



Update On January Activities

February 14, 2014

Top News

Executive Director Search

The HathiTrust Executive Director Search Committee completed interviews with final candidates, and looks forward to announcing the successful conclusion of the search in the next few weeks.

Volumes from Keio University

HathiTrust is pleased to report the ingest of more than 80,000 volumes from Keio University. Volumes in the collection can be found here: <http://bit.ly/1gmFNw2>. These materials dramatically increase HathiTrust's Japanese-language holdings. Keio University will be providing more information about the materials included in the deposit in coming weeks. The volumes represent the largest deposit in HathiTrust of materials from a non-partner institution.

CRMS Milestone

The Copyright Review Management System project team is pleased to announce a major milestone: in January 2014, staff completed review of the copyright status of all the works in HathiTrust published in the United States from 1923 to 1963. In all, of the more than 300,000 volumes in HathiTrust published during this time and presumed to be in copyright, nearly 160,000 were found to be in the public domain and are now accessible to users worldwide. A great thanks is due to Indiana University, the University of Michigan, the University of Minnesota, and the University of Wisconsin for their dedicated review work on these materials since 2008. U.S. materials published from 1923-1963 will continue to be reviewed as they come into HathiTrust, and the [16 institutions participating in CRMS-World](#) (conducting review of works published outside the United States) will continue their work. The CRMS-US and CRMS-World projects are funded by the Institute of Museum and Library Services.

Government Documents Call for Records

More than 40 institutions, including HathiTrust partners and non-partners, submitted records in response to HathiTrust's [call for U.S. federal government document records](#). The records will be sent to Google for analysis in early February. We continue to welcome the submission of records. While the records might not be included in Google's analysis, they would still be a part of subsequent analysis conducted by HathiTrust partners and support HathiTrust's efforts to create a comprehensive [registry](#) of US federal government documents.

Nominations for User Support Working Group

The User Support Working Group is seeking nominations for up to 2 new members. We are seeking staff who have expertise in providing general user support and those who have expertise in cataloging in particular. To submit nominations and for further information about the working group, please visit <http://tinyurl>.

February Forecast

Continue work to add quick links to the PageTurner to embed HathiTrust volumes in web pages.

Continue work to support indexing of JATS articles and indexing of volumes in "chunks".

Continue development of ePub and PDF generation from JATS.

Continue to explore improvements to relevance ranking in full-text search.

Papers & Presentations

J. Stephen Downie, "Unlocking the Secrets of 3 Billion Pages: Introducing the HathiTrust Research Center", University of Oxford, January 22, 2014.

Jon Rothman, "HathiTrust and Bibliographic Metadata", ALCTS CaMMS Catalog Management Interest Group Meeting, American Library Association MidWinter Convention, January 25, 2014.

Partner-specific

Heather Christenson, "Building the Research Library at Web Scale", Special Libraries Association, San Francisco and BayNet Meeting, January 23, 2014.

There's an
elephant in
the library.™





Update On January Activities

[com/m9qlyyg](http://www.hathitrust.org/m9qlyyg).

Ingest

Validation service for locally-digitized materials

HathiTrust released in beta a new [full-volume validation and packaging service](#). Information about the new service and the single-page validation tool released in December, as well as a package of code modules that can be used to validate, remediate, and package materials for ingest, is available at http://www.hathitrust.org/ingest_tools. If you are interested in receiving updates related to these tools, please subscribe to the [HathiTrust Ingest Google Group](#). We are very interested in your [feedback](#) on the tools as well.

Locally-Digitized

Several institutions tested HathiTrust's new single-image and full-volume validation tools; Emory University and the University of Illinois experimented with submitting volumes to the full-volume service. HathiTrust corresponded with Universidad Complutense de Madrid and the University of Chicago about deposit of locally-digitized materials.

Internet Archive-digitized

HathiTrust ingested new content from the University of Illinois at Urbana Champaign, Duke University, and Boston College, and began conversations about ingest with the University of Alberta. The University of Connecticut and University of Massachusetts, Amherst also prepared to submit their first batches of content.

