Participation in HathiTrust copyright projects

October 5, 2016
Kristina Eden, Copyright Review Project Manager
Mike Furlough, Executive Director
Webcast outline

• Why we do this
• Project plan for 2017-18
• High level process overview
• How participation works
Mission and Purpose

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

- A trusted digital preservation service enabling the broadest possible access worldwide.
- An organization with over 100 research libraries partnering to develop its programs and services.
- A range of transformative activities enabled by working at a very large scale.
HathiTrust in October 2016

• 14.7 million total items
  – 7.3 million book titles
  – 403,000 serial titles
  – 761,000 US federal government documents
  – 5.6 million full-view items (public domain & CC-licensed)

The collection includes (mostly) published materials in bound form, (mostly) **digitized from library collections.**
### Public domain detected through automated processes

<table>
<thead>
<tr>
<th>Pub Place</th>
<th>Pub Date</th>
<th>Other condition</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>1922 or earlier</td>
<td></td>
<td>Public domain</td>
</tr>
<tr>
<td>US</td>
<td>1923 or later</td>
<td>Not US Govt. Publication</td>
<td>In Copyright</td>
</tr>
<tr>
<td>US</td>
<td>1923 or later</td>
<td>US Govt. Publication</td>
<td>Public Domain</td>
</tr>
<tr>
<td>US or non-US</td>
<td>No date</td>
<td></td>
<td>Undetermined</td>
</tr>
<tr>
<td>Non-US</td>
<td>1875 or earlier</td>
<td></td>
<td>Public domain</td>
</tr>
<tr>
<td>Non-US</td>
<td>1876-1922</td>
<td></td>
<td>Pub domain US only</td>
</tr>
<tr>
<td>Non-US</td>
<td>Post-1923</td>
<td></td>
<td>In copyright</td>
</tr>
</tbody>
</table>
HathiTrust Volumes by View/Copyright Status - August 2016

- LIMITED VIEW / IC, 61.72%
- FULL / PD Worldwide, 19.02%
- FULL / PD US, 13.97%
- FULL-US Federal Docs, 5.09%
- FULL / Creative Commons, 0.20%
Copyright determination: Why now?

- “Bibliographic indeterminacy”
- Investment in our collections
- Low certainty on rights status
- Fulfilling the promise of the public domain
Your Questions
Copyright review

• GOAL: Improve access to limited view volumes where legally possible.

• Prior projects:
  – CRMS-US; U.S. monographs 1923-1963
  – CRMS-World; UK/Canada/Australia bef 1943/63
  – Funded by the Institute for Museum and Library Services.

• Current efforts:
  – Seeking reviewers for next cycle
  – U.S. monographs 1923-1963
  – U.S. state government documents, 1923-1977
CRMS project results to date

2008 – Sept 2016:
• 632,000 total determinations
• 336,000 (~53%) public domain
• Those volumes opened up/ now available as full text in HathiTrust

Usage stats
• 30% of top 500 accessed volumes in HathiTrust were opened by CRMS (when last checked in 2013)

• WINNER OF ALA’s 2016 L RAY PATTERSON AWARD!
Finding the Public Domain

The CRMS Toolkit

- History
- Methodology
- Best Practices

Open Access at
http://dx.doi.org/10.3998/crmstoolkit.14616082.0001.001
CRMS-US
How many volumes in candidate pool

Candidate pool scope
• Monographs
• U.S. publication
• 1923-1963
• Filtered to remove dissertations, translations, foreign language, duplicates

Increasing due to new ingest
Project goals: 2017-2018

• Relying upon voluntary, distributed reviewers.
• US publications, 1923-1963
  – Candidates: 78,000 + more???
  – Copyright renewal check, also © notice
• US state and local documents, 1923-1977
  – Candidates: 56,000
  – Research inquiries (banking reports, directories, natural resource bulletins, legislative documents)
  – Many published without © notice

5 October 2016
Why just 1923-1963 for now?

No U.S. copyright term if...

