



Update On January Activities

February 11, 2011

Top News

WorldCat Local Prototype

The HathiTrust Discovery Interface Working Group is pleased to report the availability of a prototype HathiTrust catalog. This new interface is the result of a partnership between OCLC and HathiTrust, leveraging our collective expertise to facilitate discovery of the materials held in the HathiTrust Digital Library. One of the project's main goals is to situate HathiTrust's multi-institutional holdings within the larger world of library holdings represented in WorldCat. The new prototype catalog, accessible at <http://hathitrust.worldcat.org>, is built on OCLC's WorldCat Local platform. HathiTrust and OCLC are eager to receive user feedback to inform the design of a next version of this catalog. Feedback can be submitted to HathiTrust via <http://www.hathitrust.org/feedback>. For more details about this project, see OCLC's press release at <http://www.oclc.org/news/releases/2011/20114.htm>.

Minnesota Image Ingest

From September through December 2010, HathiTrust worked with the University of Minnesota (UMN) and its partner, the Minnesota Historical Society (MHS) to add digital images from the state-wide Minnesota Digital Library and MHS collections to HathiTrust as a preservation archive. This prototype project was intended to begin addressing HathiTrust's long-term functional objective to "support formats beyond books and journals." Nearly 60,000 images and associated metadata were involved in this ingest project, providing a testbed for the evaluation of numerous technical, economic, and policy-related considerations now underway. Conclusions have yet to be drawn, but the report of one of the independent consultants for the prototype ingest effort is available at <http://eric.clst.org/wupl/MDL/MDL-HT-report-110126.pdf>. For additional information, please contact John Butler (j-butl@umn.edu).

Mobile Development

The University of Michigan Library's User Experience (UX) Department will begin work in February on the development of mobile interfaces for HathiTrust, focusing primarily on interfaces for reading volumes and bibliographic searching. The Department will contribute the time of a mobile developer and two User Experience Specialists for the next 7 months to conduct research and design and develop the interfaces. The UX Department staff will be consulting both the Discovery Interface and the Usability Working Groups throughout the development process. Anyone interested in contributing to this project should contact Suzanne Chapman (suzchap@umich.edu).

CC licenses

HathiTrust now offers rightsholders the ability to open access to their works under [Creative Commons](http://creativecommons.org/) (CC) licenses. The first CC licenses will go live in HathiTrust on

February Forecast

- Test and possibly deploy the new version of PageTurner with BookReader
- Draft a specification for Data API security enhancements
- Finalize preparations to support CC licences

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

There's an
elephant in
the library.





Update On January Activities

March 1, at which time the license designations will also begin appearing in HathiTrust's [tab-delimited metadata files](http://www.hathitrust.org/data) and OAI feed (information at <http://www.hathitrust.org/data>). The metadata files contain bibliographic and identifier information for every volume in HathiTrust.

Shibboleth

As of the end of January, users at three new partners institutions have the ability to login into HathiTrust to take advantage of additional services: the University of California-Los Angeles, the University of Utah, and the University of Washington. Current services include full-PDF download of all public domain materials and the ability to create permanent collections in HathiTrust's [Collection Builder](#) using a local sign-on. HathiTrust uses [Shibboleth](#) to enable partner authentication. In order to be configured for Shibboleth, institutions must release required attributes to the HathiTrust Shibboleth Service Provider (see <http://www.hathitrust.org/shibboleth>).

We continue to urge partners to configure Shibboleth to work with HathiTrust so that the full (and growing) array of services can be delivered to every partner institution. The institutions listed below are configured, and we are in the process of working with three other institutions (Utah State University, the University of California-Berkeley, and the University of Madrid) to enable access. If your institution is not on this list, we would appreciate your help in making the appropriate connections to enable login via Shibboleth for your institution.

