

AUTHORS GUILD, INC. v. GOOGLE, INC.
United States District Court, Southern District of New York
954 F. Supp. 2d 282

Chin, Circuit Judge:

Since 2004, when it announced agreements with several major research libraries to digitally copy books in their collections, defendant Google Inc. (“Google”) has scanned more than twenty million books. It has delivered digital copies to participating libraries, created an electronic database of books, and made text available for online searching through the use of “snippets.” Many of the books scanned by Google, however, were under copyright, and Google did not obtain permission from the copyright holders for these usages of their copyrighted works. As a consequence, in 2005, plaintiffs brought this class action charging Google with copyright infringement.

. . . For the reasons set forth below, Google’s motion for summary judgment is granted . . .

BACKGROUND

A. The Facts

[In 2004, Google announced the “Library Project,” which involved the digital scanning of books in the collections of the New York Public Library, the Library of Congress, and a number of university libraries. Google did not seek or obtain permission from the copyright holders.]

3. Google Books

In scanning books for its Library Project, including in-copyright books, Google uses optical character recognition technology to generate machine-readable text, compiling a digital copy of each book. Google 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. The index allows a search for a particular word or phrase to return a result that includes the most relevant books in which the word or phrase is found. Because the full texts of books are digitized, a user can search the full text of all the books in the Google Books corpus.

Users of Google’s search engine may conduct searches, using queries of their own design. In response to inquiries, Google returns a list of books in which the search term appears. A user can click on a particular result to be directed to an “About the Book” page, which will provide the user with information about the book in question. The page includes links to sellers of the books and/or libraries that list the book as part of their collections. No advertisements have ever appeared on any About the Book page that is part of the Library Project.

For books in “snippet view” . . . , Google divides each page into eighths—each of which is a “snippet,” a verbatim excerpt. Each search generates three snippets, but by performing multiple searches using different search terms, a single user may view far more than three

snippets, as different searches can return different snippets. 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.

Google takes security measures to prevent users from viewing a complete copy of a snippet-view book. For example, 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.

4. The Benefits of the Library Project and Google Books

The benefits of the Library Project are many. First, Google Books provides a new and efficient way for readers and researchers to find books. It makes tens of millions of books searchable by words and phrases. It provides a searchable index linking each word in any book to all books in which that word appears. Google Books has become an essential research tool, as it helps librarians identify and find research sources, it makes the process of interlibrary lending more efficient, and it facilitates finding and checking citations. Indeed, Google Books has become such an important tool 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.

Second, in addition to being an important reference tool, Google Books greatly promotes a type of research referred to as “data mining” or “text mining.” Google Books permits humanities scholars to analyze massive amounts of data—the literary record created by a collection of tens of millions of books. Researchers can examine word frequencies, syntactic patterns, and thematic markers to consider how literary style has changed over time. Using Google Books, for example, researchers can track the frequency of references to the United States as a single entity (“the United States is”) versus references to the United States in the plural (“the United States are”) and how that usage has changed over time. The ability to determine how often different words or phrases appear in books at different times “can provide insights about fields as diverse as lexicography, the evolution of grammar, collective memory, the adoption of technology, the pursuit of fame, censorship, and historical epidemiology.” Jean-Baptiste Michel *et al.*, *Quantitative Analysis of Culture Using Millions of Digitized Books*, 331 *Science* 176, 176 (2011).

Third, Google Books expands access to books. In particular, traditionally underserved populations will benefit as they gain knowledge of and access to far more books. Google Books provides print-disabled individuals with the potential to search for books and read them in a format that is compatible with text enlargement software, text-to-speech screen access software,

and Braille devices. Digitization facilitates the conversion of books to audio and tactile formats, increasing access for individuals with disabilities. Google Books facilitates the identification and access of materials for remote and underfunded libraries that need to make efficient decisions as to which resources to procure for their own collections or through interlibrary loans.

Fourth, Google Books helps to preserve books and give them new life. Older books, many of which are out-of-print books that are falling apart buried in library stacks, are being scanned and saved. . . . These books will now be available, at least for search, and potential readers will be alerted to their existence.

