



## Update On July Activities

In This Newsletter

### Top News

August 13, 2010

**Yale University Joins HathiTrust** – We are pleased to announce that Yale University Library has joined HathiTrust. Yale will initially be contributing close to 30,000 volumes digitized with support from the Yale Provost’s Office and Microsoft, and will be identifying further materials over time. HathiTrust will benefit from Yale’s deep experience and expertise in all areas of library collections and services. More information about Yale’s new partnership can be found at <http://www.hathitrust.org>.

**New Collections Committee and Usability Working Group** – Two new HathiTrust groups were charged in July: a Collections Committee, charged by the Strategic Advisory Board (SAB), and a Usability Working Group, charged by the Executive Director. The Collections Committee is a standing committee in HathiTrust, created to make recommendations about the content in HathiTrust, including the activities, policies, tools, and services needed for partners to manage collections, as well as the processes by which collection development and management decisions should be made. Members of the committee include Ivy Anderson (chair, California Digital Library), Kim Armstrong (Committee on Institutional Cooperation), Sharon Farb (University of California, Los Angeles), Bryan Skib (University of Michigan), Claire Stewart (Northwestern University), Ann Thornton (New York Public Library), and Robert Wolven (SAB liaison, Columbia University). The full charge can be found at [http://www.hathitrust.org/wg\\_collections\\_charge](http://www.hathitrust.org/wg_collections_charge).

The Usability working group is charged with coordinating and overseeing usability activities across all of HathiTrust’s public interfaces, including web and mobile devices. Working group members include Suzanne Chapman (chair, University of Michigan), Jenny Emmanuel (University of Illinois), Felicia Poe (California Digital Library), Matthew Sheehy (New York Public Library). The charge of the group is available at [http://www.hathitrust.org/wg\\_usability\\_charge](http://www.hathitrust.org/wg_usability_charge).

**Local Digitization Ingest Progress** – In July, a group of staff members at the University of Michigan drafted a policy and specifications framework to facilitate ingest of content from a variety of digitization sources into HathiTrust. Staff will be working in August to refine the framework using samples of locally digitized content from Committee on Institutional Cooperation (CIC) institutions. HathiTrust plans to begin ingest of content from some of these institutions in the fall and increase the scope and scale of local digitization ingest in the following months and year.

**Website Redesign/Usability Exercise** – Over the next couple of months staff from partner institutions, coordinated by the Communications working group and in consultation with the Usability working group, will be redesigning HathiTrust’s web presence. Their goal is to integrate the current informational ([HathiTrust.org](http://HathiTrust.org)) and access ([Catalog.HathiTrust.org](http://Catalog.HathiTrust.org)) portions of HathiTrust into a single location and interface at HathiTrust.org. Staff at the University of Michigan conducted a card sorting usability exercise in July in

### Top News

- Yale University Partnership
- Newly Charged Groups
- Local Digitization Ingest
- Site Redesign & Usability

### Ingest

- Bibliographic Ingest for Columbia, Illinois

### Working Groups

- Communications
- Development Environment

### Development Updates

- Large-scale Search
- PageTurner
- Collection Builder
- Storage Upgrade
- Improvements to Ingest

### New Growth

Number of volumes added:

|                     | Month of July | Overall   |
|---------------------|---------------|-----------|
| Indiana Univ.       | 343           | 177,676   |
| Penn State          | 331           | 23,155    |
| Univ. of California | 126,158       | 1,543,677 |
| Univ. of Michigan   | 32,307        | 4,081,191 |
| Univ. of Minnesota  | 34            | 73,620    |
| Univ. of Wisconsin  | 11,305        | 364,944   |
| Total               | 170,478       | 6,363,864 |

Public Domain (~20% of total)

|       |        |           |
|-------|--------|-----------|
| Total | 47,805 | 1,256,156 |
|-------|--------|-----------|

There’s an elephant in the library.





## Update On July Activities

August Forecast

conjunction with this redesign, to help improve the architecture of the site and general categorization. The exercise was completed by staff members across the partnership. The website redesign is targeted for completion by the end of October 2010.

