HathiTrust Digital Library

SPECIAL EDITION - 2013 Mid-Year Review

June 28, 2013

HathiTrust continued to grow in partnering institutions and digital content in the first half of 2013. Significant enhancements to HathiTrust services and strategic initiatives, including a new service for users who have print disabilities, a complete redesign of HathiTrust interfaces, the release of the HathiTrust Research Center, and an important partnership with the Digital Public Library of America, showed the vitality of an organization that continues to break new ground as we become an increasingly important part of the 21st century library landscape.

2013 marks an important juncture for HathiTrust, as John Wilkin, the founding Executive Director, steps down from his position to become the University Librarian and Dean of Libraries at the University of Illinois. John wrote a special note in the Update on May 2013 Activities. Under John's leadership, the HathiTrust partners have taken steps toward collective action on a scale that we have never before seen. We look forward to what we will be able to accomplish in the coming years together.

A summary of HathiTrust activities in the first half of 2013 is given below.

Highlighted Achievements and Activities

Details on each item can be found in the monthly updates from 2013, available at http://www.hathitrust.org/updates.

New Partners

HathiTrust grew from 78 to 84 partners in the first half of 2013. New institutions include:

- Brown University
- Tufts University
- University of Alberta
- University of Houston
- University of Oklahoma
- Wake Forest University

New Content

Partners contributed 133,771 volumes to the repository. 107,396 of these are in the public domain. Contributions of new content are shown in a table at the end of the review.

Locally-digitized content

HathiTrust provided guidance to numerous institutions in the process of preparing locally-digitized content for deposit. HathiTrust hosted a conference call to discuss development options for the next iteration of tools to be made available to aid in packaging materials for deposit, and surveyed partners to determine a development schedule for coming enhancements. Notes from the conference call are available. Anyone who is interested is encouraged to join our HathiTrust Ingest Google Group for discussions and updates related to ingest of locally-digitized materials.

Governance and Working Groups

Board of Governors

The Board of Governors met in person in April to consider a full agenda of items, including the appointment of members to the Program Steering Committee, and the ballot initiatives passed at the 2011 Constitutional Convention:

- To develop a distributed archive of print monographs
- To create an approval process for development initiatives
- To establish a new governance structure
- To make revisions to HathiTrust's mission and goals

- To expand access to United States government documents
- To develop a fee-for-service model for content deposit

The Board provided this update on the ballot initiatives in May.

HathiTrust Bylaws Accepted

HathiTrust institutions voted unanimously to accept bylaws put forward by the Board of Governors.

Program Steering Committee

The HathiTrust Board of Governors appointed the inaugural members of the Program Steering Committee (PSC). Among the first areas of work to be undertaken by the PSC are two of the ballot initiatives passed at the Constitutional Convention: expanding access to US government documents and creating infrastructure for shared monograph storage initiatives.

Nominations for the User Support Working Group

HathiTrust is accepting nominations for its User Support Working Group through July 1. Nominations can be submitted at http://tinyurl.com/m9qlyyg.

A summary of the issues received by the User Support Working Group is shown in a table at the end of the review.

User Experience Advisory Group

The UX Advisory group welcomed new member Matt Morgan of NYPL. The group reviewed and worked to prioritize elements of HathiTrust Web applications that have been identified by users or staff as being in need of improvement.

Special Initiatives

HathiTrust Partners With the DPLA

HathiTrust and the DPLA announced a formal partnership, with HathiTrust participating in DPLA as a Content Hub. From the news release: "According to HathiTrust Executive Director John Wilkin, the partnership reflects the complementary nature of the two organizations. 'The first priority of HathiTrust has always been preservation...But to fulfill the preservation mission, we must provide access: content that can't be found and used risks being forgotten.'"

Expanded Access for Users Who Have Print Disabilities

HathiTrust released a new service that allows designated proxies at partner institutions to provide access to in-copyright works in HathiTrust to users at their institutions who are certified as having a print disability. See http://www.hathitrust.org/accessibility for more information.

Website Redesign

The HathiTrust website, including all Web applications, was updated with a unified design and feature set, improving the overall look and functionality of the site. Details are available at http://www.hathitrust.org/hathitrust_new_look.

