HathiTrust Digital Library

Update On May Activities

Top News

Orphan Works Research

With funding from HathiTrust, the University of Michigan Library Copyright Office has begun work to identify orphan works – works that are known to be in copyright, but whose rights holders cannot be identified or located – in HathiTrust's growing repository. The goal of the project is to provide concrete data on the number of orphan works in HathiTrust, which could be used in the creation of legal or policy-based frameworks to allow broader access to orphan works for scholarly and research purposes. The official press release is available on the University of Michigan Library website.

New Members on the SAB

The Strategic Advisory Board welcomed two new members from the University of California in May: Todd Grappone, Associate University Librarian for Digital Initiatives and Information Technology at UCLA, and Julia Kochi, Director, Digital Libraries and Collections at UC San Francisco. Todd and Julia take the place of Bernie Hurley, UC Berkeley and Bruce Miller, UC Merced, who are stepping down from their duties on the SAB. The HathiTrust Executive Committee would like to thank Bernie and Bruce for their important contributions to the partnership on this committee.

Ingest

Local Digitization Ingest

Michigan staff continued to work out details of ingest with several institutions, including loading bibliographic metadata for additional volumes from Yale, and completing an initial pre-ingest transformation process for locally-digitized content from Universidad Complutense de Madrid.

University of Virginia

HathiTrust completed ingest of more than 47,000 volumes contributed by the University of Virginia in May.

Working Groups

Collections

The Collections Committee continues to work on its current key deliverables, including recommendations regarding duplicate volumes in HathiTrust, coordinated print management, and responding to user requests to contribute volumes to the repository. The group plans to share a draft discussion paper on duplicates with the Strategic Advisory Board in June or July for initial feedback, and will also review

June Forecast

- Begin development of Data API security features.
- Release first wave of new full-text search features
- Begin to implement improvements to the Collection Builder list of collections

Update On May Activities

its work on print management with the HathiTrust Executive Committee's print management subgroup during that same timeframe.

Communications

In May, the Communications Working Group continued planning for a HathiTrust Facebook presence, and made progress on the development of new HathiTrust promotional materials. The group also began discussions with the Usability group on working collaboratively to assemble user stories, and other potential synergies between the two groups.

Usability

Work on the development of HathiTrust personas reported in last months' update continued in May. The group continued to solicit and collect real-life user stories, including in particular those based on librarian and patron interactions. The group's liaison to the Communications group attended the May Communications call to give an update on the progress of the persona project and to discuss collaboration on, and use of, the collection of real-life user stories.

For the last few months the group has been soliciting members for a User Experience Special Interest Group (HT UX-SIG) and has now received around thirty volunteers. This new SIG will be activated in June.

Total Volumes Added

	April	Overall
Columbia University	5,423	63,906
Cornell University	121	311,231
Harvard University	1	52,710
Indiana University	838	184,719
Library of Congress	0	71,418
NYPL	0	258,691
Penn State University	135	39,151
Princeton University	2,051	239,085
University of California	47,637	2,456,364
University of Chicago	999	6,171
University of Illinois	0	14,501
University of Madrid	2,947	106,744
University of Michigan	17,516	4,355,884
University of Minnesota	1,659	86,644
University of Wisconsin	11,081	465,413
University of Virginia	47,303	47,303
Yale University	110	271
Total	137,821	8,760,206

Public Domain (~27% of total)

Total*	174,061	2,378,582

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

User Support

The User Support Working Group assumed responsibility for the wide range of inquiries and feedback received through HathiTrust interfaces and help email addresses in May. The 8-member group has established an on-call rotation throughout the week, including weekends, to address issues in a timely and efficient manner. HathiTrust's response to user feedback received several positive comments via Twitter during the last month. The working group is committed to maintaining a high level of service to address user comments, feedback, and suggestions.

