



Update On September Activities

October 12, 2012

Late Breaking News

Case Closed - HathiTrust is Fair Use

In a decision that will have broad repercussions across libraries, on October 10, 2012 Judge Baer dismissed the lawsuit filed just over a year ago by the Authors Guild et al. against HathiTrust and several participating institutions. HathiTrust has released an [official statement](#) on the ruling. [Information](#) about the lawsuit, as well as relevant [analysis and reactions](#) from around the Web are available on the HathiTrust website.

Top News

HathiTrust Research Center UnCamp

The HathiTrust Research Center held its first annual “UnCamp” in Bloomington, IN on September 10-11. 130 researchers, developers, and librarians from HathiTrust member and non-member institutions gathered in Indiana University’s new CyberInfrastructure Building for presentations, demos, and hands-on sessions with the emerging Research Center tools. These included tools both to perform research on the HathiTrust corpus and to create new or customized algorithms and processes for research. Responses to the UnCamp have been very enthusiastic, giving energy to efforts to enable computational access to the incredible body of works in HathiTrust. More information on the UnCamp, including presentations, resources, reactions and responses via tweets, and more can be found on the [HathiTrust Research Center Wiki](#). See the [press release](#) also from the University of Illinois.

Government Documents Registry

HathiTrust has initiated a project to build a comprehensive registry of U.S. federal government documents. The Registry is an emerging effort in a broader undertaking by HathiTrust partners to improve access to U.S. federal government documents. Further information and background on the project is available on the Registry [project page](#). A two-year term [Government Documents Registry Analyst position](#) for the project was posted in September.

Infrastructure Changes for Out of Print and Brittle

In the coming weeks, HathiTrust will begin making infrastructural changes to incorporate information about the holdings status and condition of volumes at partner institutions into access services. The changes will apply in particular to access on library premises to in-copyright works that fall under Section 108 provisions of the U.S. Copyright Act. One of the infrastructural changes will be altering the semantics of the “out-of-print and brittle” (“opb”) designation in HathiTrust’s [rights database](#) to “out-of-print” (“op”) only. This change will be made on November 1, 2012, and will be reflected in HathiTrust interfaces, and services such as the HathiTrust where rights information is made available.

October Forecast

Continue work on Full-text Search, including indexing of CJK languages and relevance ranking.

Begin making changes to support website redesign, including an exploration of CSS frameworks.

Papers & Presentations

Tom Burton-West, “[Practical Relevance Ranking for 10 Million Books](#)”, September 2012.

[HathiTrust Uncamp](#), September 10-11, 2012.

Heather Christenson, John Wilkin: “Intellectual Property Rights and the HathiTrust Collection” (forthcoming), September 26, 2012.

Jeremy York: “[A Preservation Infrastructure Built to Last: Preservation, Community, and HathiTrust](#)”, September 26, 2012.

There’s an
elephant in
the library.™





Update On September Activities

Ingest

Internet Archive Digitization

HathiTrust coordinated with the University of Florida on upcoming deposit of volumes, and ingested a new batch of volumes from Penn State.

Local Digitization

HathiTrust ingested a new set of volumes from Utah State University Press and began conversations with the University of Delaware about processes and requirements for deposit of locally-digitized content. HathiTrust also corresponded with the University of Iowa about use of the [new tools](#) for validating and packaging locally-digitized materials for deposit. Institutions with questions about the new tools should contact feedback@issues.hathitrust.org.

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 Working Group

The Communications Working Group continued to follow developments in HathiTrust governance, and to evaluate how the communications function in HathiTrust might be improved once the new governance structure is in place. The [survey](#) for HathiTrust training and information sessions has closed, and HathiTrust will use the results as a basis for upcoming informational events. If you did not have a chance to submit feedback and would like to, please email responses to the survey to feedback@issues.hathitrust.org.

User Experience Advisory Group

The User Experience Advisory Group continued discussions about a new home page design and provided feedback on mockups created by the University of Michigan.

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) is in the final phase of development to bring Zephir into parity with the existing bibliographic management system at the University

You can follow HathiTrust on [Twitter](#) or [Facebook](#)

[Subscribe to email updates](#) (via [Google Groups](#))

There's an
elephant in
the library.™





Update On September Activities

of Michigan. Once Zephir is in operation, institutions will submit bibliographic records for volumes they plan to deposit to Zephir, and Zephir will produce exports of bibliographic data that will be used in HathiTrust Web services. In October, as part of preparations for integration testing with HathiTrust systems, CDL staff will begin preliminary testing of the Zephir outputs to evaluate system performance and confirm the structure of outputs (that they have the correct metadata fields, etc.). CDL has been contacting institutions that are contributing records to HathiTrust on an ongoing-basis to test the process for submitting bibliographic records to Zephir. If your institution is not contributing content to HathiTrust currently but you would like to test the new submission process, please contact feedback@issues.hathitrust.org.

Copyright Review

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

	September		Overall	
	Opened	Reviewed	Opened	Reviewed
CRMS-US	4,700	9,176	174,695	330,059
CRMS-World	3,656	7,191	10,248	22,266
Total	8,356	16,367	184,943	352,325

IMLS Quality Grant

The project team finalized a catalog of commonly-seen illustration errors in HathiTrust volumes for a sub-study on illustrative error. Donald Williams, a renowned research imaging scientist, analyzed the errors and met with members of the project team to explain the sources of the errors and possibilities for correction.

