

5-1-2014

# Great Expectations: The Google Books Project and the Strength of its Fair Use Defense

Brittany Danielle Glaser

Follow this and additional works at: [http://scholarship.shu.edu/student\\_scholarship](http://scholarship.shu.edu/student_scholarship)

---

## Recommended Citation

Glaser, Brittany Danielle, "Great Expectations: The Google Books Project and the Strength of its Fair Use Defense" (2014). *Law School Student Scholarship*. Paper 479.  
[http://scholarship.shu.edu/student\\_scholarship/479](http://scholarship.shu.edu/student_scholarship/479)

# **Great Expectations: The Google Books Project and the Strength of its Fair Use Defense**

Brittany Glaser

## *Abstract*

*From the advent of Gutenberg's printing-press, which moved the written word into the printed form, technological innovation has been the impetus behind much of the development and refinement of copyright law. Today, the Google technology behind moving the printed word into a digital format, has now brought about a new debate in copyright law: whether the scanning of physical books in order to create a digital database is an infringement of copyright or is sanctioned under the Fair Use Doctrine. In an age where multiple e-reader manufacturers, authors, publishers, and distributors are vying for market share, the existence of unlicensed scans of copyrighted books is a controversial issue. Copyright policy balances upon the tension between incentivizing creativity by protecting the rights and interests of authors and providing society with access to a free flow of information. The Author's Guild v. Google Inc. case currently being litigated, provides the perfect backdrop for analysis of the traditional fair use factors in the 21<sup>st</sup> century. Google's unlicensed scanning of books in order to create their database, repurposes the creative whole of a book into metadata for research by the public without supplanting the original markets; this paper argues that this is a transformative work and thus a fair use of the copyrighted books. Despite the Author's Guild's argument for infringement, Google is likely to prevail under a fair use defense.*

I. Introduction: In the autumn of 2011, the Author’s Guild—a literary copyright advocacy group, filed a lawsuit against Google—the ubiquitous search engine company, claiming mass copyright infringement. Google had contracted with five of the world’s largest libraries to scan copies of their book collections in order to create something akin to a digital library catalogue. In the summer of 2013, Google asserted the affirmative defense of fair use of the copyrighted materials. This paper examines the application of the Fair Use Doctrine to the facts of this case, in light of the increased judicial focus on transformativeness and the advanced technology involved.

Part II presents copyright law, the rights it provides, and the policies behind it. Part III introduces the Fair Use Doctrine as an affirmative defense to infringement, its factors, its importance in relation to copyright, and its justifications. Part IV addresses transformativeness and the role it plays in a fair use determination. Judge Pierre Leval wrote a strongly influential article on the significant transformativeness must play in a fair use analysis. This guided the landmark fair use cases that followed. Part V describes the Google Books Project, its objectives, public benefits, and the process involved in its creation. Part VI iterates the Author’s Guild’s complaint against Google, regarding copyright infringement. Part VII undertakes the statutory four-factor fair use analysis of Google’s secondary use. This part is broken down into the four-factors for analysis under of the competing arguments for each one. The prolific copyright scholar, Matthew Sag, wrote two articles focusing on copyright in the Internet age. His articles provide a framework of new paradigms and terms to accommodate the analysis of new technologies in the fair use context. The subsections under the first factor make use of his

AWR  
Brittany Glaser

concepts of the expressive/nonexpressive dichotomy and reverse engineering, to examine the new genre of what he calls copy-reliant technology. Part VIII concludes that Google will likely prevail under fair use.

II. Copyright: Copyright is a branch of intellectual property law, which seeks to provide the creator of an original work exclusive rights to it, for a limited time, with the intention of enabling the creator to be compensated for his work. The utilitarian theory behind this legal concept attempts strike a balance between providing incentive to create and disabling monopolies. The latter is essential in order to facilitate the public dissemination of and access to information, which in turn spurs even more creation.<sup>1</sup>

Copyright protection provisions have been codified into statute under Title 17 of the US Code, also known as the 1976 Copyright Act.<sup>2</sup> Section 102 of this statute identifies the subject matter protectable by copyright: “Copyright protection subsists...in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device”.<sup>3</sup> Included in the realm of protectable subject matter, are literary works.<sup>4</sup> This section provides that copyright does not protect ideas, concepts, or facts; just the original, creative expression of such.<sup>5</sup> The bundle of exclusive rights<sup>6</sup> enjoyed by an owner of a copyright

---

<sup>1</sup> Matthew Sag, *The Google Book Settlement and the Fair Use Counterfactual*, 55 N.Y.L. SCH. L. REV. 19, 26 (2010).

<sup>2</sup> The Copyright Act of 1976, 17 U.S.C. §§101 et. seq.

<sup>3</sup> 17 U.S.C. §102 West (2010).

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

<sup>5</sup> *Id.*

<sup>6</sup> A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 636 (4th Cir. 2009).

AWR  
Brittany Glaser

are enumerated in section 106 and include the right to reproduce the work in copies, the right to display the work publicly, and the right to distribute or sell copies of the work.<sup>7</sup> These rights are to the exclusion of all others, unless the copyright holder agrees to license the use of the copyrighted work to a third-party for a fee.<sup>8</sup> A person who “violates any of the exclusive rights of the copyright owner, that is, anyone who trespasses into the exclusive domain by using . . . the copyrighted work without permission, is an infringer of the copyright” and subject to legal consequences.<sup>9</sup>

III. The Fair Use Doctrine: Section 107 of the Copyright Act of 1976 establishes an affirmative defense to copyright infringement, known as “fair use.”<sup>10</sup> Copyright is “intended to increase and not to impede the harvest of knowledge” and excessively broad protection would stifle, rather than advance, this objective.<sup>11</sup> Therefore, copyright owners’ rights are subject to statutory limitations.<sup>12</sup> Section 107, codifies the common-law Fair Use Doctrine, which allows the public to use not only facts and ideas contained in a copyrighted work, but also expression itself in certain circumstances.<sup>13</sup> Fair Use represents “the most important limitation on the otherwise extraordinarily broad rights granted to copyright owners under section 106.”<sup>14</sup>

---

<sup>7</sup> *Id.*

<sup>8</sup> 17 U.S.C. § 201 West (2010).

