

HathiTrust Rights Status Assignment Process Updates and Recommendations

April 12, 2010

Executive Summary

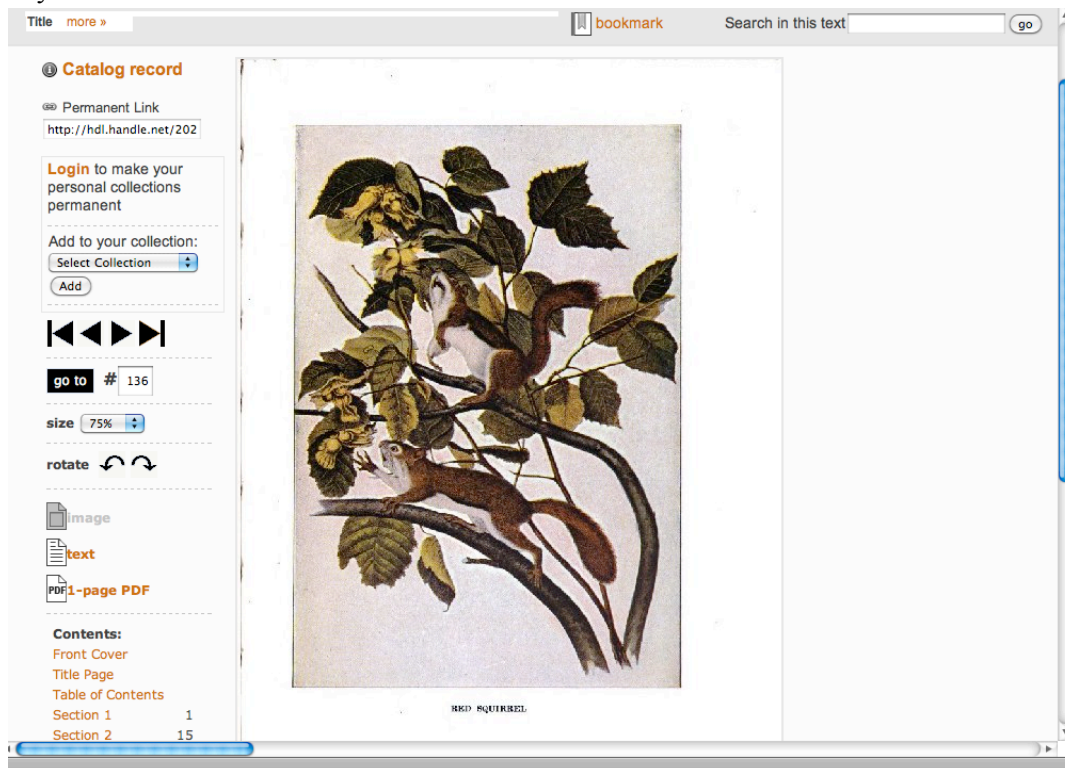
Copyright status determines whether or not a given user can see the page images of a given book in HathiTrust. Copyright determinations can sometimes be made automatically when books enter the system by Mirlyn, Michigan's catalog. This system generally works well. Rights can also be released by copyright holders via a legal release form. However, a large class of books – those published in the US between 1923 and 1963 – may or may not still be in copyright. Human review of these books is necessary, and there are many of them: UC alone has more than 200,000 such volumes in Hathi.

Michigan received an IMLS grant in 2008 to create a Copyright Review Management System, which addresses this problem with a double-blind manual review process. The system is now functional, although still under development, and has met with general success in improving the accuracy of copyright determinations. Three CIC universities are joining the CRMS project this summer, with hopes for further expansion.

