KERNOCHAN CENTER
FOR LAW, MEDIA AND THE ARTS

COLUMBIA UNIVERSITY SCHOOL OF LAW

October 8, 2015

Submitted by Online Submission Procedure

Karyn A. Temple Claggett
Associate Register of Copyrights and Director
of Policy and International Affairs
U.S. Copyright Office
101 Independence Avenue, SE
Washington, DC 20559-6000

Re: Comments Pursuant to Notice of Inquiry on
Mass Digitization Pilot Program, 80 Fed. Reg. 32614
(June 9, 2015)

Dear Ms. Claggett:

We submit these comments on behalf of the Kernochan Center for Law, Media and the Arts at Columbia Law School in response to the Notice of Inquiry cited above (NOI). The Kernochan Center is one of the leading centers for intellectual property research in the United States. Its faculty and staff dedicate their research and writing to copyright, trademarks, and related areas, as they concern traditional and emerging media, entertainment and the arts. The Center offers students an in-depth program of instruction, lectures, internships and externships while providing symposia, lectures, research studies and publications to the broader legal community. Founded as the Center for Law and the Arts, it was renamed in 1999 to honor Professor John M. Kernochan, its founder and a pioneer in teaching copyright in American law schools.

The Copyright Office's analysis and recommendations in its report on Orphan Works and Mass Digitization appropriately distinguished orphan works from mass digitization.[1] The Office's proposal for Extended Collective Licenses (ECLs) to facilitate mass digitization is a significant departure from current law, and likely to be long in gestation. But it is forward-thinking in that it avoids relying primarily on fair use, or on individual licenses (impractical for

[1] U.S. Copyright Office, Orphan Works and Mass Digitization: A Report of the Register of Copyrights (2015), available at http://copyright.gov/orphan/reports/orphan-works2015.pdf [hereinafter, "Orphan Works and Mass Digitization Report"].

Jerome L. Greene Hall 435 West 116th Street, Box A-17 New York, NY 10027
(212) 854-7424 FAX: (212) 854-9111

mass digitization), neither of which seems destined to lead to a solution optimal for both authors and users.[2] An important part of gaining acceptance of a potential ECL is the opportunity for right holders to opt out, but that opportunity must be coupled with an efficient means of making licensees aware of those works not encompassed within the ECL.

Our comments are directed toward the categories of works addressed in the NOI; ECLs for other categories of works would likely raise additional issues.

1. Examples of Projects

We are not ourselves in a position to digitize and offer access to copyrighted works and serve as a pilot project.

a. Qualifying Collections

The Office has recommended that an ECL be available for three categories of published copyrighted works: (1) literary works, (2) pictorial or graphic works published as illustrations, diagrams or similar adjuncts to literary works, and (3) photographs. It's unclear whether the Office envisions that any mass digitization project eligible for the pilot would serve as a prototype for an ECL that would embrace the full scope of one of the three categories (e.g., literary works) – perhaps modified by publication date or commercial availability – or whether the Office envisions one or more mass digitization projects, each of which would focus on a subcategory of one of these three categories. The latter approach is more realistic; it's hard to imagine that the same CMO would represent, for example, creators of books, magazines, computer programs, and ephemera such as advertisements. Even within the category of books, there may be proposed digitization activities that represent only a small subset. For example, a university library might seek to digitize scholarly monographs published before a certain date, or a culinary school library might seek to digitize its entire collection of out-of-print culinary science books.[3] A CMO composed of authors of scholarly monographs, or one of culinary science authors, might be instructive as to how a broader literary works CMO might function, but would be unlikely to lead to an ECL that could license literary works generally.

In any event, we suggest that the collections digitized for the pilot program (1) be large enough that it would be impractical to seek individual permissions, (2) have a substantive focus so that the creation of a CMO is manageable; and (3) present a range of issues that will need to be resolved for the CMO to operate effectively and be sufficiently representative – and efficient

---

[2] *See generally* Jane C. Ginsburg, "Fair Use for Free or Permitted-But-Paid," 29 BERKELEY TECH. L. J. 1383 (2014).

[3] We know of no entity seeking to digitize such collections; we created these examples merely to illustrate our larger point.

– ultimately to issue an ECL. It will have to be representative for the Copyright Office to authorize it to use the ECL framework; it will have to be efficient for its members to be willing to authorize it to issue an ECL that would necessarily embrace absent right holders. Accordingly, the project must include works by a meaningful number of right holders of works of a particular category. What that number is depends on the scope of the category of works sought to be represented.