<sup>9</sup> *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417, 433 (1984).

<sup>10</sup> Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U. PA. L. REV. 549, 558 (2008).

<sup>11</sup> Pierre Leval, *Toward A Fair Use Standard*, 103 HARV. L. REV. 1105, 1107 (1990).

<sup>12</sup> *iParadigms, LLC*, 562 F.3d at 637.

<sup>13</sup> *Id.* at 636.

<sup>14</sup> Beebe, *supra* note 10, at 558.

While section 107 offers no definition of “fair use,” it does “enumerate four broadly worded factors that courts shall consider” in determining whether a use is “fair” and thus noninfringing:<sup>15</sup>

- (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 market for or value of the copyrighted work.<sup>16</sup>

Some have said that these factors are all that “stand between us and the tyranny of copyright.”<sup>17</sup>

In addition to these four factors, the section also contains a preamble that identifies examples of fair “purposes,” that do not constitute an infringement of copyright.<sup>18</sup> These are typically for educational or illuminating purposes, including “criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research.”<sup>19</sup> Fair use adjudication requires case-by-case analysis rather than bright-line rules.<sup>20</sup> These permissible uses provide only general guidance.<sup>21</sup> The four statutory factors are to be explored and weighed together in light of

---

<sup>15</sup> *Id.*

<sup>16</sup> 17 U.S.C. §107 West (2010).

<sup>17</sup> Beebe, *supra* note 10, at 558.

<sup>18</sup> 17 U.S.C. §107 West (2010).

<sup>19</sup> *Id.*

<sup>20</sup> *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 576, (1994).

<sup>21</sup> *Id.*

AWR  
Brittany Glaser

copyright's purpose of "stimulating productive thought and public instruction without excessively diminishing the incentives for creativity."<sup>22</sup>

The importance of the Fair Use Doctrine rests in the necessity of borrowing from things of the past to create a foundation for things of the future. It allows copyright law to fulfill its purpose to "promote the Progress of Science and useful Arts."<sup>23</sup> As one Supreme Court Justice said: "[i]n truth, in literature, in science and in art, there are, and can be, few, if any, things, which in an abstract sense, are strictly new and original throughout. Every book in literature, science and art, borrows, and must necessarily borrow, and use much which was well known and used before."<sup>24</sup> Section 107 of the Copyright Act recognizes "that science and art generally rely on works that came before them and rarely spring forth in a vacuum" and fair use limits the rights of a copyright owner regarding works that build upon, reinterpret, and reconceive existing works.<sup>25</sup> There is no such thing as a wholly original thought or invention; each advance stands on building blocks fashioned by prior thinkers.<sup>26</sup> Consequently, the monopoly that copyright protection produces, has the potential to strangle the creative process and the development of new ideas out of old.<sup>27</sup> The Fair Use Doctrine protects such secondary creativity and provides subsequent authors with the breathing space to make their own contributions by adding to,

---

<sup>22</sup> Leval, *supra* note 11, at 1110-11.

<sup>23</sup> *Campbell*, 510 U.S. at 575.

<sup>24</sup> Michael D. Murray, *What is Transformative? An Explanatory Synthesis of the Convergence of Transformation and Predominant Purpose in Copyright Fair Use Law*, 11 CHI.-KENT J. INTELL. PROP. 260, 261 (2012).

<sup>25</sup> *Id.*

<sup>26</sup> Leval, *supra* note 11, at 1109.

<sup>27</sup> Murray, *supra* note 24, at 272-73.

AWR  
Brittany Glaser

reusing, or reinterpreting the facts and ideas embodied in the original work, without fear of punishment.<sup>28</sup>

Fair use benefits the public not just by maintaining the free flow of information, ideas, and facts, but also by keeping the costs to create low and helping to integrate the law with new technology. The balancing function of the fair use doctrine is critical when copyright owners are purely commercially motivated, “because licensing and other private ordering mechanisms do not provide a solution for high exchange costs, high information costs, and strategic behavior.”<sup>29</sup> Exclusive rights raise the cost for consumers to acquire those works, and for subsequent authors to create further expressive works.<sup>30</sup> Therefore, the fair use doctrine is particularly important in situations where the costs of obtaining permission outweigh the benefits of the use.<sup>31</sup> Additionally, as technology advances, it creates new vehicles for creative expression and “exposes latent ambiguities within existing doctrines.”<sup>32</sup> Fair use works to “recalibrate copyright law and help adapt it to the new technologies of reproduction and communication” as well as changes in market conditions.<sup>33</sup>

IV. Transformativeness: After years of fielding fair use cases, Judge Pierre Leval wrote an article identifying the essential role transformativeness plays in the fair use of prior copyrighted works. Referential uses of earlier works present a particular problem because “certain references are held to further the goals of copyright and the First Amendment (news, comment and

---

<sup>28</sup> Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U.L. REV. 1607, 1630 (2009).

<sup>29</sup> Sag, *supra* note 1, at 26.