Rights Status and the User

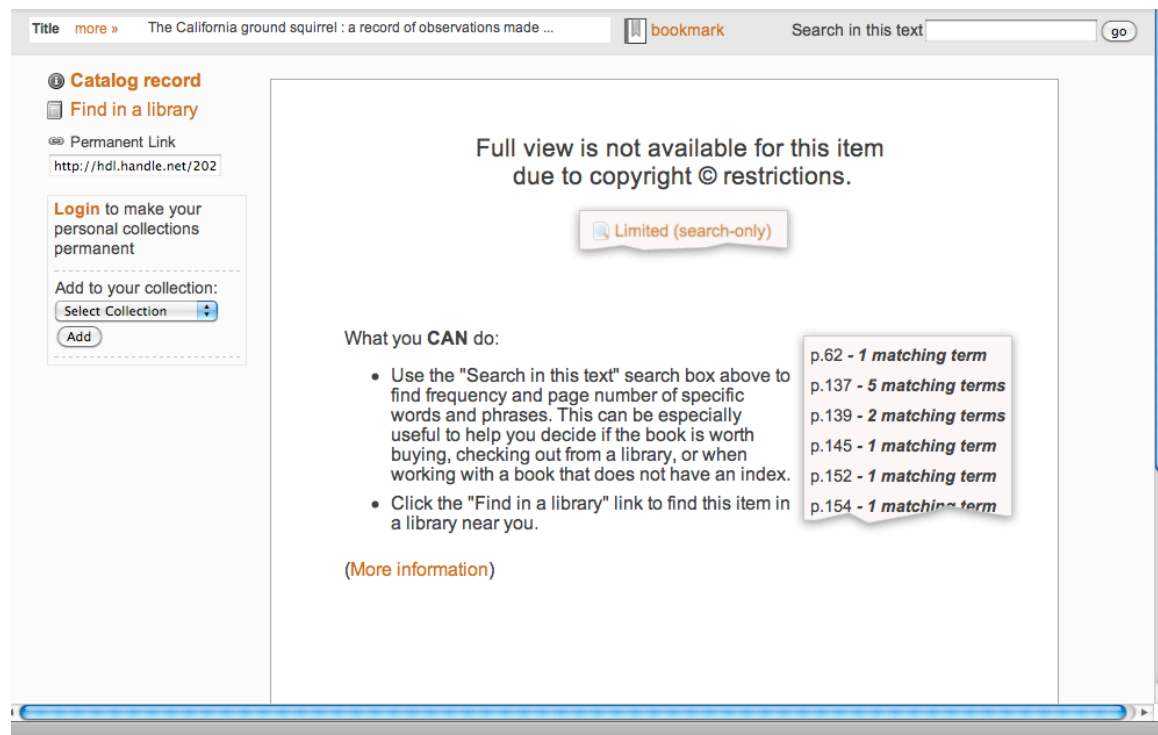
UC Libraries currently have over 1.1 million digitized books available in HathiTrust. Of these, only about 122,000 are currently available for full view because of copyright restrictions.

The difference from a user's perspective is stark. If a book is open for full view, the user sees the book in its entirety:



HathiTrust, Full View

The majority of books, however, are treated as Search-Only:



HathiTrust, Search-Only

When a user clicks through to a Search-Only book, he finds a “Find in a library” link to WorldCat, but no page images. This is still a great way to discover books, but can frustrate users hoping to read the book online.

The difference between Full View and Search-Only status makes assigning the correct rights status key. Books that are given the wrong status may be inappropriately made available or unnecessarily hidden from view. Crucially, many books printed between 1923 and 1963 may or may not be in copyright, and could potentially be put into Full View in HathiTrust. The University of Michigan has implemented an automatic process to classify books as they enter Hathi, and is working on a Copyright Review Management System to streamline the review of ambiguous books. A summary of both follows.

Hathi Rights Assignment Process

HathiTrust currently uses Mirlyn, the University of Michigan’s catalog, to process records from other institutions. After a book’s images and metadata are ingested into HathiTrust, the retrieval system sends the item’s identifier to Mirlyn to create or update the item record. When this process is complete, Mirlyn uses a simple algorithm to assign copyright status. First, the algorithm determines whether the work was published in the United States. Then, it looks for three sets of criteria:

- Works published outside the US before 1869
- Works published in the US before 1923
- Federal Government documents published anytime

If a book matches any of these three, Mirlyn flags the book with an ATTRIBUTE name of “pd” (public domain) or “pdus” (for a small set of non-US works).¹ They are given a REASON attribute of “bib”, because the decision was made automatically from bibliographic metadata. Books whose publication dates put them in copyright are marked as “ic/bib.” If Mirlyn cannot determine the item’s copyright status, it will mark it as “und/bib”. (See appendix for more information.)

After the algorithm is run, a book’s rights status may also be updated by manual review or by negotiated agreement. Hathi’s documentation provides a table of possible attribute and reason values:

ATTRIBUTES

id	name	type	dscr
1	pd	copyright	public domain
2	ic	copyright	in-copyright
3	opb	copyright	out-of-print and brittle (implies in-copyright)
4	orph	copyright	copyright-orphaned (implies in-copyright)
5	und	copyright	undetermined copyright status
6	umall	access	available to UM affiliates and walk-in patrons (all campuses)
7	world	access	available to everyone in the world
8	nobody	access	available to nobody; blocked for all users
9	pdus	copyright	public domain only when viewed in the US

Interestingly, the “out of print and brittle” classification is the only one that refers to an individual volume, rather than the larger work. This is an incongruity Michigan is aware of, but it allows the system to grant access to books for preservation reasons via Section 108 of the Copyright Act.

