

The Google Library Project: Both Sides of the Story¹

Jonathan Band

E-mail: jband@policybandwidth.com

Abstract

Google's announcement that it will include in its search database the full text of books from five of the world's leading research libraries has provoked newspaper editorials, public debates, and two lawsuits. Some of this attention can be attributed to public fascination with any move taken by Google, one of the most successful companies in the digital economy. The sheer scale of the project and its possible benefits for research have also captured the public imagination. Finally, the controversy over copyright issues has been fueled by Google's willingness to pursue this ambitious effort notwithstanding the opposition of the publishing industry and organizations representing authors. Much of the press coverage, however, confuses the facts, and the opposing sides often talk past each other without engaging directly.² This article will attempt to set forth the facts and review the arguments in a systematic manner.³ Although both sides have strong legal arguments, the article concludes that the applicable legal precedents support Google's fair use position.

The Google Book Search Project

The Google Book Search project (formerly the Google Print project) has two facets: the Partner Program (formerly the Publisher Program) and the Library Project. Under the Partner Program, a publisher controlling the rights in a book can authorize Google to scan the full text of the book into Google's search database. In response to a user query, the user receives bibliographic information concerning the book as well as a link to relevant text. By clicking on the link, the user can see the full page containing the search term, as

well as a few pages before and after that page. Links would enable the user to purchase the book from booksellers or the publisher directly, or visit the publisher's website. Additionally, the publisher would share in contextual advertising revenue if the publisher has agreed for ads to be shown on their book pages. Publishers can remove their books from the Partner Program at any time. The Partner Program raises no copyright issues because it is conducted pursuant to an agreement between Google and the copyright holder.

Under the Library Project, Google plans to scan into its search database materials from the libraries of Harvard, Stanford, and Oxford Universities, the University of Michigan, and the New York Public Library. In response to search queries, users will be able to browse the full text of public domain materials, but only a few sentences of text around the search term in books still covered by copyright. This is a critical fact that bears repeating: for books still under copyright, users will be able to see only a few sentences on either side of the search term – what Google calls a “snippet” of text. Users will *not* see a few pages, as under the Partner Program, *nor* the full text, as for public domain works. Indeed, users will never see even a single page of an in-copyright book scanned as part of the Library Project.⁴

Moreover, if a search term appears many times in a particular book, Google will display no more than three snippets, thus preventing the user from viewing too much of the book for free. Finally, Google will not display any snippets for

certain reference books, such as dictionaries, where the display of even snippets could harm the market for the work. The text of the reference books will still be scanned into the search database, but in response to a query the user will only receive bibliographic information. The page displaying the snippets will indicate the closest library containing the book, as well as where the book can be purchased, if that information is available.

Because of non-disclosure agreements between Google and the libraries, many details concerning the project are not available.⁵ It appears that Google will scan only public domain materials from Oxford and the New York Public Library, and small collections at Harvard. It will scan both public domain and in-copyright books at Michigan and Stanford. Google may make some attempts to avoid scanning the same book in different libraries (particularly Michigan and Stanford, where the overlap may be greatest); but the inaccuracy of bibliographic information in reference tools such as card catalogs makes it difficult to determine easily whether two books are, in fact, identical. For example, a card catalogue entry may not indicate whether different volumes are of the same edition. Given these inaccuracies, Google will probably err on the side of inclusion.

Google's Opt-Out Policy

In response to criticism from groups such as the American Association of Publishers and the Authors Guild, Google announced an opt-out policy in August 2005. If a copyright owner provided it with a list of its titles that it did not want Google to scan at libraries, Google would respect that request, even if the books were in the collection of one of the participating libraries.⁶ Google stated that it would not scan any in-copyright books between August and November 1, 2005, to provide the owners with the opportunity to decide which books to exclude from the Project. Thus, Google provides a copyright owner with three choices

with respect to any work: it can participate in the Partner Program, in which case it would share in revenue derived from the display of pages from the work in response to user queries; it can let Google scan the book under the Library Project and display snippets in response to user queries; or it can opt-out of the Library Project, in which case Google will not scan its book.

