CROSSING THE DIGITAL RUBICON: GOOGLE BOOKS AND THE DAWN OF AN ELECTRONIC LITERATURE REVOLUTION

*Timothy J. Busse**

* Registered Patent Attorney, specializing in patent preparation and prosecution in the chemical, life sciences, materials, and mechanical fields. For valuable feedback and guidance writing this article, I thank Professor Craig Joyce and participants in his Copyright Seminar. I thank the Houston Business and Tax Law Journal for thoughtful editorial assistance.

Abstract

The Google Books Project is an effort to digitize and index virtually all of the world's literary works with the ultimate goal to make every book on earth accessible to anyone with an Internet connection. But only three years after Google began digitizing books, copyright holders challenged the legality of mass digitization under U.S. copyright law. This paper explores the application of fair use to protect the Google Books Project and proposes that extended collective licensing and international treaty would produce a superior result that promotes the global benefit of creative works and protects copyright holders.

"In the beginning, there was Google Books."[1]

1. *Google Books History*, GOOGLE BOOKS, https://books.google.com/googlebooks/about/history.html (last visited Sept.. 30, 2017).

I. Introduction

At the turn of the twenty-first century, Google embarked on an unprecedented feat to digitize and index virtually all of the world's literary works.[2] The Google Books Project actualized a vision to make every book on earth accessible to anyone with an Internet connection.[3] But only three years after Google began digitizing books, copyright holders challenged the legality of mass digitization under U.S. copyright law.[4] As the shroud of uncertainty from ruined settlements and protracted litigation fell, Google Books stood unscathed, sheltered by the doctrine of fair use.[5] This paper explores the application of fair use to protect the Google Books Project and proposes that extended collective licensing and international treaty will produce a superior resolution that promotes the global benefit of creative works and protects copyright holders.

Part II of this paper provides a primer on the foundation of U.S. copyright law, specifically fair use, upon which the Google Books controversy was built. Part III of this paper presents a brief overview of domestic and international digitization projects, an overview of the Google Books mass digitization project, and a review of the Google Books dispute. Part IV of this paper analyzes application of fair use to mass digitization and proposes that extended collective licensing and, later, international treaty, provide a better resolution to this issue. Part V provides a brief conclusion.

This paper does not address the issue of orphan works, fair use as applied to library generated digital copies, or the infringement of foreign copyright. First, orphan works comprise texts that are in-copyright, but whose copyright holder cannot, after diligent effort, be identified or located.[6] A portion of the Google Books Project corpus are orphan works, which presents a unique, perhaps economically insurmountable, challenge in identifying copyright holders for any licensing schemes that could address the issue of mass digitization.[7] The issue of orphan works, therefore, requires separate consideration from a legislative viewpoint. Second, this paper does not address the issue of fair use as applied to library generated digital texts or use of Google's

---

2. *See id.*

3. *See id.*

4. *See id.* (beginning the digitization process in 2002); see also Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282, 284 (S.D.N.Y. 2013) (lawsuit began in 2005).

5. *See* Authors Guild, Inc. v. Google, Inc., 804 F.3d. 202, 229 (2d Cir. 2015).

6. *See* U.S. COPYRIGHT OFFICE, REPORT ON ORPHAN WORKS 15 (2006), https://www.copyright.gov/orphan/orphan-report.pdf.

7. *See* Matthew Sag, *The Google Book Settlement and the Fair Use Counterfactual*, 55 N.Y. L. SCH. L. REV. 19, 38 (2010) (explaining that the treatment of orphan works was one of the most contentious issues in the amended settlement agreement between Google and the Authors Guild).

digitized corpus by libraries.[8] This issue falls, at least partially, under the purview of §108 of the Copyright Act.[9] Yet, by necessity, some of the following discussion includes reference to the Second Circuit's opinion in *HathiTrust* because that case formed a foundation for that court to bridge the gap between a finding of fair use in the context of not-for-profit and for-profit entities engaged in mass digitization projects.[10] Finally, the scope of Google's possible infringement of foreign copyright law is beyond what can be described in the subsequent pages.[11] While the proposed scheme partly rests upon reconciling domestic and foreign copyright, the international harmonization of copyright requires evaluation of treaties and foreign law. For these reasons, the Google's possible infringement of foreign copyrights will not be considered.

II. PRIMER ON FAIR USE

A. Fundamental Protections of Copyright Law

The doctrine underlying the fair use defense—while accepting the fundamental protections of particular exclusive rights in a competitive environment—seeks to promote the utilization of works in limited ways that further benefit the public to the detriment of the copyright holder's rights.[12] Article I, section 8 of the U.S. Constitution provides the foundation of modern U.S. copyright law: "The Congress shall have Power ... to Promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[13] This clause shows an understanding that economic forces and the incentive of personal gain drive creativity and the innovation of useful works that may benefit the public.[14] Thus, works of greater public importance should carry a

---

8. *See generally* Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 94-104 (2d Cir. 2014) (discussing how fair use applies in the context of library generated digital texts).

9. *See generally* 17 U.S.C. § 108 (2016).

10. *See* discussion *infra* Section III.B.3.

11. *See, e.g.*, Pamela Samuelson, *The Google Book Settlement as Copyright Reform*, 2011 WIS. L. REV. 479, 548–49 (2011) (discussing implications of the Google Books settlement agreement on foreign copyright holders).

12. *See HathiTrust*, 755 F.3d at 95.

13. U.S. CONST. art. I, § 8, cl. 8. The preamble of England's first copyright enactment, "An Act for the Encouragement of Learning, by Vesting the Copies of Printed Books in the Authors or Purchasers of such Copies, during the Times therein mentioned," reflects a similar purpose to prevent the financial detriment of copyright infringement and encourage "Learned Men to Compose and Write useful Books ...." Copyright Act of 1710, 8 Ann. c. 19 (Gr. Brit.).

14. *See* Mazer v. Stein, 347 U.S. 201, 219 (1954) ("[T]he economic philosophy behind [art. I, § 8]... is the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors...."); *see also* Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975) (explaining that the sole interest of the United State and purpose of granting exclusive rights to copyright holders lies in the general benefits derived by the public from the labors of authors).

proportionally greater incentive.[15] And, by extension, the protections of copyright law need not apply in force where a work provides little or no additional benefit to public welfare.

To drive creativity with sufficient financial incentives, Congress granted to all copyright holders the five exclusive rights: (1) to reproduce the work, (2) to prepare derivative works, (3) to distribute copies of the work, (4) to perform the work publicly, and (5) to display the work publicly.[16] These five enumerated rights accord copyright holders "'to do and to authorize' any of the activities specified" and "comprise the so-called 'bundle of rights' that is a copyright."[17] The exclusive rights are divisible and separately enforceable, but also "cumulative and may overlap in some cases."[18] For this reason, each right is defined in "broad terms" to be read in conjunction with the "various limitations, qualifications, or exemptions in the 12 sections that follow."[19]

Many times, infringement of a copyright fits neatly into one or more of these exclusive rights, either a copyright is infringed or it is not.[20] But even when infringement is clear, there are circumstances when the underlying purpose of copyright law—to benefit the public by promoting the progress of creative works—supersedes the exclusive rights of the copyright owner.[21] One exemplary exception underlying the primary defense in numerous cases of copyright infringement is found in the body of law defining the "*sui generis* defense" of fair use.[22]

B. Fair Use

Since the enactment of the first copyright statute in 1790,[23] courts have applied the concepts tantamount to fair use in countless copyright infringement cases.[24] The fair use doctrine, however, remained a

---

15. *Cf.* Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 559 (1985) ("It is fundamentally at odds with the scheme of copyright to accord lesser rights in those works that are of greatest importance to the public. Such a notion ignores the major premise of copyright and injures author and public alike.").

16. 17 U.S.C. § 106 (2016). Section 106 includes a sixth right, "in the case of sound recordings, to perform the work publicly by means of digital audio transmission." *Id.*

17. H.R. REP. NO. 94-1476, at 61 (1976), *reprinted in* 1976 U.S.C.C.A.N. 5659, 5674.

18. *Id.*

19. *Id.*

20. *See id.*

21. *Compare* U.S. CONST. art. I, § 8, cl. 8 *with* 17 U.S.C. § 107 (2016). *See also* Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105 (1990); *but see* Lloyd L. Wienreb, *Fair's Fair: A Comment on the Fair Use Doctrine*, 103 HARV. L. REV. 1137, 1141 (1990).

22. *See* CRAIG JOYCE ET AL., COPYRIGHT LAW 809 (Matthew Bender & Co. ed., 9th ed. 2013).

23. Act of May 31, 1790, 1 Stat. 124 (1790) (current version codified at 17 U.S.C. §§ 101-1332 (2015)).

24. *See* H.R. REP. NO. 94-1476, *supra* note 17, at 65.

nebulous judge-made "equitable rule of reason"[25] until the passage of the Copyright Act of 1976.[26] The United States Supreme Court, in 1984, first attempted to provide guidance concerning the broad statutory language of the Act.[27] But the Court shifted its position merely ten years later.[28] Today, application of fair use remains in flux, including a marked shift among four considerations that originally defined the doctrine. [29]

The fair use doctrine originated with Justice Story's interpretation of fair abridgment[30] in *Folsom v. Marsh*.[31] In *Folsom*, Justice Story explained, "a fair and [bona] fide abridgment of an original work, is not a piracy of the copyright of the author."[32] Justice Story recognized that defining what he termed a "justifiable use" is a difficult and imprecise exercise in equity.[33] Even so, Justice Story outlined several factors to consider in deciding questions of fair use including "the nature and objects of the selections made, the quantity and value of the materials used, and the degree in which the use may prejudice the sale, or diminish the profits, or supersede the objects, of the original work."[34]

Twenty-eight years after *Folsom*, the term "fair use" was coined in *Lawrence v. Dana* and recognized as a defense to copyright infringement.[35] The court in *Lawrence* defined fair use narrowly as "quotations and extracts for the bona fide and avowed purpose of comment or criticism, or for the purpose of presenting the views of the writer as an authority."[36] Fair use was sparsely referenced during the next several decades.[37] But courts began to refine Justice Story's fair use

---

25. *See* H.R. REP. NO. 94-1476, *supra* note 17, at 65.