There's an elephant in the library.

www.hathitrust.org

HathiTrust Digital Library

Update On May Activities

Projects

IMLS Quality Grant

The grant project team's work in May focused on the preparation of materials to orient and bring several newly-hired reviewers at the University of Minnesota on board to data collection and review. This work included updating and streamlining the quality review Web application to allow for efficient remote operation. Training of the staff at Minnesota commenced at the end of May and the new reviewers are set to begin work in June. As data from Minnesota are collected, the project statistician will be examining inter-coder reliability among all reviewers and working to establish a final model for sampling volumes in the repository. Gathering data for analysis will be the focus of the grant team's efforts in June. Additional information on the project can be found at http://www.hathitrust.org/grants.

HathiTrust Research Center

Staff at Michigan have begun discussing mechanisms to synchronize the text and bibliographic records of public domain materials in HathiTrust to the HathiTrust Research Center. Michigan Staff developed an initial model for the data transfer (using rsync) in May, and Indiana University staff began performing tests with sample data.

Development Updates

Bibliographic Data Management

The California Digital Library (CDL) development team is preparing a demo of the Metadata Management core system for staff at the University of Michigan on June 14, 2011. As the first major component of the new HathiTrust Metadata Management System, this is a major milestone and deliverable. The next step will be for the CDL team to address feedback raised in the demo. The team continues to interview for the open Metadata Analyst position. In the meantime, a senior metadata analyst at CDL has conducted a full metadata audit, confirming the validity of the core system design. Further information on the project is available at http://www.hathitrust.org/htmms.

Data API

University of Michigan staff completed the first draft of requirements for improved security in the Data API. The draft has been made available for comment at http://bit.ly/jozHQK. We ask that interested parties submit comments to feedback@issues.hathitrust.org. Initial coding will begin as feedback is received.

Development Environment

Staff at Michigan implemented a "diff" service as part of support for administra-

There's an elephant in the library.

www.hathitrust.org

tion of the HathiTrust Development Environment (HTDE). At the time when new code is staged for testing, the code administrator can now choose to see differences between the last deployed version of the code repository and the version being staged in preparation for the next deployment.

Michigan staff also implemented topic branch staging for beta testing. This facilitates testing of code changes on a staged beta testing site without pushing the code branch to the central code repository before its desired time. Parties at partner institutions that are interested in exploring the development environment should contact feedback@issues.hathitrust.org.

Full-text Search

HathiTrust full-text search uses the Lucene-based Solr search engine to index content and provide volume-level results. However, when searches are conducted within a single volume, a different search engine known as XPat is used to dynamically index and search the volume and display page-level results. Differences between the ways that Solr and XPat work sometimes cause inconsistencies in the user's experience. To remedy this, staff at Michigan have started a process to replace XPat with Solr. The majority of this work is to be completed in June, though testing and optimizing may result in a later release date. Accomplishing the change will achieve one of the higher priority features identified by the Full-text Search Working Group: improved results display for multiword searches when searching within a book. Staff will conduct this work in parallel with other full-text search improvements currently underway, including the use of bibliographic metadata for relevance ranking and faceting of search results, and, with development contributions by CDL, a "spelling suggestion" feature. Michigan staff aim to release the relevance ranking and faceting improvements by July 1st.

HTPub

Michigan filled one of two programmer positions advertised for the new HathiTrust publishing initiative, led by the MPublishing division at the University of Michigan Library. The new hire will start on June 27th. The search continues for the second position, which is posted at http://umjobs.org/job_detail/54579/application_developer.

MPublishing recently hired an intern who will be working over the summer to explore potential archival XML schema solutions for electronic journal content.

New Auditing Process and Servers

During their last visit to HathiTrust's Indianapolis storage facility, Michigan staff installed two new servers that will perform periodic, generalized repository auditing, including checksum validation of repository content, using newly-developed auditing tools. The auditing tools can also perform ad-hoc cross-repository analysis as they run, culling information from the repository using custom one-time scripts. For example, staff may add a custom script to the next auditing run to analyze and report on a specific detail of PREMIS metadata usage.

