HATHITRUST

A Shared Digital Repository

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.*

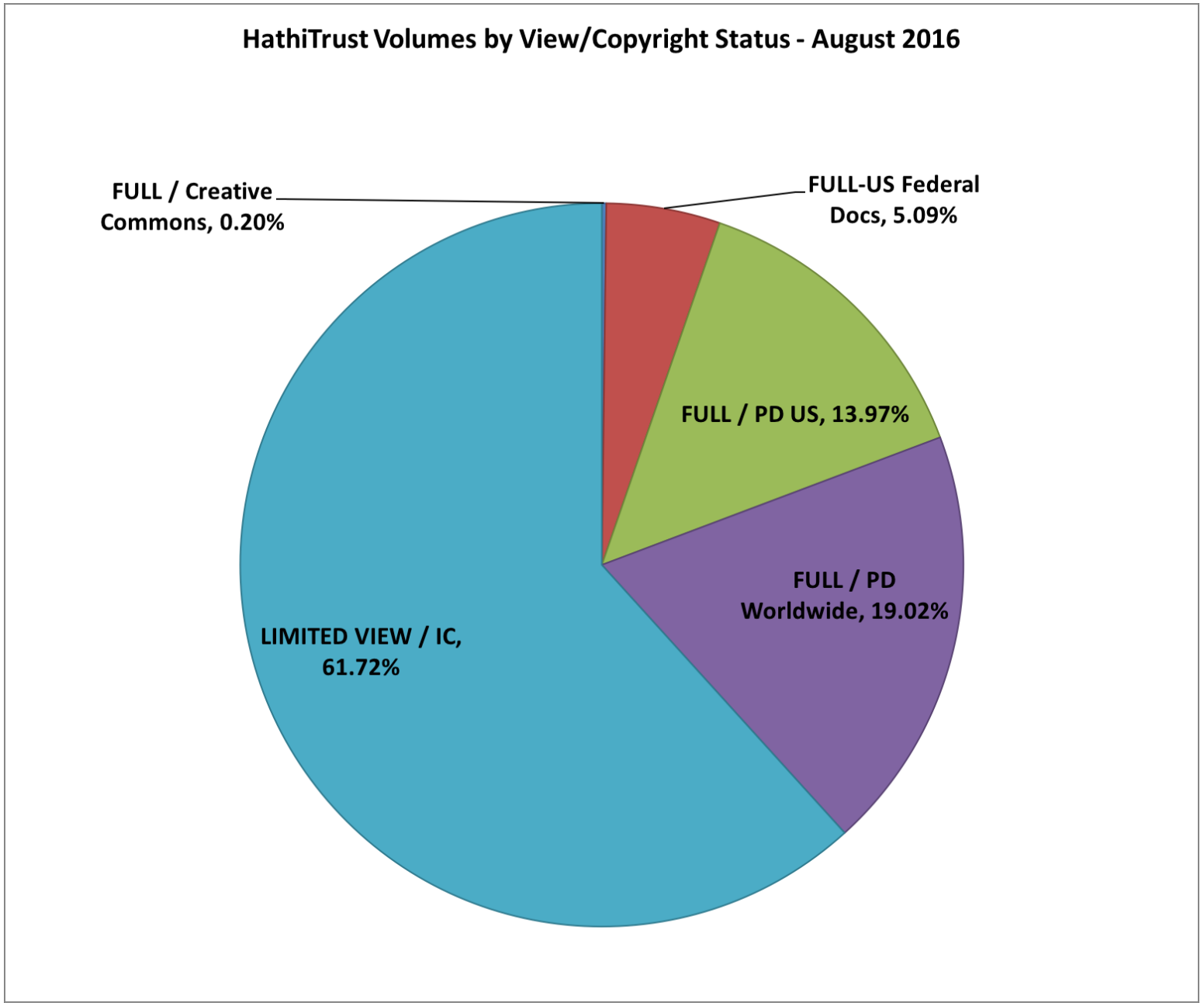
HathiTrust in October 2016

- 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)