26. *See generally* General Revision of Copyright Law, Pub. L. 94–553, 90 Stat 2541 (1976) (current version codified at 17 U.S.C. §§ 101-1332 (2015)).

27. *See* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 451 (1984) (suggesting a presumption of unfair exploitation in case of commercial use of a copyrighted work).

28. *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994) (finding that transformative use may supersede considerations of commercial use).

29. *See* Neil Weinstock Netanel, *Making Sense of Fair Use*, 15 LEWIS & CLARK L. REV. 715, 736–46 (2011) (showing by empirical analysis that today transformative use dominants fair use analysis).

30. The fair abridgment doctrine was abolished by the 1909 Copyright Act. *See* Copyright Act of 1909, Pub. L. No. 60–349, ch. 320, § 6, 35 Stat. 1075, 1077 (1909) (current version codified at 17 U.S.C. §§ 101-1332 (2015)).

31. Folsom v. Marsh, 9 F. Cas. 342 (C.C.D. Mass. 1841) (No. 4901) (establishing the seminal factors that would become the doctrine of fair use).

32. *Id.* at 345 (citations omitted).

33. *Id.* (stating that defining fair use "is one of the most difficult points . . . which can well arise for judicial discussion.").

34. *Id.* at 348.

35. *See* Lawrence v. Dana, 15 F. Cas. 26, 44 (C.C.D. Mass. 1869).

36. *Id.*

37. Only 54 cases referencing fair use in the context of copyright were reported between 1869 and 1951. During the next fifty years, that figure increased nearly twelve-fold to 651 cases. Search conducted on Westlaw, https://a.next.westlaw.com (search for "fair use"; then filter results by topic of "copyright and date range "from 1841 until 1950" ("from 1951 until 2001")).

doctrine during the mid-twentieth century.[38] Soon, fair use became a judicial tool to balance the exclusive rights of copyright holders against public welfare.[39]

Over time, the courts melded Justice Story's factors into the fair use doctrine that Congress codified as a four-factor inquiry.[40] At present, 17 U.S.C. § 107 provides:

> [T]he fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—
>
> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
>
> (2) the nature of the copyrighted work;
>
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
>
> (4) the effect of the use upon the potential market for or value of the copyrighted work.
>
> The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors.[41]

The legislative history of section 107 explains that Congress found it necessary to codify the fair use doctrine not to change judicial precedent, but rather to endorse the long history of fair use cases.[42] Congress indented the six examples in the preamble to be non-exclusive because the "endless variety of situations and combinations of circumstances that can rise in particular cases precludes the

---

38. *See, e.g.*, Berlin v. E.C. Publications, Inc., 329 F.2d 541, 543–44 (2d Cir. 1964), *cert. denied*, 379 U.S. 822 (finding that the Constitution supports "subordinat[ing] the copyright holder's interest in a maximum financial return to the greater public interest in the development of art, science and industry.").

39. *See id.*; *see also*, MCA, Inc. v. Wilson, 677 F.2d 180, 183 (2d Cir. 1981) ("The less adverse effect that an alleged infringing use has on the copyright owner's expectation of gain, the less public benefit need be shown to justify the use.").

40. *See* 17 U.S.C. § 107 (2012).

41. *Id.* The last sentence of § 107 was added in 1992 to prevent a chilling effect on publication when the Second Circuit in *Salenger v. Random House, Inc.*, 811 F.2d 90 (2d Cir. 1987), and *New Era Publ'ns, Int'l ApS v. Henry Holt and Co.*, 873 F.2d 576 (2d Cir. 1989), suggested that the *Harper & Row* decision barred application of fair use to unpublished works. *See* 138 CONG. REC. E2610 (daily ed. Sept. 14, 1992) (speech of Hon. Carlos J. Moorhead).

42. *See* H.R. REP. NO. 94-1476, *supra* note 17, at 65.

formulation of exact rules in the statute."[43] Moreover, the fair use doctrine was intended to grow independent of the statutory text, "especially during a period of rapid technological change."[44] This intended plasticity required vague language that gave little guidance to the courts.[45] Therefore, reasonable assertion of the fair use defense requires case-by-case analysis to determine if the specific circumstances warrant an exception to infringement under the doctrine.[46]

The lower courts were left without guidance as to application of fair use,[47] that is, until 1984 when the Supreme Court decided *Sony Corp. v. Universal City Studios*.[48] *Sony*, however, gave only limited guidance in interpreting section 107 because the majority—focusing on preservation of the technological niche created by video recorders—held that private copying of television programs to be fair use via time-shifting.[49] Similarly, *Harper & Row Publishers v. Nation Enterprises*, decided the following year, gave little guidance on fair use.[50] In *Harper & Row*, the Court suggested that fair use does not apply to unpublished works and echoed the dictum from *Sony,* i.e., that the effect on the market was the "single most important" factor.[51] In *Stewart v. Abend*, the Court again presumed commercial use was unfair and ignored any discussion of the first factor.[52]

In 1994, *Campbell v. Acuff-Rose Music, Inc.* gave the Court an opportunity to address the full breadth of the fair use doctrine—as

---

43. H.R. REP. NO. 94-1476, *supra* note 17, at 66; *see, e.g.*, Time, Inc. v. Bernard Geis Associates, 293 F. Supp. 130 (S.D.N.Y. 1968) (holding that the unauthorized use of still frames of President Kennedy's assassination was fair use because public interest in providing information about the assassination outweighed the copyright owner's exclusive rights).

44. H.R. REP. NO. 94-1476, *supra* note 17, at 66.

45. *See* William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659, 1692 (1988) (arguing that neither *Sony* nor *Harper & Row* provided sufficient guidance to lower courts as to which of the four statutory factors predominates in any specific case); *see also* Manali Shah, *Fair Use and the Google Book Search Project: The Case for Creating Digital Libraries*, 15 COMMLAW CONSPECTUS 569, 583 (2007) ("Unfortunately, neither *Campbell* nor the fair use statute gives courts any guidance on the relative weight each factor should be awarded.").

46. *See* H.R. REP. NO. 94-1476, *supra* note 17, at 66.

47. *See* Fisher, *supra* note 45, at 1663 (citing Benny v. Loew's, Inc., 239 F.2d 532 (9th Cir. 1956), *aff'd per curiam by an equally divided Court sub nom.* ("Prior to 1982, the Court granted certiorari in only two cases implicating the doctrine, and in both instances and equal division in the Justices' votes prevented the issuance of an opinion."); *see e.g.,* Columbia Broadcast Sys. v. Loew's, Inc., 356 U.S. 40 (1958); *see also* Williams & Wilkins Co. v. United States, 487 F.2d 1345 (Ct. Cl. 1973), *aff'd per curiam by an equally divided Court*, 420 U.S. 376 (1975)).

48. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984).

49. *See id.* at 451, 454–55 (stating in dicta that "every commercial use of copyrighted material is presumptively an unfair exploitation of the monopoly privilege").

50. Harper & Row Publishers v. Nation Enterprises, 471 U.S. 539, 560 (1985).

51. *Id.* at 564–66.

52. *See* Stewart v. Abend, 495 U.S. 207, 237–38 (1990).

fashioned in *Folsom*—under the precept of parody.[53] First, the Court made it clear that the more "transformative" the purpose and character of the use, the less weight should be given other factors.[54] In making this proposition the Court relied on Judge Pierre Leval's seminal article, *Toward a Fair Use Standard*,[55] to define transformative use as a use that "adds something new, with a further purpose or different character altering the first with new expression, meaning, or message."[56] Also, the Court discounted the weigh of the commercial or non-profit character of the use in the first-factor analysis.[57] Second, the Court explained that the nature of the work must "recognize[] that some works are closer to the core of intended copyright protection" and thereby more deserving of protection.[58] Third, the Court determined that the amount and substantiality of the portion used turns on the persuasiveness of the infringer's justification for copying.[59] Finally, with respect to the effect of the use upon the potential market for or value of the copyrighted work, the Court found it necessary to analyze both the particular acts of infringement and "'whether unrestricted and widespread conduct of the sort engaged in by the defendant . . . would result in a substantially adverse impact on the potential market' for the original."[60] The Court dismissed the lower court's presumption of harm in the commercial context, rather stating that where "the second use is transformative, market substitution is at least less certain, and market harm may not be so readily inferred."[61] Overall, the Court's analysis in *Campbell* marked distinct change in fair use analysis, making it a "true multi-factor test in which factors two, three, and four would be assessed and weighed in

---

53. *Campbell v. Acuff-Rose Music, Inc*, 510 U.S. 569, 579 (1994) ("parody has an obvious claim to transformative value. . .").

54. *Id.* (explaining that transformative works "lie at the heart of the fair use" doctrine because such uses generally further the "goal of copyright, to promote science and the arts").

55. *See* Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105 (1990). Judge Leval sat on the U.S. District Court for the Southern District of New York from 1977 - 1993 when he was appointed to the U.S. Court of Appeals for the Second Circuit, assuming senior status in 2002. *Biographical Directory of Federal Judges*, FEDERAL JUDICIAL CENTER, https://www.fjc.gov/history/judges/leval-pierre-nelson (last visited Dec. 15, 2017).

56. *See* Leval, *supra* note 55, at 1111; *see also Campbell*, 510 U.S. at 579 (utilizing Leval's article and explaining that purpose and character of the use inquiry "asks, in other words, whether and to what extent the new work is transformative.").