The Library Copies

As part of Google's agreement with the participating libraries, Google will provide each library with a digital copy of the books in its collection scanned by Google. Under the agreement between Google and the University of Michigan – the only contract disclosed to the public so far⁸ – the University agrees to use its copies only for purposes permitted under the Copyright Act. If any of these lawful uses involves the posting of all or part of a library copy on the University's website – for example, posting the full text of a public domain work – the University agrees to limit access to the work and to use technological measures to prevent the automated downloading and redistribution of the work.⁹ Another possible use described by the University is keeping the copies in a restricted (or "dark") archive until the copyright expires or the copy is needed for preservation purposes.¹⁰

Actions by Other Search Engines

Both Yahoo and Microsoft have recently announced digitization projects. Microsoft announced that it would be digitizing 100,000 volumes from the British Library. Yahoo agreed to host the Open Content Alliance, under which entities such as the University of California and the Internet Archive will post digitized works. The salient difference between these projects and Google's Library Project is that these projects will involve only works in the public domain or works where the owner has opted-in to the digitization,

while Google intends to scan in-copyright books without the owner's authorization, as well as works in the public domain.¹¹

The Litigation

On September 20, 2005, the Authors Guild and several individual authors sued Google for copyright infringement. The lawsuit was styled as a class action on behalf of all authors similarly situated. A month later, on October 19, 2005, five publishers – McGraw-Hill, Pearson, Penguin, Simon & Schuster, and John Wiley & Sons – sued Google. The authors request damages *and* injunctive relief. The publishers, in contrast, only requested injunctive relief. Neither group of plaintiffs moved for a temporary restraining order before the November 1 date on which Google announced that it would resume scanning in-copyright books. Neither group sued the libraries for making the books available to Google, nor for the copies Google is making for them.

The Library Project involves two actions that raise copyright questions. First, Google copies the full text of books into its search database. Second, in response to user queries, Google presents users with a few sentences from the stored text. Because the amount of expression presented to the user is *de minimus*, this second action probably would not lead to liability. Perhaps for this reason, the lawsuits focus on the first issue, Google's copying of the full text of books into its search database.¹²

Opt-In v. Opt-Out

As noted above, Google announced that it would honor a request from a copyright owner not to scan its book. The owners, however, insist that the burden should not be on them to request Google not to scan a particular work; rather, the burden should be on Google to request permission to scan the work. According to Pat Schroeder, AAP President, Google's opt-out procedure "shifts the responsibility for preventing infringement to the copyright owner rather than the user, turning every principle

of copyright law on its ear."¹³ The owners assert that under copyright law, the user can copy only if the owner affirmatively grants permission to the user – that copyright is an opt-in system, rather than an opt-out system. Thus, as a practical matter, the entire dispute between the owners and Google boils down to who should make the first move: should Google have to ask permission before it scans? Or should the owner have to tell Google that it does not want the work scanned?

Google's Fair Use Argument

The owners are correct that copyright typically is an opt-in system, and that Google is copying vast amounts of copyrighted material without authorization. Google responds that this copying is permitted under the fair use doctrine, 17 U.S.C. §107. The critical question in the litigation is whether the fair use doctrine excuses Google's copying.¹⁴

The U.S. Court of Appeals for the Ninth Circuit, which comprises the states on the West Coast, recently issued a decision that is directly on point. In *Kelly v. Arriba Soft*, 336 F.3d 811 (9th Cir. 2003), Arriba Soft operated a search engine for Internet images. Arriba compiled its database of images by sending out software spiders that copied thousands of pictures from websites, without the express authorization of the website operators. Arriba reduced the full size images into thumbnails, which it stored in its database. In response to a user query, the Arriba search engine displayed responsive thumbnails. If a user clicked on one of the thumbnails, she was linked to the full size image on the original website from which the image had been copied. Kelly, a photographer, discovered that some of the photographs from his website were in the Arriba search database, and he sued for copyright infringement. The lower court found that Arriba's reproduction of the photographs was a fair use, and the Ninth Circuit affirmed.

With respect to "the purpose and character of

the use, including whether such use is of a commercial nature,” 17 U.S.C. § 107(1), the Ninth Circuit acknowledged that Arriba operated its site for commercial purposes. However, Arriba’s use of Kelly’s images

was more incidental and less exploitative in nature than more traditional types of commercial use. Arriba was neither using Kelly’s images to directly promote its web site nor trying to profit by selling Kelly’s images. Instead, Kelly’s images were among thousands of images in Arriba’s search engine database.¹⁵

