

# THE ONLINE ARCHIVE: FAIR USE AND DIGITAL REPRODUCTIONS OF COPYRIGHTED WORKS

BRYAN OBERLE\*

## I. INTRODUCTION

The purpose of copyright law is to promote the progress of science and useful arts by providing sufficient protection to authors and inventors to stimulate creative activity, while permitting others to utilize protected works to advance the progress of the arts and sciences. Copyright infringement claims aim at ensuring the former, while the doctrine of fair use aims at promoting the latter. Ideally, these two forces balance each other out, but this balance can easily shift. A rising legal issue, which has the potential to cause such an imbalance, is whether the wholesale digital reproduction of copyrighted works for archival purposes should be protected by the fair use doctrine. Specifically, this Note is concerned with digital archives of books and other literary works, as these mediums have great social and intellectual impact for research and scholastic purposes.

The Copyright Act of 1976 codifies a four-factor test for what constitutes fair use: (1) the purpose and character of the work, (2) the nature of the original 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.<sup>1</sup> Before it was codified by the Copyright Act of 1976, fair use doctrine had a history of common law usage dating as far back as 1841.<sup>2</sup>

Although fair use doctrine has a long history, its boundaries remain unclear due to the context-sensitive nature of its inquiries. Potential applications of fair use are analyzed on a case-by-case basis,<sup>3</sup> making it difficult for individuals to know beforehand whether their use of a copyrighted work constitutes fair use. As technology continues to grow and develop, and more individuals are exposed to tools and techniques that give them easier access to copyrighted works, the boundaries of fair use doctrine

---

\* Class of 2016, University of Southern California Gould School of Law; B.A. Political Science and B.A. Philosophy, Politics, & Law 2013, University of Southern California College of Letters, Arts, and Sciences; Staff Member, Southern California Interdisciplinary Law Journal, Volume 24. I would like to thank the staff and executive editorial board of this Volume of the Southern California Interdisciplinary Law Journal for all of their hard work on this Note, and my parents for their love and support.

1 17 U.S.C. § 107 (1992).

2 *See generally* Folsom v. Marsh, 9 F. Cas. 342 (C.C.D. Mass. 1841).

3 Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282, 290 (S.D.N.Y. 2013).

become even more blurry. This gives rise to cases asking whether digital reproductions of copyrighted works for archival purposes satisfy fair use. These cases involve websites scanning or transcribing all or most of a copyrighted work and allowing certain individuals or institutions to have access to these reproductions for free. This issue was addressed in a recent lawsuit against one of the world's largest and most influential digital archives, Google Books.<sup>4</sup> The Google case indicated that the wholesale digital reproduction of copyrighted works could be fair use when used for archival purposes, but this idea is not universally shared among courts or legal commentators.

Due to the context-sensitive nature of fair use doctrine, any conclusion as to whether a certain type of work constitutes fair use will necessarily have certain caveats. That being said, it is imperative to analyze whether the wholesale digital reproduction of copyrighted works is more likely than not to count as fair use, given the wide-ranging implications these reproductions can have societally and financially. This Note will analyze this issue by providing background on how fair use is defined, using a recent case levied against Google Books as an example of how modern courts analyze this issue, detailing arguments for and against the proposition that fair use should be applied to digital archives, and proposing an ideal model for digital archives going forward. This Note ultimately argues that digital archives can and should be considered fair use, so long as they install limitations on the amount of access that private individuals can have to the copyrighted works, they respect the wishes of copyright holders who do not want their works to be displayed, and they are run by public—rather than private—institutions.

## II. BACKGROUND

### A. DEFINITION AND PURPOSE OF FAIR USE

This Note will begin by briefly touching on the meanings and boundaries of the four fair use factors set forth in 17 U.S.C. § 107, the fair use section of the Copyright Act of 1976.

#### *1. Purpose and Character of the Work*

The first factor of fair use analysis is the “purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes.”<sup>5</sup> The preamble to section 107 gives some indication as to what purposes are acceptable: “criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research.”<sup>6</sup> The central focus of this factor is whether the allegedly-infringing work merely supplants the use of the original work, or

---

<sup>4</sup> *Id.* at 282.

<sup>5</sup> 17 U.S.C. § 107.

<sup>6</sup> *Id.*

if it “adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message.”<sup>7</sup> When a work accomplishes the latter, it is deemed to be “transformative,” and thus lies within the heart of fair use doctrine.<sup>8</sup> For example, an author displaying pictures of licensed artwork in a book may be considered transformative when it is biographical in nature and used to illustrate the evolution of a particular artist’s style.<sup>9</sup>

*a. Transformative Use*

Although transformative use is not necessary for a holding that a work satisfies fair use, the more transformative the work, the less significance other factors, such as the commercial nature of a work, will have against a finding of fair use.<sup>10</sup> However, just because a work is transformative does not mean it will automatically be labeled fair use. For example, in the case of *Warner Bros. Entertainment Inc. v. RDR Books*, a publishing company was sued for using vast amounts of text from the *Harry Potter* books in its “*Harry Potter Lexicon*,” which was an encyclopedia for information about the *Harry Potter* universe.<sup>11</sup> Because it served reference purposes, rather than the entertainment or aesthetic purposes of the original works, the *Lexicon*’s use was transformative and did not supplant the objects of the *Harry Potter* books.<sup>12</sup> However, the *Lexicon* was deemed not to be fair use because of its verbatim copying and close paraphrasing of language from the *Harry Potter* works.<sup>13</sup>

*b. Commercial Purpose*

A secondary consideration in purpose and character analysis is whether the allegedly-infringing work was created for a commercial purpose. This consideration is not heavily emphasized by courts since “many, if not most, secondary users seek at least some measure of commercial gain from their use.”<sup>14</sup> If the commercial nature of a work were to create a presumption of unfair use, it would lead to an overly restrictive view of fair use that would swallow nearly all of the uses illustrated in the preamble of section 107.<sup>15</sup> Thus, any commercial nature of a work, as well as any nonprofit educational purposes, are merely elements of fair use analysis. That being said, the fact that a use is commercial in nature does tend to weigh against a finding of fair use.<sup>16</sup> For example, in *Los Angeles News Serv. v. KCAL-TV Channel 9*, the fact that a television news station was a for-profit institution

---

7 Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994).

8 *Id.*

9 See Warren Pub. Co. v. Spurlock, 645 F. Supp. 2d 402, 417 (E.D. Pa. 2009).

10 *Id.*

11 See generally *Warner Bros. Entm’t Inc. v. RDR Books*, 575 F. Supp. 2d 513 (S.D.N.Y. 2008).

12 *Id.* at 541.

13 *Id.* at 547.

14 *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 921 (2d Cir. 1994).

15 *Id.* at 921; *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 584 (1994).

16 *Harper & Row Publishers, Inc. v. Nation Enter.*, 471 U.S. 539, 540 (1985).

that essentially “sells” news meant that its use of thirty seconds of another station’s copyrighted footage for a story raised an inference that it was riding the other station’s “copyrighted coattails,” and was therefore not fair use.<sup>17</sup>

## 2. Nature of the Copyrighted Work

The second factor of fair use analysis recognizes that some works are closer to the core of intended copyright protection than others, and are therefore entitled to greater protection from fair use.<sup>18</sup> For example, copyright law tends to provide greater protections to fictional works, rather than non-fictional ones, as they are the product of originality and inventiveness, rather than diligence and research, and also because the law generally recognizes a greater need to disseminate factual works than works of fiction or fantasy.<sup>19</sup> Greater copyright protection is also afforded to unpublished works, rather than published ones, as the author’s right to control the first public appearance of his or her expression weighs against fair use of the work before its official release.<sup>20</sup>

## 3. Portion of the Copyrighted Work Used

The third factor of fair use analysis addresses the reasonableness of “the amount and substantiality of the portion used in relation to the copyrighted work as a whole.”<sup>21</sup> A work is more likely to be deemed fair use if the amount of the work copied is reasonable in relation to the purpose of the copying work.<sup>22</sup> In general, the less of a copyrighted work that is copied, the more likely it is that it will be labeled fair use.<sup>23</sup> However, even copying a small portion of a copyrighted work can be deemed infringing if that portion can be considered the “heart” of the original work, consisting of a substantial portion of the appeal of the work.<sup>24</sup> For example, in *Harper & Row Publishers, Inc. v. Nation Enterprises*, a magazine that published a numerically insubstantial portion of verbatim excerpts from ex-President Gerald Ford’s unpublished memoirs was held to not be protected by fair use because the portions it took were the most interesting and moving parts of the entire manuscript.<sup>25</sup> At the same time, reproducing entire works can be considered fair use if doing so falls within reason and satisfies the other fair use factors.<sup>26</sup> For example, videotaping entire television programs is considered fair use because consumers mainly use videotapes for time-shifting purposes, rather

17 *Los Angeles News Serv. v. KCAL-TV Channel 9*, 108 F.3d 1119, 1121 (9th Cir. 1997).