57. *See Campbell*, 510 U.S. at 584 ("If, indeed, commerciality carried presumptive force against a finding of fairness, the presumption would swallow nearly all of the illustrative uses listed in the preamble paragraph of § 107 . . . since these activities 'are generally conducted for profit in this country.'"); *accord Harper & Row*, 471 U.S. at 592.

58. *Campbell*, 510 U.S. at 586.

59. *See Id.* at 586–87 (recognizing overlap of the first and third factor as the copier's justification is likely tied to the purpose and character of the use, as well as overlap of the third and fourth factors as the amount of the work taken may indicate the effect on the market of the original).

60. *Id.* at 590 (citing 3 M. Nimmer, Copyright § 13.05 (1984), p. 13–102.61).

61. *Id.* at 591 (qualifying *Sony's* presumption of unfairness in the case of commercial use).

line with the degree of transformativeness of the use, rather than the market centered presumptions."[62]

Lower courts, however, largely ignored the robust discussion in *Campbell* and the new framework—continuing to rely heavily on commerciality rather than transformativeness.[63] As a result, *Sony's* commercial use presumption continued to pervade fair use opinions through 2005.[64] Around that time, however, gradual shift began toward a focus on transformative use. For example, from 2006 to 2010, "85.5% of district court opinions and 93.75%, or all but one, of appellate opinions considered whether the defendant's use was transformative," corresponding to a ten percent increase in the application of transformative use in all relevant reported opinions.[65] Thus, as of 2010, the transformativeness of the purpose and character of the use is the predominate factor in the fair use analysis.[66]

The purpose and character of mass digitization is inherently transformative because such use serves two purposes, first, to create an exact copy of the entire work and, second, produce computer searchable text that enables novel uses of the original, such as digital preservation and text mining. The next section illuminates the transformativeness of mass digitization and explains how Google Books utilizes digitized works.

---

62. Netanel, *supra* note 29 at 722–23.

63. *See* Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions*, 1975–2005, 156 U. PA. L. REV. 549, 604–05 (2008) (reporting that, post–*Campbell*, 41.2% of district court and 18.6% of circuit court fair use opinions failed to refer to transformative use).

64. *Id.* at 618–20; *see, e.g.*, Elvis Presley Enters., Inc. v. Passport Video, 349 F. 3d 622, 630–31 (9th Cir. 2003) (confusingly mixed components of each *Harper & Row*, *Campbell*, and *Sony*:

> The last, and 'undoubtedly the single most important' of all the factors, is the effect the use will have on the potential market for and value of the copyrighted works. We must 'consider not only the extent of market harm caused by the particular actions of the alleged infringer, but also whether unrestricted and widespread conduct of the sort engaged in by the defendant . . . would result in a substantially adverse impact on the potential market for the original.' The more transformative the new work, the less likely the new work's use of copyrighted materials will affect the market for the materials. Finally, if the purpose of the new work is commercial in nature, 'the likelihood [of market harm] may be presumed.').

65. Netanel, *supra* note 29, at 736–37 (noting that opinions that embraced Campbell's analysis were included even if the opinion did not explicitly use the word "transformative").

66. *See* Netanel, *supra* note 29, at 745; *see also* discussion *infra* Section III.B.3 (explaining that under the transformative use paradigm, first factor transformativeness permeates through the other three factors).

III. BACKGROUND OF MASS DIGITIZATION AND GOOGLE BOOKS

A. *Mass Digitization*

1. Defining Mass Digitization

Mass digitization is the conversion of copyright works—by the hundreds or thousands—into a computer readable format.[67] Present mass digitization methods often employ scanning stations comprising a high definition cameras fixed above a surface configured to hold an open book.[68] First, operators use those stations to capture an image of each page of the book. Then, the image is processed by optical character recognition ("OCR") software to transform the image into computer readable text (think conversion of a JPG file to a DOC file).[69] For example, Google's patented software identifies the spine and page-curvature of bound texts (imagine how pages of a book are curved when the book is open) with an infrared camera and combines that three-dimensional data with the two-dimensional image captured by an optical camera such that the resulting image text is readily identifiable by OCR software.[70] Finally, both the image and text files may be transmitted from the scanning station to a server for permanent storage.[71] In this way, mass digitization is like an industrial production-line—operators churn-out copy after copy, book after book, only limited by how fast they can turn the page and snap the picture.

Early digitization projects began as soon as computer technology would allow.[72] For example, Project Gutenberg was founded in 1971, however, its beginnings were small in scale relative to current digitization projects.[73] Yet discussion of exactly what constitutes mass digitization was absent from our case law until recently.[74] Even though no early cases arose, the publishing industry was keenly aware of

---

67. *See* MAURIZIO BORGHI & STAVROULA KARAPAPA, COPYRIGHT AND MASS DIGITIZATION: A CROSS-JURISDICATIONAL PERSPECTIVE 1-2 (Oxford Univ. Press, 2013).

68. *See* Ben Lewis, GOOGLE AND THE WORLD BRAIN (Polar Star Films and BLTV Prod. 2014) (describing the Google scanning station); *see also* U.S. Patent No. 7,508,978 (filed Sept. 13, 2006) [hereinafter "'978 Patent"] (original assignee, Google, Inc.).

69. *See* '978 Patent, *supra* note 68.

70. *See* '978 Patent, *supra* note 68.

71. *See* '978 Patent, *supra* note 68.

72. *See* Gil Press, *A Very Short History of Digitization,* FORBES (Dec. 27, 2015) https://www.forbes.com/sites/gilpress/2015/12/27/a-very-short-history-of-digitization/.

73. *See* Michael Hart, *The History and Philosophy of Project Gutenberg by Michael Hart,* PROJECT GUTTENBERG (April 8, 2010) https://www.gutenberg.org/wiki/Gutenberg:The_History_and_Philosophy_of_Project_Gutenberg_by_Michael_Hart.

74. *See* Golan v. Holder, 132 S. Ct. 873, 894 (2012) (discussing treatment of orphan-works in the context of the Berne Convention and legislation proposed by the Copyright Office).

potential issues.[75] For example, in a 2000 interview with the New York Law Journal, Kay Murray, the general counsel of the Authors Guild, alluded to the risk of mass digitization of books analogizing to the "recording industry's [then] crisis battling" peer to peer sharing of songs.[76] Murray predicted that by 2005 mass digitization in the form of print-on-demand books and e-books would merely augment the supply of traditional bound texts.[77] While Murray correctly forecast the presence of mass digitization in the publishing industry, she may have misjudged the magnitude and scope of mass digitization projects throughout the U.S. and abroad. By 2011, the prevalence and breadth of mass digitization compelled the United States Copyright Office to issue a report for preliminary analysis and discussion of issue of mass digitization under current copyright law.[78]

2. Exemplary Digitization Projects

The appreciation of a world library dates back to around 300 B.C. when the Library of Alexandria was becoming the foremost repository of written text in the ancient world.[79] In the twentieth century, computers and the Internet gave new life to the age-old dream of a common repository of knowledge.[80] The first digitization project was born on July 4, 1971, when Michael Hart transcribed the Declaration of Independence on to the Xerox Sigma V mainframe at the University of Illinois and posted the digital copy to a fifteen-computer network.[81] Twenty-five years later, the non-profit Internet Archive began creating a publicly accessible Internet repository of digital content including texts, sound recordings, films, and webpages.[82] Internet Archive hosts several projects including the Wayback Machine, archiving over 439 billion "dead" Web pages,[83] and Open Library, an open source project aimed to create "[o]ne web page for every book ever published."[84]

---

75. *See* Alan J. Hartnick, *E-Book Rights v. Traditional Publishing: Q&A with the Authors Guild*, N.Y.L.J., Oct. 20, 2000, at 24 (recounting an interview with Kay Murray, the general counsel of Authors Guild).

76. *Id.*

77. *See Id.*

78. *See* U.S. COPYRIGHT OFFICE, LEGAL ISSUES IN MASS DIGITIZATION 4 (2011).

79. *See* LIONEL CASSON, LIBRARIES IN THE ANCIENT WORLD 31–33 (Yale Univ. Press, 2001) (describing the Library of Alexandria as a center of learning facilitated by the copying of texts for research and scholarship).

80. *See* Hart, *supra* note 73.

81. *See* Hart, *supra* note 73. (explaining that Hart believed information storage would become a valuable asset); *see also* DANIEL JAMES, CRAFTING DIGITAL MEDIA: AUDACITY, BLENDER, DRUPAL, GIMP, SCRIBUS, AND OTHER OPEN SOURCE TOOLS 189 (Frank Pohlman et al. eds., 2009).

82. *See* INTERNET ARCHIVE, *About the Internet Archive*, http://www.archive.org/about/ (last visited Oct. 25, 2015).

83. *See id.*

84. Open Library, *About Us*, INTERNET ARCHIVE (Sept. 9, 2013), https://openlibrary.org/about (last visited Oct. 25, 2015).

In addition to private ventures, federal and state entities also founded digitization projects with varying concentrations.[85] For example, in 2000, the Library of Congress initiated the National Digital Information Infrastructure and Preservation Program (NDIIPP).[86] The NDIIPP partners with hundreds of domestic and foreign organizations for the preservation of "at-risk" digital collections and to "build a distributed digital preservation infrastructure."[87] Similarly, the National Archives and Records Administration has undertaken efforts in recent years to digitize records and to increase public access to important archive materials.[88] On the state level, California led early state-implemented digitization programs with Calisphere[89] and the California Digital Library.[90]

Across the Atlantic, Europeana is a major digitization project that aims to preserve European heritage and advocates for "cultural innovation" through creations of new standards and copyright reform to embrace new technology.[91] As of 2015, Europana collected over 23 million images and over 15 million texts, with Germany, the Netherlands, France, Spain, and Sweden contributing the most content, in that order.[92]

On the international stage, government sponsored digitization projects include, for example, the World Digital Library, which is led by

---

85. *See* Federal Agencies Digital Guidelines Initiative, *About,* DIGITIZATIONGUIDLINES.GOV, http://www.digitizationguidelines.gov/about/ (last visited Nov. 8, 2017).

