Big Collections in an Era of Big Copyright: Practical Strategies for Making the Most of Digitized Heritage

Melissa Levine, Lead Copyright Officer
University of Michigan

Tuesday, October 28, 2014
DLF Meeting | Session organized by John Mark Ockerbloom
Public domain?
Enter HathiTrust possibility
The HathiTrust Digital Library Today

12,364,646 total volumes
6,359,997 book titles
324,978 serial titles
4,327,626,100 pages
554 terabytes
146 miles
10,046 tons

4,500,176 volumes (~36% of total) in the public domain

Copyright Determination: Why Now?

• A response to “bibliographic indeterminacy” (Wilkin, 2011)
  – High investment in our collections
  – Low certainty about the rights status of volumes in those collections

• United States: 1923-1963
• Australia (before 1955), Canada: 1894-1964 pma 50
• Australia (after 1954), United Kingdom: 1874-1944 pma 70

• Fulfilling the promise of the public domain
• Context of orphan works
What CRMS has accomplished
Total: **312,631 determinations**
Total unique volumes exported: 228,286

Total: **105,732 determinations**
Total exported: 97,941
Partners
Indiana University • University of Minnesota • University of Wisconsin

Baylor University
University of California, Irvine
University of California, Los Angeles
University of California, San Francisco
California Digital Library
Columbia University
Duke University
Univ. Of Illinois, Urbana-Champaign
Indiana University
Johns Hopkins University
University Of Maryland
University Of Minnesota
Northwestern University
Penn State University
Princeton University
Ohio State University
McGill
The CRMS Review Approach

- Narrow the legal framework
- Translate that to an algorithm
- Express that as a working interface - candidates viewed in interface with resources
- Two independent reviewers:
  - Confirm accuracy of the catalog record
  - Assess front matter
  - Consult resources for relevant data (author death date, etc.)
  - Render a judgment
- If judgments conflict, an expert reviewer renders a final determination
Achieving legally reliable results

- identify a practical scope of inquiry
- comprehensively research the law
- use proven and auditable methods
- standardize training…questions about ‘train the trainer approach’
- expect to set many things aside because of complications and uncertainty
Works published in the United States
1923-1963
Formalities (registration and renewal)
Well-established resources
  Stanford renewal database, VIAF, LoC Authority, etc.
Reviews performed by catalogers
  Two independent reviews
  Expert resolution of conflicting reviews
Complex or unresolvable items set aside (“Undetermined”)
Response to Berne Convention
1870s – 1940s
Author death dates
Well-established resources
   VIAF, LoC, Archives Canada, NGCOBA, etc.
Two independent reviews
Expert resolution of conflicting reviews
Complex or unresolvable items set aside ("Undetermined")
CRMS – 2 & 3

New: Advisory Board (World)
  Jack Bernard, Georgia Harper, Kenneth Crews, Peter Hirtle, Kevin Smith, Jessica Litman, Bobby Glushko, Sharon Farb

New: Online training

Very new from HathiTrust: Rights and Access Committee
  Elaine Westbrook, Chair
  Members: me, Kevin Smith, Sharon Farb, Kyle Courtney, Monica McCormick, Joseph Hafner
THE BOOK OF
GLASGOW CATHEDRAL
A History and Description

EDITED BY
GEORGE EYRE-TODD

WITH SPECIAL CHAPTERS WRITTEN BY
AND STEPHEN ADAM, F.R.A.S.

ILLUSTRATED BY
DAVID SMALL, HERBERT RAILTON, J. A. DURCAN, AND OTHERS

GLASGOW
MORISON BROTHERS, 55 RENFIELD STREET
1881
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
First grant: working it all out
The CRMS Review Process
CRMS-US Decision Tree

CRMS-World Decision Tree
Second grant:

making more friends, getting more advice

moving from formalities to pma
Futurama – pma? Copyright forever?
The CRMS Wiki
Secure Review Interfaces (US)
What we are thinking about

Crowd sourcing (Author Death Dates?)
  Only if well-defined and carefully planned
  Accuracy – variations in VIAF and authority records

Ongoing improvements to training
  Online training material
  Remote training sessions
  Qualtrics surveys

Requirements for reviewers
  25% preferred - maximum of 35%
  Librarians, staff, grad students (in some cases)
  Training is an investment

Expert reviewers at UM and partner institutions
  Continuous vetting of new resources for rights research
  Copyright country-by-country
Some issues
examples
GATT

Applies to published works if:

- A work published in an eligible country more than 30 days before its US publication.
- At least one author who is a citizen or resident of non-US country (that is a member of Berne, signed WIPO, etc.)
- Still in copyright in its source country in 1996.
- Fell out of copyright in the US due to noncompliance with formalities.

Result: “GATT Restoration”

- Effective January 1, 1996
- Restored copyright period extends for the term the work would have had if it had not gone into the public domain.
Inserts

Components in a larger work—such as a photo, artwork, a foreword, a chapter—that have more than a modicum of creativity and which may have their own discrete copyrights.

As libraries, we must treat ‘books’ and other works as entire artifacts. Otherwise public domain status does not mean that much. **No redacting.**

Big issue for works with multiple creators (music recordings, film). Monographs are hard enough!

The tree on the right is a supplement for our main research process *just for dealing with inserts!*
Vulnerability to staffing changes

- Documentation
- Cross training
- Avoid knowledge silos
- Identify secondary responders
- Never allow the developer to take vacation time
- Never allow staff turnover
Foundations for any new projects

Two review process
Expert adjudicators
15-25% time commitment
Centralized training using a combination of static materials and interactive tools
third grant:
Continue World with our friends
Sustainability plan with HathiTrust
Pilots underway…

• A CRMS “Toolkit” – *new grant*
  – How to create and run a CRMS-style project
  – For institutions with sufficient resources
  – From initial design to data preservation
• Expand our scope to works beyond English
  – Spanish, German, etc.
  – partners in corresponding countries?
• Tackle the undetermined (UND) volumes.
  – Can we identify patterns?
  – Start with “easier” sub-categories?
• State government documents
• Linked data ideas
  – Sharing metadata with USCO, WATCH, etc.?
• Married women’s names as a cataloging identification project
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.
Collections Committee

- Responsible for collection development activities and policies
- Charged with making key recommendations about the content included in HathiTrust.
  - Collection development
  - Prioritization for new content types
  - Collection management tools and analytics
Rights and Access Working Group

- **Responsibilities:**
  - Become familiar with previous work to open access to HathiTrust materials.
  - Support sustainability planning for the CRMS project.
  - Clarify priorities for copyright review of specific materials, such as serials, state documents, and publications in specific languages. May propose pilot studies to assess feasibility of these reviews.
  - Develop a strategy for engaging rights holders to encourage them to open access to their materials. Working with the Collections Committee, identify sets or collections of publications to target for inclusion.
  - Develop criteria to evaluate potential open access publishing partnerships.
EDUCAUSE

Webinar, November 5
1-2 pm EST

Finding the Public Domain: 19 Institutions Making a Difference

Library at Rijksmuseum
DPLA/Europeana
Met April & Oct 2014
Knight News Challenge
Systematic framework
Rights Metadata
HathiTrust, NYPL models
-variations of IC/PD/UND
White Paper (draft)
DPLA/Europeana

Leveraging Europeana experience
Modeling Creative Commons
26,000 rights descriptions DPLA = 5-6 high level categories
Areas of common interest:
- no adding layer of © to pd items
Distinction:
- US fair use provides options na for Europeana