The project team continued work on the design of user studies to evaluate project findings, collection of data to support the user studies, and administration of the user studies themselves. Team members also discussed ways that quality review interfaces developed during the grant might be modified to support the certification of individual volumes. For more information on the project, please visit the [project website](#).

mPach

Staff at the University of Michigan created a mockup of PageTurner changes that will be needed to navigate the XML-based journal articles that will be submitted via mPach. Work also continued on modifications to PageTurner to display JATS XML and embedded media and on refinements to the METS specification for mPach Submission Information Packages. Staff completed wireframes and began coding the Dashboard module (see the [list of mPach modules](#) for more information). Michigan staff members will present on mPach at the 2012 DLF Forum.





Update On September Activities

Development Updates

Accessibility

HathiTrust has completed the first phase of improvements to enhance accessibility of HathiTrust Web applications. With a few minor exceptions that will be addressed in the second phase, HathiTrust interfaces are now compliant with [Web Content Accessibility Guidelines \(WCAG\) 2.0, Level A](#). The second phase will target compliance with WCAG 2.0 Level AA, and include usability testing by users who have print disabilities.

Data API

As of October 1, all requests to the Data API must be signed with an access key provided by HathiTrust. Details are available at http://www.hathitrust.org/data_api.

The Data API is being configured to deliver watermarked image derivatives in JPEG and PNG formats at a range of resolutions. The API currently delivers un-watermarked master images from the repository in TIFF and JP2. Enhancements to the Data API Web client were made to support image derivatives when they become available through Data API, and development-level debugging.

Full-text Search

University of Michigan staff modified the full-text search indexing process to prevent volumes from being indexed on more than one shard (section) of the full-text Solr index. Staff also began testing full-text search using Solr 4.0 Beta. Solr 4.0 offers new ranking algorithms that may provide better relevance ranking for long documents (e.g., books). A [paper](#) by Michigan developer Tom Burton-West on full-text search relevance ranking in HathiTrust was published in the INEX 2012 pre-proceedings as part of the [CLEF Labs Working Notes](#).

Following several months of informal research, Michigan staff began focused investigation into high-performance storage systems to improve full-text search response time and substantially increase search throughput capacity. An RFP for a new high-performance storage system will be issued in October.

Total Volumes Added	September	Overall
Boston College	0	1,816
Columbia University	0	64,184
Cornell University	82	408,837
Duke University	0	4,523
Harvard University	0	235,983
Indiana University	0	187,683
Library of Congress	0	89,722
North Carolina State University	0	3,196
University of North Carolina - Chapel Hill	0	8,088
Northwestern University	7	7,221
New York Public Library	0	259,571
Penn State University	113	44,131
Princeton University	0	251,644
Purdue University	2,418	40,466
Universidad Complutense	0	111,899
University of California	796	3,373,872
University of Chicago	238	24,917
University of Illinois	9	101,010
University of Michigan	22,241	4,582,544
University of Minnesota	115	102,616
University of Wisconsin	2,993	545,788
University of Virginia	0	50,790
Utah State University	27	117
Yale University	0	23,678
Total	29,039	10,524,296

Public Domain (~30% of total)

Total*	24,016	3,211,760
--------	--------	-----------

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

There's an elephant in the library.™





Update On September Activities

Imgsrv

Imgsrv is the web application that serves derivatives of HathiTrust’s master images to Web applications such as the PageTurner. HathiTrust has enhanced Imgsrv to deliver HTML derivatives of born-digital content in support of mPach and JATS XML.

PageTurner

HathiTrust implemented interface improvements designed by Michigan’s User Experience department for cases where special access to HathiTrust materials is available, such as access by users who have print disabilities. The improvements include dismissible notifications when special access is in effect, and updated explanatory text when special access that might be expected is not available (special access cases are described in HathiTrust’s [Access and Use Policies](#)). Special access is currently only available as a pilot at the University of Michigan. Extension of special access to other member institutions is still planned. More information will be forthcoming.

HathiTrust’s [embeddable](#) Pageturner is now based on the mobile Pageturner interface, which offers improved presentation and greater functionality.

HathiTrust has updated the version information displayed in the PageTurner to include the time a volume was removed from HathiTrust. Volumes may be removed from HathiTrust at the request of the rights holder, or in cases where the volume is wholly unusable or a superior copy is available.

Outages

From 1:00pm on Tuesday, September 25 to 8:30am on Friday, September 28, some bibliographic data failed to display in HathiTrust due to an outage of the system at Michigan that manages bibliographic data for HathiTrust.

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.

User Support Issues	Sept.	August
Content	248	286
Quality	242	279
Non-partner Digital Deposit Collections	0	1
	2	3
Cataloging	80	142
Access and Use	116	119
Copyright	71	62
Permissions	5	15
Takedown	2	0
Print on Demand	0	1
Inter-library loan	4	8
Full-PDF or e-copy requests	11	21
Datasets	3	7
Data Availability and APIs	0	1
Reuse of content	1	4
Web applications	12	22
Functionality problems	4	8
Problems with login specifically	0	1
General questions about login	0	1
Partners setting up login	0	4
Usability issues	0	0
Feature requests	0	2
Partner Ingest	3	4
General	55	74
Partnership	10	9
Infrastructure	0	0
Miscellaneous	45	65
Total	514	647

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

