RESPONSE TO THE U.S. COPYRIGHT OFFICE'S NOTICE OF INQUIRY ON A MASS DIGITIZATION PILOT PROGRAM

The Library Copyright Alliance (LCA) consists of three major library associations — the American Library Association, the Association of Research Libraries, and the Association of College and Research Libraries. These three associations collectively represent over 350,000 information professionals and more than 100,000 libraries of all kinds throughout the United States. An estimated 200 million Americans use these libraries over two billion times each year. These libraries spend more than $4 billion annually acquiring books and other information resources. LCA welcomes the opportunity to respond to the U.S. Copyright Office's Notice of Inquiry on a Mass Digitization Pilot Program.

From the June 9, 2015 notice of inquiry, it is clear that the Copyright Office has already decided to proceed with a pilot program for mass digitization, including the drafting of legislation that would establish an extended collective licensing (ECL) framework to enable the mass digitization. In its notice of inquiry, the Copyright Office requested comment on specific topics "regarding the practical operation of such a system," including examples of "projects that might be appropriate for licensing under the Office's proposed ECL framework;" the form of a dispute resolution process between the collective management organization (CMO) and a prospective licensee; the

appropriate timeframe for the distribution of royalties; and actions the CMO should take to diligently search for rights holders for whom royalties may have been collected.

We urge the Copyright Office to reconsider its decision to proceed with this pilot program because the program is both impractical and reflects inappropriate policy choices.

I. The Impracticability of the Pilot Program

The program is impractical in several respects. First, during the Copyright Office roundtables, there was little support for an ECL approach for the mass digitization of books. The three entities with large databases of digitized books—Google, the HathiTrust Digital Library, and the Internet Archive—have not, to our knowledge, indicated that they would be interested in participating in an ECL system. And even if they, or other entities, had some theoretical interest, it is hard to imagine how the pilot program could get off the ground.

The Google Books Settlement (GBS) provides the model for the Copyright Office's pilot program with respect to books.[1] The CMO would play the role of the settlement's Book Rights Registry (BRR), and the CMO's licensees would provide to libraries what the settlement referred to as "institutional subscriptions." An institutional subscription would have allowed all the users within an institution (*e.g.*, the faculty and students of a university) to view the full text of books in the Institutional Subscription

[1] Orphan Works and Mass Digitization Report at 85 (*hereinafter* Mass Digitization Report)("the proposed class action settlement in the Google Books litigation provides a template for an ECL system in [the] context" of literary works); 89 ("a potential model may be found in the portions of the Google Books settlement pertaining to users of Institutional Subscriptions"). For a more detailed discussion of GBS as a model for ECL, *see generally* Jonathan Band, *The Book Rights Registry in the Google Book Settlement*, 34 Colum. J.L. & Arts 671 (2011).

Database.[2] Under the Copyright Office's proposal, legislation would replace a class action settlement as the mechanism under which the CMO could authorize the licensees to provide the institutional subscriptions.

The Copyright Office, however, overlooks several features of GBS that distinguish it significantly from the Copyright Office's pilot program and call into question the economic viability of the Copyright Office's approach.

1) Under GBS, Google would have paid $34.5 million to cover the BRR's start-up costs.[3]

2) Under GBS, Google would have paid at least $45 million for distribution to rights holders whose works were scanned prior to 2009.[4]

3) The institutional subscription was just one of three primary services Google would have been able to provide under GBS. The other primary services were "Previews," under which Google could display up to 20% of a book in response to user queries;[5] and "Consumer Purchase," under which Google would sell consumers perpetual online access to the full text of a book.[6]

4) Google had the right to subsidize the purchase of institutional subscriptions by libraries that had partnered with Google in the Google Books project (*i.e*., had lent

---

[2] As a general matter, this included all the in-copyright but not commercially available books in the Google Books database.
[3] *See* Amended Settlement Agreement, *Author's Guild, Inc. v. Google Inc.*, No. 05 CV 8136-DC at § 5.2 (S.D.N.Y. Nov. 13, 2009)(*hereinafter* ASA).
[4] *Id*. at § 5.1(b).
[5] *Id*. at § 4.3(a).
[6] *Id*. at § 4.2.

