HathiTrust Digital Library

SPECIAL EDITION - 2010 YEAR IN REVIEW

January 7, 2010

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. 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.

Highlighted Achievements and Activities

Details on each item can be found in the monthly updates from 2010, available at http://www.hathitrust.org/updates.

New partners and finalization of membership for 2011 constitutional convention

26 institutions joined HathiTrust in 2010, doubling the size of the partnership and making a total of 52 institutions that will participate in a constitutional convention next year. In this convention, partners will review repository governance and sustainability and determine directions for the next phase of HathiTrust. View the press release.

New content from partners

HathiTrust partners contributed 2.6 million volumes to the repository in 2010, raising the total number of volumes to more than 7.8 million. Nearly 2 million volumes are in the public domain. New institutions to contribute content in 2010 included:

- Columbia University
- Cornell University
- New York Public Library
- Princeton University
- The University of Chicago
- University of Illinois
- University of Madrid
- Yale University

Approval of new cost model

The Executive Committee approved a new cost model for HathiTrust in February 2010, which will be the basis of costs for all partners beginning in 2013. The new model is based on the overlap of partner institutions' print collections with the digital volumes in HathiTrust. Institutions that do not have large amounts of content to deposit are able to join under the new model before 2013, and more than a dozen have already done so (view the full list of partnering institutions). A FAQ for the new model is available on the HathiTrust website.

Ingest of content from Internet Archive

Staff members at the University of California and University of Michigan worked together over a period of months to develop specifications and routines to ingest partner materials from the Internet Archive at scale. Well over 100,000 volumes from the Internet Archive have been deposited in HathiTrust by three institutions to-date, and more are on the way. This was a major step in the expansion of HathiTrust's ability to accommodate content from a variety of digitization sources.

Formation of multi-institutional groups to address key operational and strategic activities

4 new groups were formed in 2010, reflecting both the growing number of partner institutions and the expanding work of the partnership:

- Communications working group - operational, reports to the Executive Director
- Usability working group - operational, reports to the Executive Director
- Full-text search working group - strategic, reports to the Strategic Advisory Board
- Collections Committee - strategic, reports to the Strategic Advisory Board

HathiTrust Digital Library

SPECIAL EDITION - 2010 YEAR IN REVIEW

Implementation of inter-institutional authentication via Shibboleth

Authenticated users from partner institutions are able to access full PDFs of all public domain volumes in the repository, and use a local sign-on to build permanent public or private collections of volumes. More information about Shibboleth can be found on the HathiTrust website.

Expansion of copyright review work to new institutions

Over the summer, staff at Indiana University, the University of Wisconsin, and the University of Minnesota joined in work begun at the University of Michigan to review the copyright status of works in HathiTrust published from 1923 to 1963. More than 90,000 volumes have been reviewed since the project began two years ago and approximately 55% of those reviewed have been determined to be in the public domain.

Collection Builder improvements

University of Michigan staff added functionality to the Collection Builder application to enable users to add multiple items from full-text search results to public or private collections.

Full PDF download

Staff at the University of Michigan developed the capability to deliver full PDFs of all public domain materials through the HathiTrust PageTurner.

Redundancy of large-scale search

Mechanisms and servers were put in place to achieve full redundancy of the large-scale search index, with copies of the index at both the Michigan and Indiana storage sites.

Single web access portal

The Communications working group, in conjunction with the Usability working group and developers at the University of Michigan, combined existing interfaces to create a single portal at HathiTrust.org for accessing repository services and finding information about the HathiTrust partnership, infrastructure, and activities.

Collaborative Development Environment

Members of a multi-institutional working group completed the work of specifying requirements for, and developing, a collaborative environment for the development and enhancement of HathiTrust applications. Documentation of the new environment will be forthcoming in 2011.

Final report of working group on HathiTrust Storage

A multi-institutional working group was charged with exploring the value of adding a third instance of storage to HathiTrust's infrastructure. The working group's report is available the HathiTrust website.

Other Activities

Improvements to ingest

Staff at the University of Michigan made enhancements to ingest capabilities, including a general increase in processing throughput, improvements in barcode validation, preparation for PREMIS 2.0 support, cleaner integration with pre-ingest transformation processes (for non-Google-scanned materials), and new controls to automatically manage priority levels for content ingested from multiple sources.

New Bibliographic Metadata Management System

The University of Califonia began development of a new bibliographic metadata management system for HathiTrust in November 2010. The system is projected to be operational by the first quarter of 2012.