Government Documents Registry

HathiTrust hired Valerie Glenn as a Government Documents Registry Analyst to initiate work on a comprehensive registry of US federal government documents. This work is part of the larger HathiTrust effort to expand access to US government documents.

Bibliographic Data Management

California Digital Library (CDL) staff worked with staff at the University of Michigan to prepare Zephir, the new bibliographic management system for HathiTrust, to "plug in" to existing repository workflows. This included testing data exports from Zephir, working to ensure that the same records are loaded in both systems, and accommodating a new requirement that algorithmic bibliographic rights determinations on volumes take place at the University of Michigan rather in Zephir. CDL also performed a successful test of backup systems. CDL and Michigan are planning a phase later this summer when both systems will be run in parallel, prior to a full cutover from the Michigan system to Zephir.

Copyright Review

A summary of the determinations from HathiTrust copyright review activities from the first half of 2013 is given below. See CRMS-US and CRMS-World for further information.

	Jan-June 2013		Overall	
	Public Domain	All	Public Domain	All
CRMS-US	41,268	79,817	138,353	258,984
CRMS-World	13,439	23,510	29,224	54,444
Total	54,707	103,327	167,577	313,428

HathiTrust Research Center (HTRC)

The HathiTrust Research Center reached a significant milestone with its release of production infrastructure to support data mining and textual analysis of public domain volumes in HathiTrust (see the HathiTrust Updates on February and March Activities for details, and information about Getting Started with the HTRC). The release concluded Phase I of HTRC development. The HTRC immediately kicked off its second phase, which has two areas of focus: first, community engagement and community-driven enhancements to HTRC services, and second, the development of the HTRC-Sloan-Cloud – infrastructure that will allow secure, non-consumptive access to the entire HathiTrust corpus. The HTRC will hold its second annual UnCamp on September 8-9, 2013 in Champaign-Urbana, Illinois. More information about the HTRC, including links to join listservs for the UnCamp, general announcements, and user discussion, is available at http://www.hathitrust.org/htrc.

mPach

Work by University of Michigan staff focused primarily in three areas: specifying modifications to HathiTrust applications that will be needed to properly associate articles from a single journal with each another, and with information about the journal as a whole; making enhancements to the HathiTrust PageTurner to display JATS XML articles; and developing procedures to handle ingest of non-JATS content (e.g., data) that is embedded in articles or submitted with articles as supplementary material. Mockups of the new interface for viewing articles in HathiTrust are posted on the mPach website. A timeline for the project is posted on the HathiTrust project page.

Repository

Development in the first half of 2013 included the following:

New Functionality / Application Changes

Collections Application

- The addition of pagination to collection search results.
- The addition of book cover thumbnails (also added to full-text search results).
- The addition of backend functionality to batch-

remove collection items.

Full-text search

- The addition of a checkbox to the advanced full-text search page, allowing users to limit a search to items held in print by their institution. The checkbox appears only to authenticated members of partner institutions.
- Improvements to the synchronization of the full-text index from the Michigan repository instance to the instance in Indiana.
- Improvements to indexing of partner print holdings information, and optimization of indexing when maintenance or large updates affecting full-text indexing are underway.
- Design of processes to index JATS XML articles.
- Significant work was undertaken to improve relevance ranking in full-text search results and in the development of a spelling suggestion feature.

PageTurner

- Improvements to the viewing interface (larger viewing space and improved layout).
- Introduction of mechanisms to display works appropriately depending on their reading order (right-to-left versus left-to-right).
- Ability to cancel full-book downloads.
- Removal of the restriction on the number of simultaneous accesses available to users at HathiTrust partner institutions who have print disabilities, per print copy of a volume owned by the user's institution.
- Stylistic changes to messages in mobile PageTurner that appear when special access to materials is granted (e.g., access to volumes that fall under Section 108 conditions or access to users who have print disabilities).
- Updates to the way URL parameters are sent to Google Analytics in order to improve usage

reporting for full-text searches within individual volumes.
- A fix to a bug that prevented PDFs that are read from right to left from being searchable.
- The addition of a special notice to PDFs generated by proxies for users who have print disabilities;
- Staff continued work to enable the delivery of JATS XML articles as PDFs.

Print on demand

- New functionality to produce PDFs optimized for printing on Espresso Book Machines.

