Sharing Collections through Shared Stewardship

A HathiTrust Progress Report

Association of Southeastern Research Libraries
Fall 2014 Membership Meeting
Mike Furlough
Executive Director, HathiTrust
Today’s Conversation

• Do you really know HathiTrust?
  – How things work
  – Collections and data
• What are we working on now?
• How has the world changed since we began?
  – And what does that mean for HathiTrust
The Partnership
Mission

To contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge.

Efforts include, but are not limited to

...building comprehensive collections co-owned and managed by partners.
...enabling access by users with print disabilities.
...supporting computational research with the collections.
...stimulating shared collection storage strategies among libraries.
Timeline: Highlights

- Launch (2008)
- TRAC certification (2011)
- Constitutional convention (2011)
- 10 million volumes (2012)
- New governance established (2012)
- Current bylaws and fee structure (2013)
- 13 million volumes (2014)
Collective Stewardship

• Draw upon knowledge across institutions

• Distributed Functions and Services
  – Preservation repository and access services
    • University of Michigan
    • Mirror site: Indiana University
  – Metadata management services
    • California Digital Library
  – HathiTrust Research Center
    • Indiana University and University of Illinois
HathiTrust Members

Allegheny College
American University of Beirut
Arizona State University
Baylor University
Boston College
Boston University
Brandeis University
Brown University
California Digital Library
Carnegie Mellon University
Case Western Reserve
Colby College
Columbia University
Cornell University
Dartmouth College
Duke University
Emory University
Getty Research Institute
Georgetown University
Harvard University Library
Indiana University
Iowa State University
Johns Hopkins University
Kansas State University
Lafayette College
Library of Congress
Massachusetts Institute of Technology
McGill University
Michigan State University
Montana State University
Mount Holyoke College
New York Public Library
New York University
North Carolina Central University
North Carolina State University
Northwestern University
Ohio State University
Pennsylvania State University
Princeton University
Purdue University
Rutgers University
Stanford University
State University System of Florida
Syracuse University
Temple University
Texas A&M University
Texas Tech
Tufts University
Universidad Complutense de Madrid
University of Alabama
University of Alberta
University of Arizona
University of British Columbia
University of Calgary
University of California Berkeley
University of California Davis
University of California Irvine
University of California Los Angeles
University of California Merced
University of California Riverside
University of California San Diego
University of California San Francisco
University of California Santa Barbara
University of California Santa Cruz
University of Chicago
University of Connecticut
University of Delaware
University of Houston
University of Illinois
University of Illinois at Chicago
University of Iowa
University of Kansas
University of Maine
University of Maryland
University of Massachusetts Amherst
University of Miami
University of Michigan
University of Minnesota
University of Missouri
University of Nebraska-Lincoln
University of New Mexico
University of North Carolina at Chapel Hill
University of Notre Dame
University of Oklahoma
University of Pennsylvania
University of Pittsburgh
University of Queensland
University of Tennessee, Knoxville
University of Texas
University of Utah
University of Vermont
University of Virginia
University of Washington
University of Wisconsin-Madison
Utah State University
Vanderbilt University
Virginia Tech
Wake Forest University
Washington University
Yale University Library
**Five-year terms (beginning Jan 2013)**
- Betsy Wilson (University of Washington)
- Bob Wolven (University of Columbia)

**Four year terms**
- Richard Clement (University of New Mexico)
- Patricia Steele (University of Maryland)

**Three year terms:**
- Carol Mandel (New York University)
- Sarah Michalak (University of North Carolina-Chapel Hill)

**Members appointed by the founding institutions:**
- James Hilton (University of Michigan)
- Carol Diedrichs (Ohio State University)
- Laine Farley (California Digital Library)
- Wendy Lougee (University of Minnesota)
- Brian Schottlaender (UC, San Diego)
- Brenda Johnson (Indiana University)

**Ex Officio (Board, PSC, Executive Committee):**
- Mike Furlough, Executive Director

**Executive Committee**
- Chair
- Past Chair
- Treasurer
- Chair of PSC
- Executive Director

HathiTrust Board of Governors
Program Steering Committee

• Serves at the direction of the Board of Governors to...
  – Reviews HathiTrust’s development agenda, shaping initiatives and strategies.
  – Develops position papers to encourage debate or mobilize activity
  – Works with the Board of Governors to develop policies for HathiTrust and its members.
Program Steering Committee Membership

- Ivy Anderson (CDL)
- John Butler (Minnesota)
- Chris Freeland (Washington University)
- Todd Grappone (UCLA)
- Martha Hruska (UC San Diego)
- Martin Kurth (New York University)
- Erika Linke (Carnegie Mellon University)
- Robert McDonald (Indiana)
- Elaine Westbrooks (Michigan)
- Bob Wolven, Chair (Columbia)
PSC Actions 2013-2014

