

COMMENTS ON MASS DIGITIZATION PILOT PROGRAM; REQUEST FOR COMMENTS BY THE UNITED STATES COYRIGHT OFFICE

By Manny Schecter – *Chief Patent Counsel*

Terry Hardi – *Copyright Counsel*

IBM Corporation

October 9, 2015

IBM Corporation (IBM) is pleased to offer comments on the U.S. Copyright Office’s (the Office) report of its conclusions and recommendations on the issues of orphan works and mass digitization (Report).¹ With respect to mass digitization – the digitizing of large databases of material – Report proposes a new legislative framework, referred to as “extended collective licensing” (ECL).² ECL is aimed at striking an appropriate balance between facilitating those aspects of mass digitization that serve the public interest while safeguarding the rights of copyright owners.³ The Office accordingly recommends a limited “pilot program” to test and gain experience with ECL in the U.S. The pilot program (Pilot) would enable ECL for mass digitization projects serving nonprofit educational and research purposes.⁴

The following comments discuss IBM’s concerns with respect to the scope of the works to be included in the Pilot, extension of the Pilot to cover indirect commercial uses, the exclusion of works under existing licenses, restructuring of the Pilot’s five-year sunset plan and adding a fair use savings clause into the legislation. Specifically, IBM addresses the issues responsive to Questions 1 and 5 in the Office’s Notice of Inquiry dated June 9, 2015.⁵

Question 1. Examples of Projects

a. Qualifying Collections

(1) Limit the Scope of “Literary Works” by Carving Out Computer Programs

Although computer programs are considered literary works for the purposes of copyright law, they should not fall within the scope of ECL. As discussed above, the Office has suggested limiting the Pilot to specific categories of works, including literary works, as appropriate starting points for an ECL system.⁶ In this regard, the only category of literary works in the Report that the Office does address is books.⁷ For example, the Office uses the proposed Google Books settlement⁸ as an example commenting that “books are ‘the centerpiece of many cultural collections’ and as such are ‘relevant to many – perhaps the majority of – large-scale scanning initiative.’”⁹ Even as to the discussions regarding whether the ECL framework should be limited to only “out-of-commerce” works, the entire focus is on books, and the entire analysis revolves around the comparison between in-commerce books and out-of-print books.¹⁰ Properly, the Report does not refer to computer software for the ECL or Pilot.

In accordance with the purpose of the Pilot and ECL in general, and to avoid the unintended consequence of disrupting a major U.S. industry, computer software should not be part of the Pilot or ECL in general. Based on the Office's Report, the purpose behind adopting an ECL Pilot is to make it easier for users to digitally reproduce and offer full-text access to a large collection of works.¹¹ In other words, ECL is created to facilitate mass digitization projects to enhance the research and educational value of digital resources, allowing more people to gain access to, research, read and study millions of works digitally. Computer programs are different, however, from traditional literary works. They are not consumed by people for entertainment or for the information they impart. Rather they are consumed by computers to direct the computers' operations. Furthermore, in almost all instances, computer programs are made available under licenses.¹² Including computer programs in the ECL system, as will be discussed further below, will create a conflict with the licenses under which the computer programs had already been distributed.

As an exception to the forgoing exception, computer program source code that is embedded in text documents, for example software documentation or computer science educational texts, can be included in ECL since such source code performs the same functions of instruction, explanation and education as does the text of other educational literary works. Thus IBM recommends that source code embedded in educational and other software documentation be included in the ECL and Pilot, but the copyrighted software used to operate computers and other data processing devices should not.

In conclusion, IBM encourages the Office to provide a clear definition of the "literary works" that excludes computer programs with the one limited exception of source code embedded in text.

(2) Broaden Permissible Uses to Allow Certain Indirect Commercial Purposes

The Office takes a middle-ground approach by not setting limits on the categories of users who may engage in mass digitization activities, while limiting "permissible uses to those undertaken for nonprofit educational or research purposes" that do not have "any purpose of direct or indirect commercial advantage (emphasis added)."¹³ By so doing, the Office believes that the ECL framework will restrict the legislation's scope to mass digitization projects that serve the public interest by permitting entities falling outside the traditional categories of libraries and archives but that are engaged in similar activities, to utilize mass digitized works appropriately.¹⁴ The prohibition against even indirect commercial gain, however, will seriously curtail or eliminate entirely participation by commercial entities, and will stifle innovative technologies and other activities that benefit the public without harming the copyright holder.

Rather than enabling greater benefits, the indirect commercial purpose limitation will, in effect, limit users almost exclusively to traditional non-profit organizations, such as libraries or archives. In fact, even a company such as Google, whose Google Book project served as a template for the program, would be excluded since it would be difficult to show that it gained no commercial advantage from participation.