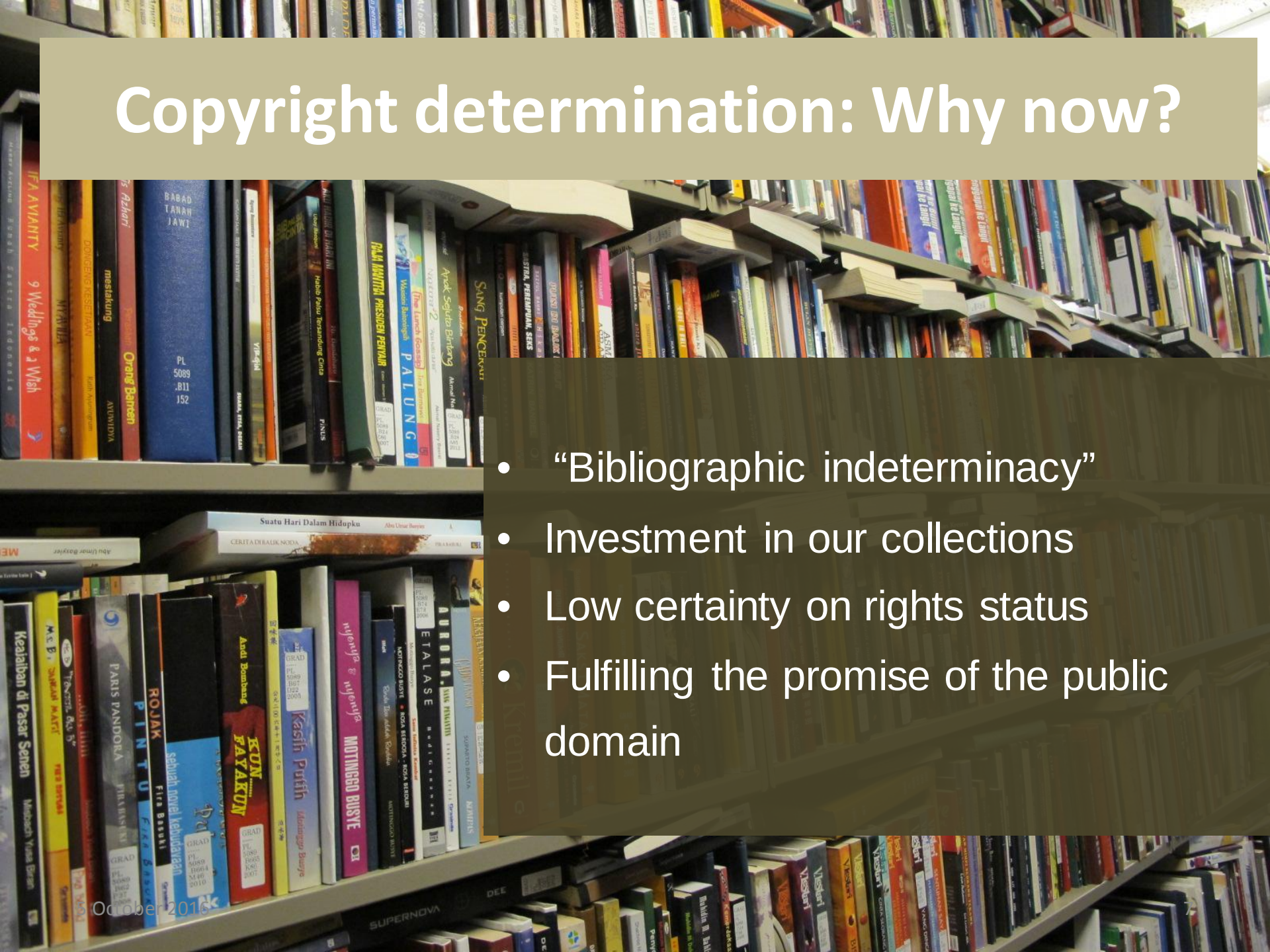
The collection includes (mostly) published materials in bound form, (mostly) **digitized from library collections.**

Public domain detected through automated processes

Pub Place	Pub Date	Other condition	Status
US	1922 or earlier		Public domain
US	1923 or later	Not US Govt. Publication	In Copyright
US	1923 or later	US Govt. Publication	Public Domain
US or non-US	No date		Undetermined
Non-US	1875 or earlier		Public domain
Non-US	1876-1922		Pub domain US only
Non-US	Post-1923		In copyright