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

19 *Warner Bros. Entm't Inc. v. RDR Books*, 575 F. Supp. 2d 513, 549 (S.D.N.Y. 2008).

20 *Harper & Row Publishers*, 471 U.S. at 564.

21 17 U.S.C. § 107 (1992).

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

23 *See id.*

24 *Harper & Row Publishers*, 471 U.S. at 565.

25 The magazine published between 300 and 400 words from the unpublished 200,000-word original manuscript, 7,500 of which *Time Magazine* had the rights. *Id.* at 539, 565, 579.

26 *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 449–50 (1984).

than commercial purposes.<sup>27</sup> This factor is viewed by some as a proxy for the fourth factor, impact on demand for the original work, since consumer demand for the original typically falls as more of it is available for free.<sup>28</sup>

#### 4. Effect of Use on the Potential Market Value

The final factor in fair use analysis is “the effect of the use upon the potential market for or value of the copyrighted work,” or how much the new work harms the market for the original work or officially-licensed derivative works.<sup>29</sup> Before *Campbell v. Acuff-Rose Music*, this factor was considered “the single most important element of fair use.”<sup>30</sup> However, the *Campbell* case instructs that all the factors be weighed together, in light of the purposes of copyright.<sup>31</sup>

This factor not only considers the harm to original copyrighted works, but also the harm to any derivative works that creators of original works would in general develop or license others to develop.<sup>32</sup> In *Twin Peaks Productions, Inc. v. Publications International, Ltd.*, a book that was meant to serve as a complete guide to the plot and characters of the television show *Twin Peaks* did not receive fair use protection because it potentially harmed the market for the show’s original and derivative works.<sup>33</sup> It interfered with the primary market for the copyrighted work because it was possible that a person who had missed an episode of *Twin Peaks* would find reading the book an adequate substitute, and would not need to rent the videotape of that episode in order to enjoy the next one.<sup>34</sup> It harmed the market for derivative works because the creators had already licensed at least two *Twin Peaks* derivative books and had plans to create more.<sup>35</sup> Another case, *Castle Rock Entertainment, Inc. v. Carol Publishing Group*, found a *Seinfeld* trivia book to be infringing because it did not respect the economic choice of the creators of *Seinfeld* to publish a trivia book themselves, even though the creators evidenced little or no interest in doing so.<sup>36</sup> Considerations on the impact towards derivative works are limited to some extent, since authors cannot bar the entire field of publishable comments, criticisms, or news reports that could cash in on the popularity of the copyrighted work, but impairment of derivative markets is a factor that weighs in favor of original authors.<sup>37</sup>

The fourth factor considers not only the effect that the allegedly-infringing works at issue have on the market, but also the effects that

---

27 *Id.*

28 *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 926 (2d Cir. 1994).

29 17 U.S.C. § 107 (1992); *Harper & Row Publishers*, 471 U.S. at 568.

30 *Harper & Row Publishers*, 471 U.S. at 566.

31 *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578 (1994).

32 *Id.* at 592.

33 *Twin Peaks Prods., Inc. v. Publications Int'l, Ltd.*, 996 F.2d 1366, 1377 (2d Cir. 1993).

34 *Id.*

35 *Id.*

36 *Castle Rock Entm't, Inc. v. Carol Publ. Group*, 150 F.3d 132, 145–46 (2d Cir. 1998).

37 *Twin Peaks*, 996 F.2d at 1377.

unrestricted and widespread derivatives of the allegedly-infringing work would have on the market.<sup>38</sup> Even if a single, minor use of a copyrighted work might be considered fair use by itself, it may still be considered copyright infringement if a great number of people doing the same thing would harm the market.<sup>39</sup> For example, in the *Harper & Row* case, one of the reasons the magazine's use of unpublished quotes was deemed infringing was because isolated instances of similar infringements, when multiplied many times, posed substantial potential for damage to the marketability of first serialization rights in general.<sup>40</sup>

B. APPLICATION OF FOUR FACTORS IN *AUTHOR'S GUILD, INC. V. GOOGLE INC.*

A recent case, *Authors Guild, Inc. v. Google Inc.*, hereinafter referred to as the "*Google Case*," provides useful insight into how modern courts analyze this issue.<sup>41</sup> This case centers around Google's digital reproduction of millions of copyrighted books, making them available for its library project partners to download and displaying "snippets" to the general public through its Google Books service. The plaintiff of this case was The Authors Guild, Inc., an organization of published authors that sued Google for alleged copyright infringement by scanning copyrighted books into the Google Books database and making the scanned copies searchable to the public.<sup>42</sup> Google Books is a program run by Google Inc. that scans and copies millions of books into an online database, ninety-three percent of which are non-fiction and seven percent of which are fiction.<sup>43</sup> Of the books in Google Books' online database, nine percent are copyrighted books still in print, seventy-five percent are copyrighted books that are out of print, and sixteen percent are books whose copyright terms have expired and are in the public domain.<sup>44</sup>

Google Books is divided into two digital book programs: the Partner Program and the Library Project.<sup>45</sup> The Partner Program contains 2.5 million books worth of material provided to Google by around 45,000 book publishers or other rights holders, displayed with the permission of the rights holders for the purpose of helping publishers sell books and helping books become discovered.<sup>46</sup> The Library Project hosts over twenty million fully-scanned copies of books borrowed from the collections of the New York Public Library, the Library of Congress, and a number of university libraries, with the agreement that participating libraries can download

---

38 *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590 (1994).

39 *Id.*

40 *Harper & Row Publishers, Inc. v. Nation Enter.*, 471 U.S. 539, 569 (1985).

41 *See generally* *Authors Guild, Inc. v. Google Inc.*, 954 F. Supp. 2d 282 (S.D.N.Y. 2013).

42 *Id.* at 288.

43 *Id.* at 285.

44 Lawrence Lessig, *Is Google Book Search "Fair Use"?*, YOUTUBE (Jan. 15, 2006), <https://www.youtube.com/watch?v=TmU2i1hQiN0>.

45 *Authors Guild*, 954 F. Supp. 2d at 285.

