



SPECIAL EDITION - 2011 MID-YEAR REVIEW

June 24, 2011

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.

Highlighted Achievements and Activities

Details on each item can be found in the monthly updates from 2011, available at <http://www.hathitrust.org/updates>.

Orphan works

HathiTrust announced a [new initiative](#) in May to identify orphan works in HathiTrust. In June, the University of Michigan announced that it will begin to make orphan works in HathiTrust that are also held in its library collections [available to University of Michigan Library users](#). Other HathiTrust partners are moving ahead with similar plans.

New Partners

HathiTrust was pleased to announce 2 new partners in the first part of 2011: Boston University, and Lafayette College, its first liberal arts college partner. Several other institutions have joined and will be announced in the coming weeks. As of June, HathiTrust has 58 partners.

New content

HathiTrust partners contributed more than 900,000 volumes to the repository between January and June 2011, raising the total number of volumes to over 8.8 million. Nearly 2.5 million volumes are in the public domain. New institutions to contribute content in 2011 include:

- Harvard University
- Library of Congress
- University of Virginia

Locally-digitized Content

HathiTrust has been working with several partners on ingest of locally-digitized volumes. These include the University of Illinois,

University of Iowa, Universidad Complutense de Madrid, Northwestern University, the University of Pittsburgh, and Utah State University Press.

MDL Images

The University of Minnesota, the Minnesota Digital Library and the Minnesota Historical Society engaged in a prototype project in 2010 to ingest nearly 60,000 images into HathiTrust. All of the images were successfully ingested in February 2011. More information is available on the HathiTrust MDL [project page](#).

TRAC Certification

HathiTrust was certified in March as a Trustworthy Repository by the Center for Research Libraries. More information can be found at <http://www.hathitrust.org/trac>.

WorldCat Local Prototype

In January, HathiTrust and OCLC announced the release of a collaborative prototype bibliographic catalog. In the next year, this catalog is planned to replace the temporary catalog HathiTrust has had in place since April 2009. More information about the catalog initiative is available in OCLC's [press release](#). Information about HathiTrust's discovery strategy more broadly is posted in the [first entry](#) in HathiTrust's new blog: Perspectives from HathiTrust.

HathiTrust Research Center

Indiana University and the University of Illinois



SPECIAL EDITION - 2011 MID-YEAR REVIEW

announced the [launch of the HathiTrust Research Center](#) in April, a cutting-edge research environment that will provide computational access to the growing body of materials in HathiTrust.

Distribution of Datasets

In conjunction with the Research Center initiative, HathiTrust has begun to distribute texts from the repository to researchers for computational analysis. Information on the available datasets and how to obtain them can be found at <http://www.hathitrust.org/datasets>.

Creative Commons Licenses

As of January, rights holders are able to open access to their works in HathiTrust under Creative Commons licenses. Several hundred volumes have already been opened in this way, including large numbers by the Brooklyn Museum and the Society of American Archivists. Creative Commons licenses can be applied using a [permissions agreement](#), which can be downloaded from the HathiTrust website.

Print Holdings Database

The University of Michigan continued work to assemble a database of the print holdings of all partner institutions. Information from the database will form the basis for [yearly cost calculations](#) beginning in 2013. It will also provide a foundation for the expansion of lawful uses of in-copyright materials held in HathiTrust by partner institutions, and facilitate collaborative collection management and collection development initiatives.

Bib Data Management

California Digital Library continued work on the new metadata management system for HathiTrust. The development team reached a major milestone in May, with the completion of the core components of the infrastructure. Information on the project and updates are posted at <http://www.hathitrust.org/htmms>.

HTPub

The University of Michigan continued work on a new initiative to enable the use of HathiTrust as a platform for publishing. An overview of the project, including development plan, design principles, and proposed architecture, is available at <http://www.hathitrust.org/htpub>.

Validating Quality in HathiTrust

Work on an IMLS grant led by University of Michigan professor Paul Conway began in January and has progressed rapidly. Background on the project is available at <http://www.hathitrust.org/grants> and updates can be found in the monthly newsletter beginning in [March](#).

3-Year Review

The Strategic Advisory Board contracted in March with Ithaka S+R to perform a formal review of HathiTrust. The results of the review will be distributed to the membership for full discussion and review prior to the Constitutional Convention of partners to occur in October 2011. An update on the review process was included in the [Update on May 2011 Activities](#).

Partner Participation

The Executive Committee charged a new User Support working group in March to respond to user inquiries to HathiTrust. The 8-member group raises to more than 40 the number of staff from partner institutions participating officially in HathiTrust working groups and committees. Many more staff, and a growing number, are participating in initiatives such as copyright review, bibliographic management system development, the HathiTrust Research Center, grant projects, pilot efforts around ingest of image and audio content, HathiTrust publishing, content quality and metadata error resolution, and listservs around communications, usability, and HathiTrust usage tracking.



SPECIAL EDITION - 2011 MID-YEAR REVIEW

Shibboleth

37 of HathiTrust's 58 partners are now configured for authenticated access to HathiTrust via Shibboleth. Authentication enables full-PDF download of all public domain materials, facilitated access to the Collection Builder feature, and will be a key mechanism for delivering additional partner services, such as those planned in limited circumstances for in-copyright materials. Information

about Shibboleth in HathiTrust can be found at <http://www.hathitrust.org/shibboleth>.

Papers and Presentations

Numerous papers and presentations were given in the first part of the year. All are available online at <http://www.hathitrust.org/papers>.

Development Highlights

PageTurner

HathiTrust introduced a number of improvements to its PageTurner in April, including new functionality to scroll and flip through volumes, a streamlined interface, and quick-copy links for individual pages.

and the specifications for the new features are available online at <http://bit.ly/jozHQK>.

Collection Builder

Enhancements made to Collection Builder in February and March allow users to create full-text searchable collections of arbitrary size from repository materials.

New Auditing Processes

HathiTrust installed new hardware in May that will be used to conduct periodic, generalized auditing procedures on the repository, such as routine checksum validation of content, as well as ad-hoc cross-repository analysis, for example investigating specific values or usage of preservation metadata elements.

Full-text Search

The Full-text Search Working Group, launched in 2010, released a list of [prioritized recommendations](#) for enhancing features of HathiTrust's full-text search. The first of these, the use of bibliographic metadata in relevance ranking and faceting of search results, will be in place in July.

Storage Replacement

HathiTrust completed its first full storage replacement cycle in May, retiring storage purchased in late 2007. Storage is replaced on a cycle of approximately 3-4 years. Replacement of storage hardware will now be an annual or semi-annual process, shadowing historical patterns of repository growth and storage purchases.

Data API Security

Staff at the University of Michigan have begun to implement new security features in the HathiTrust Data API to enable a range of new activities by partner institutions and others. These activities

HathiTrust Mobile

Staff at Michigan are on track in development of new mobile interfaces for searching and reading volumes in HathiTrust. Initial versions of the new interfaces are expected to be released by the end of the summer.