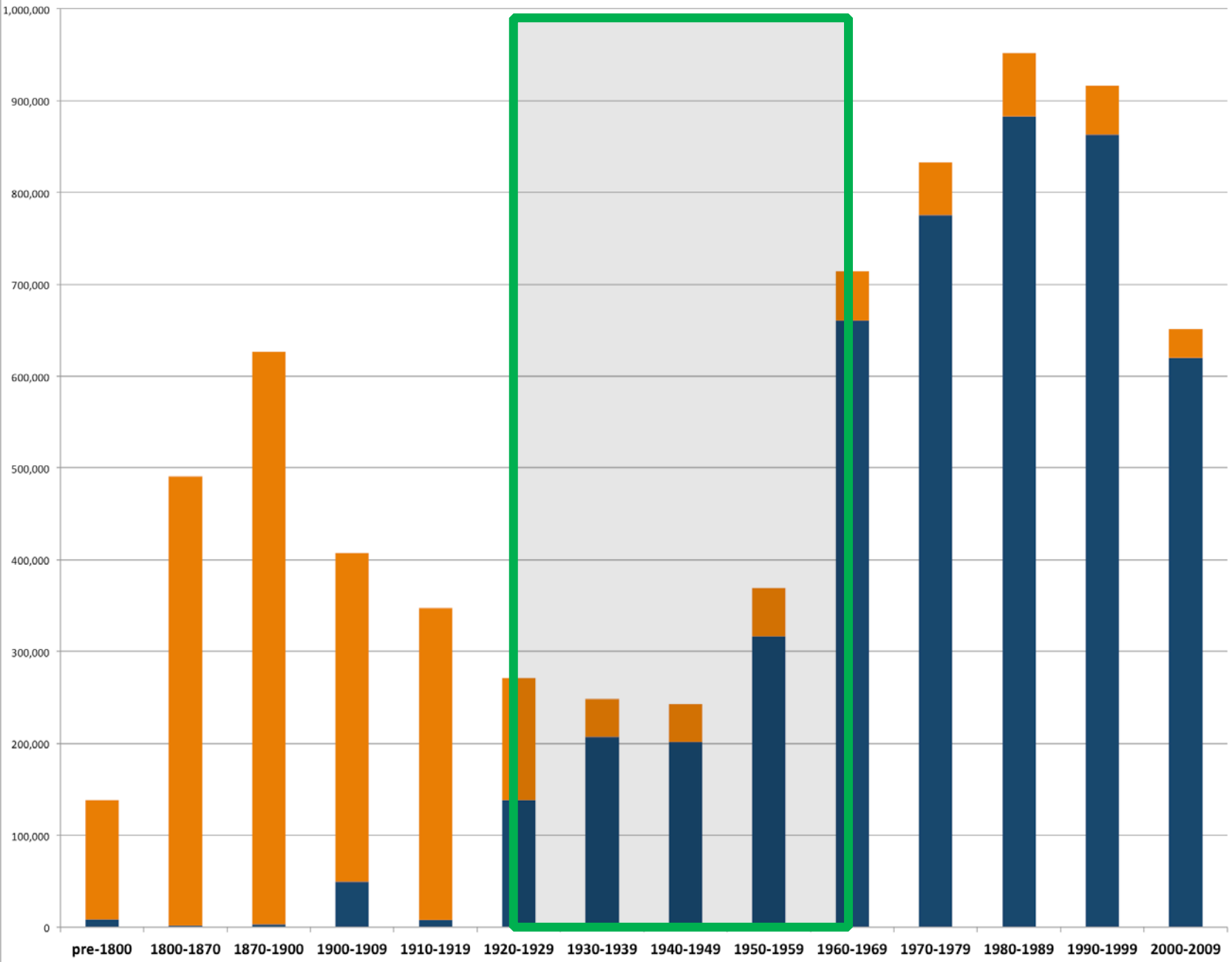

HathiTrust Volumes by View/Copyright Status - August 2016
FULL / Creative Commons, 0.20%
FULL-US Federal Docs, 5.09%
FULL / PD US, 13.97%
FULL / PD Worldwide, 19.02%
LIMITED VIEW / IC, 61.72%

Copyright determination: Why now?



- "Bibliographic indeterminacy"
- Investment in our collections
- Low certainty on rights status
- Fulfilling the promise of the public domain

HathiTrust Titles by Date and Viewing Status



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)

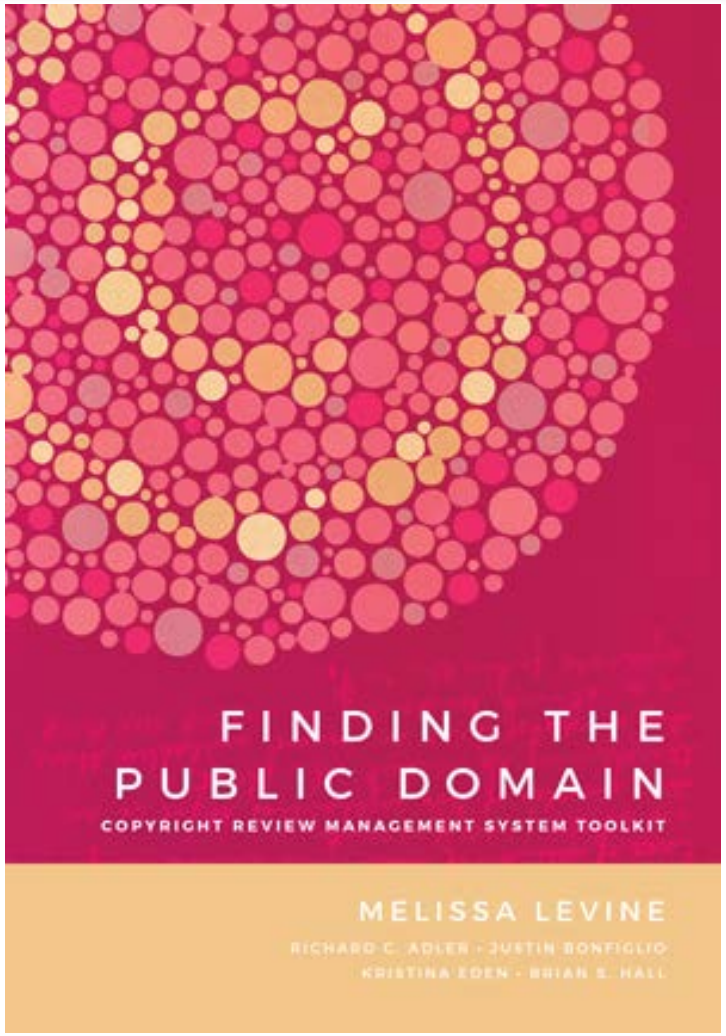
- ***WINNER OF ALA's 2016 L RAY PATTERSON AWARD!***