46 *Id.* at 285-86.

digital copies of each book scanned from their own collections, though not the collections of other libraries.<sup>47</sup> Google has not sought to compensate or obtain the permission of copyright holders in connection with the Library Project.<sup>48</sup>

After scanning a book through the Library Project, Google provides a digital copy to the participating library from which it borrowed the book and retains a copy for its own records.<sup>49</sup> Using advanced scanning technology, the text of each book scanned by Google is machine-readable and can be indexed.<sup>50</sup> This index allows Google's users to search for a particular word or phrase and obtain a list of the most relevant books in which that word or phrase is found.<sup>51</sup> Users can view information about the listed books, with links to sellers and libraries that list the books as part of their collections.<sup>52</sup> These search results can also produce several "snippets" of verbatim text from books containing the desired word or phrase, with each snippet being an eighth of a page of the book.<sup>53</sup> Google has various security measures in place to prevent users from viewing complete copies of snippet-view books, including "blacklisting" certain pages of each book and only providing the first responsive snippet on each page.<sup>54</sup> Users who try to obtain whole copies of books by piecing together different snippets would at best only generate a patchwork of snippets that would be missing at least one snippet from every page and ten percent of all pages.<sup>55</sup>

With these facts about the nature of the Google Books digital archive in mind, the court in the *Google Case* underwent the fair use analysis outlined below and ultimately held that Google was protected by fair use.<sup>56</sup>

### *1. Purpose and Character of the Work*

The preamble to section 107 mentions "scholarship" and "research" as illustrative examples of the kinds of purposes that are in line with fair use doctrine.<sup>57</sup> These are two examples that weigh in favor of archival digital reproductions as fair use, since a potential use of archives is to assist in individual research and educational efforts. The court in the *Google Case* recognized this when it said that Google's practice of displaying snippets of copyrighted works was highly transformative, in that it was transforming expressive text into a comprehensive word index that helped readers, scholars, researchers, and others find books.<sup>58</sup> Instead of supplanting the original works by serving as a tool to read books, these snippets help

---

47 *Id.* at 286.

48 *Id.*

49 *Id.*

50 *Id.*

51 *Id.*

52 *Id.*

53 *Id.* at 286–87.

54 *Id.* at 287.

55 *Id.*

56 *Id.* at 294.

57 17 U.S.C. § 107 (1992).

58 *Authors Guild*, 954 F. Supp. 2d at 291.

individuals facilitate their research by identifying books that might be of interest to them, create searchable indices of texts, increase the accessibility of texts to people with disabilities, and preserve copies of old, out-of-print books.<sup>59</sup> The act of digitizing book texts also opened up a new field of substantive research by allowing data mining to analyze the metadata of texts to determine the frequency of words and trends in their usage.<sup>60</sup> The fact that Google is a commercial enterprise did not negate the effect of these purposes, since Google only benefitted commercially indirectly, without actually selling scans or placing ads on pages with snippets.<sup>61</sup> The educational purposes outweighed any commercial incentives Google may have had in establishing this database.<sup>62</sup> Thus, the first factor was held to weigh strongly in favor of fair use.<sup>63</sup>

It is important to note that many of these valid purposes are provided less by Google itself than by the libraries that participate in its digitization efforts.<sup>64</sup> Google is not digitizing copies of books that it itself owns.<sup>65</sup> Instead, Google provides libraries with the means to make digital copies of books that they already own, and full copies are only available to partner libraries.<sup>66</sup> Thus, the purpose of Google Book's archive is to advance libraries' transformative uses of books consistent with copyright law.<sup>67</sup> The direct beneficiary of Google's archive is meant to be libraries that use copies of the digitized books for the progress of science and cultivation of the arts, rather than private individuals that might use its archive for this or any other purposes.<sup>68</sup> The outcome of this case would likely have been different if Google's purpose had been to provide unlimited access to copyrighted books to all users and leave individuals to their own devices on how they would use this information.<sup>69</sup> Instead, by only providing full copies of books to participating libraries, who already owned full copies of those books in the first place, Google was not providing access to anything libraries didn't already have.<sup>70</sup> Also, the fact that Google was working in conjunction with institutions like the New York Public Library, the Library of Congress, and university libraries likely provided Google with a sense of legitimacy that it might not have enjoyed otherwise.

---

59 *Id.* at 293.

60 *Id.* at 291.

61 *Id.* at 292.

62 *Id.*

63 *Id.*

64 *E.g.*, Nathan Chandler, *How Google Books Works*, HOW STUFF WORKS (Dec. 29, 2009), <http://computer.howstuffworks.com/google-books.htm>.

65 *Id.*

66 *Id.*

67 *Authors Guild*, 954 F. Supp. 2d at 293.

68 *See* Chandler, *supra* note 64.

69 *See Authors Guild*, 954 F. Supp. 2d at 291.

70 *See, e.g.*, Google Books, *Library Support*, GOOGLE, <https://scholar.google.com/intl/us/scholar/libraries.html> (last visited Feb. 18, 2016).



## 2. Nature of Copyrighted Work

Copyrighted works tend to enjoy greater copyright protection if they are unpublished works of fiction that are still in print, and less protection if they are published non-fiction works that are no longer in print.<sup>71</sup> These are not the only feasible attributes a copyrighted work can have that will affect its analysis under the second factor of fair use analysis, but they are the most prominent ones.<sup>72</sup> In the *Google Case*, the majority of the database in question was filled with non-fiction, out-of-print, published books, which weighed in favor of a finding of fair use.<sup>73</sup> Thus, a digital archive is more likely to fall under the scope of fair use if it consists primarily of published, out-of-print works of non-fiction.

## 3. Portion of Copyrighted Work Used

Generally, the lower the proportion of a copyrighted work that is copied, the more likely it is that its use will be labeled fair use.<sup>74</sup> The fact that Google scans the full text of books verbatim weighed against a finding of fair use in the *Google Case*.<sup>75</sup> As archives by their very nature strive for comprehensiveness, this factor weighs against granting digital archives fair use protection. Thus, the third prong of fair use analysis clashes with the first prong, as applied to digital archives. After all, an archive that only provides snippets of non-essential information from a number of texts is of little use to libraries or scholars trying to find useful texts.

There are two caveats to keep in mind when considering the conflict between the first and third elements of fair use as applied to digital archives: (1) each of these prongs is merely a factor of fair use analysis, with no one prong being conclusive on its own, so courts may favor one prong over another based on how they believe it would serve the purposes of copyright law; and (2) these factors do not exist independent of one another, and each is often analyzed in relation to the others.<sup>76</sup> The proportion of a work one is allowed to copy is not set at a specific percentage, but is evaluated based on the purposes the copies would serve.<sup>77</sup> If the purpose of an archive can only be reasonably served by creating full digital reproductions of copyrighted works, the third factor of fair use analysis may not be a death sentence.<sup>78</sup> In the *Google Case*, the third factor only slightly weighed against a finding of fair use because full-work reproduction is critical to the full-text searching functions of Google Books.<sup>79</sup> The court also noted as significant the fact that Google limits the

---

71 Warner Bros. Entm't Inc. v. RDR Books, 575 F. Supp. 2d 513, 549 (S.D.N.Y. 2008); Harper & Row Publishers, Inc. v. Nation Enter., 471 U.S. 539, 564 (1985).

72 See *Authors Guild*, 954 F. Supp. 2d at 292.

73 *Id.*

74 See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 586 (1994).

75 *Authors Guild*, 954 F. Supp. 2d at 292.

76 See, e.g., *Campbell*, 510 U.S. at 586.

77 *Id.*

78 E.g., *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014).

79 *Authors Guild, Inc. v. Google, Inc.*, 954 F. Supp. 2d 282, 292 (S.D.N.Y. 2013).

amount of text it displays in response to a search.<sup>80</sup> Thus, while the act of copying the full text of a book may be necessary to fulfill the purposes of a digital archive, it is important for an archive to be frugal in the amount of text it displays to those who use it.

#### *4. Effect of Use on the Potential Market Value*

The fourth factor of fair use analysis is another area where digital archives face trouble, as there is a risk that individuals will prefer to access sources through a free archive than to actually buy the original source itself.<sup>81</sup> The fear is that digital reproductions will become “market replacements” for their originals, thereby harming the potential market for or value of copyrighted works.<sup>82</sup> This is not an unfounded fear, as the success of peer-to-peer file sharing sites like The Pirate Bay and Megaupload shows that people will gladly acquire copyrighted works for free if given the opportunity.<sup>83</sup>

Like the third fair use factor, restraint on the part of the archive is the key to a finding of fair use. The plaintiffs in the *Google Case* made the argument that Google’s scans were hurting the market for books and that users could combine the results of multiple searches to access an entire book.<sup>84</sup> The court disagreed for several reasons: (1) the only ones with access to full books available for download are partner libraries who already owned the books in the first place; (2) it is not likely that users would take the time and energy to input countless searches through trial and error to get enough snippets to comprise an entire book; and (3) creating an entire book from snippets is impossible because certain pages and sections of the books are blacklisted and one would need a copy of the original to piece together all the snippets in a coherent fashion.<sup>85</sup> In fact, the court pointed out that Google’s archive has actually helped the market for books, since it helps books get noticed by users, thus increasing their potential audiences.<sup>86</sup> With the growth of online shopping, being noticed and convenient to access are essential to market success, thus the fourth factor weighed heavily in favor of a finding of fair use.<sup>87</sup> Once again, it was restraint on the part of Google in not allowing unfettered access to copyrighted works that permitted a finding of fair use.