### Working Groups

**Communications** – In its July meeting, the Communications Working group discussed the October website redesign, including overall goals and audience for HathiTrust.org. The group plans several new content pieces for the site: a statement on quality, and a “What is the HathiTrust” primer. Group members are planning for an in-person meeting in September where they will do substantial work on an overall communications and marketing plan for HathiTrust.

**Development Environment** – Staff at the University of Michigan continued to move active development of HathiTrust applications and services into the new development environment in July, and were able to run HathiTrust applications successfully there for the first time. Steps for completing migration to the new environment include integrating code that has been developed during the migration process, performing additional tests on migrated code, and developing scripts that will move the new code into production. UM Staff are also doing work to configure the environment for use by developers. Current areas of focus are establishing virtual web service and database resources on a per-developer basis, and establishing logical separations in the environment from which core developers will

be able to do integration testing against the full repository.

### Ingest

**Loading of Bibliographic Data from Illinois, Columbia** – Staff at Michigan finished loading bibliographic data for content digitized both by Google and the Internet Archive (IA) from Columbia University. Ingest of this content is set to begin in mid-August. Staff also received and loaded bibliographic data for IA-digitized content from the University of Illinois.

### Development Updates

**Large-scale Search** – A server dedicated to testing indexing processes and performance for large-scale search was deployed by Michigan staff in July. Specific tests mentioned in the June report remain the focus of activity and are greatly facilitated by the new testing server.

**PageTurner** – Developers at Michigan put a new page-image service into production in July. The service interfaces with master images in the repository to deliver access images on the Web in real time. It is being used initially to generate full-book PDF files, and will eventually serve individual page images directly to the HathiTrust PageTurner. Significant effort went into optimizing performance of the image service, which is important both for PDF generation and for serving images to applications such as GnuBook. Work to integrate GnuBook with PageTurner is in progress.

**Collection Builder** – Staff at the University of Michigan deployed new functionality in July that allows users to add items returned in full-text search results

- Begin ingest of Columbia University content
- Continue configuration of the new development environment and migration of current development activities
- Refine content ingest framework
- Articulate overall goals and audience for the HathiTrust website
- Install new storage at Michigan site

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

There's an elephant in the library.





## Update On July Activities

to a collection. More information on the new functionality and Collection Builder in general can be found under “Building Collections in Collection Builder” in the HathiTrust [FAQ](#).

**Storage Upgrade** – Michigan staff completed the installation of 160 terabytes of new storage at the Indiana site in July. In the process, staff also upgraded cluster interconnect switches and, as the result of a data center reorganization project, relocated and re-cabled all storage and server equipment. Similar installation work is scheduled during August in Michigan. The new storage will bring the usable storage capacity at each site to 475 terabytes.

**Improvements to Ingest** – Architectural improvements to the ingest system are in planning and early development stages. Major enhancements include a general increase in processing throughput, improvements in barcode validation, preparation for PREMIS 2.0 support, cleaner integration with pre-ingest transformation processes (for non-Google-scanned materials), and

new controls to automatically manage priority levels for content ingested from multiple sources. Suggestions and ideas in this improvement process are welcome. Please contact [hathitrust-info@umich.edu](mailto:hathitrust-info@umich.edu).

**Database Problem Resolved** – HathiTrust was unavailable for a total of 2 hours during June and July due to a database problem associated with heavy usage patterns, resulting in rapid consumption of disk space, and ultimately an outage for users of HathiTrust. A fix was developed and released in July that has resolved this problem.

**Outages** – HathiTrust services were unavailable on Wednesday, July 21 from 1:00-1:30pm EDT due to exhausted storage capacity on database servers in both data centers; services were intermittently unavailable to some users from 8:00pm EDT on Wednesday, July 21 to 10:00am on Thursday, July 22 due to a web server not being restarted properly by staff following scheduled storage work in Indiana.

There's an  
elephant in  
the library.

