# HathiTrust Digital Library

## Update On June Activities

## Top News

### Governance

The Board of Governors has begun a search for a new Executive Director to fill the position John Wilkin will be leaving in August (see the introduction to the 2013 Mid-Year Review). The Program Steering Committee is planning to hold a virtual meeting in July, and an in-person meeting in September.

### Government Documents Registry

HathiTrust completed a scope statement for the government documents registry, including project purposes, assumptions, and constraints. Project staff continued work to identify and analyze potential sources of metadata, and engaged in activities related to the registry structure and functionality.

We will soon be calling for open focus groups to provide feedback on proposed functionality, use cases, and record structure. For more information, or to volunteer for the focus groups, please contact valglenn@umich.edu.

## Ingest

### Local Digitization

Based on survey feedback from several institutions preparing locally-digitized content for deposit, HathiTrust discussed plans for two new online validation services: a simple web-based service that will validate individual TIFF or JPEG2000 files against repository specifications, and a full-volume cloud-based service that will validate entire volume packages. HathiTrust is discussing development schedules for both projects with the aim of providing the first in the late summer or early fall, and the second on a later timeframe. HathiTrust continued to correspond with several institutions about ingest of locally digitized materials, including Indiana University, North Carolina State University, Texas A&M University, and the University of Illinois.

### General

Penn State updated bibliographic records for 30,000 volumes, adding information needed to identify them as U.S. federal government documents, and making them viewable worldwide in HathiTrust.

## Working Groups and Committees

### User Support Working Group

A summary of User Support inquiries received in June is included at the end of the update.

## July Forecast

Continue work on full-text relevance ranking.

Continue exploration of the Solr experimental block-join functionality.

Continue to design a process for indexing the full text of JATS XML articles.

Continue work to enable the delivery of JATS XML articles as PDFs.

## Papers & Presentations

Robert McDonald, Yiming Sun, Miao Chen, "The HathiTrust Research Center (HTRC): An Overview and Demo", IUPUI Libraries, June 7, 2013.

Kevin Hawkins, "A Model for Integrating the Publication and Presentation of Journal Articles", Texas A&M University, June 10, 2013.

Carrick Rogers, "Generating Highly Structured XML from Word Processing Documents: An Overview of mPach", Stanford University, June 13, 2013.

There's an elephant in the library.™

www.hathitrust.org

# HathiTrust Digital Library

## Update On June Activities

## Projects

### Bibliographic Data Management

The California Digital Library (CDL) team has initiated a full load of all current HathiTrust bibliographic records to a staging instance of the new metadata management system, Zephir. Prior to loading the production system, this staging load will serve as a point of comparison with the current system and reflects recent coding changes. The CDL team continues to work with staff at the University of Michigan to ensure that Zephir includes the same records as those in the current system. Additionally, the CDL team is working with University of Michigan staff to prepare for running the two systems in parallel later this summer. The timeline for the project is available at http://www.hathitrust.org/htmms.

### Copyright Review

A summary of the determinations from HathiTrust copyright review activities in June is given below. See CRMS-US and CRMS-World for further information.

|  | June | | Overall | |
|---|---|---|---|---|
|  | Public Domain | All Determinations | Public Domain | All Determinations |
| CRMS-US | 3,774 | 7,909 | 139,201 | 260,769 |
| CRMS-World | 3,008 | 5,648 | 29,901 | 55,657 |
| Total | 6,782 | 13,557 | 169,102 | 316,426 |

### HathiTrust Research Center (HTRC)

Plans for the second annual Hathi Trust Research Center UnCamp 2013 are developing.  The September 8-9, 2013 event at University of Illinois Urbana Champaign will be as hands-on and participant-responsive as the first event was. It will feature stellar keynote speakers including Matt Wilkens, who specializes in contemporary American fiction, and digital and computational literary studies at Notre Dame, and Christopher Warren, specialist in Renaissance literature as it relates to politics, law, international political thought, and intellectual history, at Carnegie Mellon.   New this year is a Scholarly Communication Office Hours.  The office hours is a pilot for user services: participants will have the option to sign up for individual consultation sessions with members of the UIUC library.
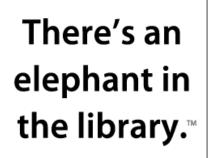
Mark your calendar to join us!  Additional information about the UnCamp will be posted to http://www.hathitrust.org/htrc_uncamp2013 as it becomes available.

### mPach

University of Michigan staff presented on mPach's publication and preservation model in separate sessions with staff from the Texas A&M University Library,

## Update On June Activities

University Press, and Integrated Ocean Drilling Program (IODP). Michigan and IODP staff discussed possible uses of mPach by IODP. Michigan staff also presented at Stanford University Libraries on the Norm tool for converting DOCX files to JATS XML. The timeline on the mPach project page was updated to show expected dates of completion for major project milestones.