Google books to scan).[7] For example, Google agreed to provide the University of Michigan with a free institutional subscription for up to 60,000 students.[8]

5) Google also agreed to provide free public access terminals to each public library and not-for-profit higher education institution, from which users could access the full text of the books in the Institutional Subscription Database.[9]

Thus, the settlement contemplated that Google would provide public libraries and higher education institutions with a free institutional subscription from a terminal on the library premises; and that it would provide its partner libraries with potentially deep discounts on institutional subscriptions that would allow all their faculty and students remote, simultaneous full text access. In other words, Google would subsidize the institutional subscription market. Google likely hoped to recoup these subsidies with profits from the Preview service, where Google would display advertising across from the responses to the search queries; and the Consumer Purchase of individual books.

Further, Google agreed to pay $34.5 million for the Registry's start-up costs, as well as at least $45 million in license fees to be distributed to rights holders. Accordingly, even if Google Books failed to generate significant revenue, Google was required to pay almost $80 million to the Registry and the rights holders.

By contrast, the institutional subscribers would bear the entire cost of the Copyright Office's proposal. There would be no Google to subsidize libraries' purchase of institutional subscriptions or to pay the CMO's start-up costs. There also would be no alternative sources of revenue such as Preview advertising or Consumer Purchase. To be

---

[7] Jonathan Band, *The Long and Winding Road to the Google Books Settlement*, 9 J. Marshall Rev. of Intell. Prop. L. 227, 272 (2009).
[8] *Id*.
[9] ASA at § 4(a)(i).

sure, some European countries have established ECL-based mass digitization programs for books in their national languages, but these programs are on a much smaller scale than would be required in the United States for English language books.[10] Moreover, these ECL-based projects have required significant government expenditure. It simply is inconceivable that the federal or state governments would come up with the appropriations necessary to support an ECL-based mass digitization program for books.[11]

Further compounding the cost of supporting an ECL system is the expense of participating in the legislative process that would be necessary to establish an ECL. The parties spent three years negotiating the GBS, and the resulting agreement was over 200 pages long, including the appendices and attachments. Google agreed to pay the plaintiffs' attorneys $45 million for their efforts.[12] Google likely spent millions of dollars on the fees of its lawyers. Although the Copyright Office would assemble the first draft of ECL legislation, using GBS as a template, many entities--including publishers, libraries, technology companies, and authors' groups--would feel compelled to participate in the legislative process because of its potential precedential value, even if they did not anticipate that the ECL ultimately would succeed. The discussions would be

[10] If an ECL system similar to Norway's were implemented in the United States for books published in the United States, the annual license fees would exceed $6 billion. Peter Hirtle, *Norway, Extended Collective Licensing, and Orphan Works*, LibraryLaw Blog (Mar. 21, 2014), http://blog.librarylaw.com/librarylaw/2014/03/norway-extended-collective-licensing-and-orphan-works.html. A system like Germany's might be less expensive, but it applies only to books published before 1966 and thus is of limit utility. So far, no entity has applied to serve as a CMO under the UK's ECL legislation, calling into question the viability if its approach.

[11] Without additional government funding, every dollar a library spends on the ECL to provide access to older works would be a dollar diverted from the acquisition of newer works.

[12] ASA at § 5.5.

contentious and protracted, and easily could last more than five years—the duration the Copyright Office recommends for the ECL regime before it sunsets.