Moreover, the court concluded that Arriba’s use was “transformative” -- that its use did not merely supersede the object of the originals, but instead added a further purpose or different character. While Kelly’s “images are artistic works intended to inform and engage the viewer in an aesthetic experience,” Arriba’s search engine “functions as a tool to help index and improve access to images on the internet.” The Ninth Circuit stressed that “Arriba’s use of the images serves a different function than Kelly’s use – improving access to information on the internet versus artistic expression.” The court closed its discussion of the first fair use factor by concluding that Arriba’s “use of Kelly’s images promotes the goals of the Copyright Act and the fair use exception.” This is because the thumbnails “do not supplant the need for the originals” and they “benefit the public by enhancing information gathering techniques on the internet.”¹⁶

With respect to the second fair use factor, the nature of the copyrighted work, the Ninth Circuit observed that “[w]orks that are creative in nature are closer to the core of intended copyright protection than are more fact-based works.”¹⁷ Moreover, “[p]ublished works are more likely to qualify as fair use because the first appearance of the artist’s expression has already occurred.”¹⁸ Kelly’s works were creative, but published. Accordingly, the Ninth Circuit concluded that the second factor weighed only

slightly in favor of Kelly.¹⁹

The court also reviewed “the amount and substantiality of the portion used in relation to the copyrighted work as a whole.” 17 U.S.C. § 107 (3). The Ninth Circuit ruled that

although Arriba did copy each of Kelly’s images as a whole, it was reasonable to do so in light of Arriba’s use of the images. It was necessary for Arriba to copy the entire image to allow users to recognize the image and decide whether to pursue more information about the image or the originating web site. If Arriba copied only part of the image, it would be more difficult to identify it, thereby reducing the usefulness and effectiveness of the visual search engine.²⁰

Finally, the Ninth Circuit decided that “the effect of the use upon the potential market for or value of the copyrighted work,” 17 U.S.C. §107 (4), weighed in favor of Arriba. The court found that the Arriba “search engine would guide users to Kelly’s web site rather than away from it.”²¹ Additionally, the thumbnail images would not harm Kelly’s ability to sell or license full size images because the low resolution of the thumbnails effectively prevented their enlargement.²²

Everything the Ninth Circuit stated with respect to Arriba can be applied with equal force to the Library Project. Although Google operates the program for commercial purposes, it is not attempting to profit from the sale of a copy of any of the books scanned into its database, and thus its use is not highly exploitative.²³ Like the Arriba search engine, Google’s use is transformative in that Google is creating a tool that makes “the full text of all the world’s books searchable by everyone.”²⁴ The tool will not supplant the original books because it will display only a few sentences in response to user queries.

Like Arriba, the Library Project involves only published works. And while some of these works will be creative, the vast majority will be non-fiction.

As in *Kelly*, Google's copying of entire books into its database is reasonable for the purpose of the effective operation of the search engine; searches of partial text necessarily would lead to incomplete results. Moreover, unlike Arriba, Google will not provide users with a copy of the entire work, but only with a few sentences surrounding the search term. And if a particular term appears many times in the book, the search engine will allow the user to view only three instances – thereby preventing the user from accessing too much of the book.

Finally, as with the Arriba search engine, it is hard to imagine how the Library Project could actually harm the market for books, given the limited amount of text a user will be able to view. To be sure, if a user could view (and print out) many pages of a book, it is conceivable that the user would rely upon the search engine rather than purchase the book. Similarly, under those circumstances, libraries might direct users to the search engine rather than purchase expensive reference materials. But when the user can access only a few sentences before and after the search term, any displacement of sales is unlikely. Moreover, the Library Project may actually benefit the market for the book by identifying it to users and demonstrating its relevance. This is particularly important for the vast majority of books that are not well publicized by their publishers. Google will encourage users to obtain a hard copy of the book by providing a link to information where the book can be borrowed or purchased.²⁵

The Owners' Response to Google's Fair Use Argument

The owners have three responses to *Kelly*. First, they note that Arriba stored a compressed, low-resolution version of each image, while Google will store the full text of each book. This seems to be a distinction without a difference, because Arriba had to make a high resolution copy before compressing it. Furthermore, the low resolution image Arriba displayed to users represents far

more of the work than the snippets Google will display to its users. In any event, neither the scanned copy nor the snippets supplant the market for the original work.²⁶

Second, they suggest that *Kelly* is distinguishable because it involved the copying of digital images on the Internet, while Google will be digitizing analog works. If an owner decides to place a work on a website, it knows that the website will be "crawled" by a software "spider" sent out by a search engine, and it knows that the spider will copy the work into its search index. Thus, by placing the work on the website, the owner has given a search engine an implied license to copy the work into its search database. By contrast, the author or publisher of a book has not given an implied license for the book to be scanned.