Furthermore, in the proposed ECL and Pilot, users of the mass digitized works will be required to pay license fees or royalties to compensate copyright owners according to the specific royalty plan developed by the relevant CMO. Therefore, it is even more unlikely that a for-profit entity will be willing to participate since it will incur fees while being precluded from realizing even indirect financial gain. As the Office concludes in its Report, the success of an ECL system will only be possible where the market wants it.¹⁵ In other words, a sufficient number of prospective users must conclude that the benefits obtainable through ECL—including legal certainty and broader permitted uses—are greater than the costs of securing a license.¹⁶ However, by restricting permitted uses to those undertaken for pure non-profit purposes and prohibiting even indirect commercial advantage, almost all for-profit entities will conclude that it is not economically reasonable to participate in a mass digitization project, and accordingly limit the success of the Pilot.

On the other hand, new uses for the information and data contained in published works are already permitting the nascent fields of data and business analytics to grow. The ability of computers to analyze massive amounts of information extracted from enormous numbers of works has the potential to lead to cures for illness and disease, minimize electrical power interruption, combat terrorism and crime and gain new insights into the complexities of human communications.

It is recommended therefore that the ECL and Pilot permit certain indirect commercial uses and that indirect commercial use be defined to include uses that do not distribute, or otherwise inject into commerce, more than minimal portions of the work in question. Such uses will not significantly reduce the demand for or revenues derived from a work to the detriment of the copyright holder but will instead have the beneficial result of increasing the public's fund of knowledge and contribute substantially to the creation of new and innovative technologies.

In conclusion, we suggest that the ECL and Pilot broaden its permissible uses to include not only those undertaken for nonprofit educational or research purposes, but also those with some purpose of indirect commercial advantage.

(3) Exclude Works under Existing Licenses from the ECL System

The Report does not discuss how to handle any of the enormous number of works that are already licensed under existing license agreements or terms. For example, over one billion works¹⁷ are already licensed under one of the family of Creative Commons licenses.¹⁸ IBM itself has licensed thousands of its own documents under IBM licenses that generally permit downloading and use as long as the users comply with the specific terms of IBM's license.

For these licensed works, it is reasonable to conclude that the rightsholders have already chosen the terms under which they wish to license their works, and, unless the terms of the ECL are completely consistent with those of their preexisting license (which is highly unlikely), they should be considered as having opted out of the ECL. Including works for which licensing terms are already available will potentially create a “battle of the forms” and uncertainty with respect to the permitted uses of the previously licensed work. Therefore, we recommend that works that

are already available to the public under existing licenses be excluded from the ECL and Pilot. It is unreasonable to require rightsholders that have already made their works available under license to monitor their works and opt out of the ECL and thereby incur substantial administrative and monetary burdens.

In summary, copyright owners that have already made their works available under license should be considered to have opted out of the ECL. Such works must not be included in the ECL and Pilot.

(4) The Scope of the Works Included Should be Clarified to Cover Works that Have Already Been Digitized

While the Report clearly addresses the collection and conversion of hard copy, printed works into digitized form, the Report should just as clearly include the collection, reformatting, or other handling of works that are already digitized for the ECL and Pilot. Conventionally, the term “digitization” is used to describe the process of converting analog information into a digital format that is useable by computers and other electronic devices.¹⁹

We note that the previous Copyright Office 2011 Report²⁰ clearly included already digitized works within the definition of mass digitization.²¹ The latest Report now generally defines mass digitization projects as those “in which the scale of digital copying is so extensive as to make the individual clearance of rights a practical impossibility.”²² If we construe “digital copying” to include digital to digital copying, as opposed to only analog (scanning) to digital copying, then this definition is sufficient.

In fact, the Report discusses databases that have been built from preexisting digital data as “examples of activities that could be described as mass digitization insofar as they involve the digital copying and storage of large numbers of works.”²³ Lastly, in order to make it clear that ECL is intended as a solution for large-scale digitization projects only, the Report recommends that the legislation require that any licensed uses be made in connection with the “creation or operation of a qualifying digital collection (emphasis added).”²⁴ The use of the disjunctive “or” suggests that mass digitization can include the integration and aggregation of already digitized works into a qualifying digital collection and the operation of such digital collection.