An important question is whether the CMO would represent only out-of-commerce works or whether it would also include in-commerce works. First, with respect to books: In its report on Orphan Works and Mass Digitization, the Office suggested that the unsuccessful Amended Settlement Agreement in *Authors Guild v. Google*, which established a kind of ECL managed by the Book Rights Registry, indicates that right holders might be willing to participate in an ECL.[4] Significantly, that settlement treated in-commerce and out-of-commerce books differently. The text of in-commerce works could be accessed only if right holders opted in; the text of out-of-commerce works could be accessed unless their right holders opted out.[5] In short, while the settlement may have demonstrated that right holders of out-of-commerce works were willing to enter into an ECL that would license full text viewing of their works, the same cannot be said for right holders of in-commerce books. One might infer from the in commerce/out-of-commerce distinction in the Amended Settlement Agreement that works with current commercial value and known right holders are poor candidates for an ECL because the right holders are likely to opt-out. The ECL would likely give them less income than their own licensing programs provide. By contrast, out-of-commerce works lack current commercial value, so mass digitization may generate income where these works currently produce no revenue stream.

The NOI suggested the possibility that in-commerce works might be included on a different basis, e.g., for purposes of full-text search but not for full-text viewing.[6] But we question whether right holders (for books, predominantly publishers) would be interested in participating in an ECL for books readily available in the marketplace, increasingly in digital form. These are not works to which it's necessary to call attention, ones that have been forgotten and are languishing on library shelves.[7] Owners of in-commerce works might be more inclined to opt-in now when they face the alternative that such digitization might be deemed fair use. On the other hand, they may have greater security concerns with a series of decentralized databases managed by entities with varying levels of sophistication than they had with Google. Our fundamental concern is that including in-commerce works could diminish the likelihood that a CMO would achieve the representation necessary to qualify to issue an ECL. Perhaps if the pilot proves successful, right holders of in-commerce works might be willing to opt in. Similar

[4] Orphan Works and Mass Digitization Report, *supra* note 1 at 83.
[5] Authors Guild v. Google Inc., 770 F. Supp. 2d 666, 672 (S.D.N.Y. 2011).
[6] Orphan Works and Mass Digitization Report, *supra* note 1 at 86-87.
[7] *Authors Guild v. Google*, 770 F. Supp. 2d at 670 (one of the benefits of the Google Books project is that "[o]lder books – particularly out of print books, many of which are falling apart buried in library stacks – will be preserved and given new life").

considerations appear to be relevant to pictorial or graphic works published as illustrations, diagrams or similar adjuncts to literary works.[8]

It will be necessary to have an efficient means of determining whether or not a work is commercially available.[9] For books, one might consult Books in Print, but the status of books is made more complicated by print on demand. In the Google Books settlement, there were specific detailed provisions for determining whether a work was commercially available, as well as for determining who owned the rights and was entitled to the proceeds.[10] Administration of a an ECL would be much easier if the scope of included works could be defined with specificity, e.g., all works published before a certain date. But this might be unfair to right holders of older works that still enjoy a lucrative market (although presumably they could opt out). But considering how this issue was dealt with in the Google Books agreement, we are skeptical whether a radically simpler approach could be acceptable to the stakeholders.

Finally, while we believe this is implicit in the NOI, it should be made explicit that an institution is limited to digitizing material in its own collections and may not digitize materials from others' collections, for example, to make a more complete collection. Also, a digitizing entity should not, through simultaneous access to a digital copy, be able, in effect, to multiply the number of copies of a work it owns. In other words, the number of people who can access a particular work simultaneously should be limited to the number of copies of that work in the licensee's collection, unless the ECL specifically authorizes otherwise.

b. Eligibility and Access

We were not entirely clear on whether the pilot ECL(s) would license full text *access* only, or would also permit full text downloads (becoming, in effect, a print on demand facility), or something in between, e.g., allowing a limited number of pages to be printed. The concerns as to eligibility and the circumstances of access could vary, depending on who are the users and what they are permitted to do.

Eligibility. Access to the ECL-licensed digital material should be limited to the defined user group for a particular entity. For example, where the digitized collection belongs to a

---

[8] Photographs and other visual materials present additional challenges. We are insufficiently familiar with the market for these works, and with the prospects outside of CMOs for reviving out of commerce visual materials, to be able to offer comments on the operation of an ECL in this sector.

