

Comments of the Triangle Research Libraries Network**in Response to the U.S. Copyright Office Notice of Inquiry Regarding a Mass Digitization Pilot Program****Library of Congress, U.S. Copyright Office****Docket No. 2015-3, 80 Fed. Reg. 32614 (June 9, 2015)**

Every year, the Triangle Research Libraries Network (TRLN) libraries¹ (Duke University, North Carolina Central University, North Carolina State University, and The University of North Carolina at Chapel Hill) collectively spend hundreds of thousands of dollars for staff and support to help us address copyright limitations on our ability to meet the needs of our students, faculty, staff, and the public. One of our focuses has been to develop strategies for addressing copyright in large digitization projects,² and we have spent considerable time and effort trying to understand how best to clear rights.³

While we appreciate the Copyright Office's interest in this issue,⁴ we believe that the Office's mass digitization licensing proposal would be ineffective and inefficient. Collectively, TRLN libraries have more than 20 million print and electronic volumes in our collections (over 11 million unique items), and we hold several times that number in special collections that include manuscripts, photographs, and other unique works. We have digitized countless in-copyright works, either with permission from rightsholders or under the doctrine of fair use or other legal defenses, coupled with a takedown mechanism, should that prove necessary.

From among our vast holdings and digitization projects, we are unable to identify a single instance in which we would benefit by pursuing a license under an extended collective licensing (ECL) system like the one proposed by the Office.

¹ Triangle Research Libraries Network (TRLN) is a collaborative organization of Duke University, North Carolina Central University, North Carolina State University, and The University of North Carolina at Chapel Hill, the purpose of which is to marshal the financial, human, and information resources of their research libraries through cooperative efforts in order to create a rich and unparalleled knowledge environment that furthers the universities' teaching, research, and service missions. See <http://www.trln.org/>.

² See *The Triangle Research Libraries Network's Intellectual Property Rights Strategy for Digitization of Modern Manuscript Collections and Archival Record Groups* (January 2011), <http://www.trln.org/IPRights.pdf>.

³ See Maggie Dickson, *Due Diligence, Futile Effort: Copyright and the Digitization of the Thomas E. Watson Papers*, 73 *Am. Archivist* 626 (2010).

⁴ Our comments are responsive to the proposal referenced in the Notice of Inquiry (NOI) and as outlined in its report *Orphan Works and Mass Digitization* (June 2015), <http://copyright.gov/orphan/reports/orphan-works2015.pdf> [hereafter *OWMD Report*]. Responses to specific questions posed in the NOI are identified throughout these comments.

Below, we outline three reasons why we think the proposed system would be ineffective and inefficient, and therefore why we would be unlikely to participate: 1) Developing a workable marketplace for ECL would be prohibitively costly; 2) There are no suitable sellers (collective management organizations or CMOs); and 3) There would be few, if any, buyers from within the library community. The cost of addressing those issues would, we believe, exceed the value of creating such a system. If the Office is interested in pursuing solutions to facilitate mass digitization, we urge it to focus on more fundamental reforms, such as creating a simple limitation or exception for nonprofit mass digitization, or more practical improvement to facilitate voluntary licensing by developing new registries and making old copyright office records fully available online.

I. Developing a workable marketplace for ECL would be prohibitively costly

We are skeptical that a workable marketplace can develop for mass digitization ECL in the United States for two reasons: 1) Resolving ownership of electronic rights in published works is an expensive and complex process; 2) An ECL system still has to grapple with the orphan works problem, which, as libraries like ours have demonstrated, would be prohibitively costly if engaging in a diligent search on a work-by-work basis.

a. Resolving ownership of electronic rights in published works is an expensive and complex process

Under current law, copyrights are infinitely divisible and can have many possible owners. Transferees and licensees have no duty to inform the public about the scope of their rights, or even the fact that they hold them. As a result, private book contracts and other transfer documents allocate rights in a wide variety of ways with no clear way for prospective licensees to know, for example, whether they need to purchase a license from an authors' rights organization, a publisher rights organization, or both. The division of electronic versus print rights has been particularly contentious, and case law resolving these disputes is murky given the focus on specific contractual language.⁵

Our experience with clearing rights make us wary of stepping into a system that tries to tackle this issue en masse. The cost of doing so is likely to be high, for three reasons. For one, factual evidence about ownership can be difficult to develop and expensive to litigate.⁶ The Google Books Settlement (GBS) resolved this rights allocation issue by establishing a dispute resolution mechanism and an initial allocation of royalties as between authors and publishers.⁷ The Office's proposal contemplates no such system.⁸ Presumably, CMOs and individual putative owners who wished to opt out would be left to

⁵ See, e.g., *Random House, Inc. v. Rosetta Books, LLC*, 150 F. Supp. 2d 613 (S.D.N.Y. 2001).