86. *See* LIBRARY OF CONGRESS, PRESERVING OUR DIGITAL HERITAGE: THE NATIONAL DIGITAL INFORMATION INFRASTRUCTURE AND PERSEVERATION PROGRAM 2010 REPORT 1–7 (2011), http://www.digitalpreservation.gov/multimedia/documents/NDIIPP2010Report_Post.pdf.

87. *See* Digital Preservation, *About,* LIBRARY OF CONGRESS, http://www.digitalpreservation.gov/ about/ (last visited Oct. 25, 2015) ("NDIIPP is based on an understanding that digital stewardship on a national scale depends on public and private communities working together").

88. *See Strategy for Digitizing Archival Materials,* NATIONAL ARCHIVES AND RECORDS ADMINISTRATION, https://www.archives.gov/digitization/strategy.html (last reviewed Aug. 15, 2016).

89. *See* Calisphere Beta, *About Calisphere,* UNIVERSITY OF CALIFORNIA, http://calisphere.cdlib.org/ about/ (last visited Oct. 25, 2015) (providing access to "over 750,000 photographs, documents, letters, artwork[s], diaries, oral histories, films, advertisements, musical recordings, and [other content]").

90. *See* California Digital Library, *Mass Digitization Overview*, THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (2015) http://www.cdlib.org/services/collections/massdig/ (last visited Oct. 25, 2015) (explaining that the California Digital Library provides project planning, coordination, and technical leadership for the mass digitization of collections at University of California system libraries in conjunction with Google and Internet Archive).

91. Europena Collections, *Our Vision*, EUROPEANA, http://www.europeana.eu/portal/ (last visited Oct. 25, 2015).

92. Imogen Greenhalgh, *Facts & Figures*, EUROPEANA, http://pro.europeana.eu/about-us/factsfigures (last visited Oct. 25, 2015); *See e.g. About us,* BRITISH DIGITAL LIBRARY http://www.bl.uk/aboutus/stratpolprog/digi/index.html (last visited Dec. 15, 2017); *About,* NATIONAL LIBRARY OF SWEDEN, http://www.kb.se/english/about/digitazation/ (last visited Dec. 15, 2017); and *Company Profile,* CHINA FOUNDER GROUP, http://www.founder.com/about/intro.html (last visited Dec. 15, 2017) (existing as foreign digitization projects) .

the U.S. Library of Congress with support from the United Nations Educational Cultural and Scientific Organization.[93] Private intentional organizations play a pivotal role as well like, for example, the Open Content Alliance that collaborates with "cultural, technological, nonprofit, and governmental organizations from around the world [to] build a permanent archive of multilingual digitized text and multimedia material,"[94] and the Universal Digital Archive who's mission is to digitally preserve and provide free global access to "all the significant literary, artistic, and scientific works of mankind" from across the globe "for our education, study, and appreciation and that of all our future generations."[95]

As of 2015, about 3.2 billion people—nearly half of the world population—has access to the internet.[96] As such, there is a tremendous amount of knowledge that can be provided to the general public through digitization projects like the ones described above. The Internet also provides a pristine platform for marketing and, with ever-increasing use, holds enormous economic potential. Inspired by this potential to spread knowledge throughout the world, or perhaps by the possibility of capitalizing an untapped online market, Google founded a digitization project that dwarfed all others before it—Google Books.[97]

B. Google Books

1. Background of Google Books

In its infancy, Google co-founder Larry Page first proposed digitizing all of the world's books.[98] Only four years later, in 2002, a small team at Google launched a secret books project.[99] By 2004, that

---

93. *See* World Digital Library, *About the World Digital Library: Background*, LIBRARY OF CONGRESS, http://www.wdl.org/en/ background/ (last visited Oct. 25, 2015) (the "principal objectives of the [World Digital Library] are to: Promote international and intercultural understanding; Expand the volume and variety of cultural content on the Internet; Provide resources for educators, scholars, and general audiences; [and] Build capacity in partner institutions to narrow the digital divide within and between countries.").

94. Open Content Alliance, *About*, INTERNET ARCHIVE, https://archive.org/details/opencontentalliance&tab=abouthttp://www.opencontentalliance.org/ab(last visited Oct. 25, 2015) (administered by the Internet Archive).

95. *See generally* Byron Spice, *Online Library Gives Readers Access to 1.5 Million Books*, CARNEGIE MELLON U. (Nov. 11, 2007), http://www.cs.cmu.edu/news/online-library-gives-readers-access-15-million-booksinternational-project-makes-complete-texts.

96. *Internet Used by 3.2 Billion People in 2015*, BBC NEWS (May 26, 2015) http://www.bbc.com/news/technology-32884867.

97. *See Google Books History*, Google, https://www.google.com/intl/en/googlebooks/about/history.html (last visited Dec. 15, 2017) [hereinafter "*Google History*"].

98. Serge Brin, *A Library to Last Forever*, N.Y. TIMES, at A31 (Oct. 8, 2009) http://www.nytimes.com/2009/10/09/opinion/09brin.html?mcubz=3(recountingLarryPage. (recounting Larry Page's vision that later became Google Books).

99. *See Google History*, supra note 97.

secret project had transformed into the "Google Print" Library Project, which aimed to increase public access to books and prevent the loss of hard copies.[100] Google Print entered bi-lateral partnerships agreements with libraries at Harvard, the University of Michigan, the New York Public Library, Oxford and Stanford.[101] Under these agreements, each partner libraries selected books for Google to digitize. [102] In return, the libraries received digital images and machine-readable versions of the books for their electronic collections.[103] One year after its public announcement, Google Print was renamed Google Books. [104]

The Google Books webpage allows users to search Google's corpus of digitized books by simply entering a search query.[105] But unlike searching a library index, users can search for any text string, be it the title, the author, or even a phrase in the middle of the book.[106] The search results are similar to that of a standard Google search, including clearly marked advertisements that may appear in the first line of results.[107] For example, in a fraction of a second, a search for "fair use" returns 325,000 results and one advertisement for LegalZoom.[108] The first result in this

---

100. *Id.; see also* Brin, *supra* note 98, at A31. In 1998, both Brin and Page, then graduate students, witnessed the loss of books when the Stanford library flooded, destroying tens of thousands of texts, ironically while working on the book-indexing algorithm Backrub for the Stanford Digital Libraries Technology Project. *Id.*

101. *See Google History, supra* note 97.

102. *See Google History, supra* note 97.

103. *See, e.g., Cooperative Agreement between Google & the University of Michigan*, UNIVERSITY OF MICHIGAN ANN ARBOR, http://www.lib.umich.edu/files/services/mdp/um-google-cooperative- agreement.pdf (last visited Sept. 11, 2017). (outlining the Cooperative Agreement between the University of Michigan ("U of M") and Google providing that:

> "'U of M Digital Copy' means a digital copy transferred by Google to U of M of Available Content that it Digitized by Google. . . . Google agrees to provide to U of M a copy of all Digitized Selected Content that has been "Successfully Processed" within thirty (30) days after the Selected Content is Digitized . . . . Unless otherwise agreed by the Parties in writing, the U of M Digital Copy will consist of a set of image and OCR [optical character recognition] files and associated information indicating at a minimum (1) bibliographic information consisting of the title and author of each Digitized work, (2) which image files correspond to that Digitized work, and (3) the logical order of those image files.").

104. See *Google History*, *supra* note 97.

105. See GOOGLE BOOKS http://books.google.com (last visited Dec. 11, 2015).

106. See How to Use Google Books, GOOGLE, https://support.google.com/websearch/answer/43729?visit_id=1-636406185253632296-3161204258&hl=en&rd=1 (last visited Sept. 11, 2017).

107. *See* Press Release, Google, Google Launches Self-Service Advertising Program (Oct. 23, 2000), googlepress.blogspot.com/2000/10/google- launches-self-service.html; *see also Where Your Ads Can Appear*, GOOGLE, https://support.google.com/adwords/answer/1704373?hl=en&ref_topic=3121763 (last visited Sept. 11, 2017).(explaining that the paid advertisements are configured with Google's AdSense program and clearly marked by the word "Ad" in a highlighted box).

108. Search for "fair use" (in quotes) performed on October 30, 2015, from the author's personal computer using the Google Chrome web browser. The same search conducted on December 11, 2015, did not include the Legal Zoom advertisement, but returned 331,000 results. (Search on file with Author).

query is *Reclaiming Fair Use: How to Put Balance Back in Copyright* by Patricia Aufderheide and Peter Jaszi.[109] Following the link in the search results opens a new page that displays information about the book, including 87 "snippet views" of where the search term "fair use" occurs in the book along with about one-eight of a page of text to put the search term in context.[110] In all, the various occurrences of "fair use" in *Reclaiming Fair Use: How to Put Balance Back in Copyright* can be viewed in 87 boxes each displaying about seven lines of text.[111]

Generally, Google provides four view options that copyright holders may request: (1) Full view—all books out of copyright, or by permission of copyright holder, can be viewed in full and, if the book is in the public domain, a full PDF copy may be downloaded and saved; (2) Limited preview—by permission of the copyright holder a limited number of pages are displayed in full; (3) Snippet view—shows information about the book and a few snippets or blocks of about seven lines of text to put the search term in context; and (4) No preview—only basic information about the book is displayed, like a traditional library card catalogue.[112] Also, links are provided to purchase the book direct from the publisher or through various retailers, obtain an electronic version of the book, find the book at a local library, and view more information in the "About the Book" link (which provides, for example, bibliographical information, reviews of the book, and common terms and phrases in the book).[113]

In addition to Google Books search, Google created a separate website that offers a function called Google Book Ngram Viewer.[114] In natural language processing (i.e., computational linguistics) *n*-gram refers to the contiguous sequence of *n* words in a text sequence.[115] Among other things, *n*-gram analysis can be used to attribute authorship (e.g., in unsigned judicial opinions) or analyze linguistic trends over time.[116] Ngram Viewer, for example, can graphically display occurrences

---

109. *Id.*

110. *Id.* (follow link to the book) (displaying no advertisements, but includes links to purchase the book).

