



## Update On July Activities

August 12, 2011

### Top News

#### New Partners

Two new partners announced membership in HathiTrust in July: the University of Notre Dame and the University of Florida. Florida announced additionally that it will be offering students, faculty, and other users of UF libraries access to orphan works in HathiTrust that UF also holds in its print collections. We are very pleased to welcome these new institutions and look forward to the ways they will enrich our partnership. News releases can be found at the following links: [University of Notre Dame](#); [University of Florida](#).

#### 3-Year Review

The 3-year review conducted by Ithaka S+R with oversight from the Strategic Advisory Board (SAB) was completed in July and is available at [http://www.hathitrust.org/constitutional\\_convention2011](http://www.hathitrust.org/constitutional_convention2011), with an introduction from Deputy Director of Libraries at the University of Wisconsin-Madison and SAB chair Ed Van Gemert. The 3-year review was planned from the time of HathiTrust's launch in 2008 to provide a meaningful assessment of the partnership's accomplishments and outlook leading up to the HathiTrust Constitutional Convention, also planned to occur in the 3rd year. Institutions and consortia that were members of HathiTrust as of October 2010 will participate in the Convention this coming October to review HathiTrust sustainability and governance, and set new directions for the partnership. Details on the Convention are available at the link above. Questions or comments regarding the 3-year review should be directed to Ed Van Gemert at [evangemert@library.wisc.edu](mailto:evangemert@library.wisc.edu).

#### Orphan Works Candidates

HathiTrust posted the first set of orphan works candidates to a [public catalog](#) in July. These are works for which, following an extensive review process, rights holders could not be found or contacted. As reported in [last month's update](#), works in the public catalog that are not claimed by the rights holder after a period of 90 days will be considered orphan works. The first 90-period will expire in October (the expiration date for each work is posted in the catalog). At that time, partner institutions that wish may begin to offer their users access to orphan works in HathiTrust. More information about the orphan works project can be found at <http://www.lib.umich.edu/orphan-works/>. Further information on how access will work is included in the Holdings Database newsletter item below.

#### Collection List and Search Enhancements

Staff at the University of Michigan released several enhancements to the HathiTrust Collections list and full-text search application in July. The enhancements to the Collections interface include improved display of collections, the ability to search collections by title and description, and the ability to filter collections by their featured status, last time of update, number of items, and whether or not they belong to the current authenticated user. New full-text search features leverage

#### August Forecast

- Release updated search-within-a-book feature
- Finalize proposal for a collaborative print archiving strategy

#### Presentations

Jeremy York "Unpacking HathiTrust's New Cost Model" Webinar for SUNY, July 15, 2011.

See <http://www.hathitrust.org/papers> for all papers, presentations, and reports.

There's an  
elephant in  
the library.™





## Update On July Activities

the addition of bibliographic metadata to the full-text search index to offer faceting (refinement) of search results, and improved search results relevance ranking. These features were the top two [prioritized by the HathiTrust Full-text Working Group](#) for implementation. Staff at Michigan and the California Digital Library will continue to work on features in the prioritized list in August. The third feature, improvements to “within book search”, will be released in the next couple of weeks. Please give these new features a try and send feedback to [feedback@issues.hathitrust.org](mailto:feedback@issues.hathitrust.org).

### Holdings Database: Update and Lawful Uses of In-Copyright Materials

Early in 2011, HathiTrust began development on a database of holdings information from partner institutions designed a) to support the new cost model that will be implemented for all partners in 2013, b) to form a foundation for the expansion of lawful uses of in-copyright materials to partner institutions (such as access to persons who have print disabilities and access to orphan works), and c) to facilitate collective collection development and management activities among the partnership.

The first iteration of this database, containing data for single part monographs at partner institutions, was put into production in July. Staff at the University of Michigan are in the process of incorporating information from the database into existing applications such as the catalog and PageTurner to begin offering partners access to orphan works in HathiTrust, as well as access to in-copyright volumes for users who have print disabilities. The systems needed to provide access in these scenarios are expected to be in place in late-summer/early-fall.

#### Access to orphan works

Beginning in October, authenticated users from HathiTrust institutions that have selected to grant their users access to orphan works will see orphan works appear as “Full view” in HathiTrust access systems. Access will only be available to orphan works in HathiTrust that are or had previously been held in the partner institution’s library system.

#### Access for users who have print disabilities

### Total Volumes Added

	July	Overall
Columbia University	95	64,001
Cornell University	17,542	345,094
Harvard University	17	52,727
Indiana University	12	184,887
Library of Congress	0	71,418
NYPL	136	258,828
Penn State University	29	39,203
Princeton University	2,489	241,595
University of California	489,975	2,983,660
University of Chicago	164	6,467
University of Illinois	0	14,501
University of Madrid	7	107,954
University of Michigan	36,362	4,404,849
University of Minnesota	376	87,645
University of Wisconsin	15,020	488,031
University of Virginia	1	47,304
Yale University	0	18,385
Total	562,225	9,416,549

### Public Domain (~27% of total)

Total*	94,470	2,508,391
--------	--------	-----------

\*Includes volumes opened through copyright review or rights holder permissions.