⁶ The Happy Birthday ownership dispute is one recent example of how messy and expensive ownership disputes can become. *Marya v. Warner/Chappell Music Inc.*, Case No. CV-13-4460, Memorandum and Order (C.D. Cal. Sept. 22, 2015).

⁷ *Authors Guild v. Google*, Case No. 05-cv-8136, Amended Settlement Agreement, Attachment A, Sec. 6.3, 6.4, 7.1-7.2 (S.D.N.Y. Nov. 13, 2009), http://www.thepublicindex.org/wp-content/uploads/sites/19/docs/amended_settlement/SettlementAttachments/Attachment-A-Author-Publisher-Procedures.pdf.

⁸ The Office's proposal contemplates the creation of a dispute resolution system to resolve disagreements as between CMOs and prospective licensees about rates, but it says nothing about a system to resolve disputes as between competing CMOs. *OWMD Report* at 96-97.

voluntarily negotiate with each other,⁹ and if necessary, litigate disputes in federal courts. Such an arrangement would likely prove expensive. Even if an alternative resolution system were created, costs for resolving factual disputes about ownership would likely be high.

Second, CMOs would have no incentive to control the costs of these disputes. Costs would be passed along to licensees. If the Office's proposal is followed, CMOs might even have incentive to prolong these disputes, given that CMOs would be entitled to deduct fees associated with such disputes from payments to rightsholders.¹⁰

Third, even assuming that rights ownership is clarified, libraries and archives will likely still need to negotiate licenses with multiple CMOs for rights that are, on agreement, split between multiple owner groups, raising transaction costs associated with licensing.

The bottom line is that resolution of ownership of electronic rights is a fundamental issue that is not resolvable within the bounds of this limited proposal. Determining ownership is not something that publishers, authors, or libraries have resolved yet, and there is no reason to think that CMOs will be able to do it at any lesser expense.

b. The orphan works problem still exists and ECL will not solve it.

For the few collections that do fit the scope of the Office's proposal,¹¹ such as the core published textual works in the HathiTrust collection (a project to which TRLN libraries have contributed but are not in a position to speak for as a whole), there is likely to be a significant number of orphan works.¹² A mass digitization license for those types of collections would only be valuable if it also covered orphan works, something the Office seems to recognize.¹³

But the ECL proposal does not resolve the orphan works problem; it merely shifts the burden of searching for owners from the digitizing library to a CMO. In doing so, it moves the responsibility for

⁹ In Norway, for example, Kopinor acts as a sort of CMO of CMOs, with 22 constituent membership organizations representing both publisher and author groups. As discussed below, we are skeptical that even a basic CMO structure can be developed in the US, much less a robust superstructure capable of facilitating agreements between CMOs.

¹⁰ *OWMD Report* at 99.

¹¹ As discussed below, we do not believe there are significant numbers of digitization-priority library collections that would fit within the scope of the Office's proposal because they include unpublished and non-textual, non-photograph works.

¹² The Office asserts without citation or support that "[i]n many cases mass digitization involves corpuses containing mostly published works, for which there is a significant likelihood of owners being found." *Id.* at 49-50. In our experience, published works are frequently orphaned as well, and what few studies have been conducted on this question align with our experience. For example, Cornell University Library attempted to identify owners for a sample of 343 in-copyright monographs from its collection. It was able to locate rightsholders for only 198 works (58%) of that sample. *Response by Cornell University Library to the Notice of Inquiry Concerning Orphan Works*, Comment OW0569, 1-2 (March 23, 2005), <http://www.copyright.gov/orphan/comments/OW0569-Thomas.pdf>. Carnegie Mellon did a similar study, attempting to identify rightsholders for a sample of 368 books in its collection. From that sample, 22% of publishers could not be found. *Response by the Carnegie Mellon University Library to the Notice of Inquiry Concerning Orphan Works*, Comment OW0537, 2 (March 22, 2005), <http://www.copyright.gov/orphan/comments/OW0537-CarnegieMellon.pdf>. More recent evidence from the UK further supports the idea that published works are often orphaned. See UK Intell. Prop. Office, *Impact Assessment (Final), Orphan Works* (June 2012) at 10-11, <http://www.ipo.gov.uk/consult-ia-bis1063-20120702.pdf>.