<sup>30</sup> Sag, *supra* note 28, at 1630.

<sup>31</sup> *Id.*

<sup>32</sup> *Id.*

<sup>33</sup> *Id.*



AWR  
Brittany Glaser

criticism, parody), and others hinder the goals of copyright (references that exploit the creative content of original works).”<sup>34</sup> If a reference to copyrighted work is made “merely to avoid the drudgery in working up something fresh” the claim of fair use would not apply.<sup>35</sup> If, however, a secondary use adds value to the original—“if [copyrightable expression in the original work] is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings”—it epitomizes the very type of activity that the fair use doctrine intends to protect for the enrichment of society.”<sup>36</sup> The use of copyrighted work can be transformative in function, meaning, or purpose without altering or actually adding to the original work.<sup>37</sup> Over the years, the transformative test has become the defining standard for fair use.<sup>38</sup> The more transformative the new work, the less will be the significance of the other factors that may weigh against a finding of fair use.<sup>39</sup> Transformative uses allow for the creation of new works from old when the function is modified from the original.<sup>40</sup>

V. The Google Books Project: The Google Books Project is a “self-appointed mission to organize the world's information and make it universally accessible and useful.”<sup>41</sup> Its creators liken it to a digital Library of Alexandria<sup>42</sup>—only one that won’t burn down. Such an archive preserves a historical record about books and ultimately about our culture. The project aims to make the contents of “over 9 million books—“the entire catalog of some of the world's most

---

<sup>34</sup> Murray, *supra* note 24, at 272.

<sup>35</sup> *Id.*

<sup>36</sup> *Id.*

<sup>37</sup> A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 639 (4th Cir. 2009).

<sup>38</sup> Murray, *supra* note 24, at 272-73.

<sup>39</sup> Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

<sup>40</sup> Sag, *supra* note 28, at 1646.

<sup>41</sup> *Id.* at 1620.

<sup>42</sup> *Id.*

AWR  
Brittany Glaser

prestigious and extensive libraries—searchable by anyone with an Internet connection.”<sup>43</sup> To create this searchable database, Google began the process of scanning vast collections of books, one page at a time, into a digital index.<sup>44</sup> This scanning is a necessary, intermediate step;<sup>45</sup> in order for Google to be able to recontextualize and transform the original information into something new, it must first *have* that original information.

Google’s optical character recognition technology generates machine-readable text and compiles a digital copy of each book.<sup>46</sup> It analyzes each scan and creates an overall index of all scanned books. The index links each word or phrase appearing in each book with all of the locations in all of the books in which that word or phrase is found.<sup>47</sup> Users simply enter search terms and an algorithm searches the database.<sup>48</sup> The indexing allows a search for a particular word or phrase to return results that include the most relevant books in which the word or phrase is found.<sup>49</sup> Google does not allow users to access the entire contents of any book, or even an entire page of any book, unless the book is known to be in the public domain.<sup>50</sup> Instead, the results provide 3 short “snippets” from the book.<sup>51</sup> A user can click on a particular result to be directed to an “About the Book” page, which provides the user with bibliographic data.<sup>52</sup> It also

---

<sup>43</sup> *Id.*

<sup>44</sup> JEFF JOHN ROBERTS, *BATTLE FOR THE BOOKS* 31 (Gigaom Books 2012).

<sup>45</sup> *Sag*, *supra* note 28, at 1608.

<sup>46</sup> Defendant Google Inc.’s Motion for Summary Judgment at 4, *Author’s Guild v. Google Inc.*, 770 F. Supp. 2d 666 (S.D.N.Y. 2011) No. 05 Civ. 8136(DC).

<sup>47</sup> *Id.*

<sup>48</sup> *Id.*

<sup>49</sup> *Id.*

<sup>50</sup> *Sag*, *supra* note 28, at 1620

<sup>51</sup> *Id.*

<sup>52</sup> Defendant Google Inc.’s Motion for Summary Judgment, *supra* note 46, at 5.

AWR  
Brittany Glaser

presents links to online booksellers and nearby libraries where the user can obtain the book.<sup>53</sup>

Google considers this project to be “a technological middle ground between the utopia of free access to the world's best libraries and the reality that authors and publishers must be paid or they will lose some incentive to write and publish books.”<sup>54</sup>

Indeed, the Google Books Project has created many public benefits. It is a powerful reference tool, and has become so important for researchers and librarians that it has been integrated into the educational system -- it is taught as part of the information literacy curriculum to students at all levels.<sup>55</sup> The Project provides a literary record of a collection of tens of millions of books, which has greatly promoted a type of research referred to as "data mining" or "text mining."<sup>56</sup> Researchers can examine word frequencies, syntactic patterns, and thematic markers to consider how literary style has changed over time.<sup>57</sup> Additionally, the digitization technology that Google has developed for the Project, has the potential for future applications for the blind.<sup>58</sup> Such digitization facilitates the conversion of print books into audio and Braille formats, increasing access for a traditionally underserved population.<sup>59</sup> It also helps to preserve older books, many of which are out-of-print and falling apart, giving new life to historical literary gems.<sup>60</sup>

---

<sup>53</sup> Sag, *supra* note 28, at 1620

<sup>54</sup> Emily Anne Proskine, *Google's Technicolor Dreamcoat: A Copyright Analysis of the Google Book Search Library Project*, 21 BERKELEY TECH. L.J. 213, 227 (2006)

<sup>55</sup> Brief for American Library Ass'n et al. as Amici Curiae Supporting Defendant-Appellant, at 7, Authors Guild, Inc. v. Google Inc., 721 F.3d 132 (2d Cir. 2013) (No.12-3200-cv).