HathiTrust Digital Library

SPECIAL EDITION - 2010 YEAR IN REVIEW

Discussions with the partnership

HathiTrust hosted several "HathiTrust 101" web- and phone-based discussions for new and existing partners in the summer and fall. More of these discussions and informational sessions are planned in 2011.

Demonstration application for the HathiTrust Data API

Staff at the University of Michigan created an application using only publicly available APIs to demonstrate how the Data API could be used to locate and download complete book packages for public domain volumes not digitized by Google (Google-digitized volumes can be accessed through the Data API one page at a time).

Participation in IMLS grant to Validate Quality

HathiTrust will serve as a testbed for research led by Paul Conway, Associate Professor at the University of Michigan's School of Information, to develop a framework and methodology for validating the quality of content in large-scale digital repositories. Details can be found in the School of Information news release.

Framework for scalable ingest of locally-digitized materials

Significant progress was made on developing policies, specifications, and technological infrastructure to facilitate the ingest of locally scanned materials from partner institutions at scale.

Search widgets

Staff at the University of California developed search widgets for HathiTrust that can be embedded in local websites to execute catalog and full-text searches. The widgets are available at http://www.hathitrust.org/widgets.

Partner initiatives

- Object Validation Tool - Staff at the University of California completed development of a tool to validate the completeness and correctness of volumes ingested into HathiTrust and retrieved through the Data API.

- SFX Target for HathiTrust - UC Staff developed an SFX target for HathiTrust monographs. The target is available to partner institutions who also license the Ex Libris SFX software. A copy of the code can be obtained from the California Digital Library: email CDL-SFX-Tech-l@ucop.edu.

Upcoming Highlights

TRAC certification report

The Center for Research Libraries' report on HathiTrust compliance with the Trustworthy Repository Audit and Certification criteria (TRAC) is expected in early 2011.

Minnesota image ingest

The University of Minnesota in conjunction with the Minnesota Digital Library (MDL) and the Minnesota Historical Society (MHS) have been working with staff at the University of Michigan to develop a prototype workflow for depositing images and associated metadata into HathiTrust for access, storage, and preservation. The prototype project, which includes tens of thousands of digital images from MDL and MHS, is nearing completion. Further details are available in the HathiTrust Update on October Activities.

OCLC catalog

A prototype of the HathiTrust-OCLC catalog will be released in beta in January.

Creative Commons licenses

HathiTrust will soon offer rights holders the option to attach Creative Commons licenses to works they wish to open access to in HathiTrust.

HathiTrust Digital Library

SPECIAL EDITION - 2010 YEAR IN REVIEW

Fulfillment of functional objectives

With the ingest of image content from Minnesota, the establishment of a HathiTrust Research Center, progress to enable HathiTrust as a platform for digital publishing, and significant steps towards compliance with TRAC, HathiTrust will fulfill all of the initial objectives set by the founding partners (see http://www.hathitrust.org/objectives).

Integration of BookReader into PageTurner

A new version of PageTurner, including the scroll and flip functionality and other features of the open source BookReader software, will be released in early 2011.

Full re-indexing for full-text search

The first full re-indexing of HathiTrust volumes will be completed in January.

Approval for HathiTrust Research Center

The Executive Committee has approved the proposal of Indiana University and the University of Illinois for the creation of a HathiTrust Research Center. Details and an announcement will be forthcoming.

Distribution of public domain texts for scholarly research purposes

The University of Michigan has finalized the terms of an agreement with Google that will allow HathiTrust to distribute the texts of public domain volumes to researchers for scholarly purposes. Details and announcement will also be forthcoming.

Future Highlights

Framework for extending access for users with print disabilities

A group of partners from CIC institutions is at work to develop the legal framework and technical implementation criteria to extend full-text access to both public domain and in copyright materials in HathiTrust to users at partner institutions who have print disabilities. Further reports on this work will be given throughout 2011.

Developing support for publishing

As reported in the HathiTrust Update on October Activities, the MPublishing division of the University of Michigan Library has engaged in a 2-year effort to create ingest, management, and presentation tools that will enable the use of HathiTrust as a publishing platform for encoded text and page-image materials. The effort will focus first on journal content, with support for books planned at a later stage.

Print holdings database

HathiTrust has begun to assemble a database containing the print holdings of partner institutions. The database will facilitate the calculation of costs under the new cost model (see the new cost model FAQ), as well as broader partner activities around cooperative collection management and development. Work on the database will continue through 2012, to be completed by the time the new model takes effect in 2013.

Constitutional Convention

The HathiTrust partnership will hold a major meeting in October 2011 to conduct a formal review of HathiTrust governance and sustainability and shape future directions for the partnership.