¹³ *OWMD Report* at 49-50.

finding owners to an entity that has little incentive to actually locate and pay rightsholders. For these works, CMOs will be able to deduct fees associated with both managing the rights and with conducting their diligent searches,¹⁴ and the longer the CMO can hold on to unclaimed funds and engage in drawn out (unsuccessful) searches, the higher those fees would be. Unlike libraries and archives that face negative consequences in the form of potential copyright liability for not thoroughly searching for owners when such searches are necessary, CMOs would face no negative repercussions for conducting a flawed search.

One way to exacerbate those bad incentives would be to give CMOs a structural advantage over potential users in exploiting orphan works. Unfortunately, it appears (though it is not clear) that the Office suggests doing that very thing. Unlike the complex and administratively burdensome search scheme proposed for users of orphan works,¹⁵ the Office has very little to say about what CMOs should do to find owners. It states that CMOs should engage in diligent searches, and suggests that this should include, at a minimum, that CMO's maintain a public list of works for which rightsholders could not be identified or located.¹⁶

While we have strong reservations about the administrative requirements of the proposed Orphan Works legislation,¹⁷ a "diligent search" should be the same for both CMOs and users of orphan works. Anything less would give CMO's an unfair competitive advantage in using orphan works, driving users into licenses for a category of works that the Office has recognized repeatedly are generally inappropriate for licensing.¹⁸

Finally, leaving aside the conflicts of interest that CMOs will face, there is no reason to think that CMOs will be able to engage in more efficient diligent searches than the end users who actually hold copies of the works and have information about acquisition and provenance of those copies. Those clues about when and where a work was created or sold are often critical facts needed to locate rightsholders.

II. There is little prospect of developing adequately representative Collective Management Organizations (CMOs)

In addition to the structural problems associated with creating an ECL system in the United States, the Office proposal has two practical obstacles that we believe are insurmountable, the first of which is that there is little prospect of developing adequately representative Collective Management Organizations (CMOs).

¹⁴ *Id.* at 99-100. The Office suggests that, after a period of time, unclaimed funds be distributed to an "educational or literacy based charity selected by its membership." *Id.* A better alternative would be to distribute funds back to those who paid the license fees.

¹⁵ *OWMD Report* at 56-63.

¹⁶ *Id.* at 100.

¹⁷ *Id.* at Appendix A.

¹⁸ *Id.* at 49. If anything, given the conflicts of interest noted above, we believe that CMOs should be held to a higher standard, and should be required to publicly post not only a list of works, but the supporting documentation for their search process.

As compared to other nations with a longer history of collective management of copyright for the types of works identified by the Copyright Office, the U.S. has few CMOs with experience that would meet this need. Developing the infrastructure to represent the vast numbers of U.S. rightsholders would be expensive. We have no way of knowing the true number of U.S. authors of published literary works, but we know there are many. The Bureau of Labor Statistics estimates that there are currently 43,500 “writers and authors,”¹⁹ 48,210 “technical writers,”²⁰ and 1.2 million “postsecondary teachers,”²¹ many of whom produce scholarly works of the type held by research libraries like us. For publishers, the Census Bureau estimates that as of 2012 there were 25,412 business enterprises operating just in the publishing industries (excluding the Internet). The turnover of publishers is also astounding; in just one year the Census estimates that 2,237 publishing industry establishments ceased to exist and that 2,440 were created.²² The rights of those large numbers of defunct publishers would need to be represented and licensed as well.