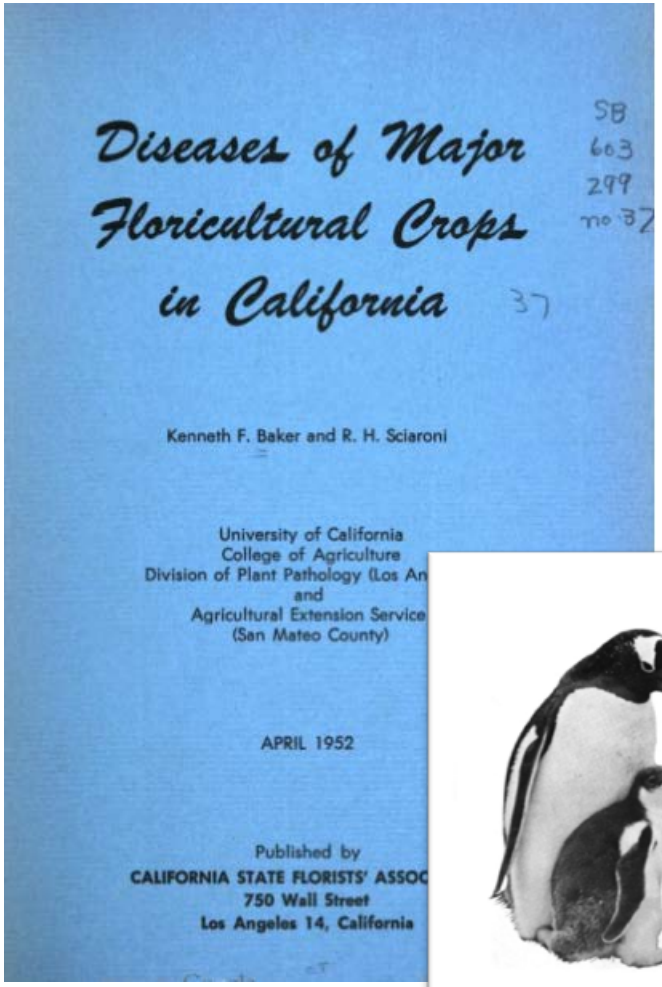
Finding the Public Domain

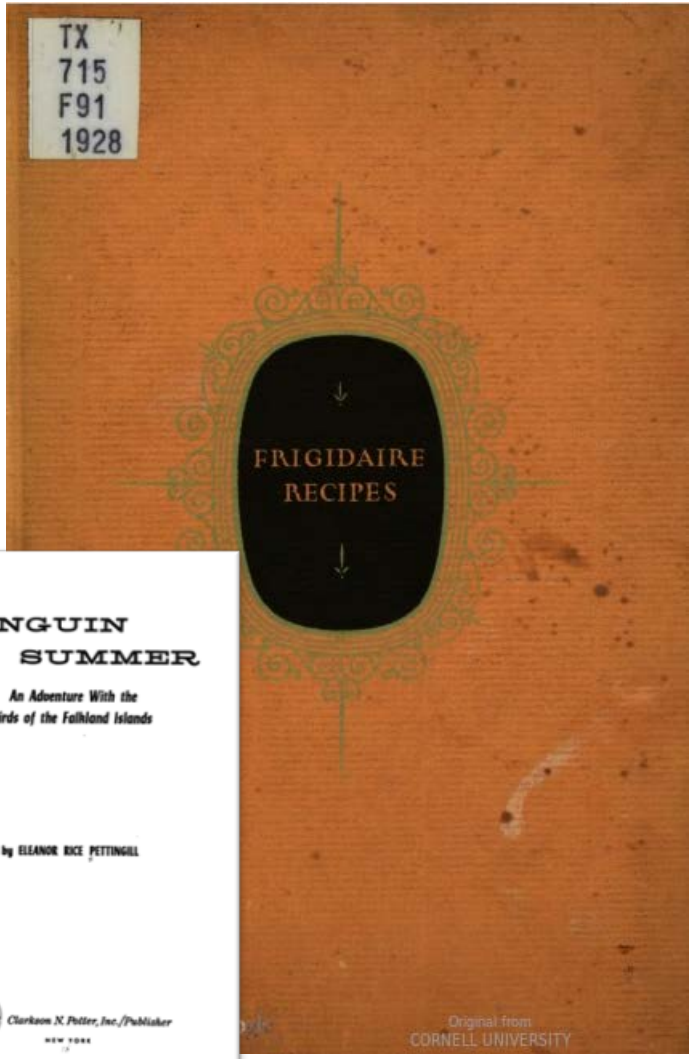
The CRMS Toolkit

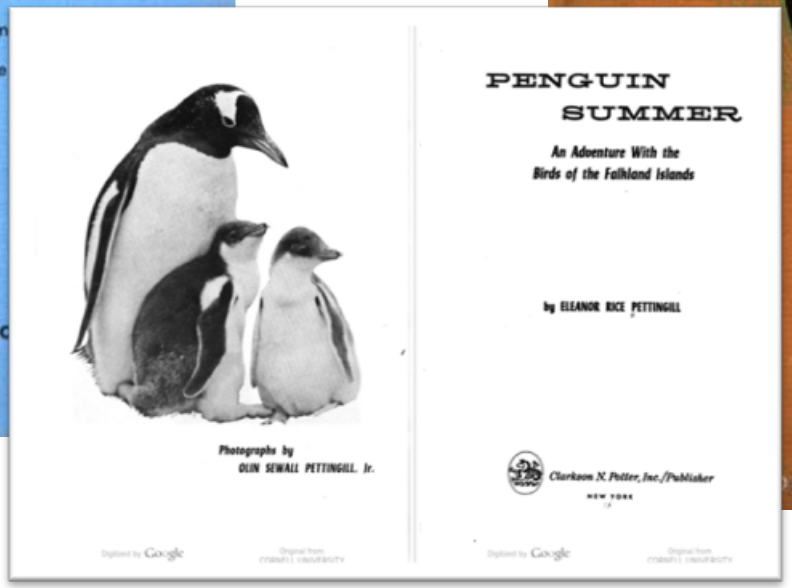
- History
- Methodology
- Best Practices

Open Access at http://dx.doi.org/10.3998/crmstoolkit.14616082.0001.001

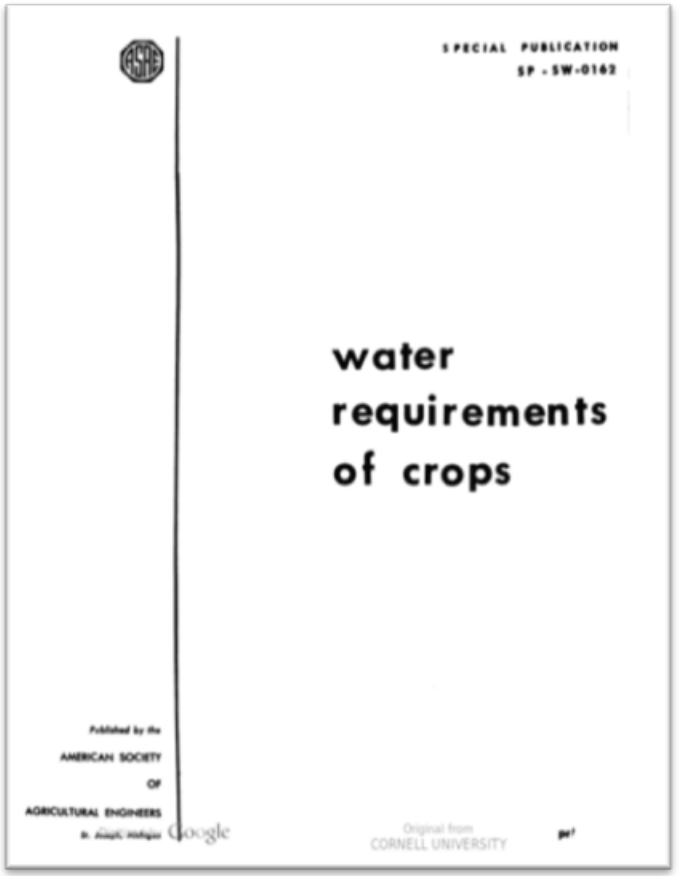



Diseases of Major Floricultural Crops in California
SB 603 299 no 37
37
Kenneth F. Baker and R. H. Sciaroni
University of California
College of Agriculture
Division of Plant Pathology (Los An
and
Agricultural Extension Service
(San Mateo County)
APRIL 1952
Published by
CALIFORNIA STATE FLORISTS' ASSOC
750 Wall Street
Los Angeles 14, California