<sup>56</sup> Brief for Digital Humanities and Law Scholars as Amici Curiae Supporting Defendant-Appellant, at 1, Authors Guild, Inc. v. Google Inc., 721 F.3d 132 (2d Cir. 2013) (No.12-3200-cv).

<sup>57</sup> Brief for American Library Ass'n et al., *supra* note 55, at 5-6.

<sup>58</sup> Defendant Google Inc.'s Motion for Summary Judgment, *supra* note 46, at 9.

<sup>59</sup> *Id.*

<sup>60</sup> JEFF JOHN ROBERTS, *BATTLE FOR THE BOOKS 27* (Gigaom Books 2012).

AWR  
Brittany Glaser

VI. The Author's Guild Complaint: The Google Book Project has been scrutinized since its conception. Recently, the Author's Guild re-launched its lawsuit against Google, on behalf of literary copyright holders. Owners of copyright in print books enjoy the exclusive rights to make and distribute copies of that work, to "prepare derivative works based upon the copyrighted work," and to display and perform the work publicly.<sup>61</sup> In order to prove infringement, "a copyright holder must demonstrate ownership of the copyright and establish that the alleged infringer copied original elements of the work."<sup>62</sup> The original work must be fixed in a tangible medium, but the copy need not be in the same medium as the original to constitute copyright infringement.<sup>63</sup> The Author's Guild alleges that the scanning of the copyrighted books qualifies as infringement.

While the libraries contracting with Google may own the physical books that are scanned and digitized by Google, the authors and—in many instances—publishers control the copyrights of the books not yet in the public domain.<sup>64</sup> Without obtaining a license from copyright holders, Google's scanning and digitizing of books for the Google Books Project constitutes infringement. However, the affirmative defense of fair use permits copying in certain circumstances, particularly those where societal value is being added to a copyrighted work.<sup>65</sup> Ultimately, the legality of the copying by the Google Books Project rests on whether Google's use constitutes a fair use.

---

<sup>61</sup> 17 U.S.C. §106 West (2010).

<sup>62</sup> Proskine, *supra* note 54, at 222.

<sup>63</sup> *Id.*

<sup>64</sup> JEFF JOHN ROBERTS, BATTLE FOR THE BOOKS 15 (Gigaom Books 2012).

<sup>65</sup> Leval, *supra* note 11, at 1111.

VII. Four-Factor Analysis: The court will likely follow the four-factor analysis of §107 of The Copyright Act in its fair use determination, keeping eye towards transformativeness and public benefit. Each are discussed below in turn;

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

The Author's Guild points to Google's intermediate step of scanning the books and contends that this scanning is not transformative in and of itself.<sup>66</sup> Without the digitizing and indexing to produce something transformative, the purpose of these copies is no different from the purpose of the original works. However, case law on the matter will not allow them to isolate Google's act of scanning from the use to which the scans are put, which is transformative of purpose.

The first factor analyzes the purpose and character of the secondary use, including whether such use is of a commercial nature or is for nonprofit educational purposes. In performing such analysis, one of the tools that courts implement is the transformative test. This test considers "transformations in the content, context, and the predominant purpose of the original work" and whether the secondary use does in fact transform any of those facets in a manner that furthers the public policies reflected in the preamble of section 107, "namely the furtherance of the progress of the arts and the promotion of the creation of new, original

---

<sup>66</sup> Defendant's Opposition to Motion for Summary Judgment, at 9, Authors Guild, Inc. v. Google Inc., 721 F.3d 132 (2d Cir. 2013) (No.12-3200-cv).

AWR  
Brittany Glaser

expression.”<sup>67</sup> Secondary uses with archival, historical, or referential purposes fall within this category because they engender education, research, and scholarship—all of which are uses explicitly favored in the statute as furthering copyright goals.

The landmark fair use case *Campbell v. Acuff-Rose Music, Inc.*, strongly encouraged the evaluation of transformativeness.<sup>68</sup> In that case, the secondary use happened to be a song parody which changed some of the original work’s lyrics with the purpose of making a social critique of the original.<sup>69</sup> However, alteration of content or the addition of significant creative expression to the original are not the only ways to transform the purpose; secondary uses can transform the predominant purpose of the original work by recontextualization of the copied material.<sup>70</sup> The non-alteration of the contents and expression of literary works still can be justified.<sup>71</sup> Even making an exact copy of a work—including literature in verbatim—can be transformative, as long as it is placed in a new context and used in such a way that the resulting work has a completely different function and purpose from the original.<sup>72</sup> This is as opposed to simply exploiting the creative aspects of the original in a new medium but with the same purpose.<sup>73</sup>

The Google Books Project copied pieces of literature in their entirety. However, this was not done to serve the same purpose of the original works. The copying was actually just an intermediate step in producing a secondary use with a different purpose. In other words, its

---

<sup>67</sup> Murray, *supra* note 24, at 261.

<sup>68</sup> *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994).

<sup>69</sup> *Id.*

<sup>70</sup> Murray, *supra* note 24, at 261.

<sup>71</sup> *Id.*

<sup>72</sup> Beebe, *supra* note 10, at 605.

<sup>73</sup> Murray, *supra* note 24, at 261.

AWR  
Brittany Glaser

purpose “was not to convey the original work's expressive qualities to the public, but rather to enable banks of microprocessors to index the content of those works and to generate metadata about the works.”<sup>74</sup> It is transformative in that the original works are used as raw material and recontextualized to create new archival, referential, and historical functions for the information.<sup>75</sup> The creation of such new insights spur progress and enrich society and as a result, their purpose is of the exact nature contemplated by the fair use statute.