111. *Id.*

112. *What You'll See When You Search on Google Books*, GOOGLE BOOKS, https://books.google.com/googlebooks/library/screenshots.html (last visited Dec. 15, 2017).

113. *See* GOOGLE BOOKS, *supra* note 105 (search and select any book and follow the "About this book" link).

114. *See Ngram Viewer*, GOOGLE BOOKS https://books.google.com/ngrams (last visited Dec. 11, 2015).

115. *See* William Li et al., *Using Algorithmic Attribution Techniques to Determine Authorship in Unsigned Judicial Opinions*, 16 STAN. TECH. L. REV. 503, 517 (2013).

116. *See id.* (examining Sebelius to gather n-gram data used to attribute authorship of every unsigned per curium opinion of the Roberts Court); *see also* GOOGLE BOOKS NGRAM VIEWER, *supra* note 163.

of the term "fair use" in Google's entire corpus of digitized texts,[117] and even show that the first occurrence of "fair use" in the context of copyright, at least in Google's corpus, occurred in the 1810 English opinion *Watkins v. Aikin*,[118] which was cited by Justice Story in *Folsom*.[119]

Since its launch 2002, Google Books has become a powerful search tool and instrument of pioneering linguistic research.[120] As early as 2005, Google had even established international partnerships in Austria, Belgium, France, Germany, Italy, the Netherlands, Spain and Switzerland.[121] But in that same year, Google Books encountered a significant obstacle— copyright infringement.

2. The Google Books Dispute—The Southern District of New York

On September 20, 2005, Authors Guild, Inc. filed a putative class action on behalf of several of its authors asserting willful copyright infringement by Google through the Google Books project.[122] Early in the suit, Google asserted a novel application of the fair use defense—that digitizing in-copyright works and displaying snippets of those works was a fair use.[123] In the face of this legal uncertainty, the parties entered into a first proposed settlement on October 28, 2008.[124] "This historic settlement [was] a win for everyone."[125] Google would receive the right to pursue its mass digitization project, sell subscriptions to its database,

---

117. *See* Ngram Viewer search of "fair use", GOOGLE BOOKS, https://books.google.com/ngrams/graph?content=%22fair+use%22&year_start=1800&year_end=2000&corpus=15&smoothing=3&share=&direct_url=t1%3B%2C%22%20fair%20use%20%22%3B%2Cc0 (last visited Sept. 25, 2017)[hereinafter Fair Use Search]

118. *See* Fair Use Search, supra note 117; see also Francis Vesey, John Beames, and John Scott Earl of Eldon, REPORT OF CASES ARGUED AND DETERMINED IN THE HIGH COURT OF CHANCERY DURING THE TIME OF LORD CHANCELLOR ELDON 420 (Edward D. Ingram, first American ed., 1822) and Wilkins v. Aikin [1810] 34 Eng. Rep. 163 (KB) ("Action directed, to try, whether a work on architecture was original; with a fair use of another work, by quotation and compilation . . . .")

119. *See* Folsom v. Marsh, 9 F. Cas. 342, 348 (D. Mass. 1841)(No. 4901).

120. *See* Google History, supra note 97.

121. *See* Google History, supra note 97.

122. *See* Google History, supra note 97; see also Authors Guild, 954 F. Supp. 2d at 284. Plaintiffs Jim Burton, author of Ball Four; Betty Miles, author of the Trouble with Thirteen; and Joseph Goulden, author of The Superlawyers: The Small and Powerful World of the Great Washington Law Firms. Id. at 285. FFive publishers and the Association of American Publishers brought a second suit. See Complaint paras. 1–4, McGraw-Hill Cos. v. Google, Inc., No. 05-CV-8881 (S.D.N.Y. Oct. 19, 2005). At this time, Authors Guild was pursuing a similar suit against the HathiTrust Digital Library, a partnership of California libraries, which compiled into a common repository books digitized under the Google Books project. See HathiTrust, 755 F.3d at 90–93, 105 (holding that HathiTrust's creation of "full-text searchable database of copyrighted works and to provide those works in formats accessible to those with disabilities" was a fair use).

123. *See* Authors Guild, 770 F. Supp. 2d at 671 (Chin, J., sitting by designation).

124. *See id.* at 671.

125. Press Release, *Authors, Publishers, and Google Reach Landmark Settlement: Copyright Accord Would Make Millions More Books Available Online*, AUTHORS GUILD (Oct. 28, 2008), https://www.authorsguild.org/wp-content/uploads/2008/10/press_release_final_102808.pdf.

sell individual books, place advertisements on online book pages, and make other commercial use of the books.[126] In exchange, Google would to pay approximately $125 million to be divided among authors who's works were digitized before the settlement's 'opt-out' deadline of May 5, 2009, to fund the launch of a Books Rights Registry to manage future licensing and revenue distribution, and to pay attorney's fees and other costs.[127]

On November 17, 2008, Judge John E. Sprizzo preliminarily approved the first settlement agreement, which "triggered hundreds of objections."[128] In an attempt to quell those objections, Google and the Authors Guild reopened negotiations and, on November 13, 2009, filed a motion for final approval of an amended settlement agreement.[129] Three months after preliminary approval of the amended settlement, Judge Denny Chin held a fairness hearing.[130] The amended settlement, however, set-off a firestorm of opposition from over 500 parties including the Department of Justice, the German and French governments, and a slew of private parties.[131] This second round of objections raised concerns that the "forward-looking" amended settlement was improper because the agreement, among other things, circumvented Congress's constitutional authority over copyright and produced anticompetitive effects by granting a private entity, Google, *de facto* compulsory licenses that narrowed the exclusive rights of copyright holders.[132] Ultimately rejecting the amend settlement on March 22, 2011, Judge Chin concluded that the agreement was "not fair, adequate, and reasonable," but that converting the opt-out framework to an opt-in form would address many of his concerns.[133]

Litigation resumed after the failure to both settlements, that is, until the fall of 2013 when Judge Chin (now appointed to the Second Circuit but sitting by designation in the Southern District of New

---

126. *See id.*

127. *Id.*; *see also* Motion for Preliminary Settlement Approval at 19–21, *Authors Guild* ,770 F. Supp. 2d 666, https://www.authorsguild.org/wp-content/uploads/2008/10/Settlement-Agreement.pdf (Oct. 28, 2008).

128. *Authors Guild*, 770 F. Supp. 2d at 671.

129. *Id.*

130. *See id; see also* Transcript of Record, Authors Guild, Inc. v. Google Inc., 770 F. Supp. 2d. 666 (S.D.N.Y. 2011) (No. 05-CV-8136) (A transcript of the February 18, 2010, fairness hearing on the amended settlement agreement can be found at http://www.thepublicindex.org/wp-content/uploads/sites/19/docs/case _order/fairness-hearing-transcript.pdf) [hereinafter "Hearing Transcript"].

131. *See Authors Guild*, 770 F. Supp. 2d at 671–73; *See also* Hearing Transcript, *supra* note 130, at 69–74 (Dr. Irene Pakuscher, Head of Copyright and Publishing at the Federal Ministry of Justice, representing the Federal Republic of Germany, arguing that the amended settlement agreement infringes the rights of foreign authors who have registered works with the U.S. Copyright Office).

132. *See Authors Guild*, 770 F. Supp. 2d at 673–74.

133. *Id.* at 686.

York)[134] granted summary judgment with respect to Google's fair use defense.[135] By this time, Google had digitized over twenty million books, many of which were still under copyright and scanned without permission from the copyright holders.[136]

Applying the four statutory factors under section 107, Judge Chin first found that the purpose and character of the use was "highly transformative" because Google's optical character recognition software enabled a scanned image to be converted into a "comprehensive word index that helps readers, scholars, researchers, and others find books."[137] Further, he found the use of Google's snippet view to be transformative in that it only "help[s] users locate books and determine whether they may be of interest."[138] Moreover, Judge Chin described the conversion of book text into data, such as *n*-gram data, as transformative because it opens the door for new areas of research in text mining.[139] Then, in accordance with the post-*Campbell* trend,[140] Judge Chin reasoned that little weight should be given to Google's potential commercialization—through capturing a greater user base—of the copyrighted works .[141] Rather, Judge Chin gave more weight to the fact that Google did not engage in direct commercialization of the copyrighted texts through, for example, sale of digital copies or snippets, or displaying advertisements on the webpages alongside snippets.[142] Through these considerations, the court concluded that the transformative purpose and character of Google's use favored a finding of fair use.[143]After finding transformativeness under the first fair use factor, Judge Chin made quick work of the remaining factors. As to the second factor—the nature of the copyrighted work—Judge Chin found in favor of fair use because the books at issue were available to the public and mostly nonfiction titles.[144] The third factor—the amount and

---

134. *See* Press Release, Office of the Press Secretary, *President Obama Nominates Judge Denny Chin for United States Court of Appeals for the Second Circuit, Judge O. Rogeriee Thompson for United States Court of Appeals for the First Circuit*, THE WHITE HOUSE (Oct. 9, 2009), https://www.whitehouse.gov/the-press-office/president-obama-nominates-judge-denny-chin-united-states-court-appeals-second-circu (nominating Judge Chin to the Second Circuit by President Barak Obama and confirmed on April 22, 2010).

135. *See* Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282, 284 (S.D.N.Y. 2013) *aff'd sub nom.* 804 F.3d 202 (2d Cir. 2015).