TX
715
F91
1928
FRIGIDAIRE
RECIPES
Original from
CORNELL UNIVERSITY


PENGUIN
SUMMER
An Adventure With the
Birds of the Falkland Islands
by ELEANOR RICE PETTINGILL
Photographs by
OLIN SEWALL PETTINGILL, Jr.
Clarkson N. Potter, Inc./Publisher
NEW YORK
Digitized by Google

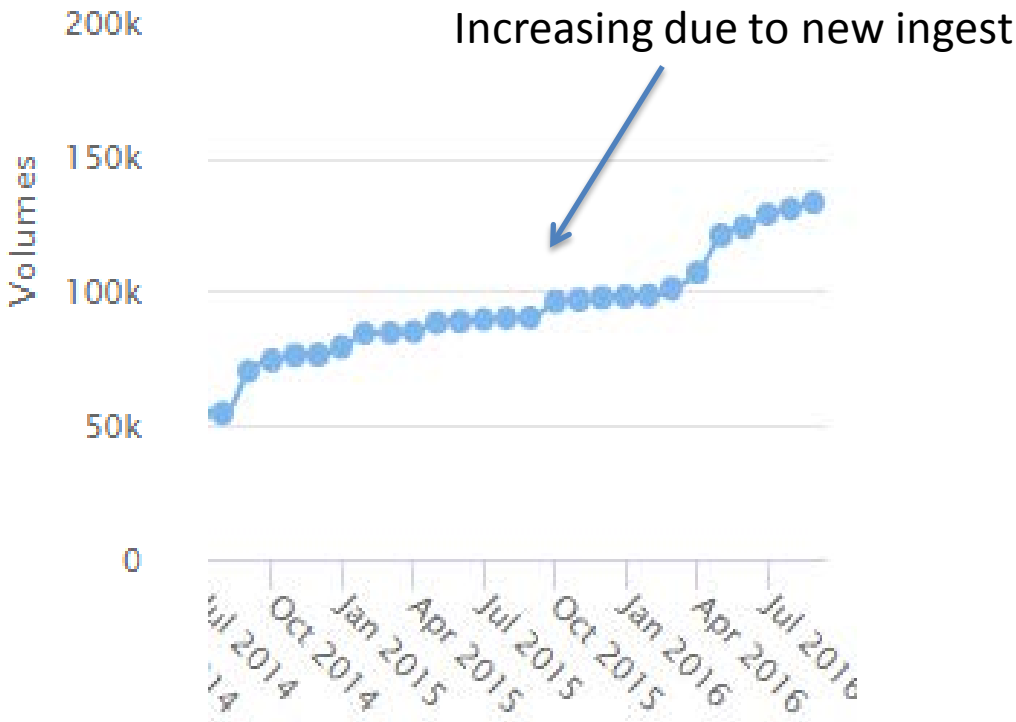
THE JEWISH
LABOR BUND
A PICTORIAL HISTORY
1897—1957
Compiled by J. S. HERTZ


SPECIAL PUBLICATION
SP - SW-0162
water
requirements
of crops
AMERICAN SOCIETY
OF
AGRICULTURAL ENGINEERS
Google
Original from
CORNELL UNIVERSITY

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



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

Why just 1923-1963 for now?

