



Update On February Activities

March 14, 2014

Top News

Executive Director Appointed

We are very pleased to announce the appointment of Mike Furlough as the Executive Director of HathiTrust. Mike will begin as Executive Director on May 19. The full announcement can be read at http://www.hathitrust.org/mike_furlough_executive_director.

11 Million Volumes

HathiTrust reached a new milestone, surpassing 11 million volumes in the digital repository. A history of HathiTrust's road to the first 10 million volumes is available on the [HathiTrust blog](#).

Updated HathiTrust Volume Identifiers

HathiTrust has made a one-time, batch change to a set of approximately 320,000 volume identifiers. These volumes were ingested with an incorrect identifier due to a vendor issue. The change involves adding a \$ symbol to affected identifiers. A [full list](#) of the updated identifiers is available at <http://www.hathitrust.org/hathifiles>. Any institutions or individuals that save links to HathiTrust volumes locally should update these identifiers to ensure working links. Please contact feedback@issues.hathitrust.org with any issues or questions.

Ingest

Locally-Digitized

HathiTrust ingested new content from the Universidad Complutense de Madrid, received content from the University of Delaware, and communicated with Emory University, University of Chicago, and University of Washington about submission of locally-digitized content.

Internet Archive-digitized

HathiTrust ingested new content from the University of Massachusetts, Amherst, and continued conversations about ingest with the University of Alberta.

Zephir

California Digital Library (CDL) loaded 71,778 new or updated bibliographic records from partners into [Zephir](#). Information about bibliographic metadata submission is available at http://www.hathitrust.org/bib_data_submission.

March Forecast

Continue development of ePub and PDF generation from JATS.

Deploy the new version of SLIP, for full-text indexing.

Continue to explore relevance ranking solutions.

Papers & Presentations

J. Stephen Downie, "Unlocking the Secrets of 3 Billion Pages: Introducing the HathiTrust Research Center", University of Tsukuba, Japan, Feb 13, 2014.

Jeremy York, "HathiTrust Overview: Partnership and Services", Wesleyan University Web presentation, Feb 18, 2014.

There's an
elephant in
the library.™





Update On February Activities

Working Groups and Committees

Program Steering Committee

The PSC continued bi-weekly meetings, focusing discussions on the [HathiTrust Distributed Print Monographs](#) proposal and a proposed HathiTrust metadata sharing and use policy.

Projects

Copyright Review

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

| | February | | Overall | |
|------------|---------------|--------------------|---------------|--------------------|
| | Public Domain | All Determinations | Public Domain | All Determinations |
| CRMS-US | 2,561 | 2,727 | 161,510 | 309,548 |
| CRMS-World | 2,670 | 5,320 | 49,832 | 96,402 |
| Total | 5,231 | 8,047 | 211,342 | 405,950 |

Government Documents Registry

Project staff continued to draft functional requirements for the registry, and are in the process of obtaining initial feedback on the requirements from selected members from HathiTrust partner and non-partner institutions. Staff also continued to develop methods for identifying duplicate and related records, and explore ways the US government documents community could contribute to the development of the registry.

HathiTrust Research Center

The HTRC invited eight finalist candidates in an [RFP](#) for [WCSA](#), a Mellon Foundation-funded project to support the prototyping of workset creation tools, to Chicago to present their proposals. Four of the candidates will be awarded grants of \$40,000 over 9 months to develop their prototypes.

mPach

University of Michigan staff began to migrate the Prepper module of mPach to a new Ruby/Rails development environment (a full list of mPach modules is available at <http://www.lib.umich.edu/mpach>). Staff added an mPach article to the HathiTrust test repository, and began to evaluate additional tools for converting articles into JATS XML that might be incorporated into the Norm component of

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





Update On February Activities

Prepper.

Development Updates

HathiTrust institutions performed the following work related to applications and infrastructure:

Full-text Search

Staff continued to test and refine the index synchronization and release process on new high-performance storage for full-text search. After stability problems were encountered during attempts to roll out the new storage in production, staff began working with the storage and network equipment suppliers to troubleshoot and optimize performance. (See Availability, below.)

Staff finished developing and testing a new version of SLIP (Solr Large-scale Indexing Processor), which is used to index the full-text of works in HathiTrust. Production deployment will occur in March. Staff added features to support the indexing of JATS XML content, and indexing of volumes into a configurable number of “chunks”. Staff have been exploring chunking volumes at indexing time in order to improve the relevance ranking of search results. Staff also added indexing support for words that are hyphenated across line breaks on pages of text. This is effective immediately for searches conducted within volumes and will take effect for volumes in cross-repository searches as volumes are indexed going forward. Approximately 4.5 million HathiTrust volumes will be re-indexed in mid-March during a regular monthly update of HathiTrust partner print holdings information; a complete re-indexing process is planned for late April. Staff additionally integrated a spelling suggester feature into a Solr request handler in development and began testing the suggester with several data sets.