The court’s analysis in the *Google Case* shows how much of a tightrope walk fair use can be: an archive has to be useful enough that it has

---

80 *Id.*

81 *Id.*

82 *Id.*

83 David Sarno, *The Internet Sure Loves its Outlaws*, L.A. TIMES (April 29, 2007), <http://www.latimes.com/la-ca-webscout29apr29-story.html> (reporting that The Pirate Bay has over five million active users, and it is likely an even higher number today than it was when the article was written).

84 *Authors Guild*, 954 F. Supp. 2d at 292.

85 *Id.* at 293.

86 *Id.*

87 *Id.*

significant educational and scholarly value, but not so useful that it makes the original works obsolete; it has to have comprehensive copies of copyrighted works for it to function as intended, but it has to limit the amount of material that users can view; it has to provide enough text from books for users to recognize which ones are relevant, but not so much that they give away the heart of the book.<sup>88</sup> If an archive, or any other kind of work, can achieve this balance, there is a good chance it will be labeled fair use. That said, there is still disagreement as to whether digital archives should be allowed to exist under fair use protection from a normative standpoint.

### III. ARGUMENTS

Even if digital archives satisfy the four factors of fair use doctrine, the question remains whether they lead fair use in the right direction. Some commentators have criticized Google's efforts as being contrary to the law of fair use, disruptive of our copyright system, and unable to achieve its lofty goals.<sup>89</sup> Others argue that a database like Google Books has enormous social and intellectual value that needs to be promoted.<sup>90</sup> As Google is a major trend-setter in this area as one of the most expansive and widely-used digital libraries, the future of digital archives will depend heavily on the extent to which courts and litigators find its pros outweigh its cons. However, whether or not Google sets a good example of how a digital archive should be structured, Google itself may not be the ideal entity to be in charge of a project with such wide-reaching social and economic impact.

#### A. COMPLIANCE WITH FAIR USE DOCTRINE

Although the court in the *Google Case* ruled in favor of Google Books, it has faced heavy criticism from commentators, and the Authors Guild has already appealed the decision.<sup>91</sup> There is a possibility that digital archives such as Google Books may be struck down as infringing in future cases.

A major criticism of digital archives, and Google Books in particular, is that its use of copyrighted texts is not truly transformative.<sup>92</sup> Since the decline of the importance of the fourth fair use factor, effect on the market of the original work, transformativeness has become the most important factor of fair use analysis.<sup>93</sup> One of the most influential articles on the subject of fair use, "Toward a Fair Use Standard," by Judge Pierre N.

---

<sup>88</sup> See generally *id.*

<sup>89</sup> See, e.g., Kathleen E. Kubis, *Google Books: Page by Page, Click by Click, Users are Reading away Privacy Rights*, 13 VAND. J. ENT. & TECH. L. 217 (2010).

<sup>90</sup> E.g., Kelvin Hiu Fai Kwok, *Google Book Search, Transformative Use, and Commercial Intermediation: An Economic Perspective*, 17 YALE J. L. & TECH. 283 (2015).

<sup>91</sup> Authors Guild, *Round One to Google: Judge Chin Finds Mass Book Digitization a Fair Use. Guild Plans Appeal*, AUTHORS GUILD (Nov. 24, 2013), <http://www.authorsguild.org/general/round-one-to-google-judge-chin-finds-mass-book-digitization-a-fair-use-guild-plans-appeal/>.

<sup>92</sup> Jonathan Band, *The Future of Fair Use After Google Books*, PROJECT DISCO (Feb. 11, 2014), <http://www.project-disco.org/intellectual-property/021114-the-future-of-fair-use-after-google-books/>.

<sup>93</sup> Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990).

Leval, emphasizes that “a quotation of copyrighted material that merely repackages or republishes the original is unlikely to pass the test.”<sup>94</sup> Leval emphasizes that a secondary use of copyrighted material should transform the expression of the original work, but a digital archive does not in and of itself express anything.<sup>95</sup> Instead, the *Google Case* judged transformativeness simply by asking if the expression of the original work is being used for a different, socially-beneficial purpose.<sup>96</sup> This can create problems, as an emphasis on socially-beneficial purpose may shift the focus of analysis from the infringer’s actions to the actions of third parties.<sup>97</sup> This focus on third party actions could make fair use protection too easy to obtain, since one would only need to point to some beneficial social purpose to succeed.<sup>98</sup>

Rather than focusing on uses created by third parties, critics of digital archives argue that more focus should be given to whether the archive itself is a transformative use. Even if libraries and users utilize Google Book’s database in different ways, Google itself is doing nothing more than copying large amounts of copyrighted works and letting others use them as they see fit.<sup>99</sup> A comparable issue arose in the case *UMG Recordings v. MP3.com*.<sup>100</sup> In *UMG Recordings*, a website called MP3.com created a “space-shifting” service that allowed people who purchased a CD to access the music on their CD from anywhere without needing the physical discs themselves.<sup>101</sup> MP3.com copied several thousand CDs into its server, then provided access to entire CDs to subscribers that demonstrated they had possessed a copy of the CD.<sup>102</sup> MP3.com argued this constituted fair use, but the court disagreed on the grounds that MP3.com added “no ‘new aesthetics, new insights and understanding’ to the original music recordings it copies, but simply repackages those recordings to facilitate their transmission through another medium. While such services may be innovative, they are not transformative.”<sup>103</sup> In the same way, it could be argued that digital archives like Google Books simply repackage physical copies of books through the medium of computer text, and are therefore not transformative. The problem with this argument is that it relies on the assumption that an archive’s use of the works it copies extends solely to scanning them, when it can provide “new understandings” of the original works by organizing their metadata and creating searchable databases for them. The distinction here is that the usefulness of archives derives not just from their ability to provide access to copyrighted works, but also from

---

94 *Id.*

95 *See id.*; Band, *supra* note 92.

96 Band, *supra* note 92.

97 *Id.*

98 *Id.*

99 *E.g.* Chandler, *supra* note 64.

100 *UMG Recordings v. MP3.com*, 92 F.Supp 2d 349 (S.D.N.Y. 2000).

101 *Id.* at 351.

102 *Id.* at 354.

103 *Id.*

their ability to show the interrelatedness of the content and subject matter of different works to each other.

The heart of the transformativeness of Google Books, and digital archives in general, is its creation of searchable indexes and the production of metadata on the works it copies. This elevates these archives from simple retransmissions to something that provides new insights and new understandings to the original works. Previous case law supports this theory—in *Kelly v. Arriba Soft Corp.*, an owner of copyrighted pictures sued the operator of a visual search engine that displayed his pictures in search results as small thumbnail pictures.<sup>104</sup> Even though the search engine was a commercial enterprise and the photographer's copyrighted photographs were displayed in its searchable database, displaying the photos as thumbnails was deemed to be transformative because the thumbnails were smaller, lower-resolution images that served an entirely different function from the originals.<sup>105</sup> While the original pieces were created for aesthetic purposes, the thumbnails were “a tool to help index and improve access to images on the internet and their related web sites.”<sup>106</sup> The thumbnails were also unlikely to replace the originals because users could not expand the images without a significant loss of clarity.<sup>107</sup> The thumbnails in *Kelly* and the snippets in the *Google Case* are highly comparable: both are smaller, limited versions of full works that are meant primarily to direct one to the original. Another case, *White v. W. Pub. Corp.*, found a commercial database of copyrighted briefs and other pleadings to be transformative fair use because the defendants used the briefs to create an interactive legal research tool for reviewing and identifying legal documents.<sup>108</sup>

Thus, case law suggests that, while simply serving as a repository for copyrighted works is not a legitimate transformative use, creating searchable indices for those works to assist in research is a legitimate transformative use. To the extent that a digital archive accomplishes this, it will likely satisfy the “transformative” element of fair use.