Website redesign

- Completion of a major project to redesign and add functionality to HathiTrust Web interfaces and services.

Infrastructure Changes

Purchase of new production web servers and new development web and index servers to replace servers scheduled to be retired (in process).

- 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, the Linux distribution in 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.

Installation of new storage at the Indiana and Michigan repository sites to accommodate 2013 volume projections and replace storage scheduled for retirement (in process).

HathiTrust Digital Library

SPECIAL EDITION - 2013 Mid-Year Review

Papers and Presentations

All papers and presentations are listed at http://www.hathitrust.org/papers.

Highlighted Web Pages

HathiTrust Resources - Looking for one-page flyers about HathiTrust or information about using the digital library? Wondering how other partners explain HathiTrust to their users? HathiTrust Resources is your one-stop shop.

What's in your collection - Ever thought about building a collection in HathiTrust? There's never been a better time. Visit this page to learn more.

Volumes Added	Jan-June	Overall
Boston College	337	2,179
Columbia University	643	65,033
Cornell University	4,401	419,836
Duke University	0	4,523
Harvard University	83	236,068
Indiana University	164	195,237
Library of Congress	2	89,724
North Carolina State University	0	3,196
Northwestern University	21,247	33,969
New York Public Library	28,769	288,343
Penn State	14,455	59,187
Princeton University	53	251,704
Purdue University	63	44,692
Universidad Complutense		111,982
University of California	5,540	3,388,795
University of Chicago	3,671	30,391
University of Florida	60	2,068
University of Illinois	4,431	109,318
University of Michigan	33,295	4,643,131
University of Minnesota	2,607	106,819
University of North Carolina - Chapel Hill	8,500	16,588
University of Wisconsin	5,353	555,733
University of Virginia	16	50,815
Utah State University	0	117
Yale University	0	23,678
Total	133,771	10,733,126

Public Domain (~31% of total)

	Jan-June	Overall
Total*	107,396	3,386,026

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

HathiTrust Digital Library

SPECIAL EDITION - 2013 Mid-Year Review

User Support Issues

User Support Issues	Jan-June 2013	Jan-June 2012
Content	**1,057**	**975**
Quality	997	979
Non-partner Digital Deposit	2	5
Collections	60	30
Cataloging	**605**	**220**
Access and Use	**822**	**771**
Copyright	463	451
Permissions	68	95
Takedown	2	7
Print on Demand	2	2
Inter-library loan	10	2
Full-PDF or e-copy requests	119	109
Datasets	34	13
Data Availability and APIs	7	7
Reuse of content	23	12
Web applications	**137**	**109**
Functionality problems	41	29
Problems with login specifically	6	6
General questions about login	12	15
Partners setting up login	10	14
Usability issues	13	6
Feature requests	11	12
Partner Ingest	**35**	**18**
General	**380**	**604**
Partnership	65	55
Infrastructure	2	4
Miscellaneous	313	545
Total	**2,976**	**2,697**

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

Most-accessed volumes

Title
Quicksand, by Nella Larsen
Startling facts for native Americans called "know-nothings"; or, A vivid presentation of the dangers to American liberty, to be apprehended from foreign influence
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
Mechanick exercises: or The doctrine of handy-works, by Joseph Moxon
The Tosa diary, by Tsurayuki Ki, translated by William N. Porter
Theoretical nuclear physics, by John M. Blatt [and] Victor F. Weisskopf
Handbook of American Indian languages, Part 1, by Franz Boas
The exploration of the Colorado River and its canyons, by J. W. Powell
What price civilization; the causes, prevention and cure of human cancer, by Charles Eliot Perkins
Air Commerce Bulletin, v.2 Jul-Jun 1930-31

About HathiTrust

HathiTrust is an international partnership of academic and research institutions dedicated to ensuring the preservation and accessibility of the vast record of human knowledge. The partnership owns and operates a digital repository containing millions of public domain and in-copyright volumes, digitized from partnering institution libraries and other sources. The preserved volumes are made available in accordance with copyright law as a shared scholarly resource for students, faculty, and researchers at the partnering institutions and as a public good to the world community. For more information, visit HathiTrust.org. You can follow HathiTrust on Facebook and Twitter.