REASONS

id	name	dscr
1	bib	bibliographically-derived by automatic processes
2	ncn	no printed copyright notice
3	con	contractual agreement with copyright holder on file
4	ddd	due diligence documentation on file
5	man	manual access control override; see note for details
6	pvt	private personal information visible
7	ren	copyright renewal research was conducted
8	nfi	needs further investigation (copyright research partially complete; an ambiguous, unclear, or other time-consuming situation was encountered)

¹ Works published outside the US between 1869 and 1908 may be in copyright in some countries. They are marked “pdus” and available only to users in US-based IP ranges.

9	cdpp	title page or verso contain copyright date and/or place of publication information not in bib record
10	cip	condition review and in-print status research was conducted
11	unp	unpublished work

In this system, “access”-type attributes take precedence over “copyright”-type attributes. (That is, works can be made available for access via negotiated contracts or for preservation reasons despite being in copyright.) Manual reason attributes override bibliographically-derived decisions, giving staff the final say. These overrides take effect at the item level and do not currently apply to duplicates from other institutions.

A problem can arise with the process when bibliographic metadata is incorrect or incomplete. For example, in CDL’s recent ingest of Internet Archive books into Hathi, we found that a book dated 1912 was marked by Mirlyn as in copyright because a trailing bracket (i.e. “1912]”) was present in the date field. Small errors like this can keep books out of public view until a person notices the problem. Overall, though, the process seems to work well.

IMLS Project Status Update

The algorithm used to review ingested books in Mirlyn only addresses works published before 1923. However, many books published in the United States between 1923 and 1963 are actually in the public domain and could be made available after human review. As of April 2010, there are at least 700,000 books in HathiTrust that were published between these dates. More than 200,000 of these come from UC. A large number of these books are likely in the public domain, but the problem of opening access to them will take systematic work to address.

In 2008, before the launch of HathiTrust, Michigan received an IMLS grant to create a collaborative Copyright Review Management System (CRMS) to systematically address this problem. The IMLS project was later integrated with Hathi. Essentially, the CRMS has created a process and tools to distribute the work of copyright review for Hathi books across partner institutions. After completing Phase One (system development) of the CRMS project last May, the project became the sole method Michigan uses for copyright review. Five staff members contribute at Michigan, and other CIC institutions will be joining the project soon.

CRMS Workflow

First, records for books with US publication dates between 1923 and 1963 from all HathiTrust institutions are ingested from Mirlyn into the CRMS database. Then, they are accessed by a reviewer in the CRMS interface:

Version 1.3.4

Copyright review management system

Hello Anne Karle-Zenith [Logout](#)

[home](#) > [admin](#) > [view all historical reviews](#)

Historical Reviews:

Currently searching by Review [Search Help](#)

Search for volumes:

Restrict selection:

Identifier: AND:

Start Date:

End Date:

Order: (Identifier)

Direction: Ascending

Display: 20 records per page.

Found 26564 reviews for 16219 volumes, page 1 of 1329
[previous 20](#) | [next 20](#) | [first page](#) | [last page](#) | [download](#) | [Use Cases](#)

ID	Title	Author	Pub Date	Review Date	Status	Legacy	User	Attr	Reason	Note Category	Note	Priority	Verdict
isu.30000000981112	Comin', round the mountain	Albert, Ned, 1896-	1938	2009-11-06	6	0	dfuimer	und	nfi	Misc	play	0	✓
isu.30000000981112	Comin', round the mountain	Albert, Ned, 1896-	1938	2009-11-06	6	0	dmow	und	nfi	Misc	play	0	✓
isu.30000000981112	Comin', round the mountain	Albert, Ned, 1896-	1938	2009-11-09	6	0	grichols123	und	nfi	Misc	Play	0	✓
isu.30000001246903	The far-off hills	Robinson, Lennox, 1886-1958.	1958	2009-11-13	6	0	sgueva	und	nfi	Misc	Play	0	✓
isu.30000001246903	The far-off hills	Robinson, Lennox, 1886-1958.	1958	2009-11-13	6	0	grichols	und	nfi	Misc	play	0	✓
isu.30000001246903	The far-off hills	Robinson, Lennox, 1886-1958.	1958	2009-11-16	6	0	grichols123	und	nfi	Misc	Play	0	✓
isu.30000001693708	Czech anthology with biographical sketches	Pospisil, Božena.	1956	2009-11-17	5	0	cwlcox	und	nfi	Collection	of translated Czech stories	0	✓
isu.30000001693708	Czech anthology with biographical sketches	Pospisil, Božena.	1956	2009-11-17	5	0	sgueva	pd	ncn			0	✗
isu.30000001693708	Czech anthology with biographical sketches	Pospisil, Božena.	1956	2009-11-18	5	0	jaheim123	und	nfi	Translation	Also a collection.	0	✓
isu.30000001707399	Complete cookbook	Vanderbilt, Amy.	1961	2009-10-29	4	0	dfuimer	ic	ren			0	✓
isu.30000001707399	Complete cookbook	Vanderbilt, Amy.	1961	2009-10-29	4	0	sgueva	ic	ren			0	✓
isu.30000001725922	Progressive Czech (Bohemian)	Mikulš, Bohumil E.(Bohumil Emil)/Czechoslovak National Council of America.	1948	2009-11-17	4	0	sgueva	pd	ren			0	✓
isu.30000001725922	Progressive Czech (Bohemian)	Mikulš, Bohumil E.(Bohumil Emil)/Czechoslovak National Council of America.	1948	2009-11-17	4	0	dfuimer	pd	ren			0	✓