Finally, the Copyright Office does not appear to appreciate the enormous complexity that would be involved in the distribution of royalties. It states that the "CMO should be required to conduct diligent searches for non-member rightsholder for whom it has collected royalties,"[13] with no recognition of how large an undertaking this would be. Over 12 million books have been published in the United States since 1923, and an estimated 11 million of them are still in copyright. Groups such as the Authors Guild and the CCC represent a small fraction of the rights holders of these millions of works. While some unrepresented rights holders may voluntarily come forward and register with the CMO, it can be assumed that many would not because they were unaware of the existence of the CMO or the fact that they were rights holders (*e.g.*, the heirs of a long-deceased author of a scholarly monograph); or because they felt the small amount of potential royalties did not warrant the effort of registration. Thus, the CMO likely would be required to search for millions of rights holders. The cost of a truly diligent search for so many rights holders would far exceed any license fees the CMO would collect.[14]

Moreover, distribution of royalties among rights holders who do register with the CMO would be no easy task. This is because it would often be unclear whether the

[13] Mass Digitization Report at 99.

[14] In the orphan works section of the Mass Digitization Report, the Copyright Office rejects an ECL for orphan works because it "would end up ultimately as a system to collect fees, but with no one to distribute them to, potentially undermining the value of the whole enterprise." *Id.* at 50. The Report suggests that in corpuses containing published works, such as books, "there is a significant likelihood of owners being found, and thus a justification for ECL representation." *Id*. However, the Copyright Office has absolutely no basis for asserting that there is "a significant likelihood" that many of the copyright owners of a large collection of books would be found.

publisher or the author would own the digitization rights for a particular book.[15] Recognizing this complexity, the Google Books Settlement included default rules for the division of royalties between the publisher and the author.[16] These default rules and the treatments of orphan works were among the settlement's most controversial provisions.

II. The Pilot Program's Inappropriate Policy Choices

The Mass Digitization Report acknowledges concerns raised by commenters that ECL "would weaken the fair use doctrine by inducing users whose activities might be protected by fair use to pay license fees rather than risk litigation."[17] The Office responds that it addresses this concern by focusing its ECL on activity "for which there is broad agreement that no colorable fair use claim exists."[18] However, it provides "Depression-era photographs"[19] as an example of a particular collection that prospective users may wish to digitize and make available through ECL. We believe that a library's digitization of a special collection of Depression-era photographs very likely would be a fair use and is precisely the sort of activity that should not be subject to a license fee.[20] The Office also suggests that ECL legislation could authorize the licensing of "narrow uses of in-

---

[15] The difficulty of determining the ownership of copyright was recently demonstrated by the decision that Warner-Chappell did not own the copyright in the lyrics to "Happy Birthday To You." *Marya v. Warner/Chappell Music, Inc*., CV 13-4460 GHK (C. D. Cal., Sept. 22, 2015).

[16] *See* ASA at Attach. A ("Procedures Governing Author Sub-Class and Publisher Sub-Class Under the Settlement Agreement").

[17] Mass Digitization Report at 101.

[18] *Id*.

[19] *Id*. at 104.

[20] Research libraries and archives currently are digitizing and making publicly available material in their special collections, including photographs. For example, the New York Public Library has digitized and provided online access to its special collection of materials relating to the New York World's Fair of 1939 and 1940, including 12,000 promotional photographs. *See New York World's Fair 1939 and 1940 Incorporated Records*, N.Y. Pub. Lib. Archives & Manuscripts, *available at* http://archives.nypl.org/mss/2233 (last visited Oct. 6, 2015).

commerce works, such as full text search and the display of short text excerpts in response to user queries."[21] But courts in numerous cases have found that mass digitization to enable full text search is a fair use.[22] And countless cases have treated quotations as fair use. Indeed, permitting quotations is mandatory under Article 10(1) of the Berne Convention.

To be sure, the applicability of fair use to the display of the full text of digitized books may be less certain than to the display of Depression-era photographs or snippets in response to search queries.[23] Nonetheless, the Copyright Office makes the wrong policy choice by advocating an ECL approach for the making available of digitized books for nonprofit educational or research purposes.