#### B. DISRUPTIVE EFFECT ON COPYRIGHT LAW

Our system of copyright law is currently going through an era of upheaval due to the evolution of digital technology. Not only do people consume media at a greater volume than ever before,<sup>109</sup> the tools for producing new media have also been made more readily accessible to the layperson.<sup>110</sup> At the same time, copyright terms are lengthening and

---

104 *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 811 (9th Cir. 2003).

105 *Id.* at 818.

106 *Id.*

107 *Id.*

108 *White v. W. Pub. Corp.*, 29 F. Supp. 3d 396, 398–99 (S.D.N.Y. 2014).

109 See Richard Wray, *Media Consumption on the Increase*, GUARDIAN (Apr. 18, 2010), <http://www.theguardian.com/business/2010/apr/19/media-consumption-survey>.

110 See Lauren Britton, *Democratized Tools of Production: New Technologies Spurring the Maker Movement*, TECH. & SOC. CHANGE GRP. (Aug. 18, 2014),

formalities for obtaining protection are being abandoned.<sup>111</sup> These factors have created an influx of both new and copied content, with the numbers rising exponentially every day.<sup>112</sup> As new technology and traditional notions of copyright clash, it has become clear that one or both must change. Google is no stranger to this struggle, as it receives tens of millions of copyright removal requests for its search results every month.<sup>113</sup>

Digital archives are certainly a new innovation on the traditional library system, and those who were responsible for the development of fair use doctrine would have likely never even considered the possibility of their existence. This raises the question of whether digital archives are within the intended scope of fair use. Cultural historian and media scholar, Siva Vaidhyanathan, says that they are not, and argues that the Google Library Project is so far beyond the scope of traditional copyright that it threatens the foundation of copyright law, going so far as to accuse Google of “exploiting the instability of the copyright system in a digital age.”<sup>114</sup>

A counter to the assertion that digital archives have a disruptive effect on copyright law is the fact that the rise of search engine websites has not horribly disrupted our copyright system (the large number of infringement claims against Google notwithstanding). Search engines are the internet’s single largest source of web traffic,<sup>115</sup> and are essentially digital archives, as they conduct vast amounts of data copying without the express permission of website authors.<sup>116</sup> Fair use protection was granted to a search engine of copyrighted pictures in the *Kelly* case,<sup>117</sup> and it could be argued that there is little reason to think a search engine of copyrighted books should be any different on a conceptual level. Unless a court distinguishes one from the other, if a court concludes that Google’s book archive is not a privileged fair use, then search engines’ scanning of millions of websites might not be considered fair use either.<sup>118</sup> Given the necessity of search engines to people’s use of the internet, a ruling that digital archives are not protected by fair use could cause greater societal disruptions than a ruling that they are protected by fair use could ever cause to copyright law.<sup>119</sup>

The problem with comparing digital archives to search engines is that, while they are related and operate similarly, their differences are significant

---

<http://tascha.uw.edu/2014/08/democratized-tools-of-production-new-technologies-spurring-the-maker-movement/>.

111 Band, *supra* note 92.

112 *Id.*

113 GOOGLE TRANSPARENCY REPORT, GOOGLE, <https://www.google.com/transparencyreport/removals/copyright/> (last visited Feb. 18, 2016).

114 Siva Vaidhyanathan, *The Googlization of Everything and the Future of Copyright*, 40 U.C. DAVIS L. REV. 1207, 1210 (2006).

115 *Why Are Search Engines So Important?*, AXEMEDIA, <http://www.axemedia.com/FAQ/Why-Are-Search-Engines-Important.html> (last visited Feb. 18, 2016).

116 Jonathan Band, *The Google Library Project: Both Sides of the Story*, 10 INFO. OUTLOOK 35, 48 (2006).

117 *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818 (9th Cir. 2003).

118 *Id.*

119 See Vaidhyanathan, *supra* note 114 at 1230.

enough to make them legally distinct. In terms of utility, search engines are the most important types of websites on the internet, but the internet would continue to function with little disruption if Google Books were to disappear tomorrow. Search engines also receive the benefit of an “implied license” to copy copyrighted works, while digital archives do not.<sup>120</sup> This implied license for search engines derives from the fact that when a copyright owner decides to place a work on a website, he or she already knows that automated “spider” programs sent out by search engines will copy the work into a search index, and has thus given the search engine an implied license to do so.<sup>121</sup> However, an author of a book does not likely anticipate the book being copied and placed on a website when published, meaning they have not granted an implied license.<sup>122</sup> The dissimilarities between digital archives and regular search engines allow them to be analyzed legally independent of each other. However, their similarities do still support the premise that they have a similar effect, or lack thereof, on copyright law in general.

Ultimately, recent upheavals in the realm of copyright law are a product of the simultaneously threatening and inspiring new technologies of the digital age. While digital archives are a product of these new technologies, their direct effect on the stability of copyright law is likely similar to their widely-embraced search engine cousins. It may be the case that the only way to reconcile copyright law with new technology is through alteration of copyright law itself.<sup>123</sup> While the nature of these changes are interesting and important to discuss, they lie outside the scope of this particular Note, which asks whether digital reproductions of copyrighted works for archival purposes are in line with copyright law as it currently stands. And to the extent that digital archives are held to be protected by fair use—and several courts have held that they are—they are in line with our current copyright regime.

### C. APPLICATION OF FAIR USE TO LARGE CORPORATIONS

Another criticism of applying fair use to digital archives like Google Books is that, although fair use is important to the functioning of our copyright system, it should only apply to small incidental activities, and not be extended to the enterprise level.<sup>124</sup> The four fair use factors do not explicitly focus on the attributes of the entity doing the copying, aside from considering the purposes of their use, but it is reasonable to argue that fair use was meant to protect private individuals from large corporations, and not the other way around. One instance of copying from a single individual is not likely to materially affect the market for a copyrighted work, but

---

120 Band, *supra* note 116 at 40.

121 *Id.*

122 *Id.* at 41.

123 See Jessica Litman, *Revising Copyright Law for the Information Age*, 75 OR. L. REV. 19 (1996).

124 Band, *supra* note 92.

copying by large entities like Google has a greater effect on the market for a copyrighted work, in part because it creates a greater risk of inspiring copycats. In fact, Yahoo and Microsoft have already become involved in their own digitizing projects following the success of Google Books.<sup>125</sup> Even though Microsoft dropped out of its partnerships with the British Library and the Internet Archive, their efforts to digitize information are still ongoing.<sup>126</sup> Even if a single digital archive structures itself in such a way that does not violate copyright, a combination of dozens, if not hundreds of other digital archives may stretch fair use too thin and open the gates for innumerable potentially-infringing archives.

This distinction between large corporations and private individuals is likely not of great significance. While the question has not been explicitly addressed by any courts or statutes, the fact that courts have granted fair use protection to corporations like Google, Arriba Soft, and West Publishing indicates that courts do not tend to assign much weight to the characteristics of the entity pleading a fair use defense.<sup>127</sup> There is also evidence that Google Books is unlikely to elicit many significant competitors.<sup>128</sup> Google has had a five-year head start to scan millions of books and form relationships with libraries and publishers, giving it an effective monopoly on the market for digitally-reproduced copyrighted books that no competitor will likely be able to compete with on the same scale.<sup>129</sup> In fact, Google would likely sue any competitor who tried to archive copyrighted works whose authors cannot be found, which covers a significant portion of Google Books's database.<sup>130</sup> There are other digital archives besides Google Books, such as the Internet Archive, HathiTrust, or the Open Content Alliance,<sup>131</sup> but these sources limit themselves to works in the public domain, which does not pose a threat to copyright law. While Google's business practices may expose them to potential antitrust litigation in the future, they do not invalidate its claim to fair use protection.

