



Update On May Activities

Top News

June 11, 2010

NYPL Partnership – We are pleased to announce New York Public Library as the newest partner in HathiTrust Digital Library. The New York Public Library is recognized around the world for its distinctive collections and services to users, and will bring valuable content and perspective to the partnership. NYPL will be contributing materials digitized in collaboration with Google, the Internet Archive, and Kirtas. The press release for the partnership announcement can be read at <http://www.nypl.org/press/press-release/2010/05/24/nypl-takes-giant-step-preserving-its-digitized-collections>.

6 Million Volumes, 1 Million Public Domain – As of May 26, HathiTrust preserves and provides access to more than 6 million volumes, over 1 million of which are in the public domain. These significant milestones draw attention to the growing value of HathiTrust as more and more volumes, representing an increasingly comprehensive collection of published literature, are contributed by partners, made available to users, and securely stored for generations to come.

Shibboleth – Implementation of authentication via Shibboleth was tested and finalized in May, and a formal release of the service scheduled for June 8. When it is released, users at partner institutions that have provided Shibboleth attributes to HathiTrust will be able to download full-PDFs of public domain volumes, and use their institutional sign-ons to access the HathiTrust Collection Builder. More information about Shibboleth in HathiTrust, including attributes, terms of use, and privacy, are available at <http://www.hathitrust.org/shibboleth>.

New Communications Working Group – The Executive Committee has formed a new working group to address an array of communication needs in HathiTrust as the partnership and user base continue to expand. Information about the communications group, including goals and specific areas of focus can be found in the formal charge at http://www.hathitrust.org/wg_communications_charge.

Partner Local Digitization – Staff at the University of Michigan continue work to establish specifications and guidelines for ingest of non-Google- and non-Internet Archive-digitized materials from partner institutions. By August, staff hope to have a clear and efficient framework defined to begin to scale up ingest of content from local digitization efforts.

Working Groups

Development Environment – Michigan staff are working on migrating active development of repository applications and services including PageTurner, Collection Builder, Large-scale Search, and Ingest, to the new development environment. The design is being adapted on an ongoing basis in response to issues encountered along the way. Michigan ordered new network hardware to enable limited access from the development environment to content in the production repository for integration testing and troubleshooting (a subsection of the repository has been copied and made available to the environment to meet the majority of development needs). The working group continues to have regular conference calls to discuss progress on the transition to the new environment.

Top News

- NYPL Partnership
- 6 Million Volumes, 1 Million Public Domain
- Shibboleth
- New Communications Working Group

Working Groups

- Development Environment
- Discovery Interface

Development Updates

- Large-scale Search
- PageTurner

New Growth

Number of volumes added:

	Month of May	Overall
Indiana Univ.	262	177,097
Penn State	5,222	22,496
Univ. of California	304,997	1,508,533
Univ. of Michigan	82,508	4,022,230
Univ. of Minnesota	348	73,413
Univ. of Wisconsin	13,522	343,566
Total	406,859	6,147,355

Public Domain (~19% of total)

Total	272,027	1,175,546
-------	---------	-----------

There's an elephant in the library.





Update On May Activities

June Forecast

Discovery Interface – On May 23, OCLC successfully installed the version 1 HathiTrust WorldCat Local instance. The catalog has been made available internally to the Discovery Interface Working Group, and is being tested and evaluated by both OCLC and HathiTrust. OCLC is now close to completing a full load of HathiTrust records into WorldCat, with just under 2.9 million records loaded. After the initial record load, OCLC will move to loading periodic HathiTrust update files.

The working group also recently drafted a charge document for its work on developing the HathiTrust Full Text Search. Some of the main goals of this project will be: charting a course of service refinement to meet scholarly need; contextualizing each of HathiTrust's search services through interface design and presentation; recommending pathways from HathiTrust search to other services essential to patterns of scholarly workflows; and evaluating the effectiveness of the HathiTrust full-text search. The group is currently working on outlining a timeline and strategy for these efforts, as well as the full-text search membership.

Development Updates

Large-scale Search – New servers were installed and configured at the Indiana site by staff from the University of Michigan, and the process for releasing daily large-scale search index updates was developed and run in a test mode. The search service running on these new servers will be put into production by Michigan staff on June 8, making the full-text search service redundant in Michigan and Indiana. Two new index building servers were put into pro-

duction in May, providing a substantial increase in index building performance and freeing one server to be repurposed for development and testing of index processes.

PageTurner – Michigan explored strategies for optimizing performance of the newly constructed image server, particularly in connection with its use in the GnuBook page viewer. Speedy extraction of image dimensions for an entire book and delivery of thumbnails are among the challenges. Performance optimization work will continue in June.

Outages – The beta* large-scale search service was unavailable on Monday, May 3 from 9:00-10:15am to apply security updates and on Thursday, May 20 from 9:00-11:25am to install new networking hardware.

* Beta services are typically non-redundant and/or volatile, and while we strive to minimize down time and report any that occurs, we do not attempt to adhere to non-peak outage windows for maintenance.

- Continue performance optimization for GnuBook
- Continue configuration of the new development environment and migration of current development activities
- Begin work on increasing the development environment's available storage

There's an
elephant in
the library.

