HathiTrust Digital Library

Update On November Activities

In This Newsletter

Top News

December 10, 2010

Extensions to Copyright Scheme – Staff at the University of Michigan outlined specifications and implementation details for supporting Creative Commons licenses in the repository's rights management scheme, including system APIs. The CC licenses are a work in progress, but HathiTrust hopes to allow these additional CC options for rights holders to open access to their works early next year. In conjunction with this work, the full range of access and use statements for HathiTrust materials was revised to be more useful to end users. Please visit HathiTrust Access and Use Policies for more details.

Bibliographic Data Management – The California Digital Library kicked off development of the new metadata management system for HathiTrust in November. The system is expected to be operational by the first quarter of 2012.

Working Groups

Communications – The group focused on work with many new partners to create press releases and announcements, and on the announcement of a major milestone: the finalization of the partnership that will participate in next year's constitutional convention.

Development Environment – The new development environment continues to function well. Work on expanding storage capacity is ongoing, and new MySQL servers are on order, to be installed in late December or early January. Although the group may schedule additional discussions if issues emerge, the primary work of the development environment working group is now complete.

Discovery Interface – Changes requested by the Discovery Interface Working Group (DIWG) to the HathiTrust-OCLC catalog were implemented by OCLC in early November and reviewed by the working group. The DIWG is working with OCLC and the Communications working group to prepare publicity on a prospective release of the prototype catalog. These preparations include a strategy for linking to the prototype from the current HathiTrust search portal page. The DIWG is planning a period of user testing, feedback gathering, and analysis following release.

Ingest

New Partner Ingest – Ingest of volumes from Cornell University began in November, and ingest testing was performed on a sample of Yale University volumes. Full ingest from Yale will begin in December.

Development Updates

Large-scale Search – Staff at the University of Michigan prepared in November to re-index the full text of all volumes in the repository, a process that is estimated to take 40 days. HathiTrust has been adding incrementally to the existing full text search index as new content has been deposited. This process will exercise for the first time the capability of the indexing system to operate in a dual-mode configuration, maintaining the currency of the production full text index while simultaneously building the new one. The new index will include several changes such as improved handling of non-Latin scripts (e.g. CJK, Thai, Devanagari) and addi-

December Forecast

- Continue work on BookReader integration, full-text re-indexing, and Data API security enhancements
- Finalize framework for ingest of locally-digitized content and establish technical systems for routine ingest
- Begin ingest of Yale volumes and content from new partner institutions

HathiTrust Digital Library

Update On November Activities

tional cataloging metadata. Re-indexing is expected to start in early December and the new index is scheduled to be available by the end of January.

Further improvements to full-text search included the implementation of optimization and integrity checks as part of the daily index-building routine. Optimization increases consistency in query response times, and integrity checks prevent a corrupted index from being released into active service.

PageTurner – In preparation for the deposit of historical photographs from the Minnesota Digital Library and Minnesota Historical Society, developers at Michigan modified PageTurner to support repository objects that do not have plain text OCR for some or all provided images. Michigan staff also made significant progress on BookReader inte-

gration with PageTurner. Next steps include additional modifications to the integrated interface, and performance testing.

Outages – There were no outages in November.

Partner News

CDL Object Validation Tool – The Object Validation tool that CDL began to develop in August is nearly complete. Next steps will be to share with the HathiTrust community and enlist partner institutions to participate in using and evaluating the tool. Interested partners should contact HOVA-L@listserv.ucop. edu.

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

New volumes added:

	November	Overall
Columbia U.	3	57,282
Cornell U.	53,807	53,807
Indiana U.	695	179,029
NYPL	1,066	257,828
Penn State	2	33,365
Princeton U.	130,648	208,142
U. of California	92,153	1,917,335
U. of Chicago	0	2,204
U. of Illinois	0	14,428
U. of Michigan	54,814	4,234,664
U. of Minnesota	141	73,864
U. of Wisconsin	6,774	407,065
Total	340,107	7,439,273

Public Domain (~24% of total)

Total	148,420	1,820,696

Presentations

Yale University	November 3

See http://www.hathitrust.org/ papers_and_presentations for links to all HathiTrust papers, presentations, and reports.

There's an elephant in the library.

www.hathitrust.org