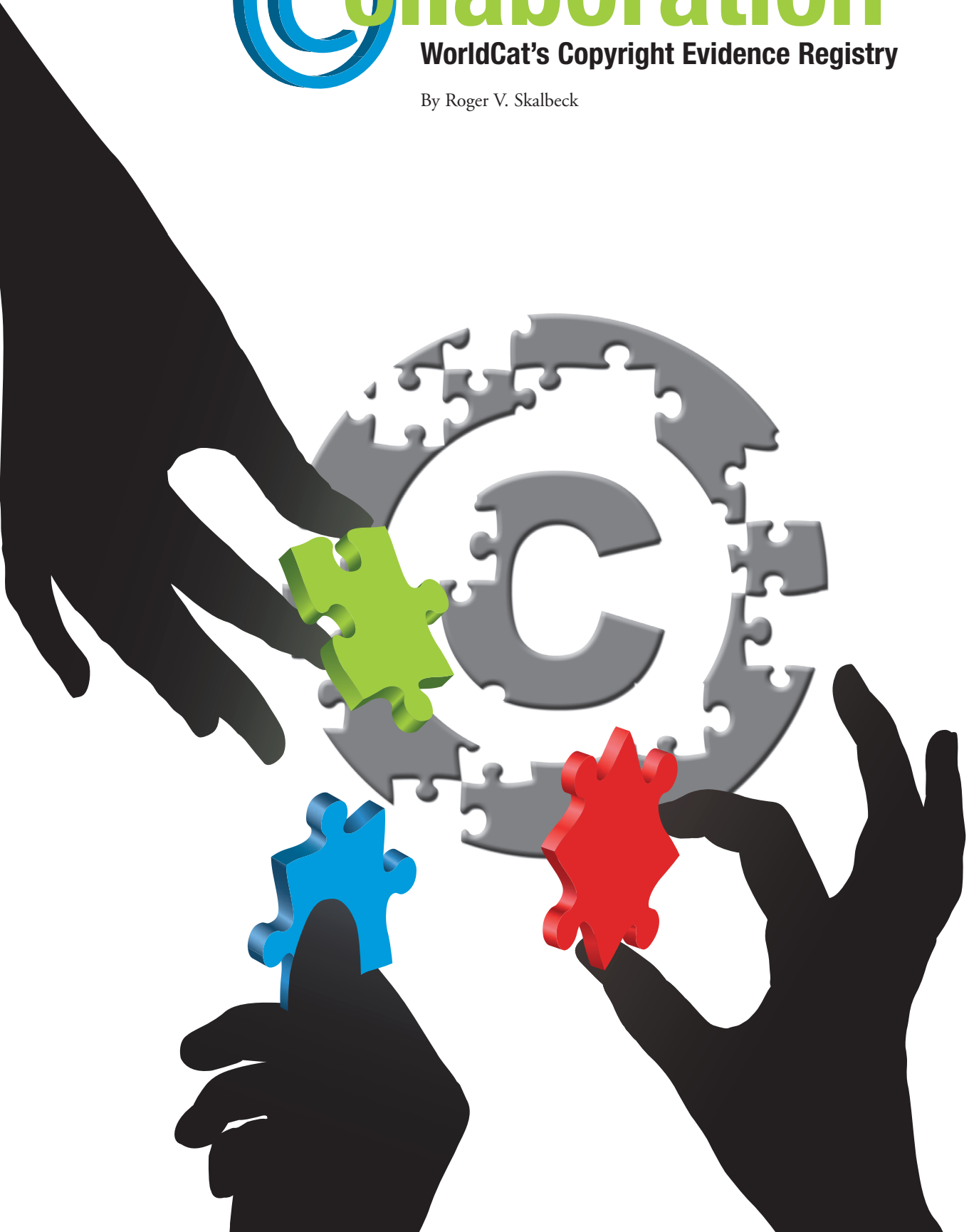


The Power of Collaboration

WorldCat's Copyright Evidence Registry

By Roger V. Skalbeck



“Books are for use.” Most librarians will recognize this as Ranganathan’s first law of library science. Unfortunately for those who want to digitize books, the time-consuming search for current copyright status information may make the project too burdensome to undertake or too fraught with legal risks. In effect, some books are less available for use than others.