Google-digitized

In addition to the volumes from Keio University, HathiTrust began ingest of Google-digitized content from Ohio State.

Working Groups and Committees

Program Steering Committee

The Program Steering Committee is in the process of forming a Government Documents Initiative Planning and Advisory Group, chaired by Mark Sandler, in accordance with [one of the ballot initiatives](#) approved at the Constitutional Convention. The group is charged to "Facilitate collective action to create a comprehensive digital corpus of U.S. federal publications including those issued by GPO and other federal agencies," and to "Initiate and carry out a planning process to coordinate operational plans and a business model to further and sustain coordinated digitization, ingest, and display of U.S. federal publications including those issued by GPO and other federal agencies." The group will coordinate its efforts with work already

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

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

There's an
elephant in
the library.™





Update On January Activities

under way in HathiTrust to build a registry of U.S. government documents. The full charge can be found at http://www.hathitrust.org/usgovdocs_planning_charge, and more information about the initiative in general at <http://www.hathitrust.org/usgovdocs>.

The PSC also completed the charge for a reconstituted Collections Committee, and for a new Rights and Access Working Group. These groups are expected to begin work shortly. The PSC has identified a core set of members to participate in each group to get the initiatives underway. Once the charges and group chairs are confirmed, the PSC will be issuing a call for nominations for additional members.

Projects

Copyright Review

A summary of the determinations from HathiTrust copyright review activities in January is given below. See [CRMS-US](#) and [CRMS-World](#), projects funded by IMLS, for further information.

	January		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	272	800	158,442	306,294
CRMS-World	2,593	5,561	46,679	90,377
Total	2,865	6,361	205,121	396,671

Government Documents Registry

The Government Documents Registry project team continued to develop and test strategies to match and identify duplicate records, and to draft functional requirements for the registry. Team members also began to identify potential processes for identifying gaps in the registry.

HathiTrust Research Center

The HTRC drafted documents covering system architecture, workflows, security measures, and data use cases in preparation for offering “non-consumptive” access to in-copyright volumes in the HathiTrust repository. The HTRC hosted its first user group meeting, with discussions focusing on the [HTRC Bookworm demo system](#) and natural language processing applications used by scholars. The HTRC received 15 proposals in response to an [open RFP](#) for [WCSA](#) (Workset Creation For Scholarly Analysis: Prototyping Project). The team has identified a shortlist of 8 candidates to present their projects at an upcoming meeting in Chicago. Final selection of the four funded prototyping projects will be announced in March. Co-director Stephen Downie delivered a lecture at Oxford University on January 22 on scholarly uses of HTRC resources.





Update On January Activities

mPach

University of Michigan staff made changes to HathiTrust indexing mechanisms to support JATS XML and prepared a poster on mPach to present at the Library Publishing Forum 2014.

Zephir

California Digital Library (CDL) loaded 348,842 new or updated bibliographic records from partners into [Zephir](#). Bibliographic records are required for volumes to be ingested into HathiTrust. Information about bibliographic metadata submission is available at http://www.hathitrust.org/bib_data_submission.

Development Updates

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

Full-text Search

Staff received and installed networking equipment to connect the new high-performance storage for full-text search at the Michigan and Indiana repository instances. Staff also completed an upgrade of storage controller modules at each site, which was recommended by the supplier, modified the full-text index synchronization and release process to accommodate the new storage, and began conducting live performance testing using the new storage.

Staff continued coding to support indexing of JATS XML content and indexing of volumes into a configurable number of “chunks” which has the potential to improve relevance ranking of large volumes.

Staff tested algorithms to index words that are hyphenated across line breaks. Production deployment of the algorithms is planned within the next few months. Staff also did preliminary investigation into processes to perform practical, automated, OCR correction. There is no timeline currently for release of these processes.

Server Replacement Cycle

Staff rebuilt and redeployed production web servers at the Michigan instance to match newly-deployed web servers in Indiana, completing the upgrade of production web servers.