• 1923-1963 = not renewed or published w/o notice
• 1964-1977 = published w/o notice
• 1978-March 1, 1989 = published w/o notice and not subsequently registered in 5 years
Your Questions
Our approach

• Define metadata parameters to assemble the candidate pool
• Legal research done
• Create a decision tree
• Working from digital scans only. No pre-digitization review.
• Two independent reviews per volume
  – Expert review if conflicting
• Set aside complicated volumes
Double review process
Process changes this time

- Switching order of operations – check for copyright renewal put first
- Will give better data on copyright status of overall volume
- Also we’re seeking to make renewal info we find available and connected to volume record

Prior results: focused on fast determination of PD. Renewal IDs weren’t searched for overall volume if it was going to be undetermined anyway because of Insert materials.

<table>
<thead>
<tr>
<th>Project Total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Public domain</td>
<td>310631 (53.1%)</td>
</tr>
<tr>
<td>In copyright</td>
<td>101440 (17.4%)</td>
</tr>
<tr>
<td>Undetermined</td>
<td>172514 (29.5%)</td>
</tr>
<tr>
<td>Total</td>
<td>584585</td>
</tr>
</tbody>
</table>
CRMS-US 1923-1963
何 much can we do in a month?

U.S. 1923-1963 candidates reached zero in 2014

PRIOR WORK: averaged 5-7K/mo; 60K+ per year

NOW:
- Have 78K candidates now
- About 3,600/mo to finish in 2 years
- Depends on how many participants
Staffing

• What makes a good reviewer?
  – Qualifications
  – Minimize distractions
  – Widescreen monitor
  – 20-30% time commitment

• How are reviewers trained?
  – Remote instruction from HathiTrust
  – Sandbox practice system
  – Decision tree and CRMS wiki
Reviewer statistics

Performance statistics can be sent to local supervisors upon request
## Timeline

<table>
<thead>
<tr>
<th>Date</th>
<th>Event Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct 15</td>
<td>Open call ends – minimal yes/no needed</td>
</tr>
<tr>
<td>mid-Nov thru Dec</td>
<td>State Gov doc training (select group only)</td>
</tr>
<tr>
<td>Dec 1</td>
<td>Need final names &amp; time commitments</td>
</tr>
<tr>
<td>Early Jan</td>
<td>Setting up authorizations and access</td>
</tr>
<tr>
<td>Jan 25 thru Mar</td>
<td>1923-1963 Training (main group)</td>
</tr>
<tr>
<td>mid-March</td>
<td>Reviewers moved into Production as ready</td>
</tr>
</tbody>
</table>

5 October 2016
How to join

• HathiTrust members only
• Reviewers commit to min. 1 year, 20-30% FTE
• Email Kristina Eden, keden@hathitrust.org
Wish List

- Searchable database of Catalog of Copyright Entries (CCE)
- Image or OCR recognition of copyright symbol in front matter
- Automated check of candidate volume titles/authors against Stanford Renewal database - string matching is hard!
Other ways to help

• Metadata corrections
  – Better duplicate detection saves us effort
  – Enum chron differences cause issues
  – Some corrections can change access even w/o a manual copyright review (U.S. fed docs...)

• What do you have authority to sign the CC license permission form and open up in HathiTrust?

• Get those 1923+ volumes deposited

• Remain engaged in the changing legal landscape
How to find out more

• Contact us: feedback@issues.hathitrust.org
  – Mike: furlough@hathitrust.org
  – Kristina: keden@hathitrust.org

• HathiTrust Copyright Review Projects
  https://www.hathitrust.org/copyright-review

• CRMS Toolkit: 2016, Michigan Publishing
  http://dx.doi.org/10.3998/crmstoolkit.14616082.0001.001

• Rights holder permission agreement
  https://www.hathitrust.org/permissions_agreement

• Newsletter: http:www.hathitrust.org/updates
• Twitter: @HathiTrust