One who wants to copy a work in print or digital form faces many hurdles to ensure unrestricted use of the work: choose works in the public domain, obtain permission for free use, or pay copyright holders for reproduction rights. Of course, there’s also the option to rely on a fair use argument, as Google essentially does with most materials included in the Google Book Search project. Those who do not want to take the risks that Google does, however, need a better way to find out what may be legally reproduced.

Mass digitization has been going on for a while, so some of the “easier” projects of interest to law libraries have already been addressed. Gale’s *The Making of Modern Law* focuses on treatises from 1800-1926, while the various LexisNexis Congressional projects include works of the U.S. government. These impressive collections are by no means easy from a technical perspective; however, neither publisher had difficult copyright status questions to face before digitizing these sets. Works published before 1923 are in the public domain, and works of the federal government cannot be copyrighted.

For anybody who wants to digitize new content, what else is available? As it turns out, there are many books that *might be* in the public domain, but determining the copyright status of these materials can be difficult. Items published between 1923 and 1963 may be found to be in the public domain if a specific set of criteria are examined, including such factors such as publication location, copyright renewal, and adherence to certain formalities.

OCLC estimates that there are more than 1.9 million records in WorldCat for books published in the United States between 1923 and 1963. A Copyright Office study from 1961 estimated that the copyrights had been renewed for only

7 percent of the books they reviewed. This represents more than one million books that could be digitized without concern for copyright status. Tempering this number slightly, the 1996 restoration of copyright in foreign works appears to have prevented a large number of items from entering the public domain.

WorldCat Copyright Evidence Registry

One of the underlying obstacles to reproducing older books is that there’s no central place to look for information about what is protected by copyright and what may have passed into the public domain. Responding to this need, OCLC recently introduced a new system for tracking various copyright details for published books. The new service, still in beta, is called the WorldCat Copyright Evidence Registry (CER). It could be a very valuable resource for recording and sharing copyright status information.

Briefly described, the WorldCat CER is a community-driven database for people to record and share information about the copyright status of books. By making this tool available to the world, OCLC seeks to use the power of mass collaboration to solve the daunting task of tracking copyright information for millions of books.

People using the WorldCat CER system can contribute data to annotate publisher status, indicate copyright registration details, or document a copyright search for a specific work. For instance, you might record that another company acquired a publisher’s works or that copyright was renewed for a book. Annotations are listed under the name of the contributor, which can be an institution’s OCLC credentials or a self-selected user name for individuals.

Positive Attributes of the Copyright Evidence Registry

The records in the WorldCat CER are very accurate because they come directly from the WorldCat database of more than 100 million bibliographic records that have been edited, revised, and verified for many years. When you search for a record on the WorldCat CER, data is pulled directly from WorldCat, so there are no issues of data synchronization. Also, by using the WorldCat CER, contributors can focus on recording copyright information without worrying about bibliographic information.

Another aspect of the CER already partially implemented is the inclusion of records from the Stanford Copyright Renewal database. Stanford’s project includes digitized renewal details for books, scanned from records of the U.S. Copyright Office at the Library of Congress. Although renewal is not currently required for copyright protection, it was necessary for maximum copyright protection for some works, in particular those published between 1923 and 1964. Under certain conditions, works automatically passed into the public domain if people did not renew copyright registrations for works published during this period. OCLC is still refining the algorithms for matching Stanford’s data with WorldCat records. As the record matching increases, so does the value of WorldCat CER records.

Criticisms of the Copyright Evidence Registry

The WorldCat CER is not a perfect solution. First and foremost, the CER is not a legal registry, and this is not something that OCLC seeks to change. Because it is not an official government

(continued on page 32)

CER Entries

To illustrate the type of information recorded in the system, below are entries from three works in the WorldCat CER where copyright details have been added by participating users.

- The world that was, by Bowman, John G. (John Gabbert), 1877-1962. (1926, OCLC # 1812021)
Entry #1: The work has a copyright statement Copyright 1926 by the Macmillan Company —Copyright statement
Citation: Examination of book
Entry #2: Copyright renewal has not been found
Citation: Not found in Stanford Copyright Renewal Database
- Mathews family record; descendants of John and Sarah Mathews of County Tyrone, Northern Ireland, by Bowman, James Ray, 1924- (1953, OCLC# 4258559)
Entry: Copyright renewal has not been found
Citation: 9/5/08 email from author to Maija Cravens, WHS regarding a title sent to Google as part of the UW-Madison library project with Google and his attempt to get GBS to display in full view, following quote in 9/4/08 email to Google which is in the email to Maija: “My genealogy book, Mathews Family Record: Descendants of John and Sarah Mathews of County Tyrone, Northern Ireland, in the collections of the Wisconsin Historical Society Library, Madison, WI, was digitized and appears in Google Book Search. The book has a copyright 1953 date. I did not renew the copyright, which has expired, and this message is to request that you open the book to the public, without limitations in any way.”
- How children learn to read. by Mackintosh, Helen K. (Helen Katherine), 1897- (1952, OCLC# 1710565):
Entry: The work does not have a copyright statement
Citation: n/a

registry, the WorldCat CER can't provide legally binding proof of copyright status. In fact, you can't currently use the WorldCat CER to search for out-of-copyright works.