Nonexpressive Use or Purpose: This distinction between expressive and nonexpressive works “is already well recognized in copyright law--novels are protected by copyright, telephone books and other uncreative compilations of data are not.”<sup>76</sup> Such exclusion of facts and ideas from the province of copyright protection, applies with equal force to nonexpressive copying in the digital age.<sup>77</sup> Prior to the digital technology of today, “it was almost inconceivable that someone could make a nonexpressive use of a copyrighted work that involved physically copying the entire work...to illustrate through absurdum, it would be both uncommon and nonsensical to photocopy ‘Gone With The Wind’ and then to use it to light a fire.”<sup>78</sup> However, that would technically be a nonexpressive use. The purpose of a book is to be read. This purpose communicates the author’s original expression. Traditionally, if books are copied, they are done so for the copies to be read—done with the same expressive purpose of communicating the author’s original expression. Copies made with this purpose are an expressive use. As copyright expert, Matthew Sag, asserts: “expressive communication to the public implicitly defines and

---

<sup>74</sup> Sag, *supra* note 28, at 1639.

<sup>75</sup> Leval, *supra* note 11, at 1111.

<sup>76</sup> Sag, *supra* note 28, at 1625.

<sup>77</sup> *Id.* at 1630.

<sup>78</sup> *Id.*

AWR  
Brittany Glaser

limits the extent of the copyright owner's exclusive rights, acts of copying that do not communicate the author's original expression to the public do not generally constitute copyright infringement.”<sup>79</sup> The purpose of Google’s initial copying of the books, does not mirror the ostensible purpose of the original works—to be read in order to communicate the author’s original expression. Rather, Google’s purpose in copying the books is to obtain information and use it as nonexpressive content to enter into a system that will ultimately create a nonexpressive result, incapable of independently communicating the author’s original expression.

The nonexpressive use of copyrighted works contained in library catalogs, topic indices, or even plot synopses are fairly uncontroversial.<sup>80</sup> Information must be available and accessible for it to be useful to anyone.<sup>81</sup> Within any given library, for instance, there are thousands of volumes worth of information, but “unless a user has some means to locate a particular book, or better yet a particular page, that may be of interest” discovery of that information becomes untenable.<sup>82</sup> The details that library catalogue cards contain are a form of “metadata.” This term simply “refers to information about information, or data about data.”<sup>83</sup> The purpose of the Google Books Project is in fact to generate just such metadata. It is well established that “metadata is not subject to copyright protection: one can extract and reproduce facts, names, and dates from a newspaper article, or ideas and processes from an instructional text, without

---

<sup>79</sup> *Id.* at 1608-09.

<sup>80</sup> *Id.* at 1625.

<sup>81</sup> *Id.* at 1643.

<sup>82</sup> *Id.*

<sup>83</sup> *Id.*



AWR  
Brittany Glaser

communicating the author's original expression and thus without infringing the author's copyright."<sup>84</sup>

Copy-Reliant Technology: The Internet has become integral to modern existence. For many, "it is the dominant medium of communication, research, entertainment, social interaction, and political participation."<sup>85</sup> Search technology drives the Internet and allows users to sift through massive amounts of data to find the specific information that is of particular interest to them.<sup>86</sup>

Search engines direct users to particular websites based on the relationship of their search term to the index of pages maintained by the search engine provider. Typically, search engines display search results in a list that features both the title of the relevant web page and a short "snippet" or extract from the targeted web page. Thus, search engines must copy web pages to generate the data that allows them to process search requests. They also must copy web pages in order to display fragments of them as search results.<sup>87</sup>

The process behind the Google Books Project falls into a category dubbed by Sag as "Copy-Reliant Technologies."<sup>88</sup> These are technologies that copy expressive works for nonexpressive ends, and there are several legal cases that address them. The intermediate copying step has largely been approved by the courts, due to its part in transforming expressive elements into

---

<sup>84</sup> *Id.* at 1630.

<sup>85</sup> *Id.* at 1616.

<sup>86</sup> *Id.*

<sup>87</sup> *Id.* at 1617.

<sup>88</sup> *Id.* at 1607.

AWR  
Brittany Glaser

nonexpressive uses. Copy-reliant technologies do not read, understand, or enjoy the expressive content of copyrighted works, nor do they deliver these works directly to the public.<sup>89</sup> They do, however, necessarily copy them in order to feed into various digital indices—a nonexpressive use.<sup>90</sup>

Prime examples of copy-reliant technology cases involve search engine caches and query results, anti-plagiarism software, and the reverse engineering of computer programs. In *Field v. Google Inc.* the plaintiff claimed that his rights were infringed when Internet users were able to access copies of copyrighted works stored in the search engine cache.<sup>91</sup> The purposes that caches serve, such as allowing users to detect changes to a web page or understanding why a seemingly irrelevant page was returned in their search results, may relate to the copyrighted expression contained in the original website but they do not replicate the expressive function of the original.<sup>92</sup> Use of a cached version of a webpage to compare to and detect changes from the original “can have important political, educational, and legal ramifications.”<sup>93</sup> These are transformative not only because they have nonexpressive purposes that are different from those of the original, but also because the copyrighted work could not serve these secondary purposes alone. For these reasons, the court held that the nonexpressive use of the works in the cache were a transformative fair use and thus did not interfere with copyright of the author.<sup>94</sup>

---

<sup>89</sup> *Id.* at 1608.

<sup>90</sup> *Id.*

<sup>91</sup> *Id.* at 1618.