**There’s an  
elephant in  
the library.™**





## Update On July Activities

Beginning in late-summer or early-fall, users at partner institutions who are certified as having a print disability will be eligible to view the full text of all in copyright volumes in HathiTrust that are or had previously been held in the partner institution's library system. In order to gain access institutions will need:

1. To be configured for authentication to HathiTrust via Shibboleth, (see <http://www.hathitrust.org/shibboleth>)
2. To have provided HathiTrust with information about their print holdings
3. To have a local process by which eligible users have been certified as having a print disability
4. To convey certification status through a new Shibboleth eduPersonEntitlement attribute

Specifics on the syntax of the attribute and any additional information will be disseminated to partners in the coming weeks.

### Call for new member of User Support Working Group

Nominations have been extended for a new member of the HathiTrust User Support working group. Please send nominations to [jjyork@umich.edu](mailto:jjyork@umich.edu) by August 19, 2011.

## Ingest

### Local Digitization Ingest

Staff at Michigan met with staff from Northwestern University to address questions related to ingest of a set of several hundred locally-digitized volumes. Staff at Universidad Complutense de Madrid began to transfer a second set of locally-digitized manuscripts and incunabula to the University of Michigan for ingest. The first set of locally-digitized materials from Madrid will be ingested in August.

## Working Groups and Committees

### Collections

The Collections Committee is putting the finishing touches on two major work items with which it has been occupied for the last several months: a ballot initiative for a Distributed Print Monographs Archive to be put forward at the Constitutional Convention, and a draft recommendation on the treatment of duplicates in HathiTrust. A draft of the print archives proposal was reviewed with a subgroup of the HathiTrust Executive Committee, which sponsored the initiative, in July; the final version will be forwarded shortly to the full Executive Committee for its approval. The draft duplicates paper will be shared with the Strategic Advisory Board in August for feedback and direction about next steps. A big thank you from the chair (Ivy Anderson) to her colleagues on the committee for terrific work in pulling these proposals together (the charge and membership of the group are available at [http://www.hathitrust.org/wg\\_collections\\_charge](http://www.hathitrust.org/wg_collections_charge)). Once these items are final-

There's an  
elephant in  
the library.™





## Update On July Activities

ized, the committee will turn its attention to other pending items on its work agenda, including a process for responding to individual requests and offers to include additional materials in HathiTrust.

### Communications

In July, the Communications Working Group focused on a number of topics including new partner announcements, a strategy to support public services staff in communicating about HathiTrust, soliciting authors and topics for the HathiTrust blog, and looking ahead to communication needs for the Constitutional Convention. The Communications group invites suggestions from partner institutions and others for topics to be covered in the HathiTrust blog. These should be directed to [heather.christenson@ucop.edu](mailto:heather.christenson@ucop.edu).

### Usability

The Usability Working Group discussed and provided feedback on the Collections list and full-text search features that were released in July. The group continued to review and track feedback received via the User Support Group on issues related to usability. The HathiTrust User Experience Special Interest Group (HT UX-SIG) has been active in discussions about feature requests and usability improvements to HathiTrust. The HT UX-SIG email group is open to anyone who is interested. Please contact Felicia Poe ([Felicia.Poe@ucop.edu](mailto:Felicia.Poe@ucop.edu)) to join.

### User Support

The adjacent table is a summary of the issues received by the User Support Working Group in July.

## Projects

### IMLS Quality Grant

In July, grant project staff at the University of Michigan and University of Minnesota started to review the first of several production-level samples of volumes in HathiTrust, conducted according to the error type and severity model developed by the grant project team. The first sample includes 1,000 randomly selected volumes published before 1923 and digitized by Google. Staff will review a set of 100 pages, chosen at evenly-distributed intervals, within each of the 1,000 volumes. A subset of volumes will be reviewed by multiple staff members as a check on inter-coder reliability. The corresponding print versions of all volumes in the sample will

### User Support Issues

July 2011

<b>Content</b>	<b>90</b>
Quality	89
Non-partner Digital Deposit	1
Collections	2
<b>Cataloging</b>	<b>20</b>
<b>Access and Use</b>	<b>81</b>
Copyright	52
Permissions	2
Takedown	0
Print on Demand	36
Inter-library loan	9
Full-PDF or e-copy requests	13
Datasets	0
Data Availability and APIs	1
Reuse of content	2
<b>Web applications</b>	<b>23</b>
Functionality problems	7
Problems with login specifically	6
General Questions about login	4
Partners setting up login	6
Usability issues	3
Feature requests	8
<b>Partner Ingest</b>	<b>2</b>
<b>General</b>	<b>23</b>
Partnership	6
Infrastructure	0
Miscellaneous	17

\*See [User Support Working Group Issue Types](#) for a description of the types of issues included in each category.