These statistics represent only a subset of the authors and publishers that would need to be represented by a CMO under the ECL scheme proposed by the Office. While the Office’s proposal says little about what factors would affect a finding that a given CMO adequately represents rightsholders, we note that the Nordic models that the Office seeks to emulate require that CMOs represent a “substantial number” of rightsholders in the given field.²³ Given their current size and orientation, the two organizations that the Report specifically identifies as potential CMOs--the Authors Registry and the Copyright Clearance Center (CCC)--are unlikely to be able to rapidly expand to adequately represent a substantial number of rightsholders in the field of literary works.

The Authors Registry itself is not a CMO and represents no authors. The Authors Guild, a related nonprofit corporation, asserts that it has only 8,500 members, a tiny subset of the millions of published U.S. authors. As shown in the Authors Guild’s prior efforts to represent U.S. authors as a class in *Authors Guild v. Google*, there are reasons to doubt their ability to represent a broader set of author interests.²⁴ As for the CCC, managing rights to “tens of millions of works”²⁵ is still only a fraction of published literary works contemplated to be licensed under the Office’s proposal. Building out these existing CMOs beyond their current, small, size would likely be costly.²⁶

¹⁹ <http://www.bls.gov/oes/current/oes273043.htm>.

²⁰ <http://www.bls.gov/oes/current/oes273042.htm>.

²¹ <http://www.bls.gov/oes/current/oes250000.htm>.

²² http://www2.census.gov/econ/susb/data/dynamic/1112/us_4digitnaics_emplchange_2011-2012.xls

²³ Johann Aximm & Lucie Guibault, Cross-Border Extended Collective Licensing: A Solution to Online Dissemination of Europe’s Cultural Heritage, Final Report for EuropeanaConnect, Institute for Information Law Research Paper 2012-19 at 30, Aug. 2011, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2001347.

²⁴ See *Authors Guild v. Google*, 721 F. 3d 132, 134 (2d Cir. 2013) (noting that objections to Authors Guild’s representativeness “may carry some force”).

²⁵ CCC Initial Comments at 10.

²⁶ Although we believe both organizations to be of modest size, neither organization has disclosed detailed information about its size. Both CCC and Authors Registry hold themselves out as nonprofit corporations, but neither is listed by the Internal Revenue Service in its list of tax-exempt non-profits, <http://www.irs.gov/Charities-&-Non-Profits/Exempt-Organizations-Select-Check>, and neither has disclosed financial details through, for example, a Form 990 disclosure.

Further, as the Office recognizes, transparency and regulatory oversight of CMOs are critical parts of an effective ECL regime. Jonathan Band and Brandon Butler have produced a helpful summary of CMO transparency and oversight issues worldwide, concluding that CMOs have “a history of corruption, mismanagement, confiscation of funds, and lack of transparency that has deprived artists of the revenues they earned.”²⁷ A large, unclaimed pot of money invites mismanagement and administrative sprawl, and the CMO model incentivizes aggressive enforcement against users that far outstrips the efforts made to locate and pay creators.

We have already experienced this type of behavior in the U.S. The Copyright Clearance Center, a voluntary license broker, spent millions of dollars - generated by the estimated 15-30% cut taken from each transaction²⁸ - to fund litigation against a nonprofit university.²⁹ A CMO with the same incentives and no clear set of rightsholders to be accounted for is unlikely to be more frugal or efficient. CMOs in Brazil, Ghana, Italy, Spain, and Sweden have been riddled with embezzlement and corruption, diverting money intended for creators into the coffers of the organization and the pockets of executives. Scholars such as Ariel Katz³⁰ and Michael Geist³¹ have written persuasively about these structural incentives for corruption and mismanagement as they have played out in Canada.

Our concerns about the costs of creating both the CMOs themselves and the necessary oversight mechanisms are only confirmed when we look at the cost of ECL models abroad. The Norwegian model in particular, which is the closest existing system to what the Office proposes, is extraordinarily expensive for what it provides. A recent article reporting on the Norwegian experience points out that the Norwegian National Library has been able to use ECL to license 250,000 Norwegian-language books for full access to all 5 million Norwegian residents.³² The cost is calculated on a per-page basis, and with an average book length of approximately 170 pages, the licensing cost is roughly NOK 60 (\$7.20) annually per book.³³ Scaling that same system up for full nationwide access in the U.S., the fees would be daunting. Norway’s population of 5 million is about 63 times smaller than the U.S. population of 318

²⁷ Jonathan Band and Brandon Butler, *Some Cautionary Tales about Collective Licensing*, 21 Mich. St. Intl. L. Rev. 689, 690 (2013).

