HathiTrust Digital Library

Update On October Activities

Late Breaking News

Constitutional Convention

HathiTrust has released a blog post on the outcomes of the October Constitutional Convention. The post includes a link to the official notes from the two-day meeting.

Top News

Government Documents Copyright

Maliaca Oxnam, an Associate Librarian from the University of Arizona, and current Chair of the Technical Report Archive and Image Library (http://www.technicalreports.org), has engaged a sabbatical research project with the goal of improving access to government documents in HathiTrust. The three primary areas of her work include 1) investigating the accurate identification of government documents, 2) analyzing the copyright status of the documents and the reasons for their copyright determinations in HathiTrust, and 3) securing permissions from government agencies to make government publications viewable to the public at large. The sabbatical work will be completed by July 2012 and a report with recommendations for future actions will be presented to the HathiTrust Executive Committee. Questions or comments about the research can be sent to Maliaca Oxnam (oxnamm@u.library.arizona.edu).

The Orphan Works Project

The Orphan Works Project (OWP) is in a pilot phase that will continue through the end of December. Researchers from the University of Michigan and the University of California - Los Angeles are conducting a parallel review of approximately 680 volumes in HathiTrust that do not have readily identifiable publisher contacts. Michigan staff have made significant changes to the research process and project tools to improve the rigor and reliability of investigation following a reevaluation of the orphan works candidate identification process in October. An overview flowchart of the new procedure is available at http://www.lib.umich.edu/orphanworks/documentation. Michigan staff will add more extensive documentation in the coming months. The pilot phase of the OWP is intended to serve as a test for an orphan works identification process, through which we will document examples and further define parameters for research.

Ingest

Google Digitization

Ingest rates for Google-digitized volumes from all Google partner libraries were low in October due to problems with Google's download mechanism. Rates are expected to pick up in November.

Update On October Activities

Internet Archive Digitization

HathiTrust began ingest of Internet Archive-digitized content from Duke University and the University of North Carolina - Chapel Hill in October, and worked with the University of Florida toward ingest of its IA-digitized volumes.

Local Digitization

Staff at the University of Michigan continued conversations with the University of Pittsburgh and University of Utah regarding bibliographic metadata for those institutions' contributed volumes. Staff at Michigan received the final set of rare manuscripts and incunabula from Universidad Complutense de Madrid and expect to finish ingest of the materials in November.

Working Groups and Committees

Working groups and committees in HathiTrust may have an operational or strategic focus. See http://www.hathitrust.org/working_groups for more information.

Operational

Communications Working Group

The Communications Working Group continued work to develop a public services-oriented communications package, highlighting ways HathiTrust can be used to address a variety of research and reference inquiries. The group also made progress on a FAQ for the HathiTrust Research Center, and worked with staff from Indiana University to prepare a presence for the Research Center on HathiTrust.org.

User Experience Advisory Group

The User Experience Working Group was pleased to welcome a new member, Darcy Duke, to the group. Darcy is the User Experience Librarian and Web Manager at MIT and has been an active member of the HathiTrust UX Special Interest Group. The group worked on finalizing the user personas it drafted over the summer and discussed details regarding a label change for the PDF download link in the Pag-

Total Volumes Added	October	Overall
Columbia University	7	64,049
Cornell University	110	368,256
Duke University	4,486	4,486
Harvard University	5	52,843
Indiana University	23	186,195
Library of Congress	2,224	73,642
North Carolina State University	0	3,194
University of North Carolina - Chapel Hill	8,087	8,087
Northwestern University	6	5,355
New York Public Library	7	259,165
Penn State University	8	40,815
Princeton University	2	248,916
Purdue University	1	1
University of California	3,646	3,144,989
University of Chicago	11	8,053
University of Illinois	2	14,503
Universidad Complutense	6	108,344
University of Michigan	195	4,446,510
University of Minnesota	163	88,595
University of Wisconsin	893	505,242
University of Virginia	3	47,330
Utah State University	0	46
Yale University	0	23,674
Total	19,885	9,702,290

Public Domain (~27% of total)

	October	Overall
Total*	13,325	2,656,160

*Includes volumes opened through copyright review or rights holder permissions.

Update On October Activities

eTurner application.

User Support Working Group

The adjacent table contains a summary of the issues received by the User Support Working Group in October. Nancy Spiegel, of the University of Chicago, stepped down from the User Support Working Group at the end of the month. The Executive Committee would like to heartily thank Nancy for her work on the group, and her contributions to establishing an ongoing body for user support in HathiTrust. Two positions on the working group are currently open. Nominations and inquiries can be sent to jjyork@umich.edu.

Strategic

Collections Committee

The Collections Committee submitted its recommendations on the treatment of duplicates in HathiTrust to the Strategic Advisory Board (SAB) in October. The recommendations will be posted to the HathiTrust website following incorporation of feedback from the SAB. The Committee will be turning its attention next to a process for responding to requests and offers to include additional materials in HathiTrust, among other pending items on its work agenda.

