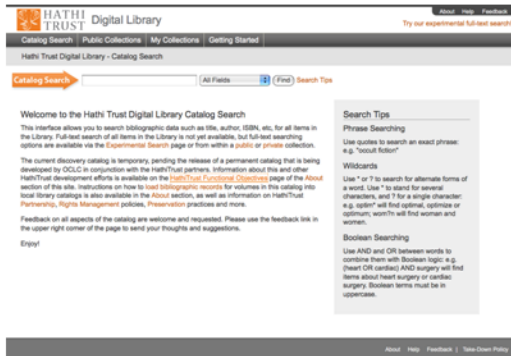




Update on April Activities

Top News

Temporary Beta Catalog Released— A major milestone for HathiTrust was reached in April, as a temporary beta catalog for HathiTrust was released on April 24. The catalog provides bibliographic search and faceted browsing of all volumes in HathiTrust, and integrates with the HathiTrust Page Turner to provide access to individual items. It can be accessed at <http://catalog.hathitrust.org>. Further integration with the HathiTrust Collection Builder, as well as other enhancements, are planned for a second phase of development. This catalog (including phases 1 and 2) is temporary, pending the release of the permanent HathiTrust catalog to be developed by OCLC in conjunction with HathiTrust partners (see the final news item on this page for details). Michigan and California also discovered a strong mutual interest in improving functionality in the HathiTrust Page Turner, which will be explored further in May.



Ingest Started From Indiana University And The University of California—In a month of significant developments for HathiTrust, ingest of content from both Indiana University and the University of California began in April. The loading of bibliographic metadata for the initial set of Indiana volumes was completed and approximately 10,100 had been ingested by May 1. Bibliographic loading continues for the University of California, and ingest started late in April. Several hundred volumes are now available in the repository.

HathiTrust-WorldCat Local Project—In April, HathiTrust team members met with OCLC to begin the process of creating a permanent bibliographic discovery interface for HathiTrust. The initial version, due out in the 1Q 2010, will build on WorldCat Local's standard functionality and will

be tailored to HathiTrust's entirely digital collection. The design of HathiTrust's recently released temporary beta catalog will also inform the WCL implementation's design. As referenced above, the WCL catalog will replace the temporary catalog upon completion. Some of the project's priorities will be:

- Achieving clarity and accuracy in the linkages between the print copy and digital copy.
- Providing specific access to multi-part digital works using local holdings records
- Clarity in the interface of access and "viewability" levels, which will vary based on copyright restrictions and user privileges
- Integration of the HathiTrust materials within other WCL catalogs

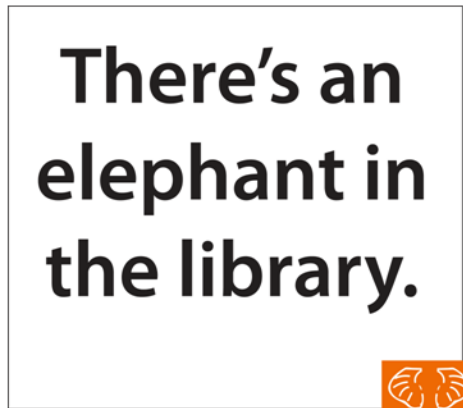
Outages in April: HathiTrust experienced reduced performance from 11:00pm EDT on Thursday, April 23 to 8:22am EDT on Friday, April 24 due to a database problem at one of the sites, and from 5:30pm to 9:00pm EDT on Thursday, April 30 due to unintended consequences from a networking configuration change.

Outages planned for May/June: No outages are planned at this time.

- **Temporary Beta Catalog Released**
- **Ingest from Indiana University and the University of California**
- **HathiTrust-WorldCat Local Project**
- **New Storage Installation**
- **Large Scale Search Update**
- **Alpha Version of Data API Completed**
- **Planning Started for Partner Development Environment**
- **New Growth**
- **May Forecast**

New Growth

- **41,927 new volumes were added in March.**
- **As of May 1st, the repository contained a total of 2,821,937 volumes.**
- **12,904 public domain volumes were added in March, bringing the total number of public domain volumes to 446,506 (16% of the total content).**
- **Ingest of Wisconsin materials continued. As of May 1, 2009, HathiTrust contained 172,160 Wisconsin volumes.**





Update on April Activities

Development Updates

Storage – New storage that was purchased in March has been installed and is operational at the Michigan site. The storage at Indiana has been received, and installation is scheduled for mid-May.

Large-scale Search – Michigan and California developers shared experiences and ideas in a fruitful discussion about Lucene-based search engines, XTF and Solr. Investigations into software solutions for improving response times for slow queries led us to add common-gram indexing and searching capabilities to Solr, significantly improving performance of slow phrase queries. Common-grams increase index size, but the difference so far seems to be manageable and worthwhile. We are continuing to refine a hardware configuration to use for Solr servers based on discussions of indexing workflows and continuing research

of different indexing algorithms, which have an impact on storage requirements.

Data API – Useful feedback was received from California Digital Library staff on the first draft of a functional specification for the HathiTrust Data API and a response is in the works. The draft is online (http://www.hathitrust.org/hathitrust_data_api). Coding of an alpha version the Data API is done, and limited use of the API will start in May; CDL will use it to validate ingest of UC content into the repository.

Development Environment – We are in the early stages of conceiving a new development environment for building and testing repository applications and services. Server hardware has been allocated to this purpose and setup will take place as design discussions progress.

PLEASE NOTE: Please contact Chris Butchart-Bailey (chrisbu@umich.edu) with email addresses of individuals or groups that should be added to our system outage mailing list to receive information about unscheduled outages.

We schedule system maintenance work that requires a system outage during time windows (in Eastern time) where academic user activity is generally lowest:

For major work, Friday evenings (8pm-1am) and Sunday mornings (5am-10am);

For minor work, weekdays from 6:30am-8am.

Advance notice for scheduled outages is given on business days, at least 24 hours in advance. Notice of unscheduled outages is given upon discovery, and additional updates are given as appropriate.

May Forecast

- **Deploy common-gram phrase searching in the HathiTrust beta large-scale search.**
- **Continue work on the HathiTrust Data API, and initiate experimental use by California to validate materials ingest.**
- **Michigan and California will explore opportunities for collaborative development and deployment of an enhanced HathiTrust Page Turner, giving special consideration to using Open Library's open source Book Reader code library. (<http://openlibrary.org/dev/docs/bookreader>)**

There's an elephant in the library.

