



## Update On March Activities

April 13, 2012

### Top News

#### Board of Governors Elections

HathiTrust has announced the members of its new Board of Governors. The [full announcement](#), as well as information about the [elections process](#), are available on the HathiTrust website. The composition of the Board, which officially begins work April 16, is as follows:

Representatives appointed from the founding partner institutions:

- Paul Courant (University of Michigan)
- Carol Diedrichs (Ohio State University)
- Laine Farley (California Digital Library)
- Wendy Lougee (University of Minnesota)
- Brian Schottlaender (UC San Diego)
- Bradley Wheeler (Indiana University)

Representatives elected at-large:

*Serving 5-year terms from 2012-2016*

- Betsy Wilson (University of Washington)
- Robert Wolven (Columbia University)

*Serving 4-year terms from 2012-2015*

- Richard Clement (Utah State University)
- Patricia Steele (University of Maryland)

*Serving 3-year terms from 2012-2014*

- Carol Mandel (New York University)
- Sarah Michalak (University of North Carolina-Chapel Hill)

#### User Support Working Group Call for Nominations

The User Support Working Group is seeking nominations from partner institutions for up to 4 new members. Nominations should be sent to Jeremy York ([jjyork@umich.edu](mailto:jjyork@umich.edu)) and include the name, title, and a short description of current job duties. Additional information that might be relevant to participation in the group may be included as well. User Support members are on call at least one day per week and follow up on inquiries throughout the week, requiring between 2-4 hours of work. Staff that participate on the group will

- Gain knowledge about HathiTrust's user base, typical problems and questions that are raised and how they are resolved.
- Become aware of new ways HathiTrust is being used, and features and functionality that users desire.
- Gain knowledge of HathiTrust organizational and technical infrastructure, and policies and procedures relating to copyright, access, collection development, deposit of materials, and preservation.

The charge for the working group is available at [http://www.hathitrust.org/wg\\_user-support\\_charge](http://www.hathitrust.org/wg_user-support_charge).

#### April Forecast

Release user interface enhancements for advanced full-text search

Continue work on relevance ranking of full-text search results

Complete work on Data API security

#### Papers & Presentations

Jeremy York, "[HathiTrust: Aspiring to Build the Universal Library](#)". UKSG Annual Conference, March 26, 2012.

Jeremy York, "[HathiTrust and the Research Library of the Future](#)". American Antiquarian Society Conference on Needs and Opportunities, March 31, 2012.

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

There's an  
elephant in  
the library.™





## Update On March Activities

### Data API Modifications

Effective May 1, support for legacy Data API URLs in the following form will be removed:

`http://services.hathitrust.org/api/htd/pathinfo-arguments`

After May 1, URLs should be submitted according to the [current Data API schema](#) without the “api” path element:

`http://services.hathitrust.org/htd/pathinfo-arguments`

**Data API Security:** Over the next several months HathiTrust will be implementing security enhancements to the Data API. The enhancements will require developers using the API to acquire an [OAuth 1.0](#) access key that identifies them, and a secret key that must be used to “sign” URLs to retrieve HathiTrust resources via the Data API. HathiTrust will also provide a Web client that employs a user’s login credentials as a proxy for these keys to facilitate non-programmatic uses. In March, staff at the University of Michigan integrated [2-legged OAuth](#) into the Data API and began to develop the Data API client. Once OAuth is released, there will be an approximately 6-month transition period, ending October 1, 2012, during which signed access to the Data API will be possible but not required. After October 1, all requests to the Data API will need to be properly signed with an access key retrieved from HathiTrust. Complete documentation of the security enhancements and methods of obtaining keys and accessing the Web client is forthcoming. OAuth is planned for release in April 2012.

### Ingest

#### Local Digitization and Internet Archive

University of Michigan staff are preparing tools that will allow partners to build complete ingest packages for materials they wish to deposit in HathiTrust. The tools will include functionality to remediate images and build METS files to HathiTrust specifications, and validate files prior to submission to HathiTrust. Sever-

#### Total Volumes Added

March

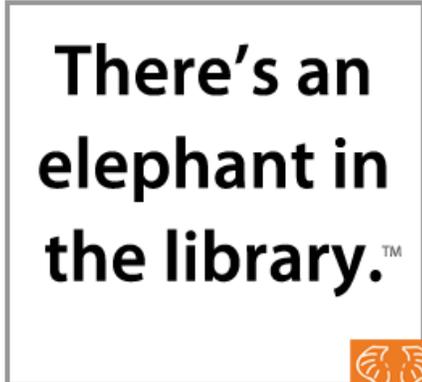
Overall

Total Volumes Added	March	Overall
Columbia University	6	64,183
Cornell University	896	392,356
Duke University	1	4,523
Harvard University	1	53,675
Indiana University	480	187,635
Library of Congress	5	89,416
North Carolina State University	0	3,196
University of North Carolina - Chapel Hill	1	8,088
Northwestern University	554	6,820
New York Public Library	31	259,537
Penn State University	18	43,280
Princeton University	171	250,789
Purdue University	41	23,981
University of California	758	3,329,769
University of Chicago	309	13,206
University of Illinois	1,001	15,504
Universidad Complutense	3,083	111,823
University of Michigan	4,124	4,529,978
University of Minnesota	2,696	95,064
University of Wisconsin	1,297	534,870
University of Virginia	0	48,921
Utah State University	0	90
Yale University	0	23,678
Total	15,473	10,090,382

#### Public Domain (~28% of total)

Total*	5,458	2,785,335
--------	-------	-----------

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





## Update On March Activities

al institutions have agreed to test the tools in the coming months. It is hoped that over time all partners and other entities that contribute content to HathiTrust will use the tools to create their submission packages, thereby distributing the effort needed to ingest materials produced from different sources.

