



## Update On June Activities

July 8, 2011

### Top News

#### Access to Orphan Works

Following the [announcement in May](#) of a new initiative to identify orphan works in HathiTrust, the [University of Michigan announced](#) last month that it would be making orphan works identified in HathiTrust that are also held in its library collections available to Michigan students, faculty, staff, and other visitors to the UM libraries. Works that are identified as orphan candidates through an extensive review process will be posted in a public catalog at UM for 90 days. Works left unclaimed by rights holders after this time will be considered orphans. Michigan expects to begin offering access to orphan works beginning in the fall. Joining Michigan at the initial release will be the University of Wisconsin-Madison; other partner institutions may begin to make these uses in the coming months.

#### 3-Year Review

[Update on the Briefing Paper on Progress and Opportunities for HathiTrust Prepared by Ithaka S+R for the HathiTrust Strategic Advisory Board \(SAB\)](#)

By Ed Van Gemert, Chair, SAB

The HathiTrust Strategic Advisory Board received the draft three-year review prepared by Ithaka S+R on 17 June 2011. The SAB along with Ithaka staff is currently working to revise that draft. Following the revision period, the final report from Ithaka is to be delivered to the SAB on 15 July 2011.

The SAB initially charged Ithaka S+R staff to challenge our collective thinking and the review has certainly done that. The final report, and portions thereof, will be broadly distributed prior to the October 2011 Constitutional Convention. Key areas of focus of the draft report suggests additional attention and work to include:

- Clearly defining objectives for the next 3-5 years, possibly mapping out a rationale for those objectives in the context of a revised mission statement.
- Enhancing information about HathiTrust's strategic priorities to partner libraries.
- Discussing the advantages and disadvantages of a membership-driven governance structure.
- Demonstrating to partner libraries the sustainability and feasibility of the new cost model for HathiTrust.
- Making decisions based on the most pressing goals and objectives for HathiTrust about how large the membership for the initiative needs to grow.

The SAB expects thorough discussions at the upcoming Constitutional Convention around these and other important questions regarding the future shape of HathiTrust and the role that current and future partner libraries will play in governing and sustaining HathiTrust.

#### July Forecast

- Release new faceting and relevance ranking features for full-text search
- SAB to receive 3-year review report from Ithaka S+R

#### Presentations

Jeremy York "HathiTrust Outside-In". June 14, 2011.

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

There's an  
elephant in  
the library.





## Update On June Activities

### HathiTrust Research Center (HTRC)

The HathiTrust Research Center hosted a reception at the Digital Humanities 2011 Conference held in Palo Alto, California June 20, 2011. The reception was sponsored by Indiana University and the University of Illinois, the institutions developing the HTRC, and by Google. Opening remarks were given by HTRC directors Beth Plale and John Unsworth, and Google Engineering Director John Orwant. The reception was well attended and well received. The HTRC stressed its receptivity to working with researchers broadly within the scope of available resources to provide computational access to the growing body of HathiTrust materials.

The day before the reception HTRC directors traveled to Oakland, CA to meet with Laine Farley, the HathiTrust Executive Committee liaison to the HTRC, and Heather Christenson, chair of the HathiTrust Communications Working Group. The group was later joined by David Greenbaum of [Project Bamboo](#). Discussions focused on interactions between the HTRC and HathiTrust and ways in which HTRC will collaborate with other projects such as Project Bamboo.

The HTRC is pleased to announce receipt of a \$606,000 three-year award from the Alfred P. Sloan Foundation to explore architectural issues around large-scale non-consumptive research. Beth Plale is the PI of the project, with co-PIs Atul Prakash of the University of Michigan and Robert McDonald of Indiana University.

The HTRC wrote letters of support for three proposals to the second round of the [Digging into Data Challenge](#).

### Call for New Member of the User Support Working Group

The Executive Committee is seeking nominations from all partner institutions for a new member of the User Support Working Group. One of the current 8 members will be stepping off the group at the end of July. User Support members are on call to answer inquiries at least one day per week and spend on average of 2-3 hours per week investigating issues and responding to users. Nominations should be sent to Jeremy York ([jjyork@umich.edu](mailto:jjyork@umich.edu)) before August 1, 2011.

