



Update On March Activities

April 12, 2013

Top News

Program Steering Committee Nominations

HathiTrust has requested nominations from partner institutions for the HathiTrust Program Steering Committee (PSC). The responsibilities of the Program Steering Committee are described in Article VII, Section 3 of the [HathiTrust Bylaws](#). Among the first areas of work to be undertaken by the PSC are the [ballot initiatives](#) passed at the 2011 Constitutional Convention, including expanding access to US government documents and creating infrastructure for shared monograph storage initiatives. Any member of a partner institution may submit nominations for the PSC until April 22 via the form at <http://goo.gl/TVoCN>.

HTRC Software Release

The HathiTrust Research Center reached a development benchmark in its release of production infrastructure to support data mining and textual analysis of volumes in HathiTrust.

The infrastructure includes an entrance portal, search and collection-building tools (using [Blacklight](#)), and access to SEASR analysis algorithms that can be run against the HathiTrust public domain corpus (more than 3 million volumes). In addition to the production services, the HTRC offers a development “sandbox”. The sandbox runs against non-Google scanned content (about 260,000 volumes) and provides a test-bed for interested researchers to experiment with writing their own algorithms for use in the HTRC infrastructure.

The production release concludes the first six month period in Phase 2 of development of the HTRC (Oct 2012-March 2014). Phase 2 will also include the development of the HTRC-Sloan-Cloud – infrastructure that will include additional mechanisms to allow secure, non-consumptive access to the entire HathiTrust corpus – and systems to accommodate the full 10.6 million HathiTrust volumes in the HTRC. For more information on HTRC services and testing of the production infrastructure, please join our HTRC-usergroup-l listserv at <https://list.indiana.edu/sympa/subscribe/htrc-usergroup-l>.

Government Documents Registry Analyst

HathiTrust is pleased to announce the hiring of Valerie Glenn to the [Government Documents Registry Analyst position](#). Valerie has served as a Federal Depository Librarian at both the University of Alabama and the University of North Texas, and has managed a variety of projects and activities related to government documents. Valerie brings deep expertise to a two-year initiative to begin to construct a comprehensive registry of U.S. federal government documents. This work is part of a larger HathiTrust effort to [expand access to US government documents](#). More information about the project is available at http://www.hathitrust.org/usgov-docs_registry.

April Forecast

Complete redesign of website, including testing and deployment.

Continue installation of new and replacement storage.

Papers & Presentations

Cory Snavely and Jeremy York, “[The HathiTrust Repository: Policy and Technical Issues in Building the Digital Archive](#)”, University of Michigan, March 1, 2013.

Stephen Downie, “[Unlocking the Secrets of 3 Billion Pages: Introducing the HathiTrust Research Center](#)”, University of Hong Kong, March 18, 2013.

Heather Christenson and John Wilkin, “[Intellectual Property Rights & the HathiTrust Collection](#)”, UNESCO The Memory of the World in the Digital Age Conference Proceedings, September 26-28, 2012.

Jeremy York, “[A Preservation Infrastructure Built to Last: Preservation, Community, and HathiTrust](#)”, UNESCO The Memory of the World in the Digital Age Conference Proceedings September 26-28, 2012.

There's an
elephant in
the library.™





Update On March Activities

HathiTrust Institution Survey

HathiTrust distributed a survey created by Syracuse University to gather information about institutions' experiences with HathiTrust. The survey includes questions about print disabilities services, special collections, digital humanities, use of HathiTrust, and technical implementation issues. The survey is available at <http://www.surveymonkey.com/s/9ZZ9KMW> until April 26. We encourage all partner institutions to participate. Results will be summarized and made available.

HathiTrust Board of Governors

During a March meeting, the Board of Governors reviewed the HathiTrust budget and planned a longer agenda for an in-person meeting in April.

Ingest

Local Digitization

HathiTrust prepared a survey to send to institutions that have indicated they intend to deposit locally-digitized materials. The purpose of the survey is to gauge interest in, and aid in determining a development timeline for, enhanced tools to assist in validating and packaging materials prior to submission to HathiTrust. The survey will be sent out in mid April. HathiTrust also provided support to several institutions making preparations to deposit locally-digitized content.

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 Experience Advisory Group

The User Experience Advisory Group was pleased to welcome a new member: Matt Morgan, Director of the Website, NYPL Office of Strategic Planning. The group continued to review elements of HathiTrust Web applications identified through user feedback and other means as being in need of improvement.

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

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

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

There's an
elephant in
the library.™





Update On March Activities

Staff from the California Digital Library (CDL) and the University of Michigan discussed implications of a new requirement that automated bibliographic rights determinations must occur at the University of Michigan rather than at the University of California. The teams expect to have revised requirements finalized in April. CDL staff are determining the impact that the change will have on the development timeline.

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in March is given below. See [CRMS-US](#) and [CRMS-World](#) for information.

	March		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	3,376	7,267	127,958	236,977
CRMS-World	3,082	5,590	21,289	39,212
Total	6,458	12,855	149,247	276,189

mPach

Staff at the University of Michigan discussed modifications that are planned to be made to the Collection Builder application in order to use it as a means to navigate from articles in a single journal to the journal’s “aboutware” (information about editorial boards, submission policies, etc.). Staff also discussed issues of discovering journal aboutware through the HathiTrust catalog, full-text search and Collection Builder interfaces, and user pathways for navigating between journal-level catalog records, article-level catalog records, and aboutware. More information about mPach is available at <http://www.hathitrust.org/mpach>.