## Development Updates

HathiTrust institutions performed the following work related to applications and Web interfaces:

### Full-text Search

As part of work to improve relevance ranking, staff opened an issue in Solr's ticketing system to allow Solr to use Lucene's BlockGroupingCollector along with Solr's result grouping feature (see the Update on May 2013 Activities for further background).

Staff continued to design a process for indexing JATS XML articles.

### PageTurner

Staff continued to develop functionality to deliver JATS XML articles as PDFs.

### Storage Hardware Replacement Cycle

Storage due for retirement at the Indianapolis site was taken offline and is undergoing the standard security wipe. Storage due for retirement at the Michigan site is halfway through the same process. The replacement process is projected to be completed in July.

### Outages

No outages were reported in June.

| Total Volumes Added | June | Overall |
| --- | --- | --- |
| Boston College | 182 | 2,361 |
| Columbia University | 0 | 65,033 |
| Cornell University | 3,600 | 423,436 |
| Duke University | 0 | 4,523 |
| Harvard University | 1 | 236,069 |
| Indiana University | 60 | 195,297 |
| Library of Congress | 0 | 89,724 |
| North Carolina State University | 0 | 3,196 |
| Northwestern University | 1,375 | 35,344 |
| New York Public Library | 11 | 288,354 |
| Penn State | 450 | 59,637 |
| Princeton University | 0 | 251,704 |
| Purdue University | 0 | 44,692 |
| Universidad Complutense | 1 | 111,983 |
| University of California | 2,496 | 3,391,291 |
| University of Chicago | 421 | 30,812 |
| University of Florida | 0 | 2,068 |
| University of Illinois | 1,807 | 111,125 |
| University of Michigan | 4,002 | 4,647,133 |
| University of Minnesota | 524 | 107,343 |
| University of North Carolina - Chapel Hill | 0 | 16,588 |
| University of Wisconsin | 11 | 555,744 |
| University of Virginia | 0 | 50,815 |
| Utah State University | 0 | 117 |
| Yale University | 0 | 23,678 |
| Total | 14,941 | 10,748,067 |
| Public Domain (~31% of total) | | |
| Total* | 19,406 | 3,405,432 |

*Includes works opened via copyright review and rights holder permissions.

# HathiTrust Digital Library

## Update On June Activities

| User Support Issues | June | May |
|---|---|---|
| **Content** | **342** | **390** |
| Quality | 329 | 386 |
| Collections | 13 | 13 |
| **Cataloging** | **81** | **135** |
| **Access and Use** | **202** | **144** |
| Copyright | 148 | 79 |
| Permissions | 10 | 10 |
| Takedown | 0 | 1 |
| Print on Demand | 1 | 0 |
| Inter-library loan | 4 | 0 |
| Full-PDF or e-copy requests | 12 | 20 |
| Datasets | 4 | 5 |
| Data Availability and APIs | 0 | 1 |
| Reuse of content | 1 | 7 |
| **Web applications** | **20** | **33** |
| Functionality problems | 9 | 13 |
| Problems with login specifically | 2 | 3 |
| General questions about login | 0 | 2 |
| Partners setting up login | 1 | 1 |
| Usability issues | 1 | 1 |
| Feature requests | 1 | 1 |
| **Partner Ingest** | **3** | **10** |
| **General** | **34** | **57** |
| Partnership | 8 | 5 |
| Infrastructure | 0 | 0 |
| Miscellaneous | 26 | 52 |
| **Total** | **670** | **769** |

*See User Support Working Group Issue Types for a description of the types of issues included in each category.

## Most-accessed volumes

| Title |
|---|
| Roster of the Confederate soldiers of Georgia, 1861-1865, v.1. |
| A treatise on money, v.1 1930, by John Maynard Keynes. |
| The nature and sources of the law, by John Chipman. |
| A history of agriculture in the state of New York, by Ulysses Prentiss Hedrick. |
| L'Algérie, ancienne et moderne depuis les premiers éstablissements des Carthaginois jusqu'à l'expedition du général randon en 1853. Vignettes par Raffet et Rouargue Frères. |
| Phantasms of the living, Vol. 1, by Edmund Gurney. |
| The human figure, by John H. Vanderpoel. |
| Air commerce bulletin, v.2, Jul-Jun, 1930-31. |
| Nouvelles lettres de la reine de Navarre: adressées au roi François Ier, son frère, by Marguerite, Queen of Navarre. |
| Heraldry in America, by Eugene Zieber. |

There's an elephant in the library.™

www.hathitrust.org