Perhaps, most importantly, it must be noted that in today’s digital economy most books are released simultaneously in digital and hardcopy form.²⁵ While hardcover sales declined slightly between 2008 and 2012 (from \$5.2 billion to \$5 billion), eBook sales grew at an astonishing rate during that period, rising from \$64 million to \$3 billion.²⁶ eBook sales in the United States had surpassed hardcover book sales by the end of the first quarter of 2012.²⁷ McPheters & Co. estimates that half of all magazine and newspaper circulation will be via digital delivery by the end of 2015.²⁸ In fact, from 2005 to 2008 U.S. libraries experienced a 60% growth in eBook collections²⁹ and by 2010 66% of U.S. public libraries held eBooks in their collections.³⁰ IBM submits it should be made clear that the ECL encompasses this large body of existing digital information which would greatly contribute to the value and success of the program.

In conclusion, the Report needs to clarify the scope of what is meant by mass digitization. While the Report's language and cited examples infer that its scope covers previously digitized works, the text should be more certain on this point.

Question 5. Other Issues

(1) Clarify the Five-Year Sunset Clause of the Pilot Program

The Office recommends that the legislation include a five-year sunset clause to give Congress the opportunity to assess the program's effectiveness and to consider whether ECL should be implemented on a long-term or permanent basis.³¹ While having a Pilot is prudent, the Office does not address what will happen to those uses of works made under the Pilot after the five-year sunset. The Report only states that a CMO's obligations regarding the maintenance and disposition of unclaimed royalties would extend beyond the sunset date until all such monies are disbursed.³² What the Report does not answer is whether uses made possible under the Pilot will be allowed to continue, and if so, under what terms? Without further clarification and a detailed plan for the maintenance and operation of the existing mass digitization activities, participants run the risk that their investment will be lost upon the end of the five year Pilot period. The issue is twofold: 1) whether a participant would be able to continue a project that it has invested in and 2) what liability it would have to rightsholders upon completion of the Pilot. Therefore, IBM encourages the addition of a savings clause granting the Pilot participants the ability to continue using the works themselves, or any derivatives of the works, created during the Pilot under the terms of the ECL.

In conclusion, we recommend against the inclusion of the five-year sunset clause without the Office first providing a clear and detailed plan as to the continued maintenance and operation of those projects initiated and developed under the Pilot. In addition, we urge the Office to include a grandfather provision or savings clause granting rights to users of works made available under the Pilot to continue to use the works and derivative works thereof in such projects in the future.

(2) Fair Use Savings Clause

The Office recommends that the legislation include a savings clause providing that nothing in the statute is intended to affect the scope of fair use.³³ We agree that such a clause should be included in the legislation. In particular, when certain uses or aspects of a mass digitization project qualify as fair use under the U.S. copyright law, we believe that the ECL system or Pilot should not be used as an excuse to limit fair use, but only to facilitate uses not available under fair use and so maintain a balanced Copyright Act.

For example, the Second Circuit has held³⁴ that mass digitization of works for the purpose of allowing the general public to search for particular terms across all copies of the digitized works in a repository to be a use that is fair. Accordingly, such use required no license or payment of fees and should remain a fair use despite the fact that legislation enacting ECL may provide an alternative.

In conclusion, we support the Office’s recommendation to include in the legislation a savings clause providing that fair use is in no way affected by the ECL or Pilot, and mass digitization users are free to forego any license and assert the fair use defense in the event litigation arises.

¹ See U.S. Copyright Office, Orphan Works and Mass Digitization: A Report of the Register of Copyrights (2015), available at <http://www.copyright.gov/orphan>.

² See Report at 75.

³ See *id.*

⁴ See Report at 89.

⁵ See U.S. Copyright Office, Library of Congress, Notice of Inquiry, Federal Register No. 32614 (June 9, 2015), available at <http://copyright.gov/fedreg/2015/80fr32614.pdf>.

⁶ See Report at 85.

⁷ See Report at 85-87.

⁸ See Report at 85. See also *Authors Guild, Inc. v. Google Inc.*, 770 F. Supp. 2d, 666, 671-72 (S.D.N.Y. 2011) (describing terms of the settlement).

⁹ Report at 85.

¹⁰ See Report at 86-87.

¹¹ See Report at 72-76.

¹² Computer programs are licensed under a variety of licenses, including proprietary licenses and open source licenses. Software licenses specify the rights granted, limitations on the use of a computer program, and possible fee arrangements among their many terms and conditions. See, for example, opensource.org for a listing of some of the most widely used open source licenses. Open Source Initiative, Licenses & Standards, available at <http://opensource.org/licenses> (last visited on Oct. 6, 2015).

¹³ See Report at 89.

¹⁴ See *id.*

¹⁵ See Report at 105.