You can follow HathiTrust on Twitter at http://www.twitter.com/hathitrust

PageTurner

Enhancements to the new PageTurner views were released in May in response to user feedback. Staff at the University of Michigan added a full-screen viewing mode, optimizing the use of screen space for content display, and improved landscape image viewing, aligning viewing controls to browser window dimensions when scrolling through the image viewport. Staff also researched ways of improving performance for larger books.

Storage Replacement Cycle

Michigan staff have completed security wipes on all recently retired storage equipment. The equipment was returned to the vendor for a credit, completing this (the first) annual replacement cycle. The next cycle is planned for the first quarter of 2012.

Outages

There were no outages in May.

Partner News

CDL Opens HathiTrust SFX Target to Broader SFX Community

In September 2010, California Digital Library's Discovery and Delivery group released an SFX target for HathiTrust monographs, which was made available to partnering libraries. In May, CDL made the target available to libraries broadly via EL Commons CodeShare, a forum hosted by Ex Libris. The formal announcement is available on the CDL website. Please contact Margery Tibbetts (Margery.Tibbetts@ucop.edu) with questions and inquiries.

Report on the HathiTrust 3-Year Review

Ed Van Gemert, for the Strategic Advisory Board

This update is a follow-on to the report on the HathiTrust 2011 Constitutional Convention given in the Update on January 2011 Activities.

HathiTrust contracted in March with Ithaka S+R to conduct a three-year review of HathiTrust's progress toward meeting the needs of libraries, scholars, students and other users. The review will inform discussion and promote participation at the October 8-10, 2011 Constitutional Convention in Washington DC. The Strategic Advisory Board (SAB) is providing oversight for the review, working closely with the Ithaka staff assigned to the project.

Ithaka's efforts include gathering and preparing research on HathiTrust's existing structure and needs, including background meetings with stakeholders, team members at the University of Michigan, and members of both the Executive Committee and the Strategic Advisory Board.

There's an elephant in the library.

www.hathitrust.org

Update On May Activities

A survey and review of user needs has followed the initial research. A survey was sent to the 52 HathiTrust Contributing Partners and Sustaining Partners (non-content contributing). Ithaka S+R is also interviewing 20 representatives from libraries that do not currently participate in HathiTrust, along with 12 scholars in the humanities and social sciences. The survey officially closed at the end of the day on 3 June and results are being formulated.

Preliminary indicators from this research process provide useful data and commentary including:

- Evidence of progress in meeting functional objectives.
- Commentary on the value of HathiTrust to the Contributing and Sustaining Partners, including perspectives on cost savings, cost avoidance, and the continued need for clarity on the new cost model which is scheduled to go into effect in 2013.
- Projected levels of contributed library staff support for development that will help to inform prioritization of projects requiring the right balance between centralized and decentralized staffing or expertise.
- As the partnership expands, who will govern? Useful data and commentary indicating a need for a clear method for input into executive decisions.
- Interest in the greater environment for HathiTrust and connections with other initiatives. The community is curious to know how HathiTrust may be connected with the Digital Public Library of America (DPLA).

Follow-up interviews by Ithaka S+R staff will probe further these and other issues.

Ithaka S+R is required to submit a draft briefing memo to the SAB on June 17, 2011. Ithaka will then take comments from the SAB and the Executive Committee until July 1. Following a two-week revision period, Ithaka will submit a final report to the SAB on July 15, 2011. The SAB will then distribute Ithaka's report to the HathiTrust membership for full discussion and comment leading up to the Constitutional Convention in October.

Questions or comments regarding the three year review can be directed to Ed Van Gemert, (evangemert@library.wisc.edu) Deputy Director of Libraries at the University of Wisconsin-Madison and Chair of the HathiTrust Strategic Advisory Board.

There's an elephant in the library.

www.hathitrust.org