HathiTrust Digital Library

SPECIAL EDITION - 2013 Year in Review

2013 was a year of significant growth and development for HathiTrust. The partnership gained more than a dozen new partners, including two in Canada and one in Australia, and forged closer ties with emerging collaborations such as the Digital Public Library of America and the Digital Preservation Network. HathiTrust continued its work to open access to publications through copyright review, licenses from rights holders, and a new arrangement with Knowledge Unlatched. The Board appointed a new Program Steering Committee to carry forward partner initiatives and began a search for a new Executive Director to lead HathiTrust in its next phase. HathiTrust revolutionized opportunities for accessing and using its collections through the release of the HathiTrust Research Center (HTRC) and expanded access for users at partner institutions who have print disabilities. HathiTrust released Zephir, a new bibliographic management system developed by the University of California, and completed a major re-design of its web interfaces. A recap of these activities and more can be found in the review 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

14 institutions joined HathiTrust in 2013:

- Allegheny College
- Brown University
- Colby College
- Temple University
- Tufts University
- University of Alabama
- University of Alberta
- University of British Columbia
- University of Houston
- University of Massachusetts
- University of Oklahoma
- University of Queensland
- University of Tennessee, Knoxville
- Wake Forest University

New Content

HathiTrust partners contributed 278,766 volumes to the repository. 263,525 of these are in the public domain. Texas A&M University was a new contributor, bringing to 26 the total number of institutions contributing content to HathiTrust.

HathiTrust made significant progress in facilitating the deposit of locally-digitized content, including running a survey about partner ingest needs, hosting a conference call with interested partners, developing a single-image validation tool, and making plans to release a full-volume validation and remediation tool in early 2014.

Executive Director Search

The Board of Governors began a search for a new executive director following the departure of John Wilkin, HathiTrust's founding executive director. The Board formed a search committee, and a job description was posted in August. The search committee reviewed applications, and held phone interviews through the fall and will be holding finalist in-person interviews in Ann Arbor, Michigan in January.

Print Disabilities Access

HathiTrust released a new service that allows designated proxies at partner institutions in the United States and Canada 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.

HathiTrust Bylaws Accepted

In early 2013, HathiTrust institutions voted unanimously

to accept bylaws put forward by the Board of Governors.

HathiTrust and DPN

HathiTrust announced its intention to become a "replicating node" in the Digital Preservation Network (DPN). The full announcement can be read at http://www.hathitrust.org/hathitrust_dpn_announcement.

HathiTrust and Knowledge Unlatched

HathiTrust announced that it would be preserving and providing access to works made available through Knowledge Unlatched, an organization that is "helping stakeholders to work together for a sustainable open future for specialist scholarly books". More information about Knowledge Unlatched is available at http://www.knowledgeunlatched.org.

HathiTrust and DPLA

HathiTrust and the DPLA announced a formal partnership, with HathiTrust participating as a Content Hub. Details are available in the news release.

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.

Assistant Director

HathiTrust appointed Jeremy York as Assistant Director.

Organization, Working Groups and Committees

Board of Governors

The Board of Governors held in-person meetings in April and October, discussing a range of issues from the appointment of the Program Steering Committee and ballot initiatives passed at the Constitutional Convention, to issues arising from the passage of the bylaws, the Ha-

thiTrust Research Center, and the search for a new executive director.

Program Steering Committee

The HathiTrust Board of Governors appointed a Program Steering Committee (PSC). The PSC kicked off its work with an in-person meeting in September and held bi-weekly phone calls throughout the fall. In early 2014 the PSC expects to appoint a new Collections Committee and Rights and Access working group, and working groups to carry forward HathiTrust's US Federal Government Documents and Shared Print Monograph Archive initiatives.

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.

User Support Working Group

The User Support Working Group welcomed 6 new members in 2013, and created a new subgroup to support corrections to bibliographic records in HathiTrust in conjunction with the move to Zephir, HathiTrust's new bibliographic management system. A summary of issues received by the User Support Working Group is given in the table at the end of the review.