¹⁶ *Id.*

¹⁷ See Creative Commons, State of the Commons – The Size of the Commons, available at <https://stateof.creativecommons.org/report/> (last visited on Oct. 6, 2015), where the methodology for arriving at this estimate is described.

¹⁸ See Creative Commons, About the Licenses, available at <https://creativecommons.org/licenses/> (last visited on Oct. 6, 2015) for a description of the different Creative Commons licenses. For more information about Creative Commons Licenses, see <http://creativecommons.org/>.

¹⁹ “Digitizing or digitization is the representation of an object, image, sound, document or signal (usually an analog signal) by generating a series of numbers that describe a discrete set of its points or samples.” Wikipedia, Digitizing,

available at <https://en.wikipedia.org/wiki/Digitizing> (last visited on Oct. 6, 2015). *See also* Techpedia, Definition - What does Digitization mean?, available at <https://www.techopedia.com/definition/6846/digitization> (last visited on Oct. 6, 2015) (“Digitization is the process of converting analog signals or information of any form into a digital format that can be understood by computer systems or electronic devices. The term is used when converting information, like text, images or voices and sounds, into binary code. Digitized information is easier to store, access and transmit, and digitization is used by a number of consumer electronic devices.”).

²⁰ *See* U.S. Copyright Office, Legal Issues in Mass Digitization: A Preliminary Analysis and Discussion Document (2011), available at http://www.copyright.gov/docs/massdigitization/USCOMassDigitization_October2011.pdf.

²¹ For example, “A more recent concern is how to apply the existing copyright framework to the capture, collection, and preservation of “born digital” works, such as electronic books (“e-books”), and the digital photographs and other visual art works that might be incorporated in those books. The number of these works available in the commercial marketplace has been expanding rapidly. For example, in 2007 the online retailer Amazon.com began challenging brick and mortar bookstores by offering electronic books for use with the Kindle and other mobile devices. Earlier this year Amazon announced that its sales of e-books now exceed its sales of hardback and paperback books combined.” 2011 Report at 16.

²² Report at 73.

²³ *See* Report at 73. In *Vanderhuy v. iParadigms, LLC*, 562 F.3d 630, 634 (4th Cir. 2009), defendant iParadigms owned and operated an online technology system designed to evaluate the originality of written student works in order to prevent plagiarism. When a school subscribes to iParadigms' service, it typically requires its students to submit their written assignment online via the iParadigms' web-based system. Defendant then performs a digital comparison of the student's work with content available on the Internet to determine whether the student work is a product of plagiarism. The student papers, in fact, may have never existed in hard copy form.

²⁴ *See* Report at 89.

²⁵ As e-readers such as tablets and smartphones have become more and more prevalent and popular among customers, traditional publishers of literary works such as books, magazines and newspapers have strived to make their new contents simultaneously available in hardcopy and digital formats. *See* Frania Hall, *The Business of Digital Publishing: An Introduction to the Digital Book and Journal Industries*, Routledge (2013). In fact, in the first quarter of 2015, worldwide tablet shipments stood at 76.1 million units, a 653% increase from the same quarter in 2010. *See* Statista, *Worldwide Tablet Shipments from 2nd Quarter 2010 to 2nd Quarter 2015 (in Million Units)*, available at <http://www.statista.com/statistics/272070/global-tablet-shipments-by-quarter/> (last visited on Sep. 30, 2015).

²⁶ Joe Hyrkin, *Issue, Digital Publishing: How it will evolve in 2014 and beyond*, available at <https://gigaom.com/2014/01/04/digital-publishing-how-it-will-evolve-in-2014-and-beyond/> (Jan. 4, 2014).

²⁷ *See* Jason Boog, *eBook Revenues Top Hardcover*, available at <http://www.adweek.com/galleycat/ebooks-top-hardcover-revenues-in-q1/54094?red=as> (Jun. 15, 2012).

²⁸ *See* Rebecca McPheters, *Magazines and Newspaper Need to Build Better Apps*, available at <http://adage.com/article/media/viewpoint-magazines-newspapers-build-apps/232085/> (Jan. 13, 2012).

²⁹ *See* Michael Saylor, *The Mobil Wave: How Mobile Intelligence Will Change Everything*, Vanguard Press (2012), at 124.

³⁰ *See* Sue Polanka, *66% of Public Libraries in US Offering eBooks*, available at <http://www.libraries.wright.edu/noshelfrequired/2010/08/18/66-of-public-libraries-in-us-offering-ebooks/> (Aug. 18, 2010).

³¹ See Report at 102.

³² *Id.*

³³ See Report at 101.

³⁴ See *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87, 97-98 (2d Cir. 2014).