• Initial focus on Constitutional Convention Ballot Initiatives (Government Documents, Shared Print)
• Re-established Collections Committee
• Created Rights & Access Working Group
• Charged Zephir Advisory Group
HathiTrust Collections and Services
Preservation with Access

• Preservation
  – TRAC-certified
  – Long-term commitments on digital content facilitate planning, decision-making

• Discovery
  – Bibliographic and full-text search of all materials
  – Mechanisms for local loading of records

• Access and Use
  – Full text search (all users)
  – Public domain and open access works (all users)
  – Collections and APIs (all users)
  – Lawful uses of in-copyright works (members)
Access: Lawful uses of in-copyright works

• Sensitive to multiple legal regimes
  – Full-text search (everyone everywhere)
  – Access to users who have print disabilities (through member proxy in US, and where law permits)**
  – Access works that are damaged or missing and also out of print and unavailable (members in US only)

**Terms and conditions at http://www.hathitrust.org/access_use#ic-access
Growth of Collection

<table>
<thead>
<tr>
<th>Year</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>2,477,871</td>
</tr>
<tr>
<td>2009</td>
<td>5,221,092</td>
</tr>
<tr>
<td>2010</td>
<td>7,836,698</td>
</tr>
<tr>
<td>2011</td>
<td>9,966,572</td>
</tr>
<tr>
<td>2012</td>
<td>10,599,355</td>
</tr>
<tr>
<td>2013</td>
<td>10,878,121</td>
</tr>
<tr>
<td>2014</td>
<td>12,960,482</td>
</tr>
</tbody>
</table>
Copyright Distribution

In Copyright or undetermined: 63%
“Public domain”: 38%
Public Domain Worldwide: 21%
US Government Documents: 5%
Public Domain (US): 11%
Creative Commons: 0.06%
Open Access: 0.06%
US Gov’t Publications by Source Library

- Michigan, 34.75%
- California, 14.30%
- Minnesota, 13.22%
- Illinois (UC), 12.71%
- Purdue, 7.30%
- Penn State, 6.92%
- Cornell, 3.61%
- Virginia, 1.41%
- Northwestern, 1.18%
- Florida, 1.16%
- Wisconsin, 0.75%
- Indiana, 0.52%
- Princeton, 0.58%
- NYPL, 0.49%
- LC, 0.25%
- Chicago, 0.24%
- Ohio State, 0.22%
- Others, 0.36%
- NYPL, 0.49%
- LC, 0.25%
- Others, 0.36%
- US Gov't Publications
82.3% of the collection was published in 1993 or before
Top Gov Docs: September 2014

US Government publications accounted for 10.65% of all HathiTrust views
Collective Action: Copyright Review

- Copyright Review Management System
  - Systematic manual review of copyright registrations to determine status of portions of the HathiTrust Collection
  - CRMS US: Published in US, 1923-1963
    - 316,396 reviewed / 166,753 PD (~53%)
  - CRMS-World: Published in UK (1874-1944), Canada, Australia (1894-1964)
    - 145,804 reviewed / 75,775 PD-world 9 (~52%)
Current Initiatives
Current Initiatives

1. Developing a shared print monographs archive
2. Expanding coverage and access to US government publications
3. Expanding support for computational (non-consumptive) research
Shared Print Monographs Archive

• Ballot Initiative passed at the 2011 HT Constitutional Convention (Con-Con)
  – “To develop a print monographs archive corresponding to volumes represented within the HathiTrust”

• Focus
  – Ensure preservation of print and digital collections
  – Catalyze national/continental collective management of collections
Why A Shared Print Archive Program

- Many regional efforts, but limited national/international coordination
- Strengthens preservation commitments
  - Connects both print and digital preservation
- Significant need and desire to reduce costs of collection management and associated footprint
Print Monographs
Archive Task Force

• Tom Teper, Chair (University of Illinois)
• Clem Guthro (Colby College)
• Robert Kieft (Occidental College)
• Erik Mitchell (University of California, Berkeley)
• Jake Nadal (ReCAP)
• Jo Anne Newyear Ramirez (University of British Columbia)
• Matthew Sheehey (Harvard University)
• Emily Stambaugh (California Digital Library)
• Karla Strieb (Ohio State University)
Questions: Monographs Archive

• What incentives will encourage participation?
• What services and access models are needed?
• What structures are needed to establish commitments from multiple programs?
• What costs will be associated with the program and how should they be allocated?
• Which libraries are most likely to benefit?
• What are the user’s needs for access?
Government Documents Initiative

• Ballot Initiative: provide “expanded coverage & enhanced access to U.S. Government Documents.”