[9] A means of distinguishing in-commerce from out-of-commerce works would be necessary even if both categories were included in the ECL, since they would in all likelihood be treated differently for purposes of the license.

[10] *See, e.g.,* Amended Settlement Agreement, Authors Guild v. Google, Inc., No. Case No. 05 CV 8136-DC (Nov. 13, 2009), available at https://www.authorsguild.org/wp-content/uploads/2008/10/Amended-Settlement-Agreement.pdf at §3.2(d) and Appendix A ("Procedures Governing Author Sub-Class and Publisher Sub-Class under the Amended Settlement Agreement").

university library, access should be limited to students, faculty, and employees of the digitizing institution. Geographically-based libraries, such as public libraries, generally have geographically-based user groups. However, some of those groups can be quite large, e.g., the entire state in which the library is located, which is potentially problematic.[11] This issue might be resolved if the scope of eligible users is appropriately taken into account in the license fee.

Specialized libraries and archives present greater challenges in defining their "user group," since they often serve geographically dispersed and unaffiliated users. Some limit access to credentialed scholars, so it might be possible to define a user group in those terms.

On-Premises Access. Eligible users should, at a minimum, be allowed to access the digitized material on the premises of the ECL-licensed institution.

Remote Access. The pros and cons of remote access were discussed in some detail in the Section 108 Group Report.[12] Remote access presents greater risks for right holders, but at the same time, the ECL approach is premised on licensing to libraries, among others, and we suspect that an arrangement that requires users to come to the library premises would be of little interest to university and other research libraries. Of course, the availability of remote access (e.g., to personal computers of faculty and students working off-campus or abroad) depends on whether security is adequate.

Some institutions of higher education provide services to their local communities, but allowing remote access by community members considerably broadens the scope of the license. A better means by which to allow a university to serve the local community would be to allow the university to permit access to the public from terminals on the premises of the library.

Allowing remote access for broadly defined user groups of public libraries, or for users of specialized libraries and archives, presents difficult issues. We suspect that for purposes of licensing certain databases, these entities and their licensors have defined user groups for remote access, and such agreements might provide guidance. The principal concerns are that the user group be defined in such a way that the license fee fairly reflect usage, and that the ECL licensee not become a general document delivery service.

---

[11] The Section 108 Study Group Report 59 n. 113 (March 2008), available at http://www.section108.gov/docs/Sec108StudyGroupReport.pdf.

[12] *Id.* at 58-59.

c. Security Requirements

Security requirements are essential so that contents of the digital database are available only to authorized users. Such security might include, for example, limiting access only to users with appropriate passwords, implementing user agreements for remote access, monitoring for downloads disproportionate in volume, and so on.[13] However, there are at least two reasons to avoid defining the requisite security in the statute. First, security standards vary over time, and embodying standards in a statute limits flexibility. Second, the degree of security may depend on the nature of the database. This may also be an area in which the practices of libraries and licensors of databases may provide guidance. For those reasons, we recommend that the security requirements be included in the statute in general terms, with details established by the Copyright Office through regulations.

2. Dispute Resolution Process

The Office has asked for views on what form a dispute resolution process should take when a CMO and a licensee are unable to agree on licensing terms. Means suggested in the NOI include informal mediation, arbitration, and resort to the CRB. While informal mediation might be a good start, it seems reasonably likely that in some cases, or on some issues, mediation is unlikely to lead to harmonious resolution. For example, licensing fees and the scope of the authorized user group may be issues on which the CMO and licensee differ. In such cases, resort to the CRB may be the best option, but we have concerns about the time and expense that might take (unfortunately, binding arbitration is not necessarily faster or cheaper). Any legislation would have to be written to avoid the possibility that the CRB would become a de facto rate court for ECLs.

3. Distribution of Royalties

The Office has recommended that law or regulations establish a specific period by which royalties must be distributed to right holders. We agree that such a deadline should be established, but suggest that more information as to how the digitized database would be managed is necessary in order to establish the appropriate timeframe. For example, who will maintain the database, and on what basis will payment be made? The Amended Settlement Agreement in *Authors Guild v. Google* envisioned that Google would maintain a centralized database, collect the fees and maintain usage information, at least in the aggregate.[14] But the pilot program seems to envision that each institution participating will digitize its own collection and maintain its own database. If so, those institutions will presumably have the usage information necessary to determine distribution, and that information (along with the license

[13] *Id.*

[14] That information may also be helpful to right holders in deciding whether to republish a work currently out-of-commerce.

fees) would have to be conveyed to the CMO before fees could be allocated and distributed. The question then is how quickly licensees can reasonably provide this information, and related license fees, to the CMO. Experience in the EU with library digitization of their collections might provide guidance.