<sup>92</sup> *Id.*

<sup>93</sup> *Id.*

<sup>94</sup> *Id.*

AWR  
Brittany Glaser

*Perfect 10 v. Amazon* also involved the Google search engine. However, in this case the issue was not about Google caching copyrighted material, but rather was about the search engine's inability to exclude from search query results, copyrighted images that had been improperly republished on third-party websites.<sup>95</sup> The plaintiff was a subscription service that sold copyrighted images of nude models via its own website.<sup>96</sup> To understand the problem:

it is first necessary to understand how image-based search technology differs from conventional text-based search technology. Instead of recognizing images themselves, image search software identifies text associated with objects identified as images. If the text associated with an image file is responsive to a user's search query, the search engine will display a small lower resolution 'thumbnail' of the image in the search results. If an Internet user selects that thumbnail, the browser takes the user to the original location to retrieve the full-scale image.<sup>97</sup>

The court held that the third-party websites might be infringing, but that Google's use of thumbnail representations in search results did not constitute copyright infringement, reasoning that "Google's use of thumbnails served a different function unrelated to artistic expression" and was thus a transformative, nonexpressive use.<sup>98</sup> Again, something that was transformative enough was deemed a fair use.

---

<sup>95</sup> *Id.* at 1619.

<sup>96</sup> *Id.*

<sup>97</sup> *Id.*

<sup>98</sup> *Id.*

AWR  
Brittany Glaser

Many types of anti-plagiarism software abound in the academic sphere. *A.V. ex rel. Vanderhye v. iParadigms, LLC* is a case that addresses one such software offered at the Turnitin.com website, which “detects improper and unaccredited copying in student papers by comparing new papers to an archive of material available on the Internet and to proprietary databases of previously submitted papers.”<sup>99</sup> Some students alleged an infringement of their copyright in the papers they submitted, which were then copied and uploaded into the proprietary database.<sup>100</sup> The court focused on whether the use of the papers was of a transformative nature and held that the anti-plagiarism software used the papers for an entirely different purpose than that of the original works, “namely to prevent plagiarism and protect the students’ written works from plagiarism...by archiving the students’ work as digital code.”<sup>101</sup> This means that the secondary use could not produce the same effect or serve the same purpose as the original paper, which was to provide a piece of written work for reading and grading. The court additionally noted that this use provides a substantial public benefit in deterring plagiarism and encouraging original and creative work in educational institutions.<sup>102</sup> Hence, it was held a fair use.

Reverse Engineering: After an amendment made in 1980, the Copyright Act now categorizes computer programs as “literary works.”<sup>103</sup> Computer programs typically work in the following way:

---

<sup>99</sup> *Id.*

<sup>100</sup> *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 635 (4th Cir. 2009).

<sup>101</sup> *Id.* at 638.

<sup>102</sup> *Id.*

<sup>103</sup> *Sag, supra* note 28, at 1637.

software sends a series of commands to a computer (usually via a software platform and an operating system) which brings about a certain result such as launching a word processor or a video game. Aspects of the word processor or the video game may themselves be expressive<sup>104</sup>

and those are the portions that are protected by copyright. However, the underlying computer program is not itself expressive in nature.<sup>105</sup> Reverse engineering of computer programs refers to the intermediate copying necessary to gain access to those functional, nonexpressive elements of the software, which are unprotected by copyright.<sup>106</sup> In the *Sega Enterprises Ltd. v. Accolade, Inc.* case, the plaintiff Sega was selling a video gaming console, called a Sega Genesis, which would only play games made or licensed by Sega.<sup>107</sup> The licensing agreement that Accolade, a video game publisher, was offered by Sega was not only prohibitively expensive, but also forbade Accolade from making games for any other system or platform.<sup>108</sup> As an alternative, Accolade copied the programming from Sega-made games and reverse engineered in order to obtain and study the nonexpressive, functional elements that allow a game to be compatible with the Sega Genesis console.<sup>109</sup> The amount copied was the entire work, but such intermediate copying was necessary for innovation and none of the expressive aspects were copied into the final product. The court held in favor of Accolade and since then, the intermediate copying required for reverse engineering has invariably been found to constitute fair use.<sup>110</sup>

---

<sup>104</sup> *Id.*

<sup>105</sup> *Id.*

<sup>106</sup> *Id.* at 1655.

<sup>107</sup> *Id.*

<sup>108</sup> *Id.*

<sup>109</sup> *Id.*

<sup>110</sup> *Id.*

The overall purpose of the Google Books Project is to provide something new—to transform something old into something new for the public benefit. In order to accomplish this, they have taken expressive content and made a nonexpressive use of it, in a process that ultimately produces nonexpressive content. Hence, like many copy-reliant technologies, Google has taken copyrighted material, gleaned the uncopyrightable information from, it and transformed that into that something that serves a completely new objective than the original. Courts have continuously deemed this type of purpose and the necessary process behind it, as fair use.

Commercial or Nonprofit Educational Character of the Use: The other inquiry under this factor examines whether the use is of a commercial or of a nonprofit educational character.

The Author's Guild argues that Google is a commercial, for-profit entity and that the Google Books Project is just a means for them to sell advertising. Thus, they aver the use to be of a commercial character.<sup>111</sup> It is first important to note that, “commercial use is no longer deemed by courts to be presumptively unfair.”<sup>112</sup> Secondly, the focus of this factor is meant to be on the secondary use, the Books Project itself, not Google, the entity behind it. Despite the fact that Google “will likely profit from advertising displayed alongside search results”<sup>113</sup> as it always does when anything is searched for with its search engine, “Google's primary objective is

---

<sup>111</sup> Defendant's Opposition to Motion for Summary Judgment, *supra* note 66, at 9.