Special Initiatives

Introducing Zephir

HathiTrust released a new bibliographic management system, Zephir, developed by the California Digital Library. See these links for the full announcement and Zephir background information and documentation.

Copyright Review Management System

A summary of the determinations from HathiTrust copy-

right review activities in 2013 is given below. See CRMS-US and CRMS-World, projects funded by IMLS, for further information.

	2013		Overall	
	Public Domain	All	Public Domain	All
CRMS-US	39,297	87,430	158,167	305,593
CRMS-World	29,768	57,414	43,872	84,524
Total	69,065	144,844	202,039	390,117

HathiTrust Research Center (HTRC)

The HTRC concluded its first phase of development in early 2013 with the release of production infrastructure to support data mining and textual analysis of public domain volumes in HathiTrust. Work began immediately on the second phase, which focuses on community engagement and community-driven enhancements to HTRC services, and development of the HathiTrust-Sloan-Cloud, to provide secure access to the entire HathiTrust corpus. Some highlighted phase 2 activities included:

- The identification of author gender information for works in the HTRC and inclusion of this information in user worksets;
- The second annual HTRC UnCamp, held at the University of Illnois at Urbana-Champaign in September;
- The initiation of monthly user group meetings;
- A call for proposals for the Workset Creation for Scholarly Analysis: Prototyping Project grant, received from the Institute of Museum and Library Services;
- Preparation of HTRC infrastructure to receive in-copyright works in the HathiTrust collection;
- Continued pursuit of grant opportunities and the preparation of a business plan for the HathiTrust Board of Governors;
- The release of version 2 of the HTRC.

Links to information about getting started with the HTRC, HTRC listservs, presentations, news, and events, can be found at http://www.hathitrust.org/htrc.

mPach

Work by 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 one another and with information about the journal; making enhancements to the HathiTrust Page-Turner to display JATS XML articles; and modifying HathiTrust ingest procedures to handle non-JATS content that is embedded in articles or submitted as supplementary material. Staff also defined preservation levels for different types of submitted content, and clarified the scope of mPach services and roles of entities using mPach to deposit materials in HathiTrust. More information about mPach is available on the HathiTrust project page.

US Federal Government Documents

HathiTrust hired Valerie Glenn as a Government Documents Registry Analyst to support work to build a public registry of US federal government documents. A registry project team held a series of focus groups in the fall with representation from a wide variety of interested groups, resulting in draft use cases and functional requirements for the registry. The team also assembled a list of known federal agencies, which is being used to review the comprehensiveness of sources for name authority records such as VIAF and the LC Name Authority Headings.

HathiTrust issued a broad call for US federal government documents records in an effort to understand the scope of the government documents corpus in the US and perform analysis to determine what portion of the corpus has been digitized. The deadline for submitting records for the initial analysis is January 31, 2014.

SPECIAL EDITION - 2013 Year in Review

Repository

Development in 2013 included the following:

New Functionality / Application Changes

Analytics

- The addition of event-tracking features to links in HathiTrust that make it possible to filter results in HathiTrust Analytics based on whether a user is logged in from a HathiTrust partner institution or a University of Michigan Friend Account.

Collections

- The addition of pagination to collection search results.
- The addition of book cover thumbnails (also added to full-text search results).
- Correction of issues related to the display of authors and titles.
- The addition of backend functionality to batch-remove collection items.

Data API

- The release of version 2 of the Data API, which included support for JATS articles, digital audio and TEI (the timeline for supporting TEI in the repository is to be determined).
- Implementation of a mechanism to automatically delete registered Data API keys that have not been activated.

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.
- Initial configuration and testing of new flash-based, high-performance storage to be used with full-text search.
- Significant work was undertaken to develop a spelling suggestion feature and to improve relevance ranking in full-text search results. Relevance ranking work included testing of Solr 4's grouping functionality and the contribution of an initial patch to Lucene to correct an issue with the ranking of long documents in the BM25 ranking algorithm. Staff began coding to implement relevance ranking improvements in late 2013.
- Design and coding of processes to index JATS XML articles.