²⁸ See Copyright Clearance Center web site, “What does it cost a rightsholder to sign an agreement.” Available at: <http://support.copyright.com/index.php?action=article&id=222&relid=14> (last visited October 2, 2015).

²⁹ Cambridge University Press v. Becker, 863 F. Supp. 2d 1190, 1213 (N.D. Ga. 2012), rev’d on other grounds, 769 F. 3d. 1232 (11th Cir. 2014).

³⁰ “Copyright Collectives: Good Solution, But For Which Problem?”, in Working Within the Boundaries of Intellectual Property (R. Dreyfuss, H. First & D. Zimmerman, eds.) (Oxford: Oxford University Press, 2010); “The Orphans, The Market, and the Copyright Dogma: A Modest Solution to a Grand Problem” (2012) 27(3) Berkeley Technology Law Journal, 1285-1346.

³¹ Michael Geist, The Economics Behind Access Copyright, May 26, 2011, <http://www.michaelgeist.ca/content/view/5819/125/>; Michael Geist, The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law, 335-403 (2013).

³² Roger Josevold, *A National Library for the twenty-first century -- knowledge and cultural heritage online*, at 5, Alexandria: The Journal Of National & Int’l Library & Info. Issues, Aug. 7, 2015, http://www.ingentaconnect.com/content/manup/alex/pre-prints/content-MUP_ALX_0037.

³³ Even without scaling up to U.S. size, the Norwegian model is expensive. Compare the annual license fee of \$7.20 to the average e-book purchase price (not an annual, recurring, license fee as in the Norwegian model) for 2013 in the U.S., at \$27.97; or the average mass-market paperback purchase price of \$7.04, or the average paperback (not mass market) price of \$43.81.

million. Adjusting for that scale difference, the per-book cost would come to something like \$458 per book (about \$114.5 million total) for nationwide access to a library of 250,000 books. Scaling up that collection to the size of something like the HathiTrust (approximately 8.4 million in-copyright volumes) would cost nearly \$4 billion per year.³⁴ As a point of comparison, that is almost 4 times the total amount spent annually by public libraries (\$1.2 billion) on all resources, and it dwarfs the fraction spent annually on electronic materials (\$200.4 million).³⁵

III. There would be few, if any, buyers (libraries)

TRLN libraries are unlikely to purchase an ECL license offered by a CMO. These organizations would not provide an equitable, practical solution to rights issues connected with the material TRLN libraries deem most important to digitize. In addition, much of the material that we prioritize for digitization is beyond the scope of the pilot project that the Copyright Office proposes. Our focus for digitization is on material that is not currently discoverable and on material that is unique, rare, or ephemeral. In short, our priorities for digitization are special collections material and, often, non-textual material. While we cannot speak for all libraries, we believe our digitization experiences and priorities are similar to those of many other U.S. research libraries.

As we have mentioned in previous correspondence, one example of these priorities is Content, Context, Capacity,³⁶ a collaborative project of TRLN members to digitize collections that chronicle the Long Civil Rights Movement. For that project we used an intellectual property assessment strategy with four components: seeking permission when practical, assessing risk, asserting fair use, and inviting the public to provide background information on the material, including rights information.³⁷

At the University of North Carolina at Chapel Hill, for example, none of our recent digitization projects have focused on conventionally published materials. Some more recent examples of the kind of material we focus on for digitization include the following:

- The Frank Porter Graham papers³⁸ are a mixture of public domain material, unpublished, and published material from both known and unknown rights holders, including a great deal of personal correspondence. About 25,000 pages from this collection have been digitized.
- We have digitized historic North Carolina postcards³⁹ with contributions from regional collections all over North Carolina. In most cases, neither we nor our partner libraries and historical societies have significant information about the postcards' publication, distribution, or the date of creation. In a number of cases, the businesses that published them have ceased operation, and there are no known successors who hold rights.

³⁴ Thanks to Peter Hirtle for his initial analysis of the costs of the Norwegian system:

<http://blog.librarylaw.com/librarylaw/2014/03/norway-extended-collective-licensing-and-orphan-works.html>

³⁵ http://www.ims.gov/assets/1/AssetManager/FY2012%20PLS_Tables_21_thru_31A.pdf (table 26).