Finally, by helping readers and researchers identify books, Google Books benefits authors and publishers. When a user clicks on a search result and is directed to an “About the Book” page, the page will offer links to sellers of the book and/or libraries listing the book as part of their collections. The About the Book page for *Ball Four*, for example, provides links to Amazon.com, Barnes & Noble.com, Books–A–Million, and IndieBound. A user could simply click on any of these links to be directed to a website where she could purchase the book. Hence, Google Books will generate new audiences and create new sources of income. . . .

B. Procedural History

Plaintiffs commenced this action on September 20, 2005, alleging, *inter alia*, that Google committed copyright infringement by scanning copyrighted books and making them available for search without permission of the copyright holders. From the outset, Google’s principal defense was fair use under § 107 of the Copyright Act.

After extensive negotiations, the parties entered into a proposed settlement resolving plaintiffs’ claims on a class-wide basis. On March 22, 2011, I issued an opinion rejecting the proposed settlement on the grounds that it was not fair, adequate, and reasonable. *Authors Guild v. Google, Inc.*, 770 F. Supp. 2d 666 (S.D.N.Y. 2011). . . .

Plaintiffs filed their class certification motion and Google filed its motion to dismiss the Authors Guild’s claims. On May 31, 2012, I issued an opinion denying Google’s motion to dismiss and granting the individual plaintiffs’ motion for class certification. *Authors Guild v. Google, Inc.*, 282 F.R.D. 384 (S.D.N.Y. 2012). . . .

On July 1, 2013, without deciding the merits of the appeal, the Second Circuit vacated my class certification decision, concluding that “resolution of Google’s fair use defense in the first instance will necessarily inform and perhaps moot our analysis of many class certification issues.” *Authors Guild v. Google, Inc.*, 721 F.3d 132, 134 (2d Cir. 2013). The Second Circuit remanded the case “for consideration of the fair use issues.” *Id.* at 135. . . .

DISCUSSION

For purposes of these motions, I assume that plaintiffs have established a *prima facie* case of copyright infringement against Google under 17 U.S.C. § 106. Google has digitally reproduced millions of copyrighted books, including the individual plaintiffs’ books, maintaining

copies for itself on its servers and backup tapes. *See* 17 U.S.C. § 106(1) (prohibiting unauthorized reproduction). Google has made digital copies available for its Library Project partners to download. *See* 17 U.S.C. § 106(3) (prohibiting unauthorized distribution). Google has displayed snippets from the books to the public. *See* 17 U.S.C. § 106(5) (prohibiting unauthorized display). Google has done all of this, with respect to in-copyright books in the Library Project, without license or permission from the copyright owners. The sole issue now before the Court is whether Google’s use of the copyrighted works is “fair use” under the copyright laws. For the reasons set forth below, I conclude that it is. . . .

1. *Purpose and Character of Use*

Google’s use of the copyrighted works is highly transformative. Google Books digitizes books and transforms expressive text into a comprehensive word index that helps readers, scholars, researchers, and others find books. Google Books has become an important tool for libraries and librarians and cite-checkers as it helps to identify and find books. The use of book text to facilitate search through the display of snippets is transformative. *See Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1168 (9th Cir. 2007) (holding that use of works—“thumbnail images,” including copyrighted photographs—to facilitate search was “transformative”); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811 (9th Cir. 2003) (same); *see also Bill Graham Archives*, 448 F.3d at 609-11 (holding that display of images of posters in 480–page cultural history of the Grateful Dead was transformative, explaining that “[w]hile the small size [of the images of the posters] is sufficient to permit readers to recognize the historical significance of the posters, it is inadequate to offer more than a glimpse of their expressive value”). The display of snippets of text for search is similar to the display of thumbnail images of photographs for search or small images of concert posters for reference to past events, as the snippets help users locate books and determine whether they may be of interest. Google Books thus uses words for a different purpose—it uses snippets of text to act as pointers directing users to a broad selection of books.

Similarly, Google Books is also transformative in the sense that it has transformed book text into data for purposes of substantive research, including data mining and text mining in new areas, thereby opening up new fields of research. Words in books are being used in a way they have not been used before. Google Books has created something new in the use of book text—the frequency of words and trends in their usage provide substantive information.