## Update On July Activities

undergo a physical assessment to identify potentially meaningful characteristics that affect quality, such as tight bindings, condition, and other physical features. A subset of the digital volumes will also be subjected to full-volume review to measure errors such as missing pages. The goals of the first production run are 1) to test the quality review system developed by the project team on a large scale; 2) to assemble a body of statistical data of sufficient size to begin to test the feasibility of sampling as a strategy to accurately describe error within a group of volumes; 3) to begin to explore the correlation of physical characteristics of books with observed errors in the digital scans. Review of the 1,000-volume sample is expected to be completed in mid-September.

### HTPub

The University of Michigan has been examining schema options for representing encoded text journal content in the HathiTrust archival package. An investigation of publisher XML formats has yielded a recommendation to use the Journal Archiving and Interchange Tag Set of JATS (an application of NISO Z39.96) as the XML format for encoded text. UM staff are currently researching Portico's use of a custom profile of an earlier version of this standard in content normalization.

### HathiTrust Research Center

The HathiTrust Research Center has received a \$600,000 award from the Alfred P. Sloan Foundation for the first investigation of non-consumptive research for a major large-scale digitized collection of content. The press release for the award is available at <http://newsinfo.iu.edu/news/page/normal/19252.html>.

The HathiTrust Research Center technical group is working on an end-to-end demonstration test of underlying infrastructure functionality. The test, which is planned to be completed in early September, is being conducted using a subset of the HathiTrust full-text Solr index and Indiana University public domain volumes deposited in HathiTrust. OCR text of the volumes was distributed to the Research Center from HathiTrust and is stored in a noSQL data store to be readily available for research purposes. The test scenario runs as follows: a user logs into the Research Center via an InCommon identity and simple algorithms are executed on the user's behalf to pull word counts out of the index and do simple pattern-matching. The algorithms and services, which are available to all users, are registered in a web services registry where they can be queried by users. Results in this simple scenario are returned to the user as a URL. The test will allow the HTRC technical group to work out issues related to the HTRC's core architecture, interfaces, and integrated security model.

Staff at the University of Michigan worked in July to prepare a dataset containing the OCR of approximately 240,000 publicly available non-Google digitized volumes in HathiTrust for distribution to the HathiTrust Research Center. The dataset will be delivered in August and also be available for public download. The HTRC is awaiting resolution on a data agreement that will allow it to host and use OCR text of the full HathiTrust public domain corpus. Pending that agreement, this da-

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

**There's an  
elephant in  
the library.™**







## Update On July Activities

taset will allow the HTRC to conduct testing of its infrastructure on a larger scale.

### Development Updates

---

#### Bibliographic Data Management

The California Digital Library (CDL) development team began the integration phase of the project in July, which focuses on adapting the new management system to the HathiTrust workflow. The team ingested bibliographic records into a virtual staging environment where integration testing with HathiTrust systems will occur. CDL has filled the second Metadata Analyst position for the project, advertised in previous updates. The new staff member will begin work in mid-September.

#### Data API

Staff at the University of Michigan continued development on security enhancements to the HathiTrust Data API. The enhancements are described at <http://bit.ly/jozHQK>. Interested parties are invited to submit comments and feedback to [feedback@issues.hathitrust.org](mailto:feedback@issues.hathitrust.org).

#### Mobile

Last February, University of Michigan staff began development on mobile interfaces to the HathiTrust catalog and PageTurner. Development of an initial version of these interfaces is nearly complete and staff hope to release beta versions for testing in September.

#### New Database and Ingest Servers

Michigan staff installed new database and ingest servers as part of the first periodic server replacement cycle, which keeps server infrastructure current on a 3-to-4-year cycle. The new database servers are a little ahead of schedule, but configured to support the higher transactional rates expected with the introduction of the print holdings database. The new ingest servers are expected to provide significantly increased throughput rates for ingesting volumes into HathiTrust.

#### PageTurner

Staff at Michigan experimented with ways to improve the speed that page images are loaded in the new views for scrolling and flipping through books that were implemented in April. Staff will continue to test the strategy, which involves estimating pixel dimensions of all pages in a volume based on a small sample and making adjustments as actual pages are retrieved, throughout August.

Michigan staff continued work on a more sophisticated throttling system to improve the experience of using HathiTrust while ensuring compliance with third-party agreements on content and offering equal access for all users to HathiTrust applications. The new system will provide throttling controls at finer levels so

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

**There's an  
elephant in  
the library.™**





## Update On July Activities

that, for example, delivering thumbnail page images to a user in PageTurner does not count as heavily against a user's access quota and limit their ability to view full-size pages.

### Outages

There were no outages in July.

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

**There's an  
elephant in  
the library.™**