- Baylor University
- Columbia University
- Cornell University
- Dartmouth College
- Indiana University
- Johns Hopkins University
- Michigan State University
- Northwestern University
- Pennsylvania State University
- Princeton University
- Purdue University
- Stanford University
- Texas A&M University
- University of California-LA
- University of California-San Diego
- University of Chicago
- University of Illinois at Urbana-Champaign
- University of Iowa
- University of Michigan
- University of Minnesota
- University of Utah
- University of Washington
- University of Wisconsin-Madison

Total Volumes Added

	January	Overall
Columbia University	98	57,414
Cornell University	29	215,639
Indiana University	655	180,006
NYPL	64	258,083
Penn State University	121	34,521
Princeton University	50	208,556
University of California	31,792	2,080,038
University of Chicago	18	2,462
University of Illinois	0	14,428
University of Madrid	1,156	79,412
University of Michigan	26,858	4,276,478
University of Minnesota	218	76,589
University of Wisconsin	9,463	423,450
Yale University	0	144
Total	70,666	7,907,220

Public Domain (~25% of total)

Total	14,127	1,973,350
-------	--------	-----------

There's an elephant in the library.





Update On January Activities

New Partner Webinars

HathiTrust will be holding informational webinars in the second half of March, geared specifically toward new partner institutions. Additional details will be disseminated soon. Please contact Heather Christenson (heather.chistenson@ucop.edu) or Julie Bobay (bobay@indiana.edu) for more information.

Print Holdings Information

As noted in the [Update on December Activities](#), partners are requested to provide information about their print holdings by the end of this month. Please contact Julia Lovett (jalovett@umich.edu) with any questions.

Working Groups

Collections

Members of the Collections Committee met with representatives from DLF, OCLC and others at ALA Midwinter to discuss the DLF/OCLC Registry of Digital Masters. The Committee has agreed to provide use cases and additional input for an assessment project that DLF is planning to mount to chart the future of the Registry. Discussions continue on several key work items, including the role of duplicates in HathiTrust and opportunities for shared print collection management.

Communications

The announcement of a number of new developments occupied the Communications working group in January; in particular, the rollout of the prototype OCLC WorldCat Local interface. The group also drafted a prioritized communications and marketing plan for 2011. Among the high priorities in the plan are repurposable materials for librarians to use in explaining HathiTrust to their constituencies, internal communications mechanisms for use among HathiTrust partners, and an introductory webinar for new partner institutions (look for an announcement soon).

Discovery Interface

In January, the Discovery Interface Working Group (DIWG) reached an important milestone in the release of the HathiTrust [WorldCat Local prototype catalog](#). Now that the prototype has been released, the DIWG's work will focus on gathering user feedback on the catalog and conducting formal usability testing.

The Strategic Advisory Board would like to take this opportunity to thank everyone in the working group for their dedication to the catalog project: John Butler, co-chair (University of Minnesota), Lee Konrad, co-chair (University of Wisconsin), Julia Lovett, project manager (University of Michigan), Suzanne Chapman (University of Michigan), Kevin Clair (Pennsylvania State University), Lisa German (Pennsylvania State University), Patti Martin (California Digital Library), Jon

There's an
elephant in
the library.





Update On January Activities

Rothman (University of Michigan), Christopher Walker (Pennsylvania State University). Adam Brin (California Digital Library) is no longer with the group but his contributions during the requirements phase were vital to the group's success.

The Strategic Advisory Board and DIWG would also like to thank OCLC's team for their very hard work, particularly Bill Carney, who served as OCLC's project manager. In addition to the creation of the prototype interface, the collaborative process itself proved to be important in helping both organizations understand the inherent benefits and challenges to working on large-scale projects across disparate types of institutions. The processes that were developed for the coordination of communication, project management, design, user testing, metadata, and systems work will serve the DIWG and HathiTrust well in future projects and partnerships.

Full-text Search

January was an important month for the newly formed Full-Text Search working group, a subgroup reporting to the DIWG. The group held its first two meetings, and will continue to meet on a weekly basis. The group is currently developing a list of features and functions that will have a high impact value for users, and can be supported in the existing technology framework.

Usability

The Usability group continues to participate in other committees via liaison roles. Two group members recently joined the Full-Text Search subgroup to discuss the future of full-text search. The group also provided feedback on proposed designs for the improvements to PageTurner. The Usability group has begun to identify areas across HathiTrust that are in need of further development, usability research, or new design solutions.

Development Updates

Bibliographic Data Management

Development at California Digital Library (CDL) on the core system for the new HathiTrust Metadata Management System progressed in January. CDL staff also consulted with staff at Michigan on documentation for the transformations involved in ingesting bibliographic records from partner institutions. CDL is in the process of hiring a Principal Metadata Analyst for the project. Ongoing project information is posted at <http://www.hathitrust.org/htmms>.

