# HathiTrust Digital Library

## Update On February Activities

## Top News

### HathiTrust Webinar

The HathiTrust Communications Working group has scheduled a second webinar, following the HathiTrust 101 webinar offered last summer, to review basic elements of the partnership (including the business model, collections and services), discuss current activities and future directions, and answer questions from participants. The webinar is targeted specifically toward new partners, but is open to members of all partner institutions. The same webinar will be held at three different times in order to provide more opportunities for participation: Wednesday March 23,1:30-3:00pm, Tuesday April 12, 12:30-2:00pm, and Friday April 15, 12:30-2:00pm (all Eastern Daylight Time). If you plan to attend, please RSVP to Jeremy York as soon as possible before each webinar: jjyork@umich.edu. Please also include any questions or issues you would like the presenters to address (a week in advance will give time to prepare, though we are interested in receiving questions and feedback at any time).

### Public Domain Distribution

HathiTrust is pleased to announce the availability of public domain texts on a large scale for computational research purposes. Approximately 120,000 texts are freely available; up to 2 million more can be obtained with institutional sponsorship through an agreement with Google. More information, including the Google agreement and directions for obtaining texts, is available at http://www.hathitrust.org/datasets. Unlocking the research potential of the collections assembled in HathiTrust is an ongoing goal of HathiTrust partners, and we are excited to take this step in enabling new forms of discovery and analysis.

### New User Support Working Group

HathiTrust is in the process of defining a new working group to respond to questions and issues received from users on a variety of topics, including searching and accessing content, copyright, quality, access to datasets, and more. A call for participants was sent to HathiTrust partner institutions in February; membership in the group will be finalized and the charge posted in the coming month.

### Minnesota Image Ingest

All of the nearly 60,000 images and associated metadata involved in the prototype project between the HathiTrust, the University of Minnesota, the Minnesota Digital Library and the Minnesota Historical Society have been successfully ingested into HathiTrust. Public access to the image content is pending approval of a formal agreement. Project members John Butler, John Weise, and Eric Celeste will give a project briefing at the upcoming CNI Sprint 2011 Membership Meeting. More information about the project can be accessed at http://www.hathitrust.org/mdl_images.

## March Forecast

- Deploy new version of PageTurner with BookReader
- Complete modifications to Collection Builder
- Draft a specification for Data API security enhancements

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

There's an elephant in the library.

www.hathitrust.org

# Update On February Activities

## Local Digitization Ingest

With the initial policies, specifications, and technical framework in place, HathiTrust is ready to begin to scale ingest of locally-digitized book and journal content from partner institutions. HathiTrust has begun working with institutions of the Committee on Institutional Cooperation (CIC) and will broaden its scope throughout the coming year. Partners with digital book and journal content should review the deposit guidelines and content deposit form available at http://www.hathitrust.org/ingest, to be apprised of ingest requirements and preparations of content that may be needed prior to submission.

## Creative Commons Licenses

HathiTrust has enabled support for Creative Commons licenses. The Brooklyn Museum has posted an entry on its blog about the volumes it has opened. If you hold the rights to a volume or volumes preserved in HathiTrust and would like to open access using a Creative Commons license, you can do this by filling out and submitting a permission form.

# Working Groups

## Collections

The Collections Committee is worki
ng on draft recommendations for the treatment of duplicate scans in HathiTrust, which it hopes to have ready for SAB consideration in late March or April. The group has also begun preliminary work on a print management proposal for the Executive Committee in advance of the Constitutional Convention. Another project the Committee will be taking up is a process for responding to requests to add specific content to HathiTrust. There has been one membership change on the Committee: Tom Teper (University of Illinois) has recently stepped in to replace Kim Armstrong (Committee on Institutional Cooperation) and will be serving as a formal liaison to the Executive Committee for the print management work item.

## Communications

The Communications Working Group is pleased to welcome 2 new members: Robin Bedenbaugh from Texas A&M University, and Oya Rieger from Cornell University. The departure of one member earlier this year left in vacancy in the group, and because of the expanding work of the group and excellent pool of nominees

### Total Volumes Added

| | February | Overall |
|---|---:|---:|
| Columbia University | 1,051 | 58,465 |
| Cornell University | 23,371 | 239,010 |
| Indiana University | 1,889 | 181,895 |
| NYPL | 482 | 258,565 |
| Penn State University | 2,653 | 37,174 |
| Princeton University | 10,910 | 219,466 |
| University of California | 224,373 | 2,304,411 |
| University of Chicago | 765 | 3,227 |
| University of Illinois | 0 | 14,428 |
| University of Madrid | 6,125 | 85,537 |
| University of Michigan | 26,878 | 4,303,356 |
| University of Minnesota | 2,909 | 79,498 |
| University of Wisconsin | 8,074 | 431,524 |
| Yale University | 0 | 140 |
| Total | 309,480 | 8,216,700 |

### Public Domain (~26% of total)

| | | |
|---|---:|---:|
| Total | 125,144 | 2,098,494 |

submitted by partner institutions, the Executive Committee decided to approve two new appointments. We are pleased to welcome Robin and Oya and add their knowledge and expertise to our communication efforts.