Image Server

- Modification of the image server for HathiTrust applications to use Unifont when embedding OCR in PDFs in cases where the language of the volume is not supported by Deja Vu Sans, allowing more PDFs to be searchable.

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.

SPECIAL EDITION - 2013 Year in Review

- 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 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.
- Reengineering of a tool to test and debug access controls.
- Tuning of heuristics that determine whether to display volumes from left to right or right to left (depending on the language).
- 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;
- Development to enable the delivery of JATS XML articles as PDFs.
- Deployment of a new robots.txt allowing search engines to crawl PageTurner and Collection Builder pages with a "noarchive" meta tag.
- Initiation of development by California Digital Library to effect a number of improvements to HathiTrust applications.

Volumes Added	Jan-Jun 2013	Overall
Boston College	521	2,363
Columbia University	646	65,036
Cornell University	22,056	437,491
Duke University	2	4,525
Harvard University	1,450	237,435
Indiana University	507	195,580
Library of Congress	2	89,724
North Carolina State University	0	3,196
Northwestern University	24,780	37,502
New York Public Library	28,796	288,370
Penn State University	23,472	68,204
Princeton University	59	251,710
Purdue University	66	44,695
Texas A&M	1,201	1,201
Universidad Complutense	113	112,014
University of California	64,915	3,448,170
University of Chicago	11,915	38,635
University of Florida	7,755	9,763
University of Illinois	8,088	112,975
University of Michigan	56,196	4,666,032
University of Minnesota	11,723	115,935
University of North Carolina - Chapel Hill	8,937	17,025
University of Wisconsin	5,544	555,924
University of Virginia	22	50,821
Utah State University	0	117
Yale University	0	23,678
Total	278,766	10,878,121
Public Domain (~33% of total)		
Total*	263,525	3,542,155

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

Print on demand

- New functionality to produce PDFs optimized for printing on Expresso Books Machines.

Website redesign

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

SPECIAL EDITION - 2013 Year in Review

Infrastructure Changes

- Replacement of servers in HathiTrust's development environment, combined with a move to a new Linux distribution to better support Ruby-based applications.
- Replacement of production web servers at the Indiana site (servers at the Michigan site will be replaced in early 2014).
- Installation of new storage at the Indiana and Michigan repository sites to accommodate 2013 volume projections and replace storage scheduled for retirement.
- Placement of order for 2014 new and replacement storage.

Papers and Presentations

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

Most-accessed volumes

Title
Quicksand, by Nella Larsen.
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.
The Five Laws of Library Science, by S. R. Ranganathan
Consumption of the Lungs and Kindred Diseases, Treated and Cured by Kerosene, by Charles Oscar Frye.
Godey's magazine, v.40-41 1850.
The Human Figure, by John H. Vanderpoel
Mechanick Exercises: or, The doctrine of Handy-Works, by Joseph Moxon.
Roster of the Confederate soldiers of Georgia, 1861-1865, v.1.
The Book of a Hundred Hands, by George Brant Bridgman.
History of wages in the United States from Colonial times to 1928.

HathiTrust Digital Library

SPECIAL EDITION - 2013 Year in Review

User Support Issues	Jan-Dec 2013	Jan-Dec 2012
Content	**1,106**	**1,038**
Quality	987	971
Collections	119	57
Cataloging	**980**	**806**
Access and Use	**950**	**969**
Copyright	997	811
Permissions	107	158
Takedown	7	11
Print on Demand	4	8
Inter-library loan	16	24
Full-PDF or e-copy requests	216	198
Datasets	48	38
Data Availability and APIs	14	9
Reuse of content	48	25
Web applications	**299**	**220**
Functionality problems	89	61
Problems with login specifically	16	9
General questions about login	24	21
Partners setting up login	20	21
Usability issues	16	20
Feature requests	21	24
Partner Ingest	**66**	**40**
General	**713**	**832**
Partnership	100	126
Infrastructure	2	4
Miscellaneous	611	702
Total	**4,114**	**3,830**

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

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.