Another possible shortcoming of the WorldCat CER project is that it may not be “sexy” enough to encourage widespread participation. The hope is that records will grow through the power of collective input, such as with Wikipedia, where enthusiasts eagerly enhance records. In the book world, it also works for projects like LibraryThing, where community value is added not only by enhanced classification, but also through book relationship references and shared networks for users to discover readers with similar interests.

For people printing records from the WorldCat CER, there are data limitations. Although one can search by ISBN, the numbers don't display in the record output. Also, there is no unified way to print all CER record elements for a single work because they appear in separate screens on the Web site.

Molly Kleinman, a librarian from Michigan, raises two additional concerns about the WorldCat CER in a post to her blog (www.mollykleinman.com):

“OCLC claims and enforces copyrights in its bibliographic records. While it grants member libraries permission to make broad use of those records, my understanding is that the same is not true for non-members. If OCLC extends that policy to the Copyright Evidence Registry, it risks becoming just another walled garden that is useful only to a select (and

paying) group of members, and less useful even to that group than it would be if it were truly open.

“Right now the registry is sparsely populated. It will take a critical mass of records and contributors to become a trustworthy source of copyright evidence. Where will that critical mass come from? What is OCLC doing to build it quickly? How will users know when the registry has reached it?”

Automated Copyright Analysis

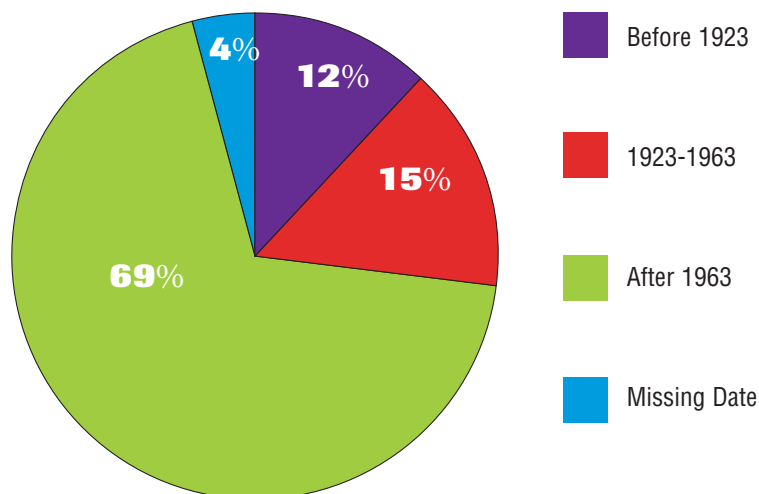
Because this is a community-sourced project without verification of entries, users will need to devise a way to evaluate contributed information. One new feature of the WorldCat CER suggests that it will soon help to inform risk assessment decisions and lead users to better copyright registration details.

This enhancement will allow subscribers to run copyright rules analysis for batches of works. Someone interested in digitizing a book could run a batch process to check for particular key words or relevant information in annotations, such as the country or year of publication. OCLC does not dictate how the analysis will run; instead the information seeker must determine the level of analysis and documentation needed to evaluate the risk of using particular materials.

Future Plans for the WorldCat CER

One feature not included in the registry is the ability to upload scanned evidence. If participants could upload scanned copies of copyright evidence, the authenticity of reported data would be

WorldCat Records Describing Books Published in the United States



enhanced. OCLC's Bill Carney says they are considering options for users to upload scanned information and linking to existing digitized works. For now, these features are not in the current system. Some contributions may have a greater degree of authenticity, but without scanned images to back up assertions, this project seems less useful for efficient copyright determinations. That said, some evidence contributions are easy to locate, such as copyright registration numbers or citation to pages in printed reference works.

