



## Update On August Activities

In This Newsletter

### Top News

September 10, 2010

**HathiTrust 101** – Members of the Communications working group and John Wilkin, the Executive Director of HathiTrust, hosted two informal “HathiTrust 101” sessions for working group members and directors of partner libraries in August. The webinars were initiated in connection with the recent growth in partnership and the deepening involvement of member institutions in new working groups and the Collections committee. The purpose was to provide an overview of foundational elements of HathiTrust, including mission, governance, finances, and collections, as well as updates on current activities and areas of focus. A third session is scheduled in September, and plans are being considered to hold similar sessions on a periodic basis to keep partners updated about recent and upcoming developments, answer questions, and receive feedback on partner activities and plans. Slides from the “HathiTrust 101” presentation are available at <http://www.hathitrust.org/documents/HathiTrust101-201008.ppt>.

**September Meeting In Chicago** – Staff from a number of partner institutions, including members of the Executive Committee, Strategic Advisory Board, and several HathiTrust working groups, will be meeting in Chicago on September 23 and 24 to discuss a broad array of issues and plans. Some of these, in addition to topics regularly reported in this newsletter, include the new cost model to be implemented in 2013, and the constitutional convention of partners to be convened in 2011. Institutions **who join HathiTrust on or before October 31, 2010** will be eligible to participate in this convention,

in which partners will conduct a formal review of HathiTrust governance and sustainability and shape new directions for the partnership.

**Local Digitization Ingest** – Staff members at the University of Michigan continued to work on the first draft of a policy and specifications framework for ingesting locally digitized content into HathiTrust. Staff have begun to use the framework to evaluate a sample of materials submitted by the University of Illinois, and the framework will go out to partner institutions for comment and further trial in September. HathiTrust plans to begin ingest of locally digitized content from Illinois and other CIC institutions in the fall.

### Working Groups

**Communications** – The Communications Working Group continued to discuss issues surrounding the redesign of the HathiTrust website, as well as plans and processes for receiving new partners.

**Development Environment** – Staff at the University of Michigan have nearly completed migration of the code for HathiTrust applications to the development environment, including establishing the methods and scripts needed to deploy applications into production. Focus has shifted from migrating code to staging, deploying, and testing applications in development and production areas of the new environment. Developers at Michigan have begun to transition to the new environment and system administrators have configured and opened access to additional servers to support this transition. Networking

### Top News

- HathiTrust 101 sessions
- Chicago Meeting
- Local Digitization Ingest

### Working Groups

- Communications
- Development Environment
- Discovery Interface
- Usability

### Ingest

- Columbia
- Yale

### Development Updates

- Bibliographic Metadata Management
- Large-scale Search
- PageTurner
- Storage Upgrade

### Partner News

- UC Validation Tool
- SFX HathiTrust Target

### Late Breaking News

**Princeton University Library Joins HathiTrust** – The full announcement can be found at <http://bit.ly/bEbkSb> and more information is available at [HathiTrust.org](http://HathiTrust.org). We are very excited to welcome Princeton University Library and look forward to the ways they will strengthen and enrich our partnership.

There's an  
elephant in  
the library.





## Update On August Activities

New Growth

changes to provide access from the integration testing area of the development environment to the full repository were also completed.

**Discovery Interface** – At the end of August, OCLC had loaded over 3.7 million HathiTrust records into WorldCat. This constitutes 98% of the available HathiTrust records. The Discovery Interface team is planning a beta release of the phase 1 HathiTrust-OCLC catalog at a date to be determined, pending some final adjustments to the interface to be completed by OCLC. The Discovery Interface team, in conjunction with OCLC, is planning usability analysis that will start before the catalog is released and continue throughout the beta release phase.

The Discovery Interface team is also looking forward to a face-to-face meeting in September, during a larger meeting of HathiTrust partners in Chicago. The agenda will include: taking stock of Discovery Interface projects and activities to date, setting the purpose and scope for future work, supporting the Discovery Interface Full Text Search subgroup, and creating a roadmap for phase 2 of the HathiTrust-OCLC catalog.