A draft of the working group's Communications and Marketing Plan for 2011 was reviewed by the Strategic Advisory Board and the Executive Committee in February, and the group is now incorporating feedback into a final version. The working group also made progress on the development of a second webinar (see announcement above) and on a handout designed to communicate the basics of HathiTrust to a broad audience.

### Discovery Interface

The Discovery Interface Working Group (DIWG) has begun to balance its efforts between advancing the full implementation of the HathiTrust WorldCat Local catalog, and enhancing HathiTrust Full-text Search. The DIWG-OCLC team is currently developing a list of desired enhancements to the functionality and interface for a second version of the HathiTrust WorldCat Local catalog. The HathiTrust Full-Text Working Group has continued to meet weekly, and is finalizing a list of features and functions to be deployed in the initial short-term phase of the Full-text Search enhancements.

User experience experts from the DIWG and OCLC have finalized a WorldCat Local Prototype usability test, which will run for about 2 weeks during March.

### Usability

The Usability Group continues to participate in other committees via liaison roles. Two group members are actively participating in the Full-text Search working group and another continues to be actively involved in the Discovery Interface Working Group.

The Usability Group is establishing a User Experience Special Interest Group (UX-SIG). Our intention is to find people at partner institutions with some experience or interest in user experience topics, including usability & interface design. In addition to being a place for user experience (UX) related discussions, this group will provide a base for the solicitation of volunteers to participate in various short-term activities related to the HathiTrust user interface (e.g., contribute to personae and use cases, provide feedback on proposed site changes, join a task force project). There is no implied commitment in joining the group unless a member chooses to participate in a project. Membership in the UX-SIG will provide an interesting opportunity to connect with your UX colleagues across the HathiTrust partnership! Please contact Suzanne Chapman (suzchap@umich.edu) if you are interested in joining this group.

There's an elephant in the library.

www.hathitrust.org

## Update On February Activities

## Development Updates

### Bibliographic Data Management

Staff at California Digital Library have completed development of the core file system, the first major component of the new HathiTrust Metadata Management System. The development team is now reviewing existing workflows for receiving bibliographic data from each HathiTrust content-contributing institution. This work includes testing record import and transformation functions and performance. Development of the next major component, the core database for the system, has begun, and CDL continues to interview candidates for a Principal Metadata Analyst position for the project. Ongoing project information is posted at http://www.hathitrust.org/htmms.

### Collection Builder

Staff at Michigan have begun modifications to Collection Builder that will allow the creation of permanent, full-text-searchable collections of HathiTrust volumes of arbitrary size. The revised design leverages the Solr index used in Full-text Search instead of relying on a dedicated Collection Builder index. In the new configuration, items added to collections of less than 1,000 volumes will be full-text searchable immediately on inclusion. Full-text indexing of collections of more than 1,000 items will be slightly delayed - generally completed within 48 hours. Very large collections of more than 20,000 items will require staff mediation. While 98% of collections contain fewer than 100 items, there has been increasing demand from users for collections with tens and potentially hundreds of thousands of items. The necessary enhancements will be completed in March.

### Data API

Work that was underway at Michigan to design and implement Data API security enhancements is temporarily on hold, with staff focusing on enhancements to Collection Builder. Michigan staff did create a simple API, however, to supply access and use statements to the HathiTrust OAI feed based on a combination of volume rights and source attribute values. This is not formally part of the Data API, and at this point is intended for internal use only.

### Full-text Search

Tests were done that confirmed the viability of the plan to make Collection Builder reliant upon the full-text search index, described above.

### PageTurner

Integration of BookReader into Page Turner was largely completed in February and the code is ready for production deployment. However, initial testing revealed that performance of the new interface could be increased significantly through the installation of the Plack (http://plackperl.org/) Perl module. Plack

There's an elephant in the library.

## Update On February Activities

is now being deployed on HathiTrust web servers and production deployment of PageTurner with BookReader is expected in April.

A bug related to proper ID representation was fixed in PageTurner's COinS implementation. COinS support was also added to PageTurner search results. COinS is an embeddable format that provides bibliographic metadata to citation tools such as Zotero.

### Storage Replacement Cycle Continues

Michigan staff have completed half of the storage replacement work at the Michigan and Indiana storage sites with no service interruptions or other issues, and are continuing replacement work in March, starting in Michigan. The process for securely purging data from retired storage nodes has been finalized and put in place.

### Outages

HathiTrust remained available during an extended scheduled outage of the main campus data center at the University of Michigan from approximately 2:00pm EST on Friday, February 18 until approximately 2:00pm EST on Sunday, February 20. There were no issues resulting from the maintenance.