The reviewer clicks through this interface to an authenticated version of the Hathi pageturner with access to all page images. He or she is checking the copyright page of the book to see whether it has a copyright notice. If not, the book is declared public domain/no copyright notice (“pd/ncn”). If the book does have a copyright notice, the reviewer checks the Stanford Copyright Renewal Database (<http://collections.stanford.edu/copyrightrenewals/>) to see whether the holder renewed copyright after the initial 28-year term. If so, the book is marked in copyright/renewal (“ic/ren”). If not, it is marked public domain/renewal (“pd/ren”).

After one reviewer makes a determination for a book, it is placed at the top of the queue for another reviewer’s analysis. If both reviewers agree, the book’s copyright status is updated in Mirllyn and HathiTrust. If they disagree, an expert reviewer (generally the Project Manager) reviews the book and makes a final determination.

Figure 2: CRMS decision interface

Findings/Accomplishments of CRMS

The CRMS has become Michigan’s sole process for copyright review. So far, over 16,000 books have been reviewed with the CRMS. Of these, about 56% have been assigned public domain status. Decisions are likely much more trustworthy than they were before. These books are now available for full view via HathiTrust.

Also, the determinations for all reviewed books are added to the tab-delimited metadata files available from HathiTrust, including the reason code and the date of

User: Anne Karle-Zenith
[Home](#) [Use Cases](#) [Logout](#)

ID: mdp.39015057097803
 Title: Solubility
 Author: Hildebrand, Joel Henry
 Pub Date: 1924

Search: [Hathi Mirllyn](#) [Mirllyn C](#)
 Stanford: [Title](#) [Author](#) [Advanced](#)

Display Options:
 Toggle: [Full/Spit](#) [Views](#)
 Review Pane: [Left](#) [Right](#)
 Size: 100%

Renewal ID:
[Get Renewal](#)
[Date:](#)

Rights/Reason:
 pd/ncn (1) ic/ren (4)
 pd/ren (2) ic/cdpp (5)
 pd/cdpp (3) und/nfi (6)
 pdus/cdpp (7)

Notes: none

[Help](#) [Use Cases](#) [Logout](#)

² Interface images from “Copyright Review and the Copyright Management System”, presentation by Anne Karle-Zenith. <http://www.lib.umich.edu/files/grants/crms/Yale-presentation-final.pdf>. Used under Creative Commons license (attribution-noncommercial 3.0 US).

last update. Determinations are also shared through an OAI feed of MARC21 and Dublin Core records and harvested by OCLC.

Problems Found So Far

As expected, accuracy comes at a price. The double-blind process of the CRMS means that Michigan must work more slowly on copyright assignments. Staff were able to make about 3,700 determinations per month prior to implementing the CRMS, while they are now producing only 1,500 (~ 40%.) Also, additional scrutiny has led to a doubling in the rate of undetermined statuses (from 8% pre-CRMS to 17% now). Inaccuracies and incomplete data in the Stanford renewal database also lead to occasional undetermined statuses.

The main constraint on throughput is expert review time – conflicts require the Project Manager's attention, so the pace of the reviewers must be throttled to match her availability. Training additional experts is key, and Michigan plans to work on this problem in the coming months. Reviewers also need extensive training, which would be made easier by having a larger pool of experts.