**Usability** – The Usability Working Group has begun regular meetings and is in the process of setting priorities and defining member roles in relation to other committees.

### Ingest

**Columbia** – HathiTrust began ingest of volumes contributed by Columbia University in August, including both Google- and Internet Archive-digitized volumes. This was the first set of Inter-

net Archive-digitized materials to be ingested since the initial deposit by the University of California in April, when specifications for Internet Archive-digitized content in HathiTrust were developed.

**Yale** – Staff from Yale and the University of Michigan have been working to determine the pre-ingest transformation steps needed for Yale’s Microsoft-digitized volumes and transfer the content to servers at the University of Michigan, where it will be ingested. Both of these tasks are nearly finished, and we hope to begin ingest of Yale’s initial set of volumes by the end of September.

### Development Updates

**Bibliographic Metadata Management** – University of California staff are collaborating with staff at the University of Michigan to produce a series of planning documents for a HathiTrust Metadata Management system to replace the system currently in use. The goal is to prepare a set of documents for in-person review at the September meeting in Chicago. Teams are at work on documents that will codify goals, success criteria, system requirements, development, integration and migration strategies, acceptance testing and project timelines and milestones.

**Large-scale Search** – Michigan staff continued tests to determine the effects of cache warming on performance. Staff also continued the tests related to scaling strategy and indexing speed that were reported in the [Update on June Activities](#).

**PageTurner** – Staff at Michigan improved the way that PDFs are created for books with landscape-oriented pages.

Number of volumes added:

	Month of August	Overall
Columbia Univ.	56,730	56,730
Indiana Univ.	286	177,962
Penn State	10,202	33,357
Univ. of California	133,900	1,769,227
Univ. of Michigan	40,866	4,130,008
Univ. of Minnesota	54	73,674
Univ. of Wisconsin	14,167	379,111
Total	199,475	6,563,339

Public Domain (~20% of total)

Total	55,132	1,311,288
-------	--------	-----------

### Presentations

HathiTrust 101	August 5 and 27
University of Iceland	August 5
IFLA 2010 (paper and presentation)	August 15

Please see <http://www.hathitrust.org/papers> for links to all HathiTrust presentations, papers, and reports.

There’s an elephant in the library.





## Update On August Activities

September Forecast

**Storage Upgrade** – Michigan staff completed the same upgrade at the Michigan storage site that was completed at the Indiana site in July: adding 160 terabytes of new storage, replacing cluster interconnect switches, reorganizing the equipment layout, and recabling all servers and storage. As reported in the [Update on July Activities](#), the usable storage capacity at each site is now 475 terabytes.

**Outages** – HathiTrust full text search was unavailable on Friday, August 20 from 2:40-2:45pm EDT due to an accidental release of a software module from the new development environment while troubleshooting a full-text indexing problem. Full text search may also have been unavailable for some users from approximately 2:30pm on Friday, August 27 to Monday, August 30 at 3:30pm due to a network file system locking problem at the Michigan site.

## Partner News

**UC Validation Tool** – Staff at the University of California are developing an automated tool to validate the completeness and correctness of objects ingested into HathiTrust and retrieved through the [Data API](#). The tool will be used initially to validate samples of ingested Google- and Internet Archive-digitized objects in comparison with their pre-ingest originals. A prototype of the tool is scheduled for demonstration by the end of September.

**SFX HathiTrust Target** – California staff are packaging code for an SFX HathiTrust target for partners who also license the Ex Libris SFX software. UC expects to announce the availability of the code to partners in late September. The target will be offered through Ex Libris EL Commons wiki later in the Fall.

- Hold committee and working group meeting in Chicago September 23-24
- Add progress bar for full-book PDF generation to the PageTurner application
- Improve PageTurner handling of volumes without OCR
- Finalize draft of policies and procedures for ingest of locally digitized content
- Test procedures with content from CIC institutions and prepare for ingest
- Continue work to redesign HathiTrust website

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

There's an  
elephant in  
the library.