Data API

Developers at the University of Michigan updated the [Data API](#) in January to support Creative Commons licenses, return [access and use statements](#) for retrieved volumes, and provide access to coordinate OCR contained within volume packages.

There's an
elephant in
the library.





Update On January Activities

Development Environment

Michigan staff made improvements to the development environment to facilitate testing of new code prior to release.

Full-text Search

Over the last 2 months, staff at Michigan worked to rebuild the entire full-text index of HathiTrust materials, composed currently of more than 8 million volumes. The new index is in production and will be updated as new volumes are ingested. The rebuilding process included an upgrade of the Solr search engine. This upgrade, coupled with a number of strategic modifications to the way the index is constructed, has resulted in faster indexing time, (staff originally estimated re-indexing would take up to 40 days but it was completed in 10), smaller index size, improved handling of non-Latin scripts (e.g., CJK, Thai, Devanagari), and the inclusion of additional catalog metadata.

PageTurner

Michigan developers made considerable progress on integrating [BookReader](#) into HathiTrust's PageTurner application. Page layout modifications specified in December were implemented, leaving performance testing as the final area of work. Performance testing will be conducted in February and the enhanced PageTurner is planned for release in early March. The current interface to PageTurner will remain the default for the initial release, with BookReader functionality introduced as a "New" feature for users to try.

Staff at Michigan also began work to include Creative Commons licensing information as RDFa in PageTurner application output. Coding will be completed in February. CC licensing information will appear in the PageTurner bibliographic metadata display.

Storage Replacement Cycle Continues

Michigan staff completed the storage replacement work described in [last month's update](#) at the Michigan site, and are beginning the replacement process at the site in Indiana. Staff expect all storage replacement to be completed by the end of March. While the process is non-disruptive and both sites remain in live service during the replacement process, staff have paused ingest and full-text indexing work at crucial moments to be prepared to respond to unexpected problems. In conjunction with this work, staff are testing a process for purging data from retired storage nodes for security purposes before those nodes are decommissioned.

Outages

There were no outages in January.

**There's an
elephant in
the library.**





Update On January Activities

Report on HathiTrust Constitutional Convention

Ed Van Gemert, for the Strategic Advisory Board

Over the past three years, HathiTrust has assisted research libraries in moving more than 8 million scanned volumes online. From an initial group of CIC libraries and the University of California System, HathiTrust has grown to include more than 50 partner libraries, including a small but growing number of international participants. Together, these contributions to HathiTrust represent a significant slice of the world's research holdings. As HathiTrust's library network and content base grows, the partnership will likely have new and different needs for governance, sustainability, and for technology.

In order to address these needs the SAB is finalizing agreement with a consultant to provide the membership with an independent, thorough review prior to the October 2011 Constitutional Convention.

The consultant's review will evaluate HathiTrust's progress to date, using the functional objectives as guideposts. The SAB is also requesting a forward-looking view of the next steps that will be needed to sustain and grow the digital library. The SAB identified these questions as the most important to address in the review:

- What do participating libraries value from HathiTrust, and what unmet needs do they have?
- What new services will draw non-participating libraries, including those that have little or no digitized content to contribute, into the HathiTrust collaboration?
- Is the digital library appropriately designed to meet the needs of end-users, including academic researchers?
- In what ways can HathiTrust differentiate itself from other digital libraries and content hosting solutions, and how should it plan its future investments accordingly?
- Does HathiTrust governance structure give partner libraries a great enough voice in the strategic direction of the digital library? What is the optimal balance between governance by a consortium of libraries and independent decision-making by HathiTrust's project team?
- Will the existing staffing and the nascent HathiTrust cost model position the initiative for growth?

The review will be completed in time to allow discussion and comment from the membership. It is anticipated that the review document will play a crucial role at the HathiTrust Constitutional Convention in October 2011.

Please direct questions or comments to any SAB member including: John Butler, University of Minnesota, Trisha Cruse, California Digital Library, Bernie Hurley, University of California-Berkeley, Bruce Miller, University of California-Merced, Sarah Pritchard, Northwestern University, Paul Soderdahl, University of Iowa, Ed Van Gemert, University of Wisconsin-Madison, (chair), John Wilkin, University of Michigan (ex-officio), and Bob Wolven, Columbia University.

**There's an
elephant in
the library.**