---

<sup>125</sup> *About: What is Open Content Alliance?*, OPEN CONTENT ALLIANCE, <http://www.opencontentalliance.org/about/> (last visited Feb. 19, 2016).

<sup>126</sup> Jeremy Kirk, *Without Microsoft, British Library Keeps on Digitizing*, NETWORK WORLD (May 30, 2008), <http://www.networkworld.com/article/2280153/data-center/without-microsoft--british-library-keeps-on-digitizing.html>; see INTERNET ARCHIVE, <https://archive.org/> (last visited Feb. 19, 2016).

<sup>127</sup> See *Authors Guild, Inc. v. Google Inc.*, 954 F. Supp. 2d 282 (S.D.N.Y. 2013); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003); *White v. W. Pub. Corp.*, 29 F. Supp. 3d 396 (S.D.N.Y. 2014).

<sup>128</sup> Geoffrey Nunberg, *Google's Book Search: A Disaster for Scholars*, CHRONICLE REVIEW (Aug. 31, 2009), <http://chronicle.com/article/Googles-Book-Search-A/48245/>.

<sup>129</sup> *Id.*

<sup>130</sup> Ryan Singel, *Critics: Google Book Deal a Monopoly, Privacy Debacle*, WIRED (June 2, 2009), [http://www.wired.com/2009/06/google\\_books/](http://www.wired.com/2009/06/google_books/).

<sup>131</sup> The Open Content Alliance has its own set of principles that it operates by that align with the factors described earlier for a digital archive to satisfy the four fair use factors. *A Call to Participate in the Open Content Alliance*, OPEN CONTENT ALLIANCE, <http://www.opencontentalliance.org/participate/> (last visited Feb. 19, 2016).



## D. COPYRIGHT AS A “COPY RIGHT”

Another criticism of the effect of digital archives on our copyright system is that it takes away the right of copyright holders to be the sole distributor of their works.<sup>132</sup> The Copyright Act of 1976 vests in copyright holders the exclusive right to “reproduce the copyrighted work in copies or phonorecords.”<sup>133</sup> By allowing digital archives to scan books and other copyrighted works, copyright may cease to exist as a “copy right” and instead morph into merely a “commercial distribution right.”<sup>134</sup>

While it is true that copyright has traditionally left the sole right to copy original works to copyright holders, the law was crafted this way because at the time of its creation it was the only useful benchmark for deciding when a copyright owner's rights had been unlawfully violated.<sup>135</sup> This is because, given the difficulty and cost of making copies at the time, the number of copies of a work in circulation was relatively easy to track.<sup>136</sup> These measures are not as useful today, as unauthorized copies are created with such ease and frequency that they are often impossible to find, let alone count, so policing their reproduction and distribution has become a herculean task.<sup>137</sup> Even when a clearly-infringing database website, such as The Pirate Bay, is identified and taken down, a new one can pop back in its place like the head of a hydra.<sup>138</sup> When it comes to creating copies of copyrighted works, digital archives might as well be one grain of sand in the beach that is the Internet.

The only way to stem the creation of unauthorized copies is to execute tight control over the technologies used to create digital copies of works, but this is not a desirable solution. First, these tools are so prevalent and easy to acquire that it would be impossible to hold a monopoly on their implementation.<sup>139</sup> Second, even if copyright holders could control access to this technology, editing and copying software has plenty of entirely legal and socially beneficial uses, not the least among them being the creation of wholly new works. Restricting access to these technologies would be an unfair monopolization of the modern tools of creation. Ultimately, the exclusive right of copyright holders to copy their works is subordinate to fair use protection,<sup>140</sup> so to the extent that digital archives are held to be protected by fair use, an argument from the exclusive right to copy will not be effective.

---

132 Litman, *supra* note 123 at 23.

133 17 U.S.C. § 106 (2002).

134 Vaidhyanathan, *supra* note 114 at 1218.

135 Litman, *supra* note 123 at 36.

136 *Id.*

137 *Id.*

138 Ernesto, *The Pirate Bay is Back Online!*, TORRENT FREAK (Jan. 31, 2015), <http://torrentfreak.com/pirate-bay-back-online-150131/>.

139 See Britton, *supra* note 110.

140 17 U.S.C. § 107 (1992).

## E. 17 U.S.C. § 108: REPRODUCTIONS BY LIBRARIES AND ARCHIVES

Aside from 17 U.S.C. § 107, which grants fair use protection, there is another statute that may be more relevant to digital archives: 17 U.S.C. § 108.<sup>141</sup> Section 108 states that it is not an infringement of copyright for a library or archive to reproduce or distribute no more than one copy of a work that is not musical, pictorial, graphic, or sculptural, so long as the reproduction is made without any direct or indirect purpose of commercial advantage, the collections of the libraries or archives are either open to the public or at least to persons doing research in a specified field, and the reproduction provides notice that the work is protected by copyright.<sup>142</sup> The section raises the allowable reproductions to three copies if the reproductions are done for the purposes of preserving copyrighted works, depositing them for research use in another library or archive (provided it is not otherwise made available to the public), or replacing damaged, lost, or stolen works (provided the library or archive determines that an unused replacement cannot be obtained at a fair price).<sup>143</sup>

At first, this section appears to be a good fit for digital archives, as it gives them permission to provide reproductions of copyrighted works under certain circumstances. However, the requirements imposed by section 108 are too restrictive for a digital archive to rely on it alone. As it only allows one to three copies of a work at a time, it would not be able to handle the massive number of requests archives receive. Granted, section 108 allows for multiple reproductions of a work when each instance is isolated, unrelated, and will not be used by multiple people,<sup>144</sup> but this means that an archive relying on section 108 would not be useful to people who are conducting research in a group, studying a work over a long period of time, or who need access to a work multiple times to get as much information as possible out of it. Ultimately, while section 108 may be a viable statute to provide some additional protection for a digital archive, its restrictions make it limited in usefulness to an archive as a whole.

## F. PERMISSION AS A REQUIREMENT

One of the arguments made by the plaintiffs of the *Google Case* is that Google should seek the permission of copyright holders before making copies of their works and displaying them. Copyright owners have three choices upon learning that their books have been scanned by Google Books: (1) participate in the Partner Program, entitling them to share in the revenue derived from displaying pages of their works in response to user queries; (2) let Google scan the book under the Library Project and display snippets in response to user queries; or (3) opt out of the Library Project, in which case Google promises that it will not scan the book and will exempt

---

141 17 U.S.C. § 108 (2005).

142 *Id.*

143 *Id.*

144 *Id.*

it from searches.<sup>145</sup> Currently, the burden lies on copyright holders to opt out of Google Books, rather than on Google to ask copyright holders for their permission before scanning their works.<sup>146</sup> Copyright holders want this burden to shift to digital archives on the grounds that, by asking for permission before copying copyrighted content, digital archives would skip having to rely on fair use doctrine, reducing litigation costs and promoting fairness to copyright holders.<sup>147</sup>

The problem with shifting the burden in this way is that it would be costly and impractical. It can be difficult to determine who owns the copyright for any particular work, especially if it is old and out of print.<sup>148</sup> Copyright holders are in a better position to determine who has the right to authorize the digitization of copyrighted works, so it is more efficient to allocate the burden of doing so to them.<sup>149</sup> Even when copyright holders could be easily found, the transaction costs of obtaining all the requisite permissions would render any such project economically impossible.<sup>150</sup> Thus, it is necessary for a digital archive to rely on fair use protection, rather than seek the permission of copyright holders, to incorporate works into their collections.