Google has three possible responses to this argument. One, the *Kelly* decision makes no reference to an implied license; its fair use analysis did not turn on an implied license. Two, this argument suggests that works uploaded onto the Internet are entitled to less protection than analog works. This runs contrary to the entertainment industry's repeated assertion that copyright law applies to the Internet in precisely the same manner as it applies to the analog environment.

Three, Google can argue that its opt-out feature constitutes a similar form of implied license. A critical element of the implied license argument with respect to material on the Internet is the copyright owner's ability to use an "exclusion header." In essence, an exclusion header is a software "Do Not Enter" sign that a website operator can place on its website. If a search engine's spider detects an exclusion header, it will not copy the website into the search index. Thus, if a website operator places content on the Internet without an exclusion header, the search engine can assume that the operator has given it an implied license to copy the website. Similarly, now that Google has announced its opt-out policy, it can argue that any owner that has not opted out has given it an implied license to scan.²⁷

The copyright owners' third response to *Kelly* is that it is wrongly decided. In other words, the Ninth Circuit made a mistake. The authors and publishers sued Google in federal court in New York, part of the Second Circuit. While the trial court in New York may look to *Kelly* for guidance, *Kelly* is not a binding precedent in the Second Circuit. Similarly, when the case is appealed to the Second Circuit, the Second Circuit will be interested in how the Ninth Circuit handled a similar case, but it is free to conduct its own analysis.

The owners suggest that the trial court in New York will be influenced by a decision by a federal trial judge in New York, *UMG Recordings v. MP3.com*, 92 F. Supp. 2d 349 (S.D.N.Y. 2000). MP3.com established a "space-shifting" service that allowed people who purchased a CD to access the music on the CD from different locations. MP3.com copied several thousand CDs into its server, and then provided access to an entire CD to a subscriber who demonstrated that he had possessed a copy of the CD. MP3.com argued that the copies it made on its server constituted fair use. The court rejected the argument and assessed millions of dollars of statutory damages against MP3.com.

Google will contend that MP3.com is easily distinguishable. It will claim that its use is far more transformative than MP3.com's – it is creating a search index, while MP3.com simply retransmitted copies in another medium. Additionally, Google will claim that its use will not harm any likely market for the books – there is no market for licensing books for inclusion in digital indices of the sort envisioned by Google. In contrast, MP3.com's database clearly could harm markets for online music, which the plaintiffs had already taken steps to enter. The issue of different licensing markets is discussed below in greater detail.

Google also will insist that the Ninth Circuit decided *Kelly* correctly. It will point to the Ninth Circuit's heavy reliance on the Supreme Court's most recent fair use decision, *Campbell v. Acuff-*

Rose, Music, Inc., 510 U.S. 569 (1994). Thus, *Kelly* noted that *Campbell* held that "[t]he more transformative the new work, the less important the other factors, including commercialism, become."²⁸ Likewise, *Kelly* cited *Campbell* for the proposition that "the extent of permissible copying varies with the purpose and character of the use."²⁹ And *Kelly* followed *Campbell*'s conclusion that "[a] transformative work is less likely to have an adverse impact on the market for the original than a work that merely supersedes the copyrighted work."³⁰

Perhaps most importantly, *Kelly* repeated the Supreme Court's articulation in *Campbell* and *Stewart v. Abend*, 495 U.S. 207, 236 (1990), of the objective of the fair use doctrine: "This exception 'permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity which that law is designed to foster'."³¹ Google will contend that the Library Project is completely consistent with this objective in that it will ensure that creative accomplishments do not fade into obscurity. Because the Ninth Circuit so closely followed *Campbell*, and because the Second Circuit is also obligated to follow *Campbell*, Google will urge the Second Circuit to conduct a fair use analysis similar to the Ninth Circuit's.

Regardless of *Kelly* and MP3.com, the issue the Second Circuit will probably be most interested in exploring is whether Google's use is transformative. On the one hand, Google is not "transforming" the text of any individual book into a new work, e.g., creating a parody. On the other hand, Google is creating something new and valuable – a search index consisting of the full text of millions of books – and this creation differs significantly from the uses offered by the owners. Weighing these arguments, the Ninth Circuit decided that Arriba's use was transformative. The Second Circuit will conduct its own analysis of this issue.