No U.S. copyright term if…

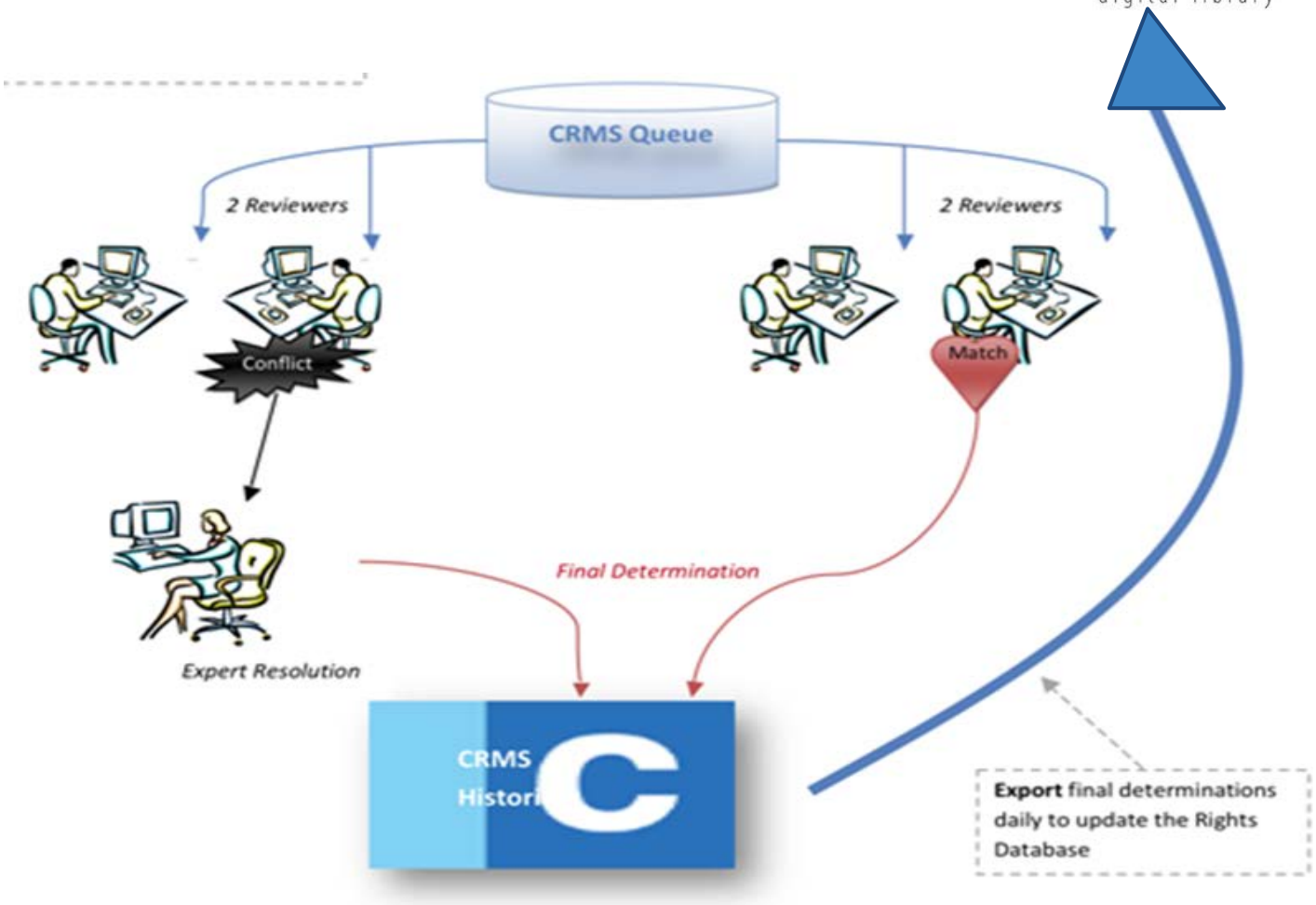
- 1923-1963 = not renewed or published w/o notice
- 1964-1977 = published w/o notice
- 1978-March 1, 1989 = published w/o notice and not subsequently registered in 5 years