In response to this argument, copyright holders can claim that the process of getting permission is far from impossible. Archives could negotiate with publishers for blanket licenses to the rights to display works whose copyright holders are known, avoiding both the difficulty of finding copyright holders and the transaction costs of obtaining permission from each one individually. Blanket licenses like these have been widely used between the radio and music industry for years,<sup>151</sup> and radios are not so different from digital archives conceptually. Both involve independent providers (radio stations and digital archives) who, among other things, help draw the attention of potential audiences towards copyrighted works (music and books) by providing samples of those copyrighted works (singles from albums and snippets of books), usually for free. If a radio station were to play a hit new song without permission, it would certainly be subject to copyright liability, so why not digital archives? Companies like Google could easily negotiate and pay for any such agreement, and the fact that they have not is a sign that they simply wish to attract visitors and advertisers without considering the original authors.

What is it that differentiates digital archives from something like radio stations that they may operate without requiring permission as they do? For one, radio stations typically play whole songs, while archives like Google Books only display snippets of whole books to users, which were held to

---

145 Band, *supra* note 116, at 36.

146 *Id.*

147 Lessig, *supra* note 44.

148 *Id.*

149 Band, *supra* note 116, at 46.

150 *Id.* at 52.

151 Marshall Brain, *How Music Licensing Works*, HOW STUFF WORKS (22 Aug. 2003), <http://entertainment.howstuffworks.com/music-licensing3.htm>.

not be market substitutes for books in the *Google Case*.<sup>152</sup> Displaying books in an archive is also useful for research and scholarly purposes, while playing songs on the radio is meant primarily for entertainment purposes. Archives, by their nature, also strive for and thrive on comprehensiveness, and having to negotiate licenses jeopardizes that feature. Ultimately, negotiating licenses is one road that digital archives could end up taking in the future, but at the moment there is nothing to indicate that they are obligated to take it now.

While archives likely do not need to ask the permission of copyright holders to display their works, it makes sense from a moral and policy standpoint for archives to respect the wishes of owners who voice their objections to their works being included, and to provide clear and practical forums for authors to voice their objections. As digital archives can have a positive effect on the market for an original work,<sup>153</sup> those who refuse to have their content copied out of fear of the financial effects should eventually recant their objections once they come to realize this, leaving only those few who object for other personal reasons. The presence of conspicuous forums for objection are important, because without them copyright holders may have little recourse other than litigation to voice their disapproval. To be fair, archives could also avoid the need for fair use entirely by limiting themselves to older works whose copyright holders are hard to find, or at least provide even more limited access to newer works whose copyright owners are known.

#### G. FAILURE TO MEET ITS GOALS

Although Google Books has become one of the world's top digital libraries, some have questioned the quality of the service it provides and whether it is truly capable of achieving the purposes the court described in the *Google Case*.<sup>154</sup> For example, a writer from the American Historical Association found the quality of Google Books' scans to be disappointing due to basic scanning errors in many of the documents and incorrect results to searches for specific pages.<sup>155</sup> Scanning errors are a problem for digital archives because their existence is exacerbated by their use as a tool for obtaining accurate information, aggregating data, and providing a searchable index. Another problem with the Google Books system is its inaccurate metadata, where incorrect dates of publication and illogical genre categorizations are recurring issues.<sup>156</sup> Once again, without accurate metadata, it becomes difficult—if not impossible—to find proper resources, which is the primary purpose of a digital archive.<sup>157</sup> In Google's defense, it

---

152 *Authors Guild, Inc. v. Google Inc.*, 954 F. Supp. 2d 282, 293 (S.D.N.Y. 2013).

153 *Id.*

154 *Authors Guild*, 954 F. Supp. 2d at 288.

155 Robert B. Townsend, *Google Books: Is it Good for History?*, AM. HIST. ASS'N (Sept. 2007), <http://www.historians.org/publications-and-directories/perspectives-on-history/september-2007/google-books-is-it-good-for-history>.

156 *Id.*; Nunberg, *supra* note 128.

157 Nunberg, *supra* note 128.

has pointed out that many of the incorrect dates and classifications originate from the providers themselves, but a number of them are also due to inadequate efforts by Google to automatically extract publication dates from scanned texts.<sup>158</sup> Google's views on copyright have also been criticized, in that it is too expansive towards recent works while being too narrow towards works actually in the public domain.<sup>159</sup> Without rigorous quality requirements and procedures, a digital archive loses its effectiveness, and correcting these mistakes will prove difficult for Google given the rapid pace at which it adds new books to its library.<sup>160</sup>

Another criticism leveled against Google Books is that its search engine functions interfere with its archival functions. Siva Vaidhyanathan argues that "the fundamental error that Google and many of its supporters make is assuming that Google's algorithms and selections are somehow neutral, that they do not betray certain biases in them."<sup>161</sup> Vaidhyanathan points out that Google Books often generates ridiculous results for simple searches and cannot screen out bad results very well.<sup>162</sup> Vaidhyanathan also criticizes Google for not offering simple information-seeking training to its customers to ensure that users are able to utilize the correct processes to generate and recognize the sources they need.<sup>163</sup> Another writer argues that the traditional "googling" process of entering strings of keywords to find specific information is not a very efficient way to search for books based on metadata.<sup>164</sup> Given the faults with Google's metadata, it would be exceedingly difficult to, for example, track the historical development of an area of literature when dates of publication are woefully inaccurate.<sup>165</sup>

As Google Books is seen as the world's top digital library, any widespread defaults present within its system are likely to reflect poorly on the quality of digital libraries as a whole. Of course, incorrect information can always be corrected and better quality assurance/control methods can be put in place, but the inaccuracies may be too many for Google's archive to ever truly work as intended. If Google, with all of its famed innovation and technological resources, is open to so many mistakes, what hope do future digital archives have of succeeding where Google has failed? There are other digital archives, such as the Internet Archive, HathiTrust, and the Open Content Alliance that may develop themselves to be more ideal services than Google Books, but only time will tell if they can accomplish this.<sup>166</sup> Accuracy of information is essential to the functioning of a digital archive, but it seems fatalistic to argue that Google or other digital archivists will never be able to adequately correct the problems they

---

158 *Id.*

159 *Id.*

160 *Id.* (explaining that Google adds thousands of books to its library every day).

161 Vaidhyanathan, *supra* note 114, at 1228.

162 *Id.*

163 *Id.*

164 Nunberg, *supra* note 128.

165 *Id.*

166 *Id.*

currently face. This service is relatively new, and there are bound to be bugs in the system as the technology develops, but it remains to be seen how or when these kinks will be fixed.

#### H. PUBLICLY VS. PRIVATELY RUN

Much of this Note has focused on what digital archives are or should be, but not enough focus has been on who should run digital archives. The fact that one of the biggest and most widely-used digital archives of copyrighted works, Google Books, is run by a private company raises some legitimate concerns.<sup>167</sup> As a private company, Google has the right to make business decisions that favor itself and its investors, rather than what is best for society and copyright holders.<sup>168</sup> Given that Google Books has already been the subject of major litigation,<sup>169</sup> Google may one day decide that the service is more trouble than it is worth and shut it down, despite the social benefits such an archive brings. Even if Google pledges to never close down the system even if it becomes unprofitable, who is to say that Google itself will never go out of business one day and have to close down Google Books?<sup>170</sup> If one of the purposes of a digital archive is the preservation of culture and knowledge, having those archives run by a public library or other government-sponsored institution may make more sense.<sup>171</sup> Even if most Americans have a very high approval rating of Google,<sup>172</sup> the fact is that public libraries are more open with their intentions and standards, more stable than private entities, and less likely to manipulate or exploit the information they preserve.<sup>173</sup>

There are also concerns about the privacy of users, as Google's practices of gathering information on its users rival that of any governmental or corporate entity.<sup>174</sup> Google retains information related to what users read, search through, and buy, which could be combined with a user's web history, general searches, and even YouTube usage.<sup>175</sup> Google is very tight-lipped about its internal data-handling practices and what it uses its data for,<sup>176</sup> but it could range anywhere from targeted advertising to surveillance.<sup>177</sup>

---

167 Vaidhyanathan, *supra* note 114, at 1220.

168 *Id.*

169 Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282 (S.D.N.Y. 2013).