Google Books does not supersede or supplant books because it is not a tool to be used to read books. Instead, it “adds value to the original” and allows for “the creation of new information, new aesthetics, new insights and understandings.” Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. at 1111. Hence, the use is transformative.

It is true, of course, as plaintiffs argue, that Google is a for-profit entity and Google Books is largely a commercial enterprise. The fact that a use is commercial “tends to weigh against a finding of fair use.” *Harper & Row*, 471 U.S. at 562; *accord Campbell*, 510 U.S. at 585. On the other hand, fair use has been found even where a defendant benefitted commercially from the unlicensed use of copyrighted works. *See, e.g., Blanch*, 467 F.3d at 253; *Bill Graham Archives*, 448 F.3d at 612. . . . Here, Google does not sell the scans it has made of books for Google Books; it does not sell the snippets that it displays; and it does not run ads on the About

the Book pages that contain snippets. It does not engage in the direct commercialization of copyrighted works. Google does, of course, benefit commercially in the sense that users are drawn to the Google websites by the ability to search Google Books. While this is a consideration to be acknowledged in weighing all the factors, even assuming Google's principal motivation is profit, the fact is that Google Books serves several important educational purposes.

Accordingly, I conclude that the first factor strongly favors a finding of fair use.

2. *Nature of Copyrighted Works*

The second factor is “the nature of the copyrighted work.” 17 U.S.C. § 107(2). Here, the works are books—all types of published books, fiction and non-fiction, in-print and out-of-print. While works of fiction are entitled to greater copyright protection, *Stewart v. Abend*, 495 U.S. 207, 237 (1990), here the vast majority of the books in Google Books are non-fiction. Further, the books at issue are published and available to the public. These considerations favor a finding of fair use. *See Arica Inst., Inc. v. Palmer*, 970 F.2d 1067, 1078 (2d Cir. 1992) (“Whether or not a work is published is critical to its nature under factor two because the scope of fair use is narrower with respect to unpublished works.”). . . .

3. *Amount and Substantiality of Portion Used*

The third factor is “the amount and substantiality of the portion used in relation to the copyrighted work as a whole.” 17 U.S.C. § 107(3). Google scans the full text of books—the entire books—and it copies verbatim expression. On the other hand, courts have held that copying the entirety of a work may still be fair use. *See, e.g., Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 449-50 (1984); *Bill Graham Archives*, 448 F.3d at 613 (“copying the entirety of a work is sometimes necessary to make a fair use of the image”). Here, as one of the keys to Google Books is its offering of full-text search of books, full-work reproduction is critical to the functioning of Google Books. Significantly, Google limits the amount of text it displays in response to a search.

On balance, I conclude that the third factor weighs slightly against a finding of fair use.

4. *Effect of Use Upon Potential Market or Value*

The fourth factor is “the effect of the use upon the potential market for or value of the copyrighted work.” 17 U.S.C. § 107(4). Here, plaintiffs argue that Google Books will negatively impact the market for books and that Google's scans will serve as a “market replacement” for books. It also argues that users could put in multiple searches, varying slightly the search terms, to access an entire book.

Neither suggestion makes sense. Google does not sell its scans, and the scans do not replace the books. While partner libraries have the ability to download a scan of a book from their collections, they owned the books already—they provided the original book to Google to scan. Nor is it likely that someone would take the time and energy to input countless searches to try and get enough snippets to comprise an entire book. Not only is that not possible as certain

pages and snippets are blacklisted, the individual would have to have a copy of the book in his possession already to be able to piece the different snippets together in coherent fashion.

To the contrary, a reasonable factfinder could only find that Google Books enhances the sales of books to the benefit of copyright holders. An important factor in the success of an individual title is whether it is discovered—whether potential readers learn of its existence. Google Books provides a way for authors' works to become noticed, much like traditional in-store book displays. Indeed, both librarians and their patrons use Google Books to identify books to purchase. Many authors have noted that online browsing in general and Google Books in particular helps readers find their work, thus increasing their audiences. Further, Google provides convenient links to booksellers to make it easy for a reader to order a book. In this day and age of on-line shopping, there can be no doubt but that Google Books improves books sales.