Projects

Bibliographic Data Management System

The California Digital Library development team worked with staff at the University of Michigan on a workflow and timeline for migrating all bibliographic data from Michigan's integrated library system to California. CDL's metadata analyst finalized the internal metadata schema to be used in Zephir, the core metadata management system. Further information about the project can be found at http://www.hathitrust.org/htmms.
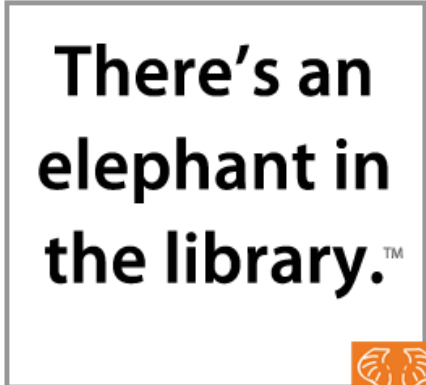
HathiTrust Publishing

MPublishing staff at the University of Michigan gathered input from colleagues in library-based publishing programs in October as they worked to finalize requirements, architecture, and design principles for the new publishing system, and archival package specifications for the published content. Michigan developers began adapting the HathiTrust PageTurner to display the new content based on initial

User Support Issues

	Sept.	October
Content	**171**	**154**
Quality	154	142
Non-partner Digital Deposit	2	0
Collections	4	1
Cataloging	**25**	**44**
Access and Use	**127**	**136**
Copyright	73	75
Permissions	12	4
Takedown	3	4
Print on Demand	17	2
Inter-library loan	5	0
Full-PDF or e-copy requests	24	23
Datasets	1	1
Data Availability and APIs	7	2
Reuse of content	5	2
Web applications	**22**	**29**
Functionality problems	5	6
Problems with login specifically	0	3
General questions about login	2	4
Partners setting up login	5	1
Usability issues	6	2
Feature requests	2	5
Partner Ingest	**0**	**1**
General	**65**	**59**
Partnership	12	8
Infrastructure	0	0
Miscellaneous	53	51

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

specifications. Details about the publishing effort are available at http://www.hathitrust.org/htpub.

IMLS Quality Grant

Data collection on the second sample of 1,000 volumes in HathiTrust continued in October; nearly 80% of the sample was reviewed by month's end. October also saw the launch of the official grant project website, available at http://hathitrust-quality.projects.si.umich.edu/. The website features an overview of the project and detailed status reports by quarter, from the project's beginning in January 2011 to the present.

Review of the physical copies of volumes included in the first 1,000-volume sample continued throughout October. The review focuses on capturing bibliographic information and physical characteristics of the volumes that may have an impact on errors observed in the digital volumes. By the end of the month, a volunteer staff of 12 students from the School of Information reviewed 476 volumes, or nearly 50% of the sample. Staff are coordinating inter-library loan requests with member libraries to facilitate efficient receipt of volumes, or on-site review of volumes by member library staff.

Initial analysis of the data from the first 1,000-volume sample was completed by the project statistician and will be available on the project website in November. The second round of data collection is expected to be complete in mid-November.

HathiTrust Research Center

Indiana University staff worked on implementing the technical security infrastructure for the Research Center in October. The first part of this involved setting up InCommon Federation security, which will allow researchers to login to the HTRC with the username and password issued by their own institution. Once logged in, researchers will have the ability to access data and analysis tools in ways not available to the public. Authenticated access to the HTRC is expected to be available on a limited basis to HathiTrust partners in spring, 2012. As the key architectural pieces of the HTRC are put in place, Indiana staff are examining the adoption of a single API by which researchers can access all pieces of the data infrastructure. The best candidate for this appears to be the HathiTrust Data API. Staff will be making proposed extensions to this API available for comment.

Development Updates

Collection Builder

Staff at the University of Michigan improved processes to synchronize bibliographic and rights metadata in the Collection Builder with metadata in the catalog and rights database.

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Update On October Activities

Full-text Search

University of Michigan staff re-indexed the full-text search index to add additional bibliographic metadata including title information that will enable title displays in full-text search results to match those in the bibliographic catalog. Staff also continued work on advanced search, prototyping several designs for the user interface and working to improve relevance ranking of results. Staff expect to release the advanced search feature in November.

Michigan developers continue to work with staff at the California Digital Library on the development of a spelling suggestion feature. Developers at CDL are investigating modifications to traditional spelling suggestion algorithms, which are generally designed for single-language corpora, to accommodate the many languages in HathiTrust. CDL developers are also testing alternative spelling suggestion algorithms against a sample index.

Staff at Michigan made minor changes to the full-text indexing process to automatically receive notifications when volumes need to be removed from the index, and to improve index monitoring.

PageTurner

Michigan staff completed enhancements to BookReader and underlying infrastructure to improve the speed that images from the repository are rendered on the Web. The image-serving application behind BookReader now estimates dimensions for images and updates them as the images are loaded in the Web browser, rather than inspecting each image prior to making the whole volume available. Further enhancements included better positioning of images in the thumbnail and scrolling views, and improved relative sizing of images when pages within a volume vary dramatically in their dimensions.

Throttling

Staff at Michigan made progress on the development of new throttling mechanisms for the PageTurner and other applications, which will enter an initial internal testing phase in early November.

Outages

No outages were reported in October 2011.

HathiTrust sends notice upon discovery and resolution of unscheduled outages and in advance of scheduled outages and maintenance work that may result in an outage. We welcome and encourage additional recipients for these notices. If your institution is not receiving outage notifications and would like to, please contact feedback@issues.hathitrust.org.