



Update On December Activities

In This Newsletter

January 15, 2010

Top News

Columbia Partnership – HathiTrust is very pleased to welcome Columbia University as its newest partner. A representative of HathiTrust will be traveling to Columbia in late January to give a full introduction to repository operations, current activities, and future plans. We look forward to the experience and expertise that Columbia will bring to the enterprise, and the new possibilities that are opening for HathiTrust as it continues to expand its membership and its collections. A full press release on the new partnership can be read at <http://www.columbia.edu/cu/lweb/news/libraries/2009/20091216.hathi.html>.

5 Million Volumes – A significant milestone was passed in December as HathiTrust exceeded 5 million volumes in digital holdings. More than 3/4 of a million of these are in the public domain. A steady rate of growth is expected to continue in 2010, and partner collections are projected to grow to more than 8 million volumes.

TRAC Audit – In early December, HathiTrust began a process with the Center for Research Libraries (CRL) to assess the digital repository in relation to the Trustworthy Repositories Audit and Certification (TRAC) criteria. The assessment is scheduled to proceed until mid-February, and the findings will be publicly available. More information about the audit can be found on the CRL website at <http://www.crl.edu/archiving-preservation/digital-archives/certification-and-assessment-portico-and-hathitrust>.

Bib API – HathiTrust has released a new bibliographic API that enables retrieval of descriptive and rights information for objects in the repository based on standard identification numbers (e.g., ISBN, ISSN, LCCN, OCLC). The API is a replacement for the (now deprecated) **Rights API** and the specification is available at http://www.hathitrust.org/bib_api.

Working Groups

Discovery Interface – OCLC is completing preparations for the import of HathiTrust data into WorldCat Local (WCL). The installation of a HathiTrust WCL instance is scheduled to be complete in late February, and loading of records into this first version of the joint catalog will begin in March 2010. Looking towards version 2 of the catalog, the HathiTrust-partner working group began reviewing its scope and membership needs as its purview expands beyond bibliographic metadata in the catalog to include the integration of features such as full-text search and the HathiTrust Collection Builder. The group was renamed the HathiTrust Discovery Interface Working Group (from HathiTrust/OCLC Catalog) to reflect this broadening scope. The HathiTrust Executive Committee approved a proposal to have the working group report to the Strategic Advisory Board (SAB) in December, ensuring stronger alignment of the development and delivery of discovery services with future directions in HathiTrust as a whole.

Collaborative Development Environment – Staff at the University of Michigan completed setup of one of the servers that will be used in the initial proof-of-concept partner development

Top News

- Columbia Partnership
- 5 Million Volumes
- TRAC Audit
- Bib API

Working Groups

- Discovery Interface
- Development Environment
- Research Center

Ingest

- Internet Archive Ingest
- Non-Google Ingest

Development Updates

- Shibboleth
- Large-scale Search

Partner News

- UC and SFX

New Growth

Number of volumes added:

	December	Total
Indiana Univ.	16,923	133,482
Penn State	233	5016
Univ. of California	263,089	1,155,367
Univ. of Michigan	230,881	3,659,874
Univ. of Wisconsin	12,137	267,353
Total	516,514	5,221,092

41,006 public domain volumes were added in December, bringing the total number to 758,947 (about 15% of total content).

There's an elephant in the library.





Update On December Activities

January Forecast

environment. The server is configured with all of the tools and software needed to support the PageTurner development that the University of California and Michigan engaged in collaboratively in 2009. A developer at UC has begun to test features of the environment and will be reporting and providing feedback to the working group when the full group is re-engaged in January.

Research Center – The RFP produced by the working group was approved by the Executive Committee in December and is available at <http://www.hathitrust.org/documents/hathitrust-research-center-rfp.pdf>.

Ingest

Internet Archive Ingest – During the month of December, staff from UC and UM finalized many of the procedures and conventions related to the ingest of Internet Archive-digitized books into HathiTrust. These included file identification, preservation and technical metadata elements, content transformation and validation processes, error logging, and exception handling. UC delivered bibliographic metadata for an initial set of IA-digitized volumes to UM, and UM worked steadily on coding the transformation and validation processes for ingest. An end-to-end pilot test, including download, ingest, and quality review of ingested items will be performed in late-January.

New Programmer For Non-Google Ingest – Applications are still being taken for a programmer to receive and prepare non-Google materials for ingest into HathiTrust. Review of applications and interviews are being conducted simultaneously. The bidding process will close in mid-January, but will be

extended again if an applicant is not selected. Full-time and part-time positions are being considered, and it is increasingly likely that one of each may be filled.

Development Updates

Shibboleth – In the near future HathiTrust will be implementing Shibboleth as a mechanism for inter-institutional authentication into HathiTrust. Distributed authentication will make it easier for users to take advantage of personalized services in HathiTrust, such as the Collection Builder. It will also enable the delivery of enhanced services to HathiTrust partner institutions. Staff at UM discussed the implementation strategy for Shibboleth in December and installed the Shibboleth service provider software on development servers to begin the work of integration. A forecast for the timeline of implementation will be included in the next update.

Large-scale Search – Staff at UM continue to refine the daily index update and release workflow, making it more resilient to problems that are sometimes encountered during indexing. New server equipment will soon be purchased for use at the Indiana site, and a schedule projected for continuous new hardware acquisition to maintain performance levels as the size of the index grows. As part of index and query response time testing, UM staff also updated and released a revised cache-warming procedure based on production log analysis. Warming (pre-populating) the cache of completed queries improves search performance.

Outages – There were no outages in December.

- Staff visit to Columbia
- Begin Internet Archive Ingest Pilot
- Discuss the development of a validation mechanism for repository content using the Data API
- Begin to explore ingest and delivery of born-digital objects
- Finalize a draft report and recommendation on a third instance of HathiTrust storage

There's an
elephant in
the library.





Update On December Activities

Partner News

(What is your institution doing with HathiTrust? Let us know!)

UC and SFX – A University of California group has started work on a project to demonstrate proof-of-concept success in exposing HathiTrust public domain books through UC's UC-eLinks service (SFX). The project is investigating the various HathiTrust APIs capable of supporting this service, and in addition to gathering usage statistics for the new target, will report on the functionality, usefulness, and viability of each of the APIs for future endeavors. The target will eventually be made available to ExLibris so that it can be added to the SFX package for all customers, but will be available to HathiTrust partners who use SFX before then.

There's an
elephant in
the library.