Hence, I conclude that the fourth factor weighs strongly in favor of a finding of fair use.

5. Overall Assessment

Finally, the various non-exclusive statutory factors are to be weighed together, along with any other relevant considerations, in light of the purposes of the copyright laws.

In my view, Google Books provides significant public benefits. It advances the progress of the arts and sciences, while maintaining respectful consideration for the rights of authors and other creative individuals, and without adversely impacting the rights of copyright holders. It has become an invaluable research tool that permits students, teachers, librarians, and others to more efficiently identify and locate books. It has given scholars the ability, for the first time, to conduct full-text searches of tens of millions of books. It preserves books, in particular out-of-print and old books that have been forgotten in the bowels of libraries, and it gives them new life. It facilitates access to books for print-disabled and remote or underserved populations. It generates new audiences and creates new sources of income for authors and publishers. Indeed, all society benefits.

CONCLUSION

For the reasons set forth above, plaintiffs' motion for partial summary judgment is denied and Google's motion for summary judgment is granted. Judgment will be entered in favor of Google dismissing the Complaint. . . .

NOTES AND QUESTIONS

(1) Google may have benefitted from the long delay caused by settlement negotiations and the appeal of the class-action certification ruling. In 2004, when the Google Books project was first announced, it was extremely controversial. In the ensuing decade, however, Google Books has demonstrated its usefulness, and there has been no indication that authors or publishers have been harmed by it. To the contrary, a large number of publishers are now participating in Google's Partner Program, allowing more generous displays (from selected pages to full text) than the "snippets" described in the opinion.

(2) *Transformative use or transformative purpose?* When Google scans and indexes entire books verbatim, how is this “transformative” of those books? The Google Books case continues a trend in the case law away from the type of “transformative use” described in *Campbell* (“altering the first with new expression, meaning, or message”) to an emphasis on whether the *purpose* of the use is different from the purpose of the original work. Is this consistent with *Campbell*? Is it consistent with the Constitutional purpose of copyright, to promote the progress of knowledge?

(3) *Commercial use.* In the Google Books opinion, the court finds that although Google “benefit[s] commercially,” it “does not engage in the *direct* commercialization of copyrighted works” (emphasis added). What makes a commercial benefit “direct” or “indirect”? *Campbell* and *Bill Graham Archives* likewise downplay the “commercial” nature of the use. Is a “commercial” use always outweighed by a “transformative” use or purpose?

(4) *Market effect.* The court also finds that Google Books “improves book sales” by helping a book “become noticed.” Suppose a movie studio includes a sound recording of a musical work in a movie. Wouldn’t that “improve sales” of the song by helping it “become noticed”? Does that suggest that the movie studio’s use is a fair use? If not, what is the difference between the hypothetical and the Google Books case?

(5) The Google Books project also spawned satellite litigation involving HathiTrust, a non-profit “digital library” to which the various libraries involved in the project donated their digital copies, which were used for three purposes: 1) enabling full-text searches that return only page numbers, without revealing any text; 2) providing access to print-disabled persons; and 3) preserving existing materials;. The first two uses were held to be “transformative” uses by non-profit educational institutions; although entire works were copied, it did not significantly affect the market for those works, because the market for blind readers is very small, and a market for full-text search does not exist. *Authors Guild, Inc. v. HathiTrust*, 2014 U.S. App. LEXIS 10803 (2d Cir. 2014). The court remanded the “preservation” issue to determine whether the plaintiffs had standing. A fourth purpose, a proposed “Orphan Works Project” in which HathiTrust would make the full text of orphan works available to patrons, was held not yet ripe for review.