### Total Volumes Added

	June	Overall
Columbia University	0	63,906
Cornell University	16,321	327,552
Harvard University	0	52,710
Indiana University	156	184,875
Library of Congress	0	71,418
NYPL	1	258,692
Penn State University	23	39,174
Princeton University	21	239,106
University of California	37,321	2,493,685
University of Chicago	132	6,303
University of Illinois	0	14,501
University of Madrid	1,203	107,947
University of Michigan	12,603	4,368,487
University of Minnesota	625	87,269
University of Wisconsin	7,598	473,011
University of Virginia	0	47,303
Yale University	18,114	18,385
Total	94,118	8,854,324

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

Total*	35,670	2,413,921
--------	--------	-----------

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

There's an  
elephant in  
the library.





## Update On June Activities

### Ingest

---

#### Local Digitization Ingest

HathiTrust began ingest of the first large set of locally-digitized volumes from Yale University in June. More than 18,000 had been ingested as of July 1.

### Working Groups

---

#### Collections

A draft ballot initiative for a print management proposal intended to be voted on at the Constitutional Convention will be shared with the HathiTrust Executive Committee's print management subgroup in July. The Committee also expects to submit its draft discussion paper on duplicate volumes in HathiTrust to the Strategic Advisory Board in July for initial feedback. Recommendations for a process for responding to user-initiated requests has been put on temporary hold while the first two deliverables are finalized.

#### Communications

The Communications Working Group launched a new HathiTrust blog in May, "Perspectives from HathiTrust", with its [inaugural post](#) by HathiTrust Executive Director John Wilkin. The blog will feature authors from among the partner institutions writing on a variety of topics. The group also released a [mid-year update on HathiTrust activities](#) in conjunction with the ALA annual conference.

#### Discovery Interface

After careful consideration and consultation with the HathiTrust Strategic Advisory Board, the Discovery Interface Working Group (DIWG) has officially disbanded. The DIWG, initially convened in spring 2009, fulfilled its charge to accomplish the implementation of the HathiTrust WorldCat Local Prototype interface. One important aspect of this project was working with OCLC to get all of the HathiTrust records loaded into WorldCat. Along the way, the DIWG also supervised the first phase of the HathiTrust Full-Text Search Subgroup and delivered a set of requirements to OCLC for the next phase of HathiTrust WorldCat Local catalog development in FY 2012.

At this point, the focus will shift from the group's original charge to the ongoing maintenance and development of the HathiTrust WorldCat Local catalog. Julia Lovett of the University of Michigan will be the project manager for this effort, and will draw on the expertise of HathiTrust partner colleagues as needed. The DIWG executive team—John Butler, Lee Konrad, and Julia Lovett—would like to thank all the DIWG members for their contributions: Adam Brin, Patricia Martin, Christopher Walker, Lisa German, Kevin Clair, Suzanne Chapman, and Jon Rothman. Special thanks to John Wilkin and to the HathiTrust SAB for providing

There's an  
elephant in  
the library.





## Update On June Activities

valuable guidance and input, and to Bill Carney and the OCLC WorldCat Local team for their very hard work on this project.

### Usability

Work on the development of HathiTrust personas reported in [April's update](#) continued in June. The group has also begun reviewing feedback received via the User Support Group to help discover and track usability issues.

### User Support

The User Support Working Group and staff at the University of Michigan fielded more than 750 user inquiries from April through June 2011. The break-down of issues received during that time is shown in the adjacent table. We will continue to report these statistics on a monthly basis.

### Projects

#### IMLS Quality Grant

The grant project team's work in June focused on preparations for production level data collection to begin in early July. These preparations included continuing work to examine and improve inter-coder consistency, incorporating new data from the University of Minnesota review team, and undertaking several small sampling exercises to guide development of a model for systematic random sampling of HathiTrust volumes, and pages within volumes, for quality review. The project team, under the guidance of the Principal Investigator and team statistician, completed a draft of this model in June. The first large sample for production level analysis will be drawn in early July. Additional information on the project can be found at <http://www.hathitrust.org/grants>.