136. *See Id.*

137. *Id.* at 291

138. *Id.* (comparing snippet view to the transformative thumbnail images (citing Kelly v. Arriba Soft Corp., 336 F.3d 811 (9th Cir. 2003)

139. *See id.*

140. *See* discussion *supra* Section II.B.

141. *See Authors Guild*, 954 F. Supp. 2d at 291–92.

142. *See id.*

143. *See id.* at 292.

144. *See id.* (stating fiction titles may deserve more protection (citing *Stewart*, 495 U.S. at 237)).

substantiality of the portion used— only "slightly against a finding of fair use," Judge Chin reasoned, because the entire work by necessity would need to be copied to achieve the purpose of digitization.[145] Regarding the fourth factor—the effect of use upon potential market or value—Judge Chin found weighed strongly in favor of fair use because Google's snippets could not serve as a replacement of the original works, even by creative searching of the snippets, since certain snippet views and pages were blacklisted.[146]

Based on an evaluation of the four statutory factors and considerations of public benefits—including advancing copyrights constitutional underpinnings, providing value to academia, preserving books, and facilitating access to books—Judge Chin concluded that Google's mass digitization project was fair use.[147]

3. The Google Books Dispute—The Second Circuit

On appeal in the Second Circuit, Authors Guild argued five points.[148] First, that Google's digitization of in-copyright books and display of snippets on its website are not transformative under *Campbell*.[149] Second, that Google's commercial intentions preclude a finding of fair use, specifically because Google utilizes digitized books to expand its dominance in the Internet search engine market.[150] Third, that Google's digitization and display of snippets infringe the copyright holder's derivative rights.[151] Fourth, that Google's use expose the copyright holder's to a significant risk of infringement if the digitized copies were ever lost to hackers or the like.[152] Finally, Google's distribution of digitized copies to its member-libraries is not transformative and exposes copyright holders to loss of sales to libraries.[153]

---

145. *Id.* ("copying of the entire work is sometime necessary to make a fair use of a [work]." (*quoting* Bill Graham Archives v. Dorling Kindersley Limited, 448 F.3d 605, 613 (2d Cir. 2006)).

146. *See id.* at 292–93 (describing that one of each eight snippets per page and one in every ten pages are not displayed to any user in any search).

147. *See id.* at 293–94 (dismissing Authors Guild's theory of secondary liability).

148. *See* Authors Guild, Inc. v. Google, Inc., 804 F.3d 202, 207 (2d Cir. 2015). Authors Guild was also plaintiff and appellate, however, the Second Circuit determined in *Authors Guild, Inc. v. HathiTrust* that Authors Guild lacked standing to sue for copyright infringement on behalf of its members. *Id.* at n.1; *see also* Authors Guild, Inc. v. HathiTrust, 755 F.3d 87, 94 (2d Cir. 2014).

149. *See Authors Guild*, 804 F.3d at 207.

150. *See id.*

151. *See id.*

152. *See id.*

153. *See id.*

Writing for an unanimous panel, Judge Pierre N. Leval[154] framed his fair use analysis to assess "the crucial question: how to define the boundary limit of the original author's exclusive rights in order to best serve the overall objectives of the copyright law to expand public learning while protecting the incentives of authors to create for the public good."[155]

With respect to the first statutory factor—the purpose of character of the use—Judge Leval suggested that *Campbell's* transformative use was a threshold question rooted in constitutional foundation of copyright law.[156] But Leval also warned against extending transformative use to "any and all changes made to an author's original text" or confusing "transformed" as applied in the statutory definition of "derivative work" with transformativeness under the fair use analysis.[157] Thus, Leval defined transformative use as a use "that communicates something new and different from the original or expands its utility."[158]

Applying this definition to the Google Books search function, the court, relying on its own decision in *HathiTrust*, found that digitization of entire books was "highly transformative" because Google's search feature serves a different function than the original and merely provides information about the books.[159] Then, expanding on *HathiTrust*, the court went on to explain that display of snippets was also transformative because snippet view adds important functionality to the transformative search function by placing search terms in context while not making available the expression of the original.[160] Finally, the court determined that Google's possible commercial benefit from digitization of books should not outweigh the "highly convincing transformative purpose" of Google Books because Google provides no "significant substitutive competition with the originals."[161]

---

154. *Accord Campbell*, 510 U.S. at 579 and Leval, *supra* note 55. (Judge Pierre Leval's seminal work—arguing that transformative use should be a critical factor in fair use analysis—was pivotal in *Campbell*.)

155. *Authors Guild*, 804 F.3d at 213–14.

156. *See id.* at 214

157. *See id.*; *see also* 17 U.S.C. § 101 (2015) (defining a "derivative work" as "a work based upon one or more preexisting works ... or any other form in which a work may be recast, transformed, or adapted.").

158. *Authors Guild*, 804 F.3d at 214.

159. *See id.* 216–17.

160. *See id.* at 217–18.

161. *Id.* at 218–19 (recognizing that *Campbell* culled *Sony's* presumption of unfair commercial use). The court ignored the statutory text, which states that consideration of the purpose and character must include "whether such use is of a commercial nature or is for nonprofit educational purposes," § 107, though the court did note that commercial use should not be preclude fair use, nonprofit educational use should not categorically be accepted as fair use. *Authors Guild*, 804 F.3d at n.20.

The highly transformative purpose and character of the Google Books search function and snippet views was the most significant factor for the court.[162] This finding allowed the court to quickly dismiss the remaining three statutory factors as having no substantial influence to sway the court in the opposite direction.[163]

The court noted that the second factor—the nature of the copyrighted work—"rarely plays a significant role in the determination of a fair use dispute" but, even so, the prior finding of highly transformative use indicated that the nature of the copying was not to duplicate the original's expression but rather to utilize the original's discrete words for an entirely different purpose via the search function and snippets.[164]

Regarding the third factor—the amount and substantiality of the portion used—the court found that copying of the works in their entirety was justified because in order to make the search function reliable "not only is copying of the totality of the original reasonably appropriate to Google's transformative purpose, it is literally necessary to achieve that purpose."[165] As to the snippet view, the court found that display of snippets was fair because only "the amount and substantiality of *what is thereby made accessible* to the public," not that "'used *in making a copy*,'" is relevant and those publicly available snippets could not serve as a "significantly competing substitute" given blacklisting of one out of eight snippets from each page and one page out of every ten pages.[166]

The court found that the fourth factor—the effect on the market for the original work—favored fair use because snippets, as presently configured on Google Books with blacklisting of one out of eight snippets from each page and one out of every ten pages, could not revealed a "sufficiently significant portion" of the work as to constitute a "significant competing substitute."[167]

---

162. *See id.* at 215.

163. *See id.* at 230.

164. *Id.* at 220. Dismissing the dictum in *Harper & Row*, which called for greater protection of fictional works, the court explained:

> To the extent that the 'nature' of the original copyrighted work necessarily combines with the 'purpose and character' of the secondary work to permit assessment of whether the secondary work uses the original in a 'transformative' manner, as the term is used in *Campbell*, the second factor favors fair use not because Plaintiffs' works are factual, but because the secondary use transformatively provides valuable information about the original, rather than replicating protected expression in a manner that provides a meaningful substitute for the original.

Id.; see also Harper & Row, 471 U.S. at 563.

165. *Authors Guild*, 804 F.3d at 221.

166. *Id.* at 222 (emphasis in the original).

167. *Id.* at 223 (recognizing that even a highly transformative use may be unfair if copying "results in widespread revelation of sufficiently significant portions of the original as to make available a significantly competing substitute"). "Snippet view, at best and after a large

"[C]onsidering the four fair use factors in light of the goals of copyright," the court held that Google's copying in full texts for the purpose of providing the public with a searchable collection of books, wherein the results of the search may display one or more snippets of the actual text, is a fair use and, thereby, not an infringement of the owners' copyright.[168]

The court also rejected the plaintiffs' other claims. First, as to infringement of the exclusive right to produce derivative works, the court held that the type of information provided by Google's search or snippet view were outside the scope of the owners' copyright.[169] The fact that the opinion effectively collapsed the potential market for licensing copyrights for digitization projects was of no consequence to the court.[170] Second, regarding the risk of loss of data by Google to hackers, the court found such risk a real concern.[171] With no evidence and an exceedingly small probability, however, the Plaintiffs' failed to carry their burden.[172] Finally, the court found that Google's providing digital copies to its library partners fell within fair use.[173] The court even expanded this point stating that "[i]f the library had created its own digital copy to enable its provision of fair use digital searches, the making of the digital copy would not have been infringement. Nor does it become an infringement because, instead of making its own digital copy, the library contracted with Google that Google would use its expertise and resources to make the digital conversion for the library's benefit."[174]

In all, the Second Circuit endorsed wholesale digitization of copyrighted books for the purpose of enabling searches of the content of those books, so long as searches return only small segments of the full text such that the expressive heart of the work is adequately protected

---

commitment of manpower, produces discontinuous tiny fragments, amounting in the aggregate to no more than 16% of a book. This does not threaten the rights holders with any significant harm to the value of their copyrights or diminish their harvest of copyright revenue." *Id.* at 224.

168. *Id.* at 225.

169. *See id.* ("The copyright resulting form the Plaintiffs' authorship of their works does not include an exclusive right to furnish the kind of information about the works that Google's programs provide to the public. For substantially the same reasons, the copyright that protects Plaintiffs' works does not include an exclusive derivative right to supply such information through query of a digitized copy.").

170. *See id.* at 226 (explaining that potential licensing arrangements like that provided in the earlier rejected settlement agreements "have no bearing on Google's present programs, which, in an non-infringing manner, allow the public to obtain limited data about the contents of the book, without allowing any substantial reading of its text.").

171. *See id.* at 227–28.

172. *See id.* at 228.