Storage Replacement Cycle

Staff received new storage for the annual growth and replacement cycle, and completed installation at the Michigan site. Installation at the Indiana site is scheduled for February. Storage due to be retired will be taken offline in March.

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





Update On January Activities

Total Volumes Added

	January	Overall
Boston College	323	2,686
Columbia University	0	65,036
Cornell University	3,720	441,211
Duke University	1,339	5,864
Harvard University	0	237,435
Indiana University	0	195,580
Keio University	80,125	80,125
Library of Congress	0	89,724
North Carolina State University	0	3,196
Northwestern University	78	37,580
New York Public Library	0	288,370
Penn State	1,219	69,423
Ohio State University	6	6
Princeton University	0	251,710
Purdue University	3	44,698
Texas A&M	0	1,201
Universidad Complutense	0	112,014
University of California	6,028	3,454,198
University of Chicago	357	39,992
University of Florida	0	9,763
University of Illinois	2,640	115,615
University of Michigan	1,406	4,667,438
University of Minnesota	2,685	118,620
UNC - Chapel Hill	0	17,025
University of Wisconsin	2	555,926
University of Virginia	0	50,821
Utah State University	0	117
Yale University	0	23,678
Total	99,931	10,978,052

Public Domain (~33% of total)

Total*	73,668	3,615,823
--------	--------	-----------

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

Most-accessed volumes

- Godey's magazine. v.40-41 1850.
- Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.
- The Sunlight Book of Knitting and Crocheting, by Adelaide Gray.
- The Human Figure, by John H. Vanderpoel.
- History of wages in the United States from Colonial times to 1928, United States Department of Labor.
- Bradshaw's handbook for tourists in Great Britain & Ireland. Sec 1, 1866.
- The making of the University of Michigan, 1817-1992, by Howard H. Peckham.
- The Five Laws of Library Science, by S. R. Ranganathan.
- Highway safety, design, and operations; freeway signing and related geometrics. 4.P 96/11:90-39. Hearings, Ninetieth Congress, second session.
- Freight car distribution and car handling in the United States, by Eugene W. Coughlin.

Availability

Repository

Cumulative 12-month availability of repository access: 99.827%*

Users were not able to submit feedback using HathiTrust's Feedback link from approximately 11:00pm on Sunday, January 5 to 3:30pm on Tuesday, January 7 due to a software problem.

There's an elephant in the library.™





Update On January Activities

HathiTrust may have been inaccessible to some users on Monday, January 13 from 9:23-9:30am, on Tuesday, January 14 from 1:40-1:50pm, and on Thursday, January 16 from 8:25-9:30am due to temporarily exhausted scratch storage space on newly-deployed web servers.

University of Michigan users may not have been able to log in to HathiTrust on Tuesday, January 14 from 6am - 12:37pm due to a configuration error on a newly-deployed web server.

* Repository access refers to page viewing and full-text search functionality, i.e., user-facing applications. It does not refer to preservation or storage infrastructure, which is under continual operation.

Zephyr

A maintenance outage is planned on the Zephyr FTPS server on February 19, 2014 from 6:00-6:30am PST. Zephyr systems other than the FTPS server will not be affected. During the maintenance outage, contributors will not be able to submit bibliographic records.

User Support Issues	January	December
Content	102	188
Quality	86	179
Collections	15	8
Cataloging	142	151
Access and Use	114	130
Copyright	59	88
Permissions	8	7
Takedown	2	0
Print on Demand	0	0
Inter-library loan	2	0
Full-PDF or e-copy requests	22	11
Datasets	6	2
Data Availability and APIs	0	1
Reuse of content	2	2
Web applications	22	21
Functionality problems	9	7
Problems with login specifically	2	0
General questions about login	2	3
Partners setting up login	1	0
Usability issues	0	0
Feature requests	1	1
Partner Ingest	8	4
General	75	77
Partnership	10	2
Infrastructure	0	0
Miscellaneous	65	75
Total	462	571

*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.™