³⁶ <http://www2.trln.org/cc/>

³⁷ See *The Triangle Research Libraries Network's Intellectual Property Rights Strategy for Digitization of Modern Manuscript Collections and Archival Record Groups* (January 2011), <http://www.trln.org/IPRights.pdf>.

³⁸ http://www2.lib.unc.edu/mss/inv/g/Graham,Frank_Porter.html

³⁹ http://www2.lib.unc.edu/dc/nc_post/index.php

- Recently we have digitized an unusual collection of North Carolina films that chronicle many aspects of life in the state during the twentieth century.⁴⁰ They include oral histories, films about local arts events, and North Carolina Film Board works about race relations and other social issues in the South during the 1960s. Many of these films were unpublished, and in many other cases we have not been able to verify the publication status completely. For some films, we were able to get permission to digitize from the rights holder, but in many other cases, neither we nor our liaisons in local communities have been able to identify who made the films.
- The Southern Folklife Collection has embarked on a large digitization project of rare sound recordings and motion pictures, “Expanding the Reach of Southern Audiovisual Resources.”⁴¹ The focus of this effort is to preserve and make accessible material such as field recordings and films of traditional music and musicians. Aside from the complicated rights issues with pre-1972 sound recordings, many rightsholders are unknown or can no longer be located. Even when record companies and other distributors can be identified, the chain of ownership is often broken or unclear.

These projects illustrate the areas where we are most willing to put our time and money for digitization. These are materials that, for the most part, no one else holds and that are of great interest to researchers. In many cases, the authors of these works did not contemplate commercializing their creations. In many other cases, the creators are unknown. We cannot compensate rightsholders whose names are lost and whom we cannot find. But we can honor them and the work they did by preserving it, attributing it to the greatest extent possible, and making it available to the public. When it is foreseeable that they cannot be found, it does these rights holders no service to send payments to a collective management organization for a fruitless search. Consequently, we would not take advantage of the pilot program the Copyright Office proposes, because it does not fit with our needs and priorities.

The only part of our collections that we can see would reasonably fit within the scope of the Office’s proposal is our core, print collection. Much of that core collection is already replicated in digital form and hosted by the HathiTrust, a digital library of which we are members, and we would be delighted with a mechanism that would allow us to open up fully digital access to that corpus at a reasonable cost. However, the conflict of interests and the likely expense, coupled with competing collections priorities, lead us to believe that the Copyright Office’s extended collective licensing pilot project proposal is not a feasible means to that end. Instead, we suggest that the Office focus on other solutions, such as finding a way to re-introduce formalities in some situations, digitizing and providing access to past Copyright Office records, and considering a more straightforward exception to the problems of orphan works and mass digitization, such as a notice and takedown system that libraries are already utilizing.

⁴⁰ <http://www.digitalnc.org/blog/welcome-to-the-movies>

⁴¹ <http://blogs.lib.unc.edu/news/index.php/2015/06/sfc-mellon-grant/>

Thank you again for your interest in this issue and for soliciting comments. If requested, we would be pleased to answer any questions you might have. You can contact TRLN Executive Director Lisa Croucher at lisa@trln.org.

Sincerely,

A handwritten signature in black ink that reads "William M. Cross". The signature is written in a cursive style with a long horizontal flourish at the end.

William M. Cross, J.D., M.S.L.S.
Director, Copyright & Digital Scholarship Center
NCSU Libraries
North Carolina State University

A handwritten signature in black ink that reads "Anne Gilliland". The signature is written in a cursive style with a long horizontal flourish at the end.

Anne Gilliland JD, MLS
Scholarly Communications Officer
University Libraries
Associate Law Librarian
The University of North Carolina at Chapel Hill

A handwritten signature in blue ink that reads "David R. Hansen". The signature is written in a cursive style with a long horizontal flourish at the end.

David R. Hansen
Clinical Assistant Professor & Faculty Research Librarian
UNC School of Law

A handwritten signature in black ink that reads "Kevin L. Smith". The signature is written in a cursive style with a long horizontal flourish at the end.

Kevin L. Smith
Director, Copyright & Scholarly Communication
Duke University Libraries