• Activities:
  – Developing a registry of US Federal Government Documents
  – Locate materials for inclusion in the collections
  – Improve search and discovery
The Registry

• Goal: “...include metadata for the comprehensive corpus of U.S. federal documents. This will include materials produced at U.S. government expense, in all formats, at the item level, from 1789 to the present.”
Working Group Membership

- Mark Sandler (CIC)
- Prue Adler (Association of Research Libraries)
- Ivy Anderson (California Digital Library)
- Joni Blake (Greater Western Library Alliance)
- Kirsten Clark (Minnesota)
- Rick Clement (New Mexico)
- Elizabeth Cowell (Santa Cruz)
- Mark Phillips (North Texas)
- Jon Rothman (Michigan)
- Judy Russell (Florida)
- Barbie Selby (Virginia)
- Jeremy York (HathiTrust)
Initial Recommendations

• Bibliographic and collections analysis
  – Registry and holdings work

• Focus first on known and cataloged materials
  – Prioritize print, post-1976 materials
  – Identify collections for inclusion (and get them)
  – Digitize where needed

• Publicize the efforts
  – Within the library community
  – To the general public
Computational Access

• HathiTrust distributes public domain datasets

• HathiTrust Research Center
  – Developed collaboratively by Indiana University and University of Illinois; launched July 2011
  – Funding from the Sloan Foundation, Andrew W. Mellon Foundation, and NEH Office of Digital Humanities.
  – Partially Funded by HathiTrust (2014-2018)
HTRC Governance

• Reports to the HathiTrust Board of Governors
  – Advisory Group in formation

• HTRC Executive Committee
  – **J. Stephen Downie** (Co-director), Professor and Associate Dean for Research, University of Illinois GSLIS
  – **Beth Plale** (Co-director and Chair), Director Data To Insight Center and professor in the School of Informatics and Computing at Indiana University
  – **Robert H. McDonald**, Associate Dean of Libraries/Deputy Director Data to Insight Center at Indiana University
  – **Beth Sandore Namachchivaya**, Associate University Librarian for Information Technology Planning & Policy at the University of Illinois
  – **John Unsworth**, Vice Provost for Library & Technology Services and Chief Information Officer at Brandeis University
Goals for HTRC

• Provide a persistent and sustainable structure to enable original and cutting edge research.
  – Leverage data storage and computational infrastructure at Indiana & Illinois
  – Stimulate community development of new functionality and tools
  – Use tools to enable discoveries that would not be possible without the HTRC

• Enable scholars to fully utilize content of HathiTrust Library while preventing intellectual property misuse within U.S. copyright law.
  – Provision secure computational and data environment for scholars to perform research using HathiTrust Digital Library.
Example Projects Supported by HTRC

• Muñoz, Trevor, University of Maryland. “Distributed Metadata Correction and Annotation.”
  – Correction, annotation and enhancement of HT records and export as linked data

• Page, Kevin, Oxford University. “EIEPHãT: Early English Print in HathiTrust, a Linked Semantic Workset Prototype”
  – Development of secondary worksets based on both HT and the Early English Books Online Text Creation Partnership (EEBO-TCP).

• Burton, Vernon. “The South as ‘Other,’ the Southerner as ‘Stranger.’”
  – Explore how attitudes expressed in print about slavery, southerners, and non-southerners have changed over both time and space.

• Ted Underwood, Associate Professor of English at the University of Illinois, Urbana-Champaign.
  – Using public domain texts received from HathiTrust to explore changing relationships in literary genres from 1700-1899.
HTRC Opportunities

• Request for Proposals: Advanced Support
  – Competitive research award with dedicated HTRC support time
  – HathiTrust Members have priority
  – Due: January 8, 2015

• Uncamp #3: MAR 30-31, 2015, Ann Arbor, MI
  – Workshops, speakers, demonstrations
Some Thoughts on the Present and Future
How are we positioned?

- Our mission, collection, and the repository operations are all strong.
- Our brand reputation is outstanding.
- Our work is solidly supported by the law.
- We have expanded access in unprecedented ways.
- The partnership provides a solid base for action.
- We have very important programs underway.
What needs thought?

• Strategy, mission, and role in the future
  – Membership growth
  – Collections program
  – Public policy
  – (Inter)National digital infrastructure
  – Services for members and the public

• Organizational
  – Engagement with researchers and libraries
  – Enabling more participation in plans and action
  – Standing on our own
Assumptions

• Our actions must align with the mission, goals, and purpose across our partnership.

• A few additional assumptions
  – We should pursue complementarity and cooperation, not competition and duplication.
  – Scale will continue to drive our strategies
  – Potential partners are not just other libraries and library organizations, but also readers, authors, publishers.
How to find out more

• About: http://www.hathitrust.org/about
• Resources: http://www.hathitrust.org/resources
• Twitter: http://twitter.com/hathitrust
• Facebook: http://www.facebook.com/hathitrust
• Monthly newsletter:
  – http://www.hathitrust.org/updates
  – RSS http://www.hathitrust.org/updates_rss
• Contact us: feedback@issues.hathitrust.org
• Blogs: http://www.hathitrust.org/blogs
  – Large-scale Search
  – Perspectives from HathiTrust
Thank you!

furlough@hathitrust.org
@MikeFurlough