### Working Groups and Committees

Working groups and committees in HathiTrust may have an operational or strategic focus. See [http://www.hathitrust.org/working\\_groups](http://www.hathitrust.org/working_groups) for more information.

#### Operational

##### User Experience Advisory Group

The UX Advisory Group conducted informal usability testing to evaluate the impact of changes proposed to the PageTurner interface to incorporate a volume version (date of last ingest). The group plans to discuss the results and make recommendations on the changes in April, with implementation to follow shortly thereafter.

##### User Support Working Group

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

#### Strategic

##### Collections Committee

The Collections Committee's [report on duplicate volumes](#) in HathiTrust is now available. As described in last month's update, the report recommends that HathiTrust retain all duplicate copies ingested into the repository for the time being, with periodic reassessment. The Strategic Advisory Board has requested that the Committee make further recommendations about the criteria that should be applied in future assessments and identify the future costs and risks of retaining duplicates in the corpus. The Committee also hopes to finalize its recommendations concerning a process for responding to requests and offers within the next several months.

### User Support Issues

	March	February
<b>Content</b>	<b>203</b>	<b>106</b>
Quality	193	97
Non-partner Digital Deposit	0	3
Collections	9	2
<b>Cataloging</b>	<b>49</b>	<b>24</b>
<b>Access and Use</b>	<b>195</b>	<b>131</b>
Copyright	137	73
Permissions	17	20
Takedown	1	1
Print on Demand	0	1
Inter-library loan	2	0
Full-PDF or e-copy requests	19	17
Datasets	2	1
Data Availability and APIs	2	0
Reuse of content	6	0
<b>Web applications</b>	<b>11</b>	<b>22</b>
Functionality problems	4	7
Problems with login specifically	1	0
General questions about login	3	5
Partners setting up login	3	3
Usability issues	0	1
Feature requests	0	0
<b>Partner Ingest</b>	<b>5</b>	<b>5</b>
<b>General</b>	<b>101</b>	<b>152</b>
Partnership	7	11
Infrastructure	0	2
Miscellaneous	94	139

\*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 March Activities

### Projects

---

#### **Bibliographic Data Management**

California Digital Library achieved a milestone in March, loading all bibliographic records submitted by HathiTrust contributing institutions into the Zephir production environment. The goals of this dry run load were to test the functionality of the new metadata management system (Zephir), to test the production infrastructure, and to compare the production loading time with a previous load on a development server. The metadata management team continued to reconcile bibliographic records in Zephir with those in the current system at the University of Michigan to assure all data was accounted for, addressing record discrepancies and ingest errors as they were encountered.

#### **jPach (formerly HathiTrust Publishing)**

Staff of the University of Michigan formally named the journal publishing platform Michigan will use in conjunction with HathiTrust: jPach. Design principles and requirements for jPach, plus a description of the platform's modules, are posted on the [University of Michigan Library website](#). The [project page](#) on the HathiTrust website now includes a full project timeline.

Michigan staff continued work to generate valid JATS XML from DOCX files, render JATS XML files in PageTurner, and create a METS profile for the jPach Submission Information Package.

#### **HathiTrust Research Center (HTRC)**

The HathiTrust Research Center released a [report of its activities](#) over the last 6 months. More information about the research can be found on the [HTRC web page](#).

#### **IMLS Quality Grant**

Project staff continued whole-volume review of digital volumes in the first production sample (pre-1923 English-language Google-digitized volumes), looking for errors such as missing, duplicate, and out-of-order pages, as well as generally "bad" pages, defined in relation to the severity scale established for page-level review. Staff also continued page-level review of the project's 4th 1,000-volume sample, consisting of non-Roman language volumes. Physical review of Michigan volumes sampled in the second production run (post-1923 Google-digitized English-language volumes) continued in March. Students have completed review of 543 of the 600 Michigan volumes present in the 1,000-volume sample. Further information about the grant project is available from the [project website](#).

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





## Update On March Activities

### Development Updates

---

#### Full-text Search

Staff at the University of Michigan completed work on the next iteration of advanced full-text search, which will allow users to build queries with greater Boolean complexity and enhance the ability to revise advanced searches. The new features will be released in early April. Staff made significant progress on plans to improve search results relevance ranking.

#### Storage Hardware Replacement Cycle

Michigan staff installed new storage at the Indiana and Michigan sites that will both accommodate 2012 volume projections and replace storage scheduled for retirement. Storage due for retirement will be taken offline starting in April.

#### Web Hosting Infrastructure Changes

Developers and system administrators at Michigan began preparations to move HathiTrust's Drupal-based informational website and VuFind-based catalog from their initial hosting environments, currently on Michigan library infrastructure, to dedicated HathiTrust hardware, where they will run alongside other HathiTrust applications. This move will simplify application integration.

#### Outages

No outages were reported in March 2012.

HathiTrust sends notice upon discovery and resolution of unscheduled outages and in advance of scheduled outages and maintenance work that may result in an outage. We welcome and encourage additional recipients for these notices. If your institution is not receiving outage notifications and would like to, please contact [feedback@issues.hathitrust.org](mailto:feedback@issues.hathitrust.org).

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

There's an  
elephant in  
the library.™