170 Vaidhyanathan, *supra* note 114, at 1220–21.

171 *Id.* at 1220 (In the words of Siva Vaidhyanathan: “Should we entrust our heritage and collective knowledge to a company that has been around for less than a decade?”).

172 Damla Ergun, *Americans Find Google Favorable, Twitter Not So Much: Study*, ABC NEWS (Apr. 5, 2012), <http://abcnews.go.com/blogs/politics/2012/04/searching-for-the-king-of-tech-in-popularity-google-lays-claim/> (stating in 2012, the public had an 82% favorable opinion of Google, with 53% expressing strongly favorable opinions).

173 Vaidhyanathan, *supra* note 114, at 1220–21.

174 Amir Efrati, *Google's Data-Trove Dance*, WALL ST. J. (July 30, 2013), <http://www.wsj.com/articles/SB10001424127887324170004578635812623154242>.

175 Singel, *supra* note 130.

176 Efrati, *supra* note 174.

177 Vaidhyanathan, *supra* note 114, at 1219.

Given these concerns about privacy and privatization, a publicly-run digital archive would be preferable to a privately-run one. Public libraries are more trustworthy and less profit-seeking than private corporations. Granted, post-Snowden America likely expects that the government would be just as likely to gather and use user information, but another ideal aspect of an archive would be that it refrains from misusing data on its users in the first place.

#### I. THE IDEAL DIGITAL ARCHIVE SYSTEM

This Note has placed a lot of focus on Google Books, due to it being one of the largest and most influential digital archives of copyrighted material currently on the Internet. However, as some of the criticisms outlined above have shown, the system it employs is far from perfect. By taking into account the nature and purpose of digital archives, the factors of fair use, and positive and negative aspects of currently-existing digital archives, I believe a more effective archive system can be put in place that accomplishes all of its desired goals, while minimizing its faults.

One viable option for an ideal digital archive would avoid the need for fair use protection all together and only limit itself to works already in the public domain. This is the realm that archives like the Internet Archive, HathiTrust, or the Open Content Alliance occupy, and it is easy to see why this would be a much easier option. By limiting oneself to only works in the public domain, no copyrights are violated, meaning no litigation is necessary. However, easier is not always better, as there are plenty of benefits from having a digital archive that includes copyrighted material. First, comprehensiveness facilitates the research and metadata-gathering functions of archives. Second, a well-structured archive can draw attention to relevant works that are for sale and still in print, turning the archive's users into potential consumers, and increasing the sales of those works. Thus, an ideal archive would strive for comprehensiveness in the works that it includes, regardless of factors like time of publication, subject matter, or whether the work is copyrighted or in the public domain. However, this drive must be tempered by the bounds of copyright law.

For an archive that includes copyrighted material to be within the bounds of copyright law, it must satisfy the elements of fair use protection. As indicated by cases like the *Google Case*, *UMG Recordings v. MP3.com*, *Kelly v. Arriba Soft Corp.*, and others, an archive is likely to satisfy the elements of fair use protection if it implements a transformative use for the works it indexes, such as creating searchable indices to assist in research, rather than simply serving as a repository for those works; consists primarily of published, out-of-print works of non-fiction; limits the portions of a work that is viewable to a general user to the smallest amount it can display while still helping the user determine whether a source is relevant to what they are looking for; and serves as a tool for identifying relevant works, with links of outlets where users can purchase said works, rather than serving as a market replacement for said works.

Of the elements described above, the one I believe requires the most diligence for an archive is the way it limits the portions of copyrighted works that are viewable to a general user, because it is this element that truly sets a proper digital archive apart from a mere repository of books. Providing the bare minimum amount of text from a copyrighted work that is sufficient to be useful for a user conveys a sense that the archive's intention is simply to direct potential consumers to a work, while displaying the entire text of a work—while possibly more helpful for research purposes—oversteps the bounds of what an archive should do. Google Books allows member libraries to download whole texts of copyrighted works, but only the ones that they specifically owned and contributed in the first place, not ones contributed by other libraries. Assuming that the court in the *Google Case* was right in saying that it would be impossible for a general user to piece together the whole text of a copyrighted work from the snippets it provides, that should be sufficient to pass this element of fair use analysis. However, to the extent that an equally effective, yet more limited, system of providing snippets is possible, an ideal archive should strive for such a system.

Of course, whether something qualifies for fair use protection is a question of fact that relies on the context of an alleged dispute, but in general an archive with the features described above is likely to enjoy fair use protection. However, there are other features that go beyond the mere floor of fair use protection that an ideal archive should adopt.

First and foremost, an ideal digital archive would be run by a governmental or public entity, rather than a private one. Not only does privatization bring concerns about privacy of the information of users, but there are also longevity concerns. A private entity could easily choose to charge its users (which would reduce the number of people willing or able to use its services), discontinue its archive service if it becomes too costly or difficult to maintain, or be forced to close it down in the event of its going out of business. Public archives bring less danger of data manipulation and ensure a sense of stability that private archives do not.

Second, although an archive is not obligated to seek the permission of every copyright holder whose works it indexes, respect for the wishes of copyright holders who do not wish for their works to be included should be a priority. This means that archives should have formal and clear forums for copyright holders to object to their inclusion in an archive, and should subsequently honor the wishes of those copyright holders. However, if we should be so concerned about the wishes of copyright holders that do not want their works to be archived, why would an ideal archive include copyrighted works in the first place? The answer is simple: not all copyright holders will be opposed to having their works archived because an archive can actually be a useful tool for potential consumers to be directed towards the author's work. The primary purpose of an ideal archive would be to direct users towards sources that may be relevant for them, rather than serve as repositories or market replacements for them. Thus, the only copyright holders who would likely request their works be



removed from an archive are those that object to it for personal or philosophical reasons, rather than financial ones. These are likely to be a minority of total copyright holders, so societal efficiency is better served by shifting the burden onto that number of holders who wish for their works to be excluded.

It should be noted that the model described in this Note has been formulated in relation to digital archives of books and other literary works, rather than other forms of media. This is because books have been viewed as the ultimate repositories for knowledge and information for hundreds of years, and I wanted to contrast this somewhat with a new potential model for the preservation of knowledge and information. However, the model described in this Note likely would not be as good of a fit for a digital archive of something like music, paintings, or some other form of art, because each is a fundamentally unique medium. An ideal model for each of these mediums, if one could exist, is a subject outside the topic of this Note, but is no less interesting or important to consider.

#### IV. CONCLUSION

While digital archives bring with them the promise of greater access to information, more effective means of research, and the preservation of culture, these potential benefits may be offset by the disruptive effects they have on copyright law, the inaccuracies of their information, and the wide-reaching social and economic impacts the technology creates. Courts have held digital reproductions of copyrighted works for archival purposes to be protected by fair use doctrine, but only under certain circumstances: the access given to users must be limited to the amount necessary for users to recognize what sources are relevant without supplanting the markets for the original; the archive should consist primarily of published, out-of-print works or works in the public domain; and it must do more than simply serve as a repository for information by providing some useful data or additional tools for research.

In light of concerns about privacy and privatization, the social and scholarly benefits of digital archives would best be served by publicly-run archives, rather than archives run by private entities, but fair use law does not necessitate this. Of course, fair use cases are analyzed on a case-by-case basis, so an infinite number of factors could affect whether a digital archive will be granted fair use protection, and two otherwise similar archives may face different levels of success in courts. However, a defining feature of this new digital age we live in is increased ease and availability of information to regular people. So long as researchers are looking for new source materials or scholars are studying new subjects, there will always be a use for vast repositories of information. Whether it is accomplished with the help of fair use or some other legal doctrine, it is my prediction that complete digital archives will eventually become as common as traditional libraries are today. One can only guess how this shift will affect copyright law in general, if at all, but one can only hope that that the end result is an

778                      *Southern California Interdisciplinary Law Journal*      [Vol. 25:753

ideal balance between rewarding original creators and inspiring creative freedom in others.