173. *See id.*

174. *Id.* at 229.

from unauthorized pubic dissemination of a significant market substitute.[175]

IV. ANALYZING THE APPLICATION OF FAIR USE IN THE CONTEXT OF MASS DIGITIZATION

The doctrine of fair use can be tailored to suit mass digitization of books under certain circumstances, for example, in niche digitization projects or the specific case of Google Books. In applying fair use to Google's mass digitization, however, the fair use doctrine may actually impede a great public benefit that could be derived in guiding Google's mass digitization project toward establishment of a freely accessible global digital library. Applying fair use to mass digitization results in these impediments because the doctrine was fashioned to carve out specific exceptions to the exclusive rights in favor of progressing creativity and innovation, but not to shelter fundamental shifts in use of copyrighted works. Mass digitization ushered in a fundamental shift in the way the public will use books, so this type of rapid technological change is likely beyond the scope that considered Congress some forty years ago.[176] Thus, because the tide of social progress embodied in mass digitization only swells with the surge of immense capital and technological resources, such as from Google and other corporate giants, the public benefit is not promoted when fair use stems the tide by constraining use of digitize books to bibliographies, *n*-grams, snippet views and the like. The better course is to establish a statutory framework whereby commercial entities are guided to facilitate mass digitization projects while simultaneously compensating authors and providing the most benefit to the public via widespread access to all digitized literature.

A. Fair Use May Work For Niche Digitization Projects

Certain niche or highly constrained digitization projects may be conducted in accordance with, and respect for, the principles of copyright law. For example, a project that digitizes only books that are in the public domain or with the permission of copyright holders will not violate copyright law.[177] Most digitization projects likely abide by these constrains for two reasons. The first is that limited resources, either total allocable resources in physical copying of texts or fiscal resources in seeking copyright holder permissions, may prohibit

---

175. *See id.* at 229–230.

176. *See* H.R. REP. NO. 94-1476, *supra* note 17, at 66.

177. *See, e.g., No Cost or Freedom?,* PROJECT GUTENBERG (Aug. 15, 2014), http://www.gutenberg.org/wiki/Gutenberg:No_Cost_or_Freedom%3F (explaining that most of the projects eBooks are in the public domain).

digitization of larger collections. For example, Project Gutenberg is constrained on both fronts because the project is operated nearly entirely by volunteers and funded only by private donations.[178] The second reason is that the scope of the digitization project may be intentionally limited to a certain collection of materials. For example, the National Archives preservation project aims to digitize public records of national significance to provide greater public access to those records.[179] Therefore, most digitization projects are constrained by limited resources or project scope.

When an entity, such as Google, has abundance resources and the vision to digitize nearly every book ever published, then that project certainly steps outside the bounds of the standard digitization venture and directly challenges the exclusive rights of certain copyright holders. But even then, such projects may fall within the protections of fair use under certain limited circumstances.

B. Extension of Fair Use to Mass Digitization Injures the Public

In *Authors Guild*, the district court and Second Circuit carefully followed the trend in post-*Campbell* fair use analysis to reasonably extend the doctrine of fair use to the specific circumstances presented by Google's mass digitization project.[180] Yet, in both opinions, the courts ignored several substantive components of a robust fair use analysis.[181] Those omissions, in turn, spawned opinions that place artificial constraints on Google Books mass digitization. And those constraints injure the greater public interest.

The primary weakness in both the district court and Second Circuit opinions stems from allowing a proper finding of transformativeness to pervade through other significant aspects of the fair use analysis.[182] Most notably, both courts dismissed the commerciality of Google's use, instead finding that the transformative purpose of the digitization significantly outweighed Google's prospective commercial use.[183] Yet Google's advertising business thrives on data, especially data that reveals characteristics or habits of individual users.[184] In gifting Google

---

178. *See* William L. Hosch, *Project Gutenberg*, ENCYCLOPEDIA BRITANNICA ONLINE, http://www.britannica.com/topic/Project-Gutenberg (last visited Dec. 15, 2015).

179. *See Strategy for Digitizing Archival Materials for Public Access*, NATIONAL ARCHIVES AND RECORDS ADMINISTRATION, , 2015 - 2024 1 - 2 (2014), https://www.archives.gov/files/digitization/pdf/digitization-strategy-2015-2024.pdf.

180. *See* discussion *supra* Section III.B.2 and III.B.3.

181. *See Authors Guild,* 954 F. Supp. 2d at 289–294; *see also Authors Guild,* 804 F. 3d at 212-225.

182. *See Authors Guild,* 954 F. Supp. 2d at 291; *see also Authors Guild,* 804 F. 3d at 220.

183. *See Authors Guild*, 954 F. Supp. 2d at 291–92; *see also Authors Guild*, 804 F.3d at 218–19.

184. *See* Greg McFarlane, *How Does Google Make Its Money?*, INVESTOPEDIA (Nov. 5, 2012) http://www.investopedia.com/stock-analysis/2012/what-does-google-actually-make-money-

with free reign over the wealth of data held in the world's books, the courts discounted the very real possibility that Google will use that information to improve its revenue generating advertisement programs.[185] Allowing a for-profit company to capture each word in millions of copyrighted works for the purpose of indirect commercial benefit through adverting revenue should not be considered fair use.[186] Therefore, by allowing transformativeness to taint the fair use analysis, the courts did not give due weigh to the commerciality of Google's use.

A second significant shortcoming in *Authors Guild* was failing to analyze whether the court's endorsement of Google's mass digitization could have a substantial adverse impact on the rights of copyright holders if more entities enter the playing filed.[187] In *Campbell*, the Court made clear that the fourth fair use factor requires consideration of "whether unrestricted and widespread conduct of the sort engaged in by the defendant . . . would result in a substantially adverse impact on the potential market for the original."[188] Neither the district court nor the Second Circuit addressed this issue directly. Rather, the courts fixated only on the transformativeness and present configuration of the Google Books search engine and snippet views.[189]

If pervasive use similar to that of Google Books were unrestricted, then copyright holders may be exposed to significant risks of infringement through the dissemination of their works by various established channels such as BitTorrent file sharing software.[190] In *Authors Guild*, the Second Circuit correctly noted that there is little risk of Google or its library partners losing digitized books to hackers.[191] The

---

from-goog1121.aspx. Google derives nearly all of its revenue from advertisements. *Id.* Google collects and uses individual user data—such as personal demographics, search histories, location, Internet cookies, and even the content of emails—to improve its search engine, AdSense program, and other services. *See Privacy Policy,* GOOGLE https://www.google.com/policies/privacy/#infocollect (last modified Oct. 2, 2017)

185. *See Authors Guild*, 954 F. Supp. 2d at 291–92; *also Authors Guild*, 804 F.3d at 218–19.

186. *Cf.* Sony, 464 U.S. at 481 (creating a library of movies that are analogous to this generation of a private digital library of books. The Court in *Sony* found that even small scale library-building by individuals raised the potential for harm to a copyright holder).

187. *See Authors Guild*, 804 F.3d at 223–225.

188. *Campbell*, 510 U.S. at 590 (*citing* 3 M. Nimmer Copyright § 13.05[A][4] (1984), p. 13–102.61). *Accord Harper & Row*, 471 U.S. at 569; S. Rep. No. 94–473, at 65; *Folsom*, 9 F. Cas. at 349. The Court in *Campbell* dismissed this argument because plaintiff did not provide significant evidence of a market impact. 510 U.S. at 590. With respect to Google Books, however, the novelty of mass digitization requires relaxation of this evidentiary burden because the economic potential impact of mass digitization and associated emergent technologies is not yet well understood.

189. *See Authors Guild*, 954 F. Supp. 2d at 292–93; *also Authors Guild*, 804 F.3d at 223.

190. *See* Matthew Sag, *Copyright Trolling, an Empirical Study*, 100 IOWA L. REV. 1105, 1108 (2015) (showing that BitTorrent peer-to-peer file sharing copyright infringement in multi-defendant John Doe cases comprised approximately 43% of 3817 copyright law suits filed in 2013. This shows that file sharing remains a significant risk for copyright holders of all types of media including music, movies, video games, software and print media).

191. *Authors Guild*, 804 F.3d at 227–28.

court did not recognize, however, that such risk grows exponentially if other players are allowed to undertake similar digitization projects because as the market expands those new players may not have the security protocols necessary to protect their servers from breach. Furthermore, other players may choose to display snippets that more easily allow copying of full texts. The theft of digital copies or recompilation of snippets could quickly turn into a flurry of pirating similar to that experienced by the music industry beginning in the mid-nineties with Napster and continuing today with BitTorrent software.[192]

As demonstrated Judge Chin and Judge Leval, the four factors of fair use may be manipulated to encompass mass digitization. Yet, overall, applying fair use to mass digitization understates the potential benefit that the public could derive from fully embracing mass digitization technologies. That fully realized benefit—truly promoting the progress of creative works—is a global digital library. Therefore, the courts, legislature and stakeholders must work in concert toward a viable statutory framework to guide entities practicing mass digitization toward the goal of a global library to provide the public with widespread access to all digitized literature, while simultaneously compensating authors to encourage continued composition of creative works.[193] One such system may take the form of extended collective licensing programs overseen by collective management organizations.

C. Extended Collective Licensing as a Potential Solution to Mass Digitization

As discussed above, copyright law must simultaneously promote the rights of copyright holders and the progress of creativity that benefits the public.[194] When fair use is applied to mass digitization, as in *Authors Guild*, the rights of the copyright holder yield to private interests and a limited public benefit. In the case of Google Books, that public benefit is limited because even though Google Books enables new search capabilities, the benefit is not as great as providing full unbridled access to the digitized books. As such, the decision in *Authors Guild* may, over time, have a net negative effect on the public by hindering development of a licensing framework that could provide the necessary foundation to achieve the ultimate goal of constructing a global digital

---

192. *See, e.g.*, Bryan H. Choi, *The Anonymous Internet*, 72 MD. L. REV. 501, 509–16 (2013) (describing the struggle of the music industry to regulated file sharing networks as the technology and the law developed throughout the 1990s well into the 2000s).