Other Uses for the WorldCat CER

In conclusion, the WorldCat CER is a great starting point for finding or recording notes on the copyright status of works. Even without extensive annotations, searching the WorldCat CER for copyright details is a good way for people to help document search efforts before making use of a work that may require permission.

In addition to tracking books that may no longer be protected by copyright, WorldCat CER could also be valuable for ascertaining the status of orphan works, i.e., works whose copyright owner cannot be found. In a report on orphan works, the U.S. Copyright office suggested that legislation might be necessary to limit liability for those who use copyrighted works after performing a diligent search for a copyright owner who cannot be found. There is a fair amount of debate as to what constitutes a reasonably diligent search, but searching WorldCat CER should help to strengthen a claim of due diligence. To date, the United States has passed no orphan works legislation.

Not the Only Game in Town

OCLC is not the only organization working to build an expansive database suitable for tracking the copyright status of works. A similar project is OpenLibrary, a collaboration of the Internet Archive and the Boston Public Library, which aims to create "one Web page for every book." The service is freely accessible to anybody on the Internet.

The aim of OpenLibrary is to create bibliographic records for books, incorporating direct links for users to buy, borrow, and browse them. They link to sites like Amazon, WorldCat, and Google Books, as well as book trading sites like BookMooch and Title Trader. The content is presented in a wiki format, so anybody can alter existing entries.

In addition to bibliographic data, OpenLibrary includes scanned versions of books in the public domain that users can download and search in full-text. This gives OpenLibrary a distinct advantage over WorldCat CER. Not only can you find out about works that may be out of copyright, you can also read and search the scanned works. Because the OpenLibrary ecosystem has a broader range of features, it is likely to appeal to a wider audience, although librarians may question the value of records that anyone can edit.

OpenLibrary plans to include copyright status information in their records. According to sources familiar with the project, this will come in the form of a computational algorithm to determine what is or is not in the public domain. This approach sounds similar to the copyright rules engine developed for WorldCat CER, with one important distinction. As described, the OpenLibrary approach appears to be based on a mathematical analysis with uniform rules integrated into the system. In contrast, WorldCat CER rules are meant to be user-defined and fully customizable.

With this feature, OpenLibrary potentially would have the advantage of being fully integrated as a free feature. However, users will probably want flexibility in applying any automated analysis because strict mathematical analysis isn't possible.

Conclusion

Projects like WorldCat CER are wonderful resources for collectively recording and sharing the copyright status of books. If publishers and libraries contribute large amounts of data to the system, it will become an invaluable resource. In addition, WorldCat CER may also become a resource for finding copyright owners. Anybody starting a digitization project should consider the CER as a place to share information discovered when investigating the copyright status of books. Without a collective and shared resource such as this, digitization will remain too risky for many to undertake without fears of liability for unauthorized reproduction. ■

Roger V. Skalbeck (rvs5@law.georgetown.edu) is associate law librarian for electronic resources and services at Georgetown Law Library in Washington, D.C., and a member of the AALL Copyright Committee.

Links and Additional Resources

Press Release: OCLC pilots WorldCat Copyright Evidence Registry

www.oclc.org/us/en/news/releases/200832.htm

WorldCat Copyright Evidence Registry

www.worldcat.org/copyrightevidence

Copyright Term and the Public Domain in the United States

www.copyright.cornell.edu/public_domain

Stanford University Copyright Renewal Database

<http://collections.stanford.edu/copyrightrenewals>

U.S. Copyright Office Report on Orphan Works

www.copyright.gov/orphan/orphan-report.pdf

LibraryThing

www.librarything.com

Open Library

www.openlibrary.org

Google Book Search Bibliography

www.digital-scholarship.org/gbsb/gbsb.htm

OCLC's new Copyright Evidence Registry

mollykleinman.com/2008/08/26/oclc-new-copyright-evidence-registry

Hirtle, Peter B. "Copyright Renewal, Copyright Restoration, and the Difficulty of Determining Copyright Status." *D-Lib Magazine* Volume 14 Number 7/8 (July/August 2008) 2 Oct 2008 www.dlib.org/dlib/july08/hirtle/07hirtle.html

Barbara Ringer, "Study No. 31: Renewal of Copyright" (1960), reprinted in Library of Congress Copyright Office. *Copyright law revision: Studies prepared for the Subcommittee on Patents, Trademarks, and Copyrights of the Committee on the Judiciary, United States Senate, Eighty-sixth Congress, first [-second] session.* (Washington: U.S. Govt. Print. Off, 1961), p. 220.