Pageturner

Staff at California Digital Library developed an “Embed this Book” feature that is now available in the “Share” section of the PageTurner sidebar. Users can copy the HTML for embedding either 1up or 2up views into websites and blogs.

Storage Replacement Cycle

Staff completed installation of new and replacement storage for the 2014 cycle. Retired storage will undergo security wiping in March and be returned to fulfill trade-in credit obligations.

Availability

Repository

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

There's an
elephant in
the library.™





Update On February Activities

HathiTrust was unavailable for some or all users on Monday, February 3 from 12:05-12:10pm and Tuesday, February 4 from 1:45-1:55am and 6:45-7:00am due to stability problems encountered during attempted production rollouts of new high-performance storage for full-text search.

HathiTrust was unavailable for some or all users on Thursday, February 20 from 2:53-3:07pm due to a temporary network issue at the Michigan instance that occurred while the Indiana instance was out of service for routine maintenance.

* 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 occurred on the Zephyr FTPS server on March 6, 2014 from 6:00-6:30am PST. During the brief maintenance outage, contributors were not able to submit bibliographic records. Zephyr systems other than the FTPS server were not affected and maintenance was conducted successfully.

| User Support Issues | February | January |
|----------------------------------|------------|------------|
| Content | 220 | 120 |
| Quality | 200 | 86 |
| Collections | 18 | 15 |
| Cataloging | 165 | 142 |
| Access and Use | 130 | 114 |
| Copyright | 82 | 59 |
| Permissions | 16 | 8 |
| Takedown | 0 | 2 |
| Print on Demand | 0 | 0 |
| Inter-library loan | 0 | 2 |
| Full-PDF or e-copy requests | 21 | 22 |
| Datasets | 7 | 6 |
| Data Availability and APIs | 0 | 0 |
| Reuse of content | 2 | 2 |
| Web applications | 29 | 22 |
| Functionality problems | 13 | 9 |
| Problems with login specifically | 13 | 9 |
| General questions about login | 2 | 2 |
| Partners setting up login | 3 | 1 |
| Usability issues | 0 | 0 |
| Feature requests | 2 | 1 |
| Partner Ingest | 2 | 8 |
| General | 112 | 75 |
| Partnership | 5 | 10 |
| Infrastructure | 0 | 0 |
| Miscellaneous | 107 | 65 |
| Total | 658 | 462 |

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





Update On February Activities

| Total Volumes Added | February | Overall |
|--------------------------------------|---------------|-------------------|
| Boston College | 110 | 2,796 |
| Columbia University | 1 | 65,037 |
| Cornell University | 3,120 | 444,331 |
| Duke University | 1,394 | 7,258 |
| Harvard University | 0 | 237,435 |
| Indiana University | 0 | 195,580 |
| Keio University | 8,829 | 88,954 |
| Library of Congress | 18,205 | 107,929 |
| New York Public Library | 2 | 288,372 |
| North Carolina State University | 0 | 3,196 |
| Northwestern University | 21 | 37,601 |
| Ohio State University | 19,439 | 19,445 |
| Penn State | 1,906 | 71,329 |
| Princeton University | 0 | 251,710 |
| Purdue University | 0 | 44,698 |
| Texas A&M | 0 | 1,201 |
| Universidad Complutense | 133 | 112,147 |
| University of California | 7,725 | 3,461,923 |
| University of Chicago | 85 | 39,077 |
| University of Florida | 2 | 9,765 |
| University of Illinois | 10,988 | 126,603 |
| University of Massachusetts, Amherst | 8,731 | 8,731 |
| University of Michigan | 1,043 | 4,668,481 |
| University of Minnesota | 1,148 | 119,768 |
| UNC - Chapel Hill | 0 | 17,025 |
| University of Virginia | 0 | 50,821 |
| University of Wisconsin | 21 | 555,947 |
| Utah State University | 0 | 117 |
| Yale University | 0 | 23,678 |
| Total | 82,903 | 11,060,955 |

Public Domain (~33% of total)

| | | |
|--------|--------|-----------|
| Total* | 59,381 | 3,675,204 |
|--------|--------|-----------|

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

Most-accessed volumes

- [Organized crime in America: hearings before the Committee on the Judiciary, United States Senate, Ninety-eighth Congress, Pt. 1.](#)
- [Quicksand, by Nella Larsen.](#)
- [The Cosmopolitan, v.72 \(1922\).](#)
- [The Utopia of Sir Thomas More, ed. with introduction, notes, and glossary by William Dallam Armes.](#)
- [Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.](#)
- [Quintus Curtius \[History of Alexander\] with an English translation by John C. Rolfe.](#)
- [The Human Figure, by John H. Vanderpoel](#)
- [The making of the University of Michigan, 1817-1992 / Howard H. Peckham.](#)
- [Concepts in Calculus, III : Multivariable Calculus](#)
- [Roster of the Confederate soldiers of Georgia, 1861-1865, v.3.](#)

There's an elephant in the library.™