In short, the logistical set up – in particular, who will hold the data relevant to distribution and how quickly the CMO will get the data and the payments, is material to setting the time frame for making payments to right holders.

Fees payable to right holders will likely be low, at least at the outset, so having quarterly reporting periods may place an unrealistic burden on the CMO. It might be wiser to start with a requirement of semi-annual reporting periods, and revisit this issue after ECLs are up and running. Also, there should be no requirement to make a payment more than once a year if the royalties fall below a certain threshold (perhaps $10). Otherwise, the transaction cost of making the payment could well exceed the amount payable.

4. Diligent Search

The NOI seeks input on the scope of the CMO's responsibility to locate absent right holders. First, we suggest that the CMO's responsibility should not be defined in the same terms as the requirement for a potential user to qualify for an orphan works limitation of liability under the proposal in the Orphan Works and Mass Digitization Report. It is likely to be expensive to establish and run an ECL, and revenues are uncertain. The Google Books search settlement allocated $34.5 million (less the costs of notice to the class) for the startup of the Book Rights Registry;[15] the ECLs established in this pilot program will not have similar seed capital.

We agree with the NOI that at a minimum, the CMO must maintain a publicly available list of works for which one or more right holders can't be found. It would also be helpful if the Copyright Office website could list authorized ECLs and indicate that right holders potentially owed money should consult the relevant websites. The CMO might also use social media and crowdsourcing to good effect.

Beyond that, there are all kinds of resources that might be used, e.g., contacting the publisher, running online searches for the author or the author's heirs, etc. But in many cases the CMO will be holding only a very small amount of money for a particular right holder; perhaps the scope of the search for that right holder should vary with the amount of money that has been collected on his/her behalf. There could be a reasonable limit as to the total amount per year the CMO should expend on searching for absent right holders so that actively participating right holders can still get some recompense for the licensing of their works.

[15] Amended Settlement Agreement, *supra* note 10 at §2.1(c).

5. Other Issues

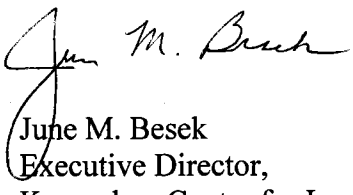
While a resolution process for disputes between the CMO and its ECL licensees is important, so too is a dispute resolution process that the CMO can use when there is disagreement on ownership rights with respect to a work and the CMO is uncertain to whom to remit revenues collected. The CMO should not be responsible for resolving such disagreements, but perhaps could facilitate arbitration or mediation between the disputants.

Finally, we wondered whether the ECL licensees should have any obligations with respect to preservation of the digital database that they create. As noted above, the ECL framework in the NOI seems to envision a series of decentralized databases, in contrast to the Google Books database envisioned in the Amended Settlement Agreement. Preservation is an important aspect of the Hathitrust database as well as others. [16] It is clearly in the public interest to ensure long-term access to these works.

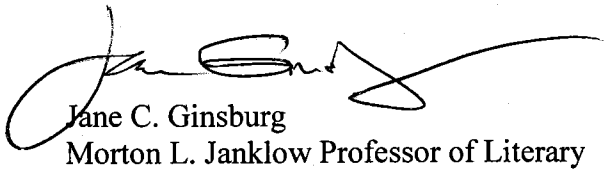
****

We appreciate the opportunity to submit these comments, and look forward to working with the Copyright Office on issues concerning the mass digitization pilot program.

Respectfully submitted,



June M. Besek
Executive Director,
Kernochan Center for Law,
Media and the Arts
Columbia Law School



Jane C. Ginsburg
Morton L. Janklow Professor of Literary
and Artistic Property Law
Columbia Law School

[16] *See* Authors Guild v. Hathitrust, 755 F.3d 87, 103 (2d Cir. 2014); JSTOR, Preserving Scholarship, http://about.jstor.org/content/preserving-scholarship (last visited Oct. 8, 2015).