,343
University of Chicago	6	8,042
University of Illinois	0	14,501
Universidad Complutense	183	108,338
University of Michigan	14,018	4,446,315
University of Minnesota	181	88,432
University of Wisconsin	6,810	504,349
University of Virginia	19	47,327
Utah State University	0	46
Yale University	5,289	23,674
Total	144,748	9,682,405

Public Domain (~27% of total)

Total*	49,797	2,642,832
--------	--------	-----------

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





Update On September Activities

of Michigan development team. The improvements are:

- Redesigning the page turner “landing page” for Limited (search-only) items to better communicate available options
- Revising PDF download link labels in page turner to better communicate when a full PDF is available without login
- Adding explicit page numbers or page status to page turner interface

User Support

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

Projects

Bibliographic Data Management

The California Digital Library development team continued to work on improvements to Zephir, the core metadata management system, and adaptations of system components to HathiTrust ingest and management workflows. As part of these improvements, project staff developed a program that doubles the speed of ingest for normalized bibliographic records. The team also worked with University of Michigan staff to identify modifications that have been made to records in HathiTrust over time, part of a broader strategy for managing updates to records in the new system.

HTPub

A project manager from the University of Michigan joined the team working on HTPub, a two-year project to develop a system that will enable MPublishing at the University of Michigan Library to use HathiTrust as a publishing platform for its journals. The team has refined the project [goal and requirements](#) and is formulating design principles, a use case specification, and the system architecture. A full-time software developer has joined MPublishing, focusing on the content ingest and publication management components of this system.

HathiTrust Research Center

The Communications Working Group began working with staff at the University of Indiana to create a presence for the HathiTrust Research Center on [HathiTrust.org](#). The new portion of the website is expected to be released in the next several

User Support Issues

	August	Sept.
Content	110	171
Quality	96	154
Non-partner Digital Deposit	3	2
Collections	8	4
Cataloging	26	25
Access and Use	111	127
Copyright	58	73
Permissions	23	12
Takedown	2	3
Print on Demand	6	17
Inter-library loan	0	5
Full-PDF or e-copy requests	14	24
Datasets	1	1
Data Availability and APIs	1	7
Reuse of content	7	5
Web applications	27	22
Functionality problems	5	5
Problems with login specifically	1	0
General Questions about login	3	2
Partners setting up login	4	5
Usability issues	11	6
Feature requests	7	2
Partner Ingest	2	0
General	59	65
Partnership	13	12
Infrastructure	1	0
Miscellaneous	45	53

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





Update On September Activities

weeks.

IMLS Quality Grant

In September, staff at the University of Michigan and University of Minnesota completed quality review of a sample of 1,000 public domain volumes selected at random from HathiTrust (the sampling strategy is described in the [July newsletter](#)). Data for more than 110,000 pages in all were collected. Two reviewers coded 10% of the sampled volumes as a check on inter-coder reliability. The project statistician is analyzing the data and initial findings will be available in October.

In addition to review of the digital volumes, the project team launched a process to perform physical review on all volumes in the sample. The project programmer created a data collection interface for this review and a volunteer staff of students as well as project staff began to retrieve and evaluate the physical volumes according to a list of specific criteria. The volunteer staff reviewed approximately 10% of the physical volumes by the end of September.

The project team also prepared for and began review of a second sample of 1,000 digital volumes. The second sample focuses on volumes published after 1922 and employs a different within-book sampling methodology. Whereas in the first run 100 pages at most were sampled from each volume, this run will review a number of pages in each volume proportional to the size of the volume. The second round of data collection is expected to be complete in mid-November. Background information on the project can be found at <http://www.hathitrust.org/grants>.

Development Updates

Collection Builder

Staff at the University of Michigan implemented a new process for updating rights information for items saved to personal and private collections.

Full-text Search

University of Michigan staff made modest modifications to full-text search indexing as part of a revised re-indexing strategy. Re-indexing of the full-text and bibliographic metadata for the entire corpus of 9+ million books began in late September and will be completed in early October. The re-index updates the full-text index to Unicode 6, and includes metadata changes that will improve title displays and provide the metadata needed to support access mechanisms that depend on holdings information (e.g., print disabled users).

Michigan staff developed a prototype for advanced full-text search and performed a preliminary user interaction/usability walkthrough. Michigan developers provided query logs, N-gram data, and term frequency information to staff at the California Digital Library for use in developing and testing a spelling suggestion feature.

There's an
elephant in
the library.™





Update On September Activities

PageTurner

University of Michigan staff worked on improvements to the algorithm used to estimate and update page image sizes for display with BookReader, resulting in a faster time for image display. Staff also included the “missing page” placeholder that appears in traditional views of volumes when pages are known to be missing to the thumbnail view. Pages may be missing from volumes for a variety of reasons, including the pages not being present in the physical volumes that were scanned, and errors in post-scan processing.

Developers at Michigan made progress on new throttling mechanisms that will be implemented at the web application level. Once completed, these mechanisms will make it possible to adjust throttling thresholds depending on the type of content delivered and ultimately reduce the likelihood of users being throttled during normal use.

Michigan staff put additional access controls into place in PageTurner, in anticipation of offering access to orphan works. The controls include limiting access to:

- One simultaneous user per print copy held by the user’s institution
- One page at a time download
- Only authenticated users on US soil

Interface changes were also made to improve display of the copyright status of each work.

Outages

No outages were reported in September 2011.

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 at <http://www.twitter.com/hathitrust>

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