Development Updates

HathiTrust institutions performed the following work related to applications and Web interfaces:

Collection Builder

Staff corrected issues in the display of authors and titles, added an option to remove collection items to a batch Collection Builder tool, and discussed ways of supporting very large collections. Staff also worked on the development of new features to be implemented as part of the Website Redesign (see below).

Digitization Sources

Staff planned a new back-end strategy for recording content digitization sources and associated access parameters, which are expressed in HathiTrust interfaces.





Update On March Activities

Full-text Search

Staff continued research to improve relevance ranking.

PageTurner

Staff re-engineered a tool for testing and debugging volume access controls.

Website Redesign

Staff continued work to implement a redesign of HathiTrust Web interfaces, using a unified framework for application code. Release of the new design is expected in April. Other improvements to be made in conjunction with the redesign include:

- Pagination of results in Collection Builder.
- The addition of book cover thumbnails to Collection Builder and full-text search results.
- Improved viewing interface in PageTurner and differential display of works depending on their reading order (right-to-left versus left-to-right).
- Ability to cancel full-book downloads.

A preview of the new redesign is available in the [online version](#) of this newsletter.

Storage Replacement Cycle

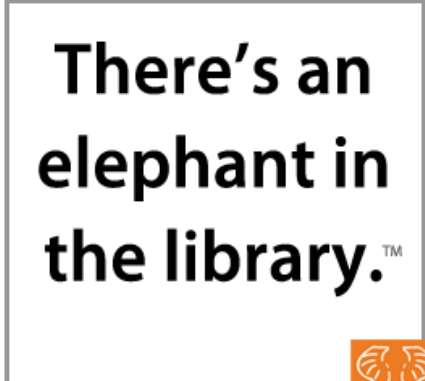
HathiTrust began to install new and replacement storage hardware at the Michigan repository instance as part of its regular purchase and replacement cycle. Installation of new storage and retirement of storage to be replaced will continue in April.

Server Replacement Cycle

HathiTrust purchased and received new production web servers and new development web and index servers to replace servers scheduled to be retired. The new development servers will make use of virtualization to improve resource utilization and availability, and to reduce acquisition and operational costs. In concert with this upgrade, which is planned for the second quarter of 2013, the Linux distribution in

Total Volumes Added	March	Overall
Boston College	337	2,179
Columbia University	0	65,033
Cornell University	1,563	418,525
Duke University	0	4,523
Harvard University	14	236,041
Indiana University	10	195,212
Library of Congress	0	89,723
North Carolina State University	0	3,196
Northwestern University	2,446	15,394
New York Public Library	8	259,680
Penn State	10	45,425
Princeton University	1	251,701
Purdue University	10	44,692
Universidad Complutense	0	111,982
University of California	1,152	3,387,448
University of Chicago	368	28,908
University of Florida	0	2,068
University of Illinois	5	109,311
University of Michigan	5,134	4,634,958
University of Minnesota	435	104,685
University of North Carolina - Chapel Hill	0	16,588
University of Wisconsin	752	555,707
University of Virginia	10	50,815
Utah State University	0	117
Yale University	0	23,678
Total	12,255	10,657,589
Public Domain (~31% of total)		
Total*	12,076	3,321,707

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





Update On March Activities

User Support Issues

	March	February
Content	382	430
Quality	373	421
Non-partner Digital Deposit	1	1
Collections	8	6
Cataloging	87	82
Access and Use	149	96
Copyright	77	51
Permissions	16	5
Takedown	0	1
Print on Demand	1	0
Inter-library loan	4	0
Full-PDF or e-copy requests	32	16
Datasets	5	7
Data Availability and APIs	2	3
Reuse of content	3	3
Web applications	11	15
Functionality problems	4	3
Problems with login specifically	1	0
General questions about login	2	3
Partners setting up login	2	2
Usability issues	1	0
Feature requests	0	4
Partner Ingest	13	1
General	87	74
Partnership	9	12
Infrastructure	0	0
Miscellaneous	78	62
Total	729	698

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

Most-accessed volumes

Title	#Visits*
The United States Strategic Bombing Survey: over-all report (European war)	5,006
Quicksand, by Nella Larsen	3,545
Investigation of Korean-American relations: Report of the Subcommittee on International Organizations of the Committee on International Relations, U.S. House of Representatives, October 31, 1978	2,618
Perfume and flavor materials of natural origin, by Steffen Arc-tander	2,044
Godey's magazine, v.40-41, 1850	1,680
Bradshaw's handbook for tourists in Great Britain & Ireland, sec.1 1866	1,112
McClure's magazine. v.12 1898/1899 Nov-Apr	
Shocking life, by Elsa Schiaparelli	1,304
Noblesa catalana: cavallers y burgesos honrats de Rossello y Cerdanya, v.2., by Philippe Laz-erme	1,030
Coffee processing technology, v.1., by Michael Sivetz and H. El-liott Foote	963

* Approximate due to a system configuration change.

use for the entire server infrastructure is being changed from Red Hat to Debian, to provide better and more manageable infrastructure for deploying Ruby-based applications.

Outages

No outages were reported in March.

There's an
elephant in
the library.™