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

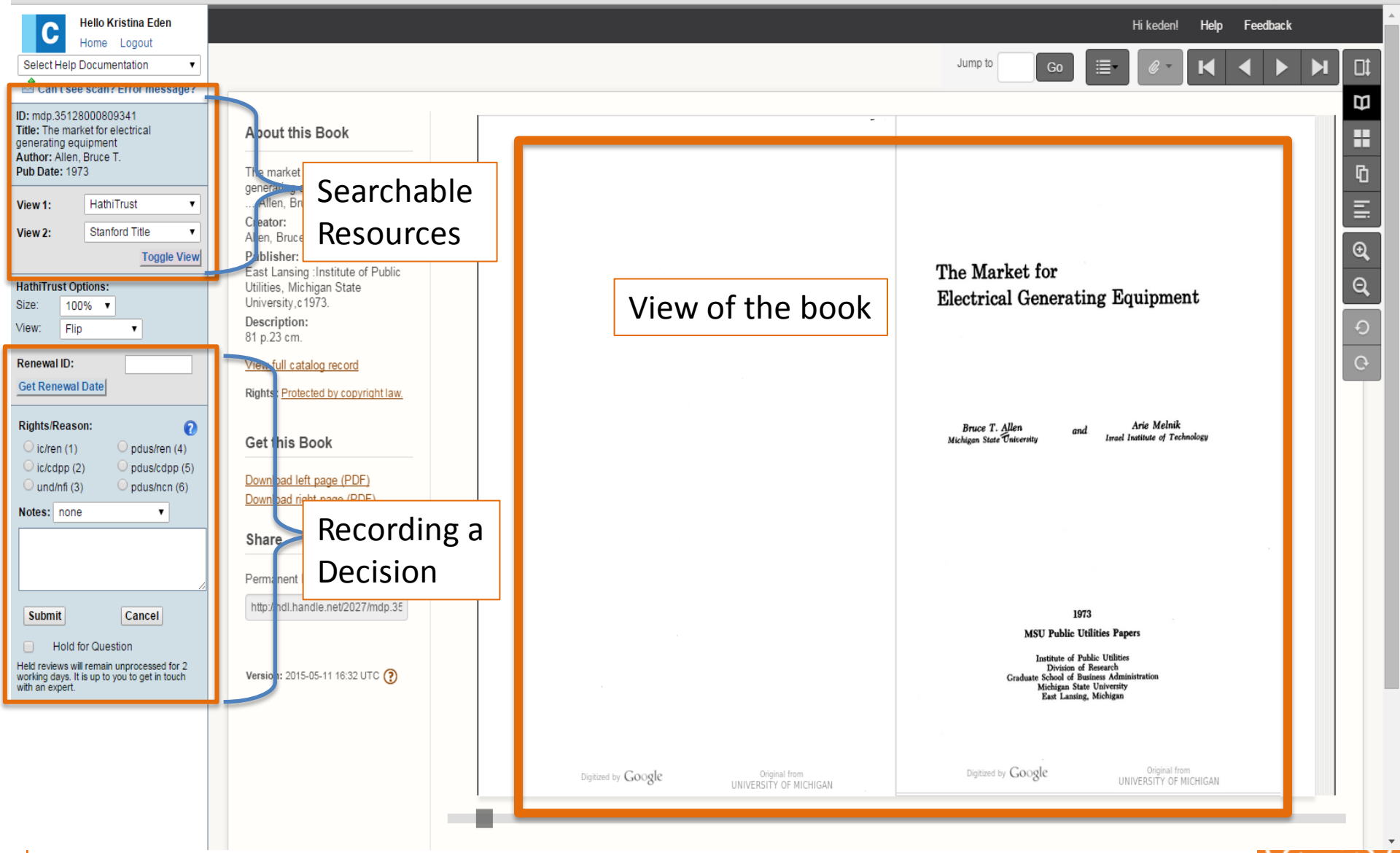


Process changes this time

- 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

	Project Total
Public domain	310631 (53.1%)
In copyright	101440 (17.4%)
Undetermined	172514 (29.5%)
Total	584585

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.


Hello Kristina Eden
Home Logout
Select Help Documentation
ID: mdp.35128000809341
Title: The market for electrical generating equipment
Author: Allen, Bruce T.
Pub Date: 1973
View 1: HathiTrust
View 2: Stanford Title
Toggle View
HathiTrust Options:
Size: 100%
View: Flip
Renewal ID:
Get Renewal Date
Rights/Reason:
ic/ren (1)
pdus/ren (4)
ic/cdpp (2)
pdus/cdpp (5)
und/nfi (3)
pdus/ncn (6)
Notes: none
Submit
Cancel
Hold for Question
Held reviews will remain unprocessed for 2 working days. It is up to you to get in touch with an expert.
Hi keden!
Help
Feedback
Jump to
Go
About this Book
Creator:
Publisher:
East Lansing :Institute of Public Utilities, Michigan State University,c1973.
Description:
81 p.23 cm.
View full catalog record
Rights: Protected by copyright law.
Get this Book
Download left page (PDF)
Share
Searchable Resources
Recording a Decision
View of the book
The Market for Electrical Generating Equipment
Bruce T. Allen
Michigan State University
and
Arie Melnik
Israel Institute of Technology
1973
MSU Public Utilities Papers
Institute of Public Utilities
Division of Research
Graduate School of Business Administration
Michigan State University
East Lansing, Michigan
Digitized by Google
Original from UNIVERSITY OF MICHIGAN

CRMS-US 1923-1963

How 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

My Review Stats:

My Review Stats Help

Reviewer name obscured for privacy

	Project Total	Total 2015	Jan 2015	Feb 2015	Mar 2015	Apr 2015	May 2015	Jun 2015	Jul 2015
PD Reviews	38577 (64.9%)	4445 (83.9%)	593 (76.9%)	730 (80.7%)	753 (79.4%)	503 (79.3%)	166 (90.7%)	331 (96.2%)	438 (79.9%)
IC Reviews	8505 (14.3%)	13 (0.2%)	0 (0.0%)	0 (0.0%)	2 (0.2%)	3 (0.5%)	0 (0.0%)	0 (0.0%)	4 (0.7%)
UND/NFI Reviews	12371 (20.8%)	840 (15.9%)	178 (23.1%)	175 (19.3%)	193 (20.4%)	128 (20.2%)	17 (9.3%)	13 (3.8%)	106 (19.3%)
Total Reviews*	59453	5298	771	905	948	634	183	344	548
Valid** Reviews & Rate	56415 (94.9%)	4959 (93.6%)	728 (94.4%)	856 (94.6%)	897 (94.6%)	591 (93.2%)	178 (97.3%)	341 (99.1%)	507 (92.5%)
Validation** Rate (all reviewers)	97.4%	97.4%	96.5%	97.6%	97.8%	97.2%	99.0%	99.7%	96.8%
Time Reviewing (mins)	66533	6703	992	976	1004	732	281	624	843
Time per Review (mins)	1.2	1.4	1.4	1.1	1.1	1.2	1.7	2.1	1.8
Reviews per Hour	51.8	43.6	42.7	52.3	55.1	49.4	35.7	28.3	33.2
Outlier Reviews	2029	426	65	54	26	31	16	50	81

Performance statistics can be sent to local supervisors upon request

Timeline

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

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