Under the Copyright Office proposal, full-text access to literary works would be limited to commercially unavailable works.[24] As a practical matter, this means books for which there no longer is a market and from which the rights holders no longer derive any

---

[21] Mass Digitization Report at 86. The Copyright Office also implies that licensing could be necessary for the text and data mining of works. *Id*. at n. 355.

[22] *See Authors Guild, Inc. v. HathiTrust,* 755 F.3d 87 (2d Cir. 2014); *Authors Guild, Inc. v. Google Inc.,* 954 F. Supp. 282 (S.D.N.Y. 2013); *White v. West Publ'g Corp*., 1:12-cv-01340-JSR (S.D.N.Y. July 3, 2014); *A.V. v. iParadigms, LLC*, 562 F.3d 630 (4th Cir. 2009).

[23] The Mass Digitization Report and the Notice of Inquiry are ambiguous about whether the pilot program would apply to literary works generally, or the much narrower class of books. GBS, of course, concerned books as defined in the agreement. *See* ASA at § 1.19. Even if the ECL were limited to books, it should be noted that libraries could have strong fair use arguments for making available the full text of certain types of books, such as books with time-limited markets. If that period has long since expired, the original market for that work no longer exists and subsequent uses would likely be considered fair and not a market substitution for the original work. Examples include factual works published annually such as travel guides or directories. Additionally, fair use probably would allow libraries to provide digital access to the many books that are distributed by their rights holders for free (e.g., religious or policy books).

[24] The Mass Digitization Reports suggests that the class of eligible works could be limited to those published before a certain date as a way of avoiding resolution of questions about works' commercial availability. *See* Mass Digitization Report at 87.

royalties. Mass digitization presents the technological means for creating a new market for these books, once Congress amends the Copyright Act to eliminate the barrier posed by the cost of clearing the rights in these books. If Congress established a system that enabled third parties to profit from this new market, for example by selling ebooks, Congress as a matter of equity could require that the rights holders receive some compensation from these sales. This is so even though the new market does not harm the rights holders by diminishing the sales in any existing market.

However, the equity argument for rights holder compensation is much weaker when Congress dedicates the new market to nonprofit educational or research uses. The rights holders would have already exhausted the intended markets for the books. Any additional compensation they would receive from the new market would be a windfall profit to them at public expense.

Moreover, given the orphan works issue identified above, as well as CMOs' long history of corruption, mismanagement, confiscation of funds, and lack of transparency,[25] there is a high probability that much of the revenue collected would not actually reach the rights holders. To its credit, the Copyright Office recognized this concern when it recommended that a CMO administering an ECL should be required by regulation "to demonstrate its adherence to transparency, accounting, and good-governance standards."[26] But regulation is effective only if it is enforced by a governmental body, which in turn requires significant resources. Further, regulation will not reduce the cost of identifying and locating the copyright owners of orphan works.

---

[25] *See* Jonathan Band and Brandon Butler, *Some Cautionary Tales About Collective Licensing*, 21 Mich. St. Int't L. Rev. 687 (2013), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2149036.

[26] Mass Digitization Report at 92.

Accordingly, sensible public policy would favor an exception permitting free access to digitized, commercially unavailable books for nonprofit educational and research purposes. Congress has adopted other specific exceptions for nonprofit uses as it has sought to achieve a balance in Title 17 among the interests of the diverse stakeholders in the copyright system. *See, e.g.*, 17 U.S.C. §§ 108, 109(b)(2), 110(1), 110(2), 110(3), 110(4), 110(6), 110(8), 110(9), 112(b), 112(c), 112(d), 121, and 1201(d). Therefore, we strongly suggest that the Copyright Office move away from an ECL, with all its impracticalities and costs, and instead study a framework for a true exception permitting access to digitized, commercially unavailable literary works for nonprofit educational and research purposes.

October 9, 2015