Next Steps/Ongoing

Work on the CRMS project continues. Michigan will be bringing staff CIC partners Indiana University, University of Wisconsin, and University of Minnesota in Q2-Q3 of 2010. Version 2.0 of the CRMS software is slated for April 2010, and should include the following improvements:

- Administrative and reviewer UI views
- Database redesign
- Clearer instructions for reviewers
- Mechanism for reviewer feedback
- Alternative to using the Hathi pageturner to view images

Soon, Michigan plans to hire the U.S. Copyright office to check its determinations. The Project Manager traveled to Washington, D.C. in November 2009 to begin this process. Michigan also plans to begin reconciling records in Hathi where multiple copies of a given volume have different copyright statuses. (Usually, this problem occurs when Michigan's copy has been opened by the CRMS process and another institution's has not.) Hopefully, this will make available many books from UC and other Hathi partners.

Rightsholder Agreements

Michigan's process for opening access to copyrighted works by agreement is simple. Typically, a copyright holder finds his or her work online and contacts Michigan to make it accessible. The rightsholder fills out a one page form, available on the Hathi website. This agreement permits the University to make digital copies for its library collection and for HathiTrust. It also allows Hathi to make these copies available to the public "without restrictions." An optional clause also permits reprints. Michigan has not made a concerted effort to attract rightsholders to the program.

The only verification Michigan performs of the rights themselves is to have the signer of the form warrant that they hold copyright to the work and that it does not infringe on the rights of others. This form is mailed to the Copyright Review Project Librarian at Hatcher Library. If she approves it, the form is filed in a cabinet and the work's record is updated with a "world/con" attribute and reason. Rarely, Anne will forward unusual agreements to the Lead Copyright Officer, AUL for IT, or the Head of Collections.

Interestingly, Google also has a process for authors and publishers to make their works available in full view, but it requires the rightsholder to sign up as a Google partner. This process is called “library uplift.” In the process, the book is removed from the control of the scanning library and associated with the account of the rightsholder.

Avenues for Further investment

There are a number of ways we can engage with these processes to increase the number of public domain volumes available within the HathiTrust. One way is to create local processes for rightsholders to release their works into Full View in HathiTrust. Following the example of Michigan, a legal form permitting the override of default rights status in Hathi could be created in consultation with local counsel and relevant academic units, and utilized locally. This idea could be potentially taken a step further by publicizing the benefits of this to the local scholarly community.

We could also contribute to the CRMS by exploring the errors we’ve found in bibliographic metadata (e.g. the “1912]” problem) with Michigan. If there are consistent errors we can identify and eliminate, we may be able to open a large number of books without a massive effort.

Finally, HathiTrust partners can work together to identify and eliminate rights discrepancies across duplicate books in Hathi. In some cases, a Michigan volume has been opened to Full View, while the corresponding exact UC duplicate is still Search-Only. For example, the 1960 biography of Hamlin Garland below has the Michigan copy in Full View and the UC copies in Search-Only:

HATHI TRUST Digital Library About Help Feedback

Search Public Collections My Collections

Catalog Search

Catalog Search All Fields Full view only Find Advanced Catalog Search Search Tips

← Previous record Showing record 7 of 1081 Next record → Email this Cite this Export to Endnote

Similar Items

- Edward Everett Hale : a biography.**
By: Holloway, Jean.
Published: (1950)
- Hamlin Garland : a life /**
By: Newlin, Keith.
Published: (2008)
- Hamlin Garland /**
By: McCullough, Joseph B.
Published: (1978)
- Hamlin Garland Memorial /**
By: South Dakota Federal Writers Project.
Published: (1939)
- Henry Blake Fuller and Hamlin Garland : a reference guide /**
By: Silet, Charles L. P.
Published: (1978)
- Selected letters of Hamlin Garland /**
By: Garland, Hamlin, 1860-1940.
Published: (1998)
- The beginnings of naturalism in American fiction; a study of the works of Hamlin Garland, Stephen Crane, and Frank Norris, with special reference to some European influences, 1891-1903.**
By: Ahnebrink, Lars.

Hamlin Garland; a biography.

Main Author: Holloway, Jean.
Language(s): English
Published: Austin, University of Texas Press [1960]
Subjects: Garland, Hamlin, > 1860-1940.
Physical Description: xii, 346 p. illus., ports. 24 cm.
Original Format: Book
Biography
Original Classification Number: 828 G233O H74
Locate a Print Version: Find in a library

Viewability:

- Full view (original from University of Michigan)
- Limited (search-only) (original from University of California)
- Limited (search-only) (original from University of California)

Resolving these discrepancies would bring the work of the CRMS to bear on a wider set of books and make it easier to share copyright determinations with across institutions. It would be a good place to get started.