HathiTrust Digital Library

Update On September Activities

October 13, 2014

Late Breaking News

HathiTrust Member Meeting

Representatives of HathiTrust institutions gathered for the first annual Member Meeting on October 10 in Washington, D.C. The agenda for the meeting, and presentations given, are posted online. Meeting notes and further information will be available soon.

Top News

HathiTrust Member Meeting

Representatives of HathiTrust institutions will gather for the first annual Member Meeting on October 10 in Washington, D.C. The agenda and presentations given during the meeting are posted online, and a notes and further information on the meeting will be posted soon.

Statement on "Shellshock" Bash Vulnerability

HathiTrust released a statement on the "Shellshock Bash Vulnerability. In short:

- HathiTrust infrastructure was only negligibly vulnerable, as there was only one user interface function in HathiTrust that employed bash, and that function required authenticated access.
- Developers resolved this limited vulnerability by removing the use of bash for this user interface function on September 25 at 11:20am ET, approximately 25 hours after the vulnerability was widely announced.
- Per standard security practice, bash was updated on HathiTrust systems later the same day, at 5:35pm ET, when a fix was made available.

Third Annual HathiTrust Research Center UnCamp

Save the date! The third annual HTRC UnCamp will be held at the University of Michigan, March 30-31, 2015. Additional details will be posted at http://www.hathitrust.org/htrc_uncamp2015 as they become available.

Reminder about Print Disabilities Services

This is a reminder that member institutions are able to gain access to in-copyright works in HathiTrust for users at their institutions who are certified as having a print disability. Details about the service are available at http://www.hathitrust.org/accessibility. Please contact us if you have any questions.

October Forecast

Continue work on new Image Server capabilities for continuous text content.

Reassess accessibility features of PageTurner with particular attention to supporting new content types.

Migrate to Solr 4.10 and re-index the collection.

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Ingest

General

HathiTrust ingested nearly 350,000 volumes in September, including large amounts of content from the Getty Research Institute, the University of Alberta, the University of Illinois, and Indiana University. Many hundreds of thousands more are expected in the coming months, as more Google-digitized content, previously held in escrow, is ingested from Committee on Institutional Cooperation institutions, and ingest from other HathiTrust members continues.

Locally-digitized Content

HathiTrust began to process content submitted for ingest by the University of Illinois and the University of Delaware. HathiTrust was also in communication with Boston College, University of Iowa, New York University, Yale University, Texas A&M University, Virginia Tech, Princeton University, Columbia University, and the University of Maryland about upcoming content deposits.

Bibliographic Data Management

In September, the California Digital Library loaded 275,833 new or updated bibliographic records into Zephir.

Projects

Copyright Review

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

	September		Overall	
	Public Domain	All Determinations	Public Domain	All Determinations
CRMS-US	312	518	166,753	316,396
CRMS-World	3,579	6,630	75,775	145,804
Total	3,891	7,148	242,528	462,200

For many years, the University of Michigan Copyright Office has provided an invaluable service for HathiTrust, performing copyright reviews on works in the repository in response to user inquiries from around the world on a wide variety of materials. A large number of requests fall outside the scope of those reviewed in the IMLS-funded CRMS-US and CRMS-World projects. However, due to the volume of inquiries received, we have decided to pause reviews of such materials in order to strategize more efficient means of handling ad-hoc copyright reviews. Part
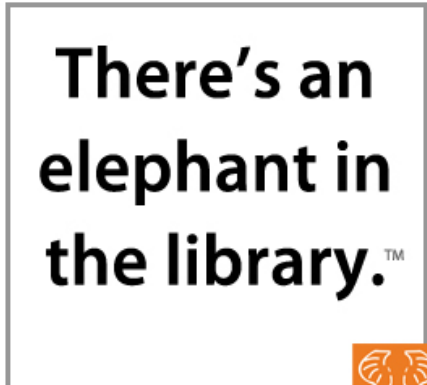
Papers & Presentations

Mike Furlough, "Sharing Collections through Shared Stewardship: A HathiTrust Progress Report". Multiple venues, September 8, 10, 12, 19, 2014.