(6) *“Non-expressive” uses.* Is the Google Books case *sui generis*, or is there a more general principle to be gleaned from it? One scholar has proposed that courts should distinguish between “expressive” uses of copyrighted works — uses that communicate the author’s original expression to the public — and “non-expressive” uses — acts of copying in which the digital data representing the work (or large numbers of works) is “processed” to useful and valuable ends, but the author’s original expression is not communicated to the public. His thesis is that “non-expressive uses of copyrighted works . . . should not generally be regarded as infringing.” Sag, *Copyright and Copy-Reliant Technology*, 103 Nw. U. L. Rev. 1607, 1625 (2009). While courts have not yet adopted this terminology, several courts have held such “non-expressive” uses to be fair uses. *See, e.g., Field v. Google, Inc.*, 412 F. Supp. 2d 1106 (D. Nev. 2006) (search-engine “caching” and indexing of authorized websites); *A.V. v. iParadigms, LLC*, 562 F.3d 630 (4th Cir. 2009) (plagiarism detection software).

(7) Before signing off the World Wide Web, consider *Los Angeles Times v. Free Republic*, 54 U.S.P.Q.2d (BNA) 1453 (C.D. Cal. 2000), in which the defendants operated a “bulletin board” website, where they posted news stories from mainstream media (including the plaintiffs L.A. Times and Washington Post) and invited commentary from visitors to the site. The defendants lost on summary judgment, primarily because they could not persuade the court that their use of copyrighted news articles was “transformative” — and because the court credited the plaintiffs’ claims that the Free Republic had the potential to reduce traffic to their own web sites.

Fair use was raised in a number of cases filed by serial copyright plaintiff Righthaven LLC, which searched the internet for news articles, obtained assignments of copyright from the newspapers in which those articles originally appeared, and sued bloggers and websites that posted the articles. Critics claimed that Righthaven’s business model amounted to near-extortion, because the cost of litigating a fair use claim vastly exceeds the nuisance value of a settlement (especially given the threat of statutory damages). *See, e.g., Righthaven LLC v. Choudhry*, 99 U.S.P.Q.2d (BNA) 1225 (D. Nev. 2011) (denying defendant’s motion for summary judgment despite finding that fourth factor weighed in defendant’s favor). After losing several cases on fair use grounds, however, Righthaven went out of business. *See Righthaven, LLC v. Hoehn*, 792 F. Supp. 2d 1138 (D. Nev. 2011) (summary judgment for defendant), *vacated on other grounds*, 716 F.3d 1166 (9th Cir. 2013); *Righthaven, LLC v. Jama*, 2011 U.S. Dist. LEXIS 43952 (D. Nev. Apr. 22, 2011) (same); *Righthaven LLC v. Realty One Group, Inc.*, 96 U.S.P.Q.2d (BNA) 1516 (D. Nev. 2010) (granting motion to dismiss). Does copyright law need some kind of “small claims” procedure to deal with mass infringement lawsuits? *See Lemley & Reese, A Quick and Inexpensive System for Resolving Peer-to-Peer Copyright Disputes*, 23 *Cardozo Arts & Ent. L.J.* 1 (2005).

Today, of course, posting entire articles is no longer necessary. Instead, one links to the news article on the publisher’s website. Google News, for example, typically copies the headline and the first sentence of a news article and generates a link to that article. Assuming the amount of copying is not *de minimis*, should it be considered a fair use? *See Associated Press v. Meltwater U.S. Holdings, Inc.*, 931 F. Supp. 2d 537 (S.D.N.Y. 2013) (unlike search engines, defendant’s unlicensed news monitoring service substitutes for licensed news sites, rather than facilitating public access to those sites). Note also the possibility of a state-law misappropriation claim, which may survive preemption by federal copyright law. *See Associated Press v. All-Headline News Corp.*, 608 F. Supp. 2d 454 (S.D.N.Y. 2009) (denying motion to dismiss). Misappropriation is discussed in more detail in § 11.02.

(8) For good or ill, the Google Books case represents an attempt to adapt copyright law to the rich and unruly electronic information environment. Back in § 9.04, we reviewed one of the other legal initiatives designed to accomplish a similar end: the insertion of “anti-circumvention” provisions into Title 17, courtesy of the 1998 Digital Millennium Copyright Act. Of course, the DMCA’s title was, in part, a misnomer. As we already noted, the new prohibitions and penalties in Chapter 12 of Title 17 are not expansions of copyright as such, but something else: “paracopyright.” And that sets up some interesting conflicts between these new provisions, on the one hand, and traditional copyright doctrines like fair use, on the other.