Intermediate Copying

Google's supporters contend that the "intermediate copying" cases also demonstrate the fair use nature of the Library Project. In these cases, courts found that fair use permitted the translation of machine-readable object code into human-readable source code as an essential step in the development of non-infringing interoperable computer programs. See, for example, the following cases: *Sega Enterprises v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992); *Atari Games Corp. v. Nintendo*, 975 F.2d 832 (Fed. Cir. 1992); *Sony Computer Entertainment v. Connectix Corp.*, 203 F.3d 596 (9th Cir.), *cert. denied*, 531 U.S. 871 (2000). Thus, Google's scanning of books should be excused because it is a necessary step in the development of a search index that displays non-infringing snippets to users.

The owners respond that the intermediate copying cases are distinguishable because they address a problem specific to software: translation of the programs is the only means of accessing ideas unprotected by copyright that are contained within the program. This problem, of course, does not exist with books. Furthermore, in the intermediate copying cases, the software developer discarded the translation once it developed its new, non-infringing program. Google, conversely, will retain the scanned copy in its search index. While acknowledging these factual differences, Google's supporters stress the underlying principle of the intermediate copying cases: that copying may be excused if it is necessary for a socially useful non-infringing end use.

The Equities

Although courts typically focus on the four fair use factors and technical questions such as whether a use is transformative, the Supreme Court has stressed that fair use is an "equitable rule of reason which permits courts to avoid rigid application of the copyright statute when, on occasion, it would stifle the very creativity

which that law is designed to foster."³² In the public debate concerning the Library Project, supporters and opponents have made a wide variety of equitable arguments that may ultimately factor into the Second Circuit's analysis. Some of these equitable arguments overlap with factors discussed above with respect to *Kelly*.

The Social Benefit of the Library Project

Google's supporters stress that by assembling a searchable index of the full text of millions of books, Google is creating a research tool of historic significance. The Library Project will make it easier than ever before for users to locate the wealth of information buried in books. Moreover, by including this information in its search index, Google will be directing students to sources of information far more reliable than the websites they so often frequent, and reacquaint a new generation with books and libraries. Additionally, by helping users identify relevant books, the Library Program will often increase demand for these works.

The Owners' Desire for Control

The owners by and large agree that the Library Project has significant social utility. Indeed, authors participating in the Authors Guild lawsuit acknowledge that the Library Project will provide them with a helpful research tool. Their objection is not that Google is creating a full text search index; it is that Google is creating the index without their permission. To be sure, the Supreme Court has stated that "[c]reative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting the broad public availability of literature, music, and the other Arts."³³ Nonetheless, some courts have viewed copyright as a mechanism for providing an author with control over the use of his creative expression. Thus, some owners believe that Google's opt-out option is insufficient because they, and not Google, should decide whether their works are digitized.

The Owners' Desire for Compensation

Additionally, many owners do not want to be left out of the search index; they want to be included, just on better terms than Google is offering. Most major U.S. publishers have joined the Partner Program, at least on a trial basis with respect to some of their titles. But the fact that the copyright owners have sued Google notwithstanding the three choices Google has given them (the Partner Program, the Library Project, or opting out) indicates that some owners want a better deal than Google is offering. One obviously better deal than the revenue sharing under the Partner Program is an up-front payment by Google for each title in the search index.

This insistence on remuneration seems to have two related bases. First, the owners repeatedly point to Google's financial success. They argue that given a market capitalization and level of profitability that may exceed that of the entire publishing industry, Google can afford to pay for the right to index their works. Second, the owners claim that Google will profit from including their works in its index, presumably by the selling of advertising. Google should not be permitted to profit from their labor without sharing more of the revenue than Google is offering under the Partner Program.

The Economics of the Library Project

Google has not disclosed how much it estimates it will spend scanning books into its search index. Microsoft announced that it will spend \$2.5 million to scan 100,000 volumes in the British Library. Assuming similar scanning costs, Google will spend \$750 million to scan the 30 million volumes contained in the collections of the five participating libraries. Google will not display advertisements on the page displaying the snippets from a particular book. Moreover, at present, advertisements will not appear on the search results page listing the responsive items

contained in the Google search index. Thus, Google will receive no advertising revenue directly attributable to the inclusion of books in the search index, at least in the short term. Instead, it appears that Google hopes that by including a large number of books in its search index, it will differentiate itself from its competitors and attract more "eyeballs," which in turn will lead to more advertising revenue.