#### HTPub

The University of Michigan hired the first of two programmers to work on the HTPub project. Interviews will take place in July for the second opening. Meanwhile, Michigan continued to examine schema options for representing journal content in the HathiTrust archival package, and questions surrounding in-

### User Support Issues

April - June 2011

<b>Content</b>	<b>347</b>
Quality	302
Non-partner Digital Deposit	2
Collections	21
<b>Cataloging</b>	<b>54</b>
<b>Access and Use</b>	<b>246</b>
Copyright	139
Permissions	14
Takedown	2
Print on Demand	16
Inter-library loan	3
Full-PDF or e-copy requests	59
Datasets	19
Data Availability and APIs	10
Reuse of content	11
<b>Web applications</b>	<b>86</b>
Functionality problems	30
Problems with login specifically	9
General Questions about login	8
Partners setting up login	3
Usability issues	13
Feature requests	19
<b>Partner Ingest</b>	<b>12</b>
<b>General</b>	<b>68</b>
Partnership	30
Infrastructure	5
Miscellaneous	33

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

There's an elephant in the library.





## Update On June Activities

teroperability of the envisioned HTPub software components with the HathiTrust repository. Details on the project can be found at <http://www.hathitrust.org/ht-pub>.

### Development Updates

---

#### Bibliographic Data Management

The California Digital Library team completed development of the major functionality for the core metadata management system, and on June 14, 2011, demonstrated the core system to staff at the University of Michigan. For initial testing, the system was loaded with approximately 200,000 metadata records from HathiTrust partner institutions. When it is implemented in 2012, the system will manage initially close to eight million.

The next major development effort is to adapt the new system to the HathiTrust workflow. This includes integrating the system with the HathiTrust rights management database and developing batch export functionality for metadata records. CDL is working with University of Michigan staff to understand the particulars of the HathiTrust workflow.

CDL continues to interview for the open Metadata Analyst position: [http://www.cdlib.org/services/d2d/d2d\\_mda2.html](http://www.cdlib.org/services/d2d/d2d_mda2.html).

#### Collection Builder

University of Michigan staff began to code enhancements to the Collection Builder interface in June. The enhancements will allow users to explore the list of collections more easily using new filtering and searching options. Deployment of the new interface is expected in July.

#### Data API

Michigan staff began development of security enhancements to the HathiTrust Data API in June. The enhancements are described at <http://bit.ly/jozHQK>. We invite interested parties to submit any comments or feedback to [feedback@issues.hathitrust.org](mailto:feedback@issues.hathitrust.org).

#### Development Environment

Michigan deployed a timestamp-based sentinel file in the development environment to make it easier for the Plack Perl module, which was implemented to support the new PageTurner functionality, to stay up-to-date when changes to Plack-based applications are deployed to production.

#### Full-text Search

Staff at Michigan completed development to replace the XPat search engine with Solr as the mechanism for searching inside individual volumes from Pageturner (details on the change were reported in the [Update on May 2011 Activities](#)). Use

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

There's an  
elephant in  
the library.





## Update On June Activities

of the Solr back-end will eliminate differences between the ways that Solr and XPat work currently, which can interfere with searching activities, and improve relevance ranking of page-level results. Michigan staff have begun to test the current Solr configuration and search performance to optimize indexing and query response times. The code supporting the new functionality will undergo final testing for production deployment after the release of the new faceting and relevance-ranking features for full-text search, which is projected for mid-July. The coding for these features, the [top two identified by the HathiTrust Full-text Working Group](#), was completed in June, and usability and internal tests are underway in preparation for the mid-July release.

### PageTurner

HathiTrust has throttling protections in place to prevent systematic download of materials in the repository for which, due to third-party agreements, this type of activity is not allowed (see the Message from John Wilkin in the [Update on September 2010 Activities](#)). Staff at Michigan have started a process to add more sophisticated capabilities to HathiTrust applications (for instance, optimization of thumbnail presentation) that will ensure compliance with such agreements while offering fewer interruptions to use.

### New Storage

Michigan staff upgraded software on both Michigan and Indiana storage instances and added 100TB of new capacity with no service interruption.

### Outages

There were no outages in June.

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

There's an  
elephant in  
the library.