<sup>112</sup> Proskine, *supra* note 54, at 226.

<sup>113</sup> *Id.*

AWR  
Brittany Glaser

to make books accessible by making them easier to find.”<sup>114</sup> It is not attempting to profit from the sale or distribution of full-text copies of the books scanned and digitized into its database. The search results provided by the Google Books Project will serve the nonprofit educational purpose of aiding users in their research and determining “whether they are interested in pursuing the work further.”<sup>115</sup> Additionally, Google does not run any advertisements on “About the Book” pages associated with the Project and does not receive any profits from books that are purchased.<sup>116</sup>

The issue of commercial gain from advertisements was argued in *Perfect 10*, however, the court held that “the transformative nature of Google's use was more significant than any minor commercial aspects of Google's search engine and website.”<sup>117</sup> Therefore, it is likely that the transformative nature of the Google Books Project, which provides an educational service, would outweigh the commercial nature of advertising revenue.

As courts de-emphasize the importance of commercial use and highlight the importance of the transformative nature of a use, Google's case is strengthened. The Google Books Project is transformative enough to be of the proper purpose and character to fulfill the first fair use factor.

(2) the nature of the copyrighted work

---

<sup>114</sup> *Id.*

<sup>115</sup> *Id.*

<sup>116</sup> Defendant Google Inc.'s Motion for Summary Judgment, *supra* note 46, at 5.

<sup>117</sup> *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1167 (9th Cir. 2007).

The Author’s Guild concedes that the nature of the individual book has little relevance to the analysis in this case.<sup>118</sup> Under this factor, focus tends toward the factual versus the creative in a copyrighted work, with creative works—or at least the creative elements of works—being given the greater protection.<sup>119</sup> The Google Books Project presents a difficult scenario to analyze under this factor, “due to the fact that Google is scanning and digitizing millions of books ranging from creative to fact-based.”<sup>120</sup> However, the courts have said that:

if the disputed use of the copyrighted work ‘is not related to its mode of expression but rather to its historical facts,’ then the creative nature of the work is mitigated...therefore, even if the original works were highly creative in nature...the secondary use constitutes fair use as long as it is not related to its creative expression but rather to its historical facts.<sup>121</sup>

It is important to remember that fair use is an *affirmative* defense to infringement—meaning that infringement *has* taken place, but is of a type that is regarded as justifiable. Some of the books being scanned by Google are of a creative nature, and such copying on its own would certainly be more likely to constitute infringement. However, the actual purpose and use of these copies is not for the enjoyment of the creative expression of the original, but rather to obtain the same kinds of historical facts that might be found on a library catalogue card in order to facilitate research and disseminate ideas; the creative aspects are not distributed. Despite the strong

---

<sup>118</sup> Defendant’s Opposition to Motion for Summary Judgment, *supra* note 66, at 16.

<sup>119</sup> Proskine, *supra* note 54, at 227.

<sup>120</sup> *Id.* at 227-28.

<sup>121</sup> A.V. ex rel. Vanderhye v. iParadigms, LLC, 562 F.3d 630, 640 (4th Cir. 2009).



AWR  
Brittany Glaser

copyright protection of creative works, facts about books derived from scanning, are no more copyrightable than any other facts.<sup>122</sup>

In *Campbell*, the Supreme Court admitted that this second factor is not “likely to help much in separating the fair use sheep from the infringing goats” especially where the secondary use was judged to be transformative.<sup>123</sup> The significance of the second factor is likely to be overwhelmed by the other factors, which focus more on the secondary use itself rather than the original.<sup>124</sup> The nature of the copyrighted works in question, do not necessarily undermine the ability of the Google Books Project to fulfill the second factor. However, if the court were to find otherwise, the overall balancing of the factors could still find for fair use.

(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole;

The Author’s Guild asserts that simply because Google has made a copy of the entire work, the third factor will automatically weigh against fair use. However, full-text copying does not weigh against fair use when it is necessary to facilitate a use that is otherwise fair. This factor seeks to determine whether “the amount and substantiality of the portion used in relation to the copyrighted work as a whole, are reasonable in relation to the purpose of the copying.”<sup>125</sup>

Copying of a work in its entirety does not necessarily preclude a finding of fair use; the amount permissible is really based on the justification behind the copying. It has been observed that:

---

<sup>122</sup> Sag, *supra* note 28, at 1622.

<sup>123</sup> *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 586, 114 S. Ct. 1164, 1175, 127 L. Ed. 2d 500 (1994).

<sup>124</sup> Sag, *supra* note 1, at 32.

<sup>125</sup> *Campbell*, 510 U.S. at 586.

There is simply no magic number in fair use cases because the extent of permissible copying varies according to the purpose of the copying and the effect of the copying on the rights-holder. The key inquiry under the third factor is not simply how much of the copyright owner's work was used by the defendant, but rather what proportion of the work's expressive value was appropriated.

Transformative uses in general and nonexpressive uses in particular, do not tend to substitute for the value of the author's original expression and are thus regarded favorably under the third factor.<sup>126</sup>

In the case at hand, "without full-text copying, Google would not be able to produce its robust database."<sup>127</sup> It would be impossible to match accurate and germane results to a user's search query. There is simply no other way to obtain full and complete information of everything contained within the books, for database search purposes. As has been previously established, the Google Books Project is a nonexpressive use and therefore, none of the work's expressive value is appropriated or put towards an expressive purpose.