Stock analysts have questioned the wisdom of this \$750 million investment. What the owners seek will render this already questionable investment an economic impossibility. The transaction cost of determining who owns the copyright, locating the copyright owner, and negotiating a license would be overwhelming, even to an entity like Google. Most books published in the United States include a copyright notice, but that notice does not specify whether the author, the publisher, or a third party has the right to authorize digitization. Books published outside the United States often have no copyright notice. Moreover, there is no registry of current copyright ownership, with current contact information for the owner. Thus, Google could easily spend more than a thousand dollars per volume to identify, locate, and contact the owner – even if the owner had no objection to Google scanning its work for free. The transaction costs alone could easily reach over \$25 billion (\$1000/book x 25,000,000 in-copyright books).

Google might be willing to take a \$750 million gamble, but almost certainly will not be willing to take a \$25 billion gamble, which does not even include the license fees some owners might demand. If Google were required to obtain permission to scan the in-copyright books, it probably would scan only public domain works and works whose owners affirmatively requested to be included. Only a small percentage of owners are likely to take this step. As noted above, it often is not clear whether the publisher or the author has the right to authorize digitization of a work. If the author is deceased, his heirs might not be aware that they own the copyright. And

for out of print books where the publisher controls the copyright, the publisher – if it is still in business – might not have any economic incentive to request Google to scan the book (more than 75% of the in-copyright books are out-of-print). In sum, most books probably would not be included in the search index.

Some have suggested that the transaction costs could be reduced by a collective license, along the lines of the licenses ASCAP and BMI provide for the public performance of musical compositions. While such an arrangement could theoretically work on a going forward basis – for books published after 2005 – it would not work for the 25,000,000 existing, in-copyright books. Getting a significant share of the copyright owners of these 25,000,000 works to agree to participate in a collective license system would be as costly as Google getting their permission directly. Moreover, even on a going forward basis, it is unlikely that the hypothetical newly established authors' collection society would reach an agreement on license terms with Google. The collection society and Google likely would have very different perceptions on what would be a reasonable license fee for including a book in the search index.

Harm to the Owners

It is easy to see the harm to the public flowing from an incomplete search index – the public will not find as complete a universe of relevant books. And an incomplete search index is the inevitable result from placing on Google the burden of obtaining permission from the owners.

But it is much more difficult to identify the harm to the owners deriving from allocating to them the burden of opting out. The cost of the owners opting out is much less than Google's cost of seeking permission. An author and her publisher are much better placed than Google to determine who has the right to authorize digitization. And whoever the owner proves to be, it obviously knows where it is located.³⁴

Moreover, an owner's failure to opt-out probably will not harm the market for the work. As noted above, because Google will display only snippets of the work, a book's inclusion in the search index will not displace sales. Google will display no more than three snippets per work with respect to a particular search term. Further, Google will not display any snippets from reference works such as dictionaries where the display of snippets arguably could harm the market.

The owners argue that the Library Project restricts owners' ability to license their works to search engine providers. The existence of the Partner Program, which involves licensing, demonstrates that the Library Project does not preclude lucrative licensing arrangements. By participating in the Partner Program, publishers receive revenue streams not available to them under the Library Project. Google presumably prefers for publishers to participate in the Partner Program because Google saves the cost of digitizing the content if publishers provide Google with the books in digital format. And Google has made clear that it is willing to upgrade a book from the Library Project to the Partner Program upon the owner's request.

Furthermore, Yahoo announced the formation of the Open Content Alliance, which will include works licensed by their owners, nearly a year after Google announced the Library Project. Google's Library Project obviously did not deter Yahoo from adopting a different business model based on licensing.

Significantly, the Library Project will not compete with a business model involving licensed works because such a model will probably show more than just snippets. While the Library Project will help users identify the entire universe of relevant books, a model with licensed works will provide users with deeper exposure to a much smaller group of books.³⁵ Each business model will satisfy different needs. Stated differently, the Library Project targets the indexing market, while other online digitization projects aim at the

sampling market. See *BMG Music v. Gonzalez*, 430 F.3d 888 (7th Cir. 2005). By concentrating on the indexing market, the Library Project will not harm the sampling market.