193. *See* Office of the Register of Copyrights, *Orphan Works and Mass Digitization*, U.S. COPYRIGHT OFFICE 78 (June 2015) https://www.copyright.gov/orphan/reports/orphan-works2015.pdf. ("[S]hould Congress wish to encourage or facilitate mass digitization projects providing substantial access to the expressive contents of copyrighted works, it would need to look beyond fair use to a licensing model, either voluntary or statutory.") [hereinafter *Orphan Works*].

194. *See* discussion *supra* Section II.A.

library. Extended collective licensing can both foster that goal and protect the interests of all stakeholders.

Generally, in the context of mass digitization, effective licensing agreements may be either voluntary direct license agreements or collective licensing agreements. Voluntary direct licensing can provide the copyright holders with greater control over the terms of agreements, but the transactional cost associated with such agreements, especially in the context of mass digitization, is prohibitive.[195] In contrast, collective licensing allows individual authors to opt into aggregated groups, which substantially reduce transaction costs in facilitating negotiation and rights management.[196] But collective licensing requires oversight because antitrust concerns must be addressed before such systems can be effectively implemented.[197]

Collective licensing itself may be subdivided into three groups: (1) mandated statutory licensing, i.e., a draconian model permissible in the U.S. only if a compelling public need and market frustration compels such a drastic play of Congressional power,[198] (2) voluntary opt-in collective licensing, which is currently available in the U.S. as a central private organizations managing the rights of a group of copyright holders,[199] and (3) extended collective licensing (ECL), under which legislation extends the effect of collective agreements in specific areas of copyright to effectively apply provisions of such agreements to copyright holders who are not members of the contracting organization.[200]

Generally, ECL may take the form of a government-based system under which approved collective management organizations "negotiate licenses for a particular class of works (e.g., textbooks, newspapers, and magazines) or particular uses (e.g., reproduction of published works for

---

195. *See* Office of the Register of Copyrights, *Legal Issues in Mass Digitization: A Preliminary Analysis and Discussion Document*, U.S. COPYRIGHT OFFICE 29–31 (Oct. 2011) https://www.copyright.gov/docs/massdigitization/USCOMassDigitization_October2011.pdf [hereinafter *Legal Issues in Mass Digitization*].

196. *See id.* at 29–30.

197. *See id.*

198. *See id.* at 30.

199. *See, e.g.*, *About Us*, COPYRIGHT CLEARANCE CENTER, https://www.copyright.com/about/ (last visited Dec. 15, 2015) (describing the Copyright Clearance Center as a not-for-profit copyright management organization, assisting clients license copyrights across the globe); *see also* Vicenç Feliú, *Orphans in Turmoil: How A Legislative Solution Can Help Put the Orphan Works Dilemma to Rest*, 12 RUTGERS J.L. & PUB. POL'Y 107, 135 (2015) (discussing licensing schemes for orphan works).

200. Henry Olsson, *The Extended Collective License as Applied in the Nordic Countries* (Aug. 5, 2005) https://web.archive.org/web/20110302080731/http:/www.kopinor.no/en/copyright/extended-collective-license/documents/The+Extended+Collective+License+as+Applied+in+the+Nordic+Countries.748.cms.

educational or scientific purposes)."[201] For example, the terms and provisions of an extended collective license for a particular class of work apply to all members of that class by operation of law.[202] Furthermore, a typical ELC only applies to particular uses, e.g., digitization for use by libraries, digitization for use by private entity search engines, or the like.[203] It is the duty of the collective management organization[204] to determine the terms of those licenses and applicable royalty rates for members of each class. Once the collective organization and user (e.g., the entity undertaking a mass digitization project) come to an agreement, the user can freely use any copyrighted material under the purview of the collective organization. Also, copyright holders may, if allowed under specific terms of the legislation, opt out of particular uses by users.[205] Thus, ECL arrangements are not a limitation on the rights of copyright holders, but rather an arrangement for efficiently manage the rights of all copyright holders in a particular class of works.[206]

In the context of mass digitization of books, ECL is better than mandated licensing because copyright holders retain some degree of control, as well as better than voluntary opt-in licensing because the agreement captures to all copyright holders in a particular category of works.[207] Also, under an ECL regime, all relevant stakeholders are adequately represented.[208] For example, the exclusive rights of the copyright holder can be protected through contractual protections.[209] Such a system also promotes the progress of creative works by

---

201. U.S. COPYRIGHT OFFICE, *supra* note 195, at 35. The amended settlement agreement that was rejected by Judge Chin "could be characterized as an [ECL] regime, because the settlement would have established a Book Rights Registry to collect and distribute royalties for the benefit of rights holders – regardless of whether rights holder authorized the Registry to act on their behalf." U.S. COPYRIGHT OFFICE, *supra* note 195, at 36.

202. *See Orphan Works, supra* note 193, at 19.

203. *See* Olsson, *supra* note 200.

204. *See* Olsson, *supra* note 200 (defining the requirements of the collective management organization and including, for example, the proportion of copyright holders in a class of work that must be represented (e.g., a majority or substantial number of copyright holders), the scope of the representation (i.e., only national copyright holder or any copyright holder with work exploited in the country)).

205. *See* Olsson, *supra* note 200 (explaining that in the Nordic countries, guarantees for copyright holder, including foreign rights holders, include (1) a right to prohibit use of their work, i.e., opt out, and (2) a right to remuneration by (a) the collative agreement, or (b) individual remuneration outside the collective agreement).

206. *See* Olsson, *supra* note 200 (describing the Infosoc Directive negotiations. Specifically that the Nordic delegations were somewhat nervous about the status of the extended license system in the context of a closed list of possible exceptions that the Directive prescribes in its Article 5.2 and 5.3. The end result of these deliberations is documented in Preamble 18 which states that: "This Directive is without prejudice to the arrangements in the Member States concerning the management of rights such as extended collective licenses." This makes it very clear that the Nordic system by nature is not to be seen as a limitation but as a management arrangement.)

207. *See* Olsson, *supra* note 200.

208. *See* Olsson, *supra* note 200.

209. *See* Olsson, *supra* note 200.

providing copyright holders with reliable royalty streams.[210] As to the public, ECL delivers the most public benefit by making digitized works widely available and advancing the creation of a global digital library. Finally, ECL can provide incentives for private entity funding by granting limited contractual rights, such advertising or sales.

The ECL scheme has proven effective in the Nordic countries of Denmark, Finland, Iceland, Norway and Sweden since the 1960's.[211] Sweden, in particular, provides ECL arrangements in retransmission of television broadcasts, reproduction for educational or internal purposes, recording of television or radio for educational purposes, and library distribution of digitized material.[212] In Sweden, the ECL system "solve[s] the need for copyright clearance in mass use situations where there is a need for appropriate and efficient solutions. . . . to satisfy both right-owners, users and the society at large."[213] Therefore, ECL may be the most a viable means to provide an efficient, predictable and practical solution to mass digitization.

In June 2015, the U.S. Copyright Office recommended an ECL pilot program to address mass digitization issues.[214] In its report, Office recognized that fair use is an unpredictable standard for entities engaged in digitization, which "will slow the development of future mass digitization projects by dissuading litigation-averse users from undertaking such activities."[215] To solicit input for formal legislation, the Office outlined several aspects of its ECL pilot program including: (1) limiting the scope to, *inter alia*, literary works; (2) prohibiting direct or indirect commercialization of digitized texts, e.g., advertising on the digitized text or charging access fees; (3) providing authorization and oversight of collective management organizations; (4) granting copyright holders the right to opt-out; (5) providing anti-trust exemption to collective management organizations and user groups; (6) requiring adequate security measures to protect digitized copies; (7) providing for equitable distribution of royalties; (8) confirming that the ECL program does not diminish traditional application of the fair use doctrine; and (9) satisfying international treaty obligations.[216] This type of ECL framework "could do much to further the objectives of the copyright system by providing legal certainty to users, establishing

---

210. *See* Olsson, *supra* note 200 (explaining that the author permanently retains the right to individual remuneration).

211. *See* Olsson, *supra* note 200.

212. *See* Olsson, *supra* note 200.

213. *See* Olsson, *supra* note 200.

214. *See Orphan Works, supra* note 193, at 1-2 (recognizing that a U.S. ECL program would be consonant with recently established ECL frameworks in France, Germany, and the United Kingdom).

215. *See Orphan Works, supra* note 193, at 77.

216. *See Orphan Works, supra* note 193, at 84–104.

reliable mechanisms for the compensation of authors, and making vast numbers of long forgotten works available for the public good."[217]

D. Toward Freedom of Literature: The Global Library

The Copyright Office's proposed ECL pilot program is a necessary first step in properly addressing mass digitization. An ECL regimen capable of capturing the global scope of Google Books, however, will require international cooperation. One possible solution is to establish an international collective management organization by international treaty proposed by, for example, the World Intellectual Property Organization. The treaty can also provide the foundation to establish a global digital library. Member countries, and perhaps through limited roles of private entities like Google, could fund the global library and royalty pool managed by the international collective management organization. Such international cooperation would provide an unrivaled global benefit through near-instant access to all literature.

VII. CONCLUSION

By examining the development of mass digitization through the lens of Google Books, it is apparent that the world's literature is undergoing a digital revolution. Also, by reflecting on the current state of copyright law through the decisions in *Authors Guild*, we see that the fair use doctrine cannot, in the face of that revolution, optimally advance the interests of copyright holders and the public. The digital literature revolution has come too far for the traditional notions of fair use to suppress the transformation—the literary world has crossed the digital Rubicon. Looking back to the far bank, to the doctrine of fair use, will only delay the inevitable. Hence, the U.S. must enact a statutory extended collective licensing framework to address mass digitization. Such legislation is the first step along the path toward an international treaty that will provide the foundation for the global digital library of the future.

---

217. *See Orphan Works*, *supra* note 193, at 106.