Jeremy York, "Today's Needs, Tomorrow's Necessities: Future Practitioner Skills", Digital Cultural Content Forum, September 11, 2014.

Downie, J. Stephen, Kirstin Dougan, Sayan Bhattacharyya, Colleen Fallaw (2014). The HathiTrust Corpus: A Digital Library for Musicology Research? In Proceedings of The 1st International Digital Libraries for Musicology workshop (DLfM 2014), ACM/IEEE Digital Libraries Conference 2014, London, September 12, 2014. Forthcoming DOI: http://dx.doi.org/10.1145/2660168.2660173.

Mike Furlough, "Preservation and Sustainability at HathiTrust", Beyond NDNP Symposium, Washington, DC, September 16, 2014.

Mike Furlough, "Getting More from HathiTrust: Resources, Tools, and Services", Carnegie Mellon University, September 12, 2014.

There's an elephant in the library.™

www.hathitrust.org

of the University of Michigan's third IMLS grant for copyright review involves an exploration of sustainability strategies with HathiTrust, and support for reviews outside the current CRMS projects will be considered in conjuction with that work. We would like to express our deep gratitude to the University of Michigan Library for its work in this area, and through the CRMS projects we will continue our efforts to make as many volumes in the HathiTrust repository available as legally possible.

Government Documents Registry

Staff continued to test and refine a relationship detection process, working with sets of known duplicate and related bibliographic records in HathiTrust. Work also continued to develop and improve processes for normalizing bibliographic metadata such as enumeration and chronology information, and merging duplicate bibliographic records. Applications are still being accepted for a developer position to support the work of the HathiTrust registry. Applications can be submitted online through the University of Michigan Jobs site.

HathiTrust Research Center

Miao Chen, Robert H. McDonald, and Zong Peng from the IU Data to Insight Center gave a series of presentations at The Ohio State University on September 4, 2014 on the HathiTrust Research Center that included a 2 hr hands-on tutorial using an OSU computer lab. Many thanks to the OSU Libraries for hosting the HTRC set of lectures in the new wing of their Thompson Library. Below are the details:

- Public lecture about the HathiTrust Research Center (Robert McDonald, Associate Dean of Libraries, Indiana University) (approx. 50 attendees)
- Hands-on with data from the HathiTrust (Miao Chen and Zong Peng) - (approx. 22 attendees)
- HTRC Community discussion session (Robert McDonald, Miao Chen, and Zong Peng) - (approx. 30 attendees)
- Miao Chen and Robert H. McDonald led a breakfast discussion session at the IU Statewide IT Meeting on October 8, 2014.

The HTRC team* delivered an HTRC Data Capsule hands-on workshop on Sep 15 at Scholars Commons of IU Library. 8 participants from different backgrounds, including computer science, education, and digital library, attended the session.

*Workshop hosts were: Robert McDonald, Miao Chen, Guangchen Ruan, Jiaan Zeng, Peng Zong

Development Updates

Development activities by HathiTrust institutions included the folowing:

Authentication, Authorization and Access:

- Added functionality to automatically expire access keys that are configured

to allow special access to content via the HathiTrust Data API.

- Began to add support for "access profiles", which will associate materials with the same access and use restrictions together, facilitating the management of access control parameters.
- Made enhancements to the way authentication and access are handled for institutions that are members of consortia.

Full-text Search

- Investigated numerous Solr 4 configuration issues in preparation for migration from Solr 3 to Solr 4.
- Prepared to incorporate item-level date information into full-text search (e.g. for serial and multi-volume publications) to improve the accuracy of date searches.
- Received and installed long-awaited pre-release software for the high-performance storage system and confirmed that the software resolved previously observed performance and stability problems. An additional software release, expected to make the storage suitable for production, is forthcoming. In the meantime, staff will conduct preliminary system benchmarking using the storage in October.