In terms of the effect on the rights of the copyright owner under this factor, Judge Pierre Leval, in his seminal article on fair use and transformativeness, offers this:

As to the relationship of quantity to the market, presumptively, of course, the more taken the greater the likely impact on the copyright holder's market, and the

---

<sup>126</sup> Sag, *supra* note 1, at 32.

<sup>127</sup> Proskine, *supra* note 54, at 228.

more the factor favors the copyright holder. Too mechanical a rule, however, can be dangerously misleading. One can imagine secondary works that quote 100% of the copyrighted work without affecting market potential. Consider, for example, a lengthy critical study analyzing the structure, symbolism and meaning, literary antecedents and influences of a single sonnet. Fragments dispersed throughout the work of criticism may well quote every word of the poem. Such quotation will not displace the market for the poem itself. If there is strong justification and no adverse market impact, even so extensive a taking could be a fair use.<sup>128</sup>

Similarly, Google's use transforms the original works and uses them in a different manner that neither displaces nor adversely impacts the market for the original books. It does not provide access to the actual books by the public. It actually has the potential to improve the market for the original books by bringing old or forgotten books back into prominence and guides users to booksellers, thus increasing sales. Hence, despite copying the works in their entirety, the transformative nature of the Google Books Project's use does not substitute for the original, and is a fair use of the material.

(4) the effect of the use upon the market for or value of the copyrighted work.

The Author's Guild asserts that the Google Books Project could usurp the book market by serving as a market substitute.<sup>129</sup> This is a completely unfounded fear. The Project does not provide print books or e-books. The service it provides cannot substitute for reading a book or

---

<sup>128</sup> Leval, *supra* note 11, at 1123.

<sup>129</sup> Defendant's Opposition to Motion for Summary Judgment, *supra* note 66, at 13.

AWR  
Brittany Glaser

compete with the purchase of a book. It only furnishes a brief snippet to help the user evaluate whether that book is pertinent to his search. The underlying book itself is not distributed, whatsoever, unless it happens to be in the public domain.

The fourth factor requires a court to consider whether the alleged infringer's use, has an adverse impact on the market for the original. When a secondary use "amounts to a mere duplication of the entirety of an original, it clearly 'supersedes the objects' of the original and serves as a market replacement for it, making it likely that cognizable market harm to the original will occur."<sup>130</sup> An example of a use that did not satisfy this fair use factor can be found in *On Davis v. The Gap, Inc.*<sup>131</sup> In that case, the plaintiff created a fanciful form of eyewear. The defendant, a chain of clothing and accessories stores, used identical eyewear in an advertisement without being granted a license by plaintiff.<sup>132</sup> Since the eyewear in the advertisement was presented as a market substitute to buying the plaintiff's product, it was likely to harm the market for the original work.<sup>133</sup> The Google Books Project does not act as a substitute to buying a book and it does not replace the need for books; it merely indexes them. It offers a completely different function than a book itself.

The Author's Guild has voiced concerns a user could piece together an entire book from the snippets.<sup>134</sup> In their scenario, by performing multiple searches using different search terms, a single user may view far more than three snippets, as different searches can return different

---

<sup>130</sup> Beebe, *supra* note 10, at 618.

<sup>131</sup> *On Davis v. The Gap, Inc.*, 246 F.3d 152 (2d Cir. 2001).

<sup>132</sup> *Id.*

<sup>133</sup> *Id.*

<sup>134</sup> Defendant Google Inc.'s Motion for Summary Judgment, *supra* note 46, at 4-6.

AWR  
Brittany Glaser

snippets.<sup>135</sup> For example, by making a series of consecutive, slightly different searches of the book a single user can view many different snippets from the book.<sup>136</sup> However, this is unrealistic “due to the difficulty and time-consuming nature of copying a work in such a manner,” not to mention the security measures Google has implemented to prevent users from viewing a complete copy of a snippet-view book.<sup>137</sup> For example, print, copy, and email functions have been disabled for snippets.<sup>138</sup> Moreover:

a user cannot cause the system to return different sets of snippets for the same search query; the position of each snippet is fixed within the page and does not ‘slide’ around the search term; only the first responsive snippet available on any given page will be returned in response to a query; one of the snippets on each page is "black-listed," meaning it will not be shown; and at least one out of ten entire pages in each book is black-listed. An "attacker" who tries to obtain an entire book by using a physical copy of the book to string together words appearing in successive passages would be able to obtain at best a patchwork of snippets that would be missing at least one snippet from every page and 10% of all pages. In addition, works with text organized in short "chunks," such as dictionaries, cookbooks, and books of haiku, are excluded from snippet view.<sup>139</sup>

---

<sup>135</sup> *Id.*

<sup>136</sup> *Id.*

<sup>137</sup> Proskine, *supra* note 54, at 230.

<sup>138</sup> *Id.*

<sup>139</sup> Defendant Google Inc.’s Motion for Summary Judgment, *supra* note 46, at 4-6.

AWR  
Brittany Glaser

Therefore, Google is not providing a market substitute, whatsoever. It is recontextualizing the works in way that provides a completely different service. The only impact that this different service has on the market of the original works is a positive one. It benefits the copyright owners enormously, by stimulating demand for their books and guiding users to libraries, publishers, and booksellers; this produces a money-making opportunity for authors as their book sales increase. Since the fruits of the Google Book Project do not fulfill demands for the original works and improve rather than impair the market for the originals, the fourth factor weighs in favor of a fair use finding.

VIII. Conclusion: Based on the four-factor analysis, it is likely that Google will prevail with its fair use defense, as its scanning of books has the transformative purpose of creating a searchable digital database that acts as something other than a mere substitute for the original copyrighted works and benefits society.