Image Server

- Configured applications (PageTurner, Collection Builder, bibliographic and full-text catalogs) to display thumbnail images in search results from local image files when thumbnails are not returned by the Google Books API.

Server replacement

- Completed the installation of new full-text search servers in Michigan, and scheduled installation (in early October) for new full-text search servers in Indiana.

Availability

Repository

Cumulative 12-month availability of repository access: 99.844%*

HathiTrust service was interrupted briefly on Wednesday, September 17 from 11:41-11:42am when a manual maintenance activity was accidentally started on full-text search servers at the Michigan instance while the Indiana instance was out of service. The Indiana instance was put into service immediately when the issue was detected.

An intermittent disc issue caused degraded performance of the Zephir FTPS server on September 23, 2014 (the server used by content contributors to submit bibliographic records). The issue was resolved by early afternoon on September 24.

* Repository access refers to page viewing and full-text search functionality, i.e., user-facing applications. It does not refer to preservation or storage infrastructure, which is under continual operation.

HathiTrust Digital Library

Update On September Activities

User Support Issues	September	August
Content	**172**	**154**
Quality	161	145
Collections	10	9
Cataloging	**223**	**181**
Access and Use	**110**	**172**
Copyright	61	115
Permissions	8	5
Takedown	1	0
Print on Demand	1	1
Inter-library loan	2	2
Full-PDF or e-copy requests	16	18
Datasets	4	4
Data Availability and APIs	1	1
Reuse of content	5	5
Web applications	**22**	**30**
Functionality problems	10	12
Problems with login specifically	1	0
General questions about login	2	0
Partners setting up login	2	5
Usability issues	0	0
Feature requests	0	2
Partner Ingest	**12**	**28**
General	**101**	**99**
Partnership	14	8
Miscellaneous	87	91
Total	**640**	**664**

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

Most-accessed volumes

Quicksand, by Nella Larsen.

Mitchell's Modern Atlas: A Series of Forty-Four Copperplate Maps.

The lion monument at Amphipolis, by Oscar Broneer.

An Oration, Commemorative of the Late Major-General Alex R. Hamilton, by J. M. Mason.

The Five Laws of Library Science, by S. R. Ranganathan.

Industrial Instrumentation, by Donald Eckman.

Godey's magazine. v.40-41 1850

The Human Figure, by John H. Vander-poel

Modern California Houses: Case Study Houses, 1945-1962, by Esther McCoy.

Bad Boys: Public Schools in the Making of Black Masculinity, by Ann Arnett Ferguson.

There's an elephant in the library.™

www.hathitrust.org

HathiTrust Digital Library

Update On September Activities

Total Volumes Added	September	Overall
University of Alberta	75,974	75,974
Boston College	0	3,210
Columbia University	0	65,166
Cornell University	4,397	502,467
Duke University	0	7,775
Getty Research Institute	16,121	16,121
Harvard University	0	238,065
Indiana University	196,136	392,262
Keio University	0	90,080
Knowledge Unlatched	0	27
Library of Congress	0	108,883
McGill University	0	893
New York Public Library	0	294,818
North Carolina State University	0	3,196
Northwestern University	21	56,642
Ohio State University	20	50,569
Penn State	30	91,527
Princeton University	19	252,800
Purdue University	0	46,913
Sterling & Francine Clark Art Institute	0	358
Texas A&M	0	1,201
Universidad Complutense	0	113,378
University of California	7,154	3,581,318
University of Chicago	72	51,903
University of Connecticut	0	4629
University of Delaware	9	37
University of Florida	0	9,866
University of Illinois	117,784	295,036
University of Massachusetts, Amherst	0	11,115
University of Michigan	2,031	4,703,633
University of Minnesota	90	138,580
UNC - Chapel Hill	0	17,025
University of Virginia	0	51,206
University of Wisconsin	628	559,312
Utah State University	0	117
Yale University	0	23,678
Total	344,512	11,783,806

Public Domain (~35% of total)

	September	Overall
Total*	143,875	4,155,433

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

There's an elephant in the library.™

www.hathitrust.org