Finally, as discussed above, the enormous transaction costs involved in compiling a comprehensive full text search index with the owners' authorization preclude the creation of such an index in that manner. Thus, Google's index does not deprive owners of potential revenues from "traditional, reasonable, or likely to be developed markets" for the work. See *American Geophysical Union v. Texaco Inc.*, 60 F.3d 913 (2nd Cir.), cert. denied, 516 U.S. 1005 (1995).³⁶

The Value of an Indexing License

Assuming that the transaction costs were not an insurmountable barrier to the existence of a licensing market for indexing rights for the universe of published books, the value of a license with respect to any particular book would be relatively small. For the vast majority of users, an index to the vast majority of books is more than adequate. Thus, from the perspective of Google and its users, the marginal importance of the inclusion of any particular book is small, and Google would be willing to pay at most an extremely modest fee for the indexing rights to any single book. Even for a publisher that owned the rights to a large backlist of books, the total license fees it would receive would probably be significantly less than the legal fees the litigation over the Library Project will generate. Although the aggregate value of all the licenses in this hypothetical market would be enormous, copyright ownership is dispersed among so many authors and publishers that any one owner could reasonably expect only trivial license fees.

The Definition of Snippets

The owners have argued that "snippet" is not a legal term. Therefore, at some point in the future Google could start displaying larger portions of

the indexed books, which could displace sales. Google responds that if it does change its policy in a manner that hurts sales, the owners can sue at that time. Since displaying some of a book's text in response to a search query implicates both the reproduction right and the display right, an owner will be able to bring an infringement action against Google when it changes its policy, even if that occurs long after the original scanning of the book. Accordingly, there is no reason to prevent Google from proceeding now, when its practices do not harm owners. It is unlikely that these fees would increase authors' incentive to write.

Security

The owners have expressed concern about the security of the digitized copies in Google's search index. They fear that someone would be able to hack into the index and upload the digitized books onto the Internet, where they would be publicly available.³⁷ Google, however, has a significant incentive to protect the security of its index: it would not want to see its \$750 million investment evaporate. Moreover, given the ease of digitizing any single book bought in a bookstore or checked out of a library, it is far from clear why anyone would bother to hack into the Google index to access digitized books. And even if someone were to hack into the Google search index, the information would be formatted in a manner that facilitates word search, not distribution of full text, *i.e.*, the search index does not consist of pdf files.

Finally, the Second Circuit has made clear that if an entity lawfully extracted information from another company's database, the entity is not liable for a third party's use of that information to infringe the other company's copyright in its database. See *Matthew Bender & Co. v. Hyperlaw, Inc.*, 158 F.3d 693 (2nd Cir. 1998). Thus, the Second Circuit would not hold Google responsible for hackers' unlawful uses of the contents of its search index, unless the owners can show that

Google somehow encouraged or induced the hackers to infringe. Notwithstanding the absence of direct or secondary liability on its part for the infringing actions of hackers, Google would still have substantial business reasons for maintaining the security of its search index, as discussed above.

Floodgates

The owners suggest that if Google is “able to get away” with its Library Project, other search engines will also digitize their works without authorization.³⁸ But it is not clear how more digitization will harm the owners, so long as the other search engines confine their display of text to snippets. And if the other search engines display more than snippets, in a manner that interferes with the sale of works or their licensing to business models such as the Partner Program, the owners can sue those search engines at that time.

The owners also use the floodgates argument to attack the utility of Google’s opt-out. If other search engines engage in mass digitization projects with opt-out features, owners would have to opt-out repeatedly -- a burdensome process, especially for individual authors. As a practical matter, however, only a small number of search engine firms have the resources to engage in digitization programs on the scale of Google’s Library Project. And even if many specialized indices emerge, the number of indices that likely will include any specific book is small. Also, if this does become a problem at some point in the future, groups like the Authors Guild could maintain a general opt-out register that search engines could honor.

The Impact on Search Engines

The court’s analysis of the Library Project could affect the operation of Internet search engines generally. A search engine firm sends out software “spiders” that crawl publicly accessible websites and copy vast quantities of data into the

search engine’s database. As a practical matter, each of the major search engine companies copies a large (and increasing) percentage of the entire World Wide Web every few weeks to keep the database current and comprehensive. When a user issues a query, the search engine searches the websites stored in its database for relevant information. The response provided to the user typically contains links both to the original site as well as to the “cache” copy of the website stored in the search engine’s database.

Significantly, the search engines conduct this vast amount of copying without the express permission of the website authors. Rather, the search engine firms believe that their activities constitute fair use. In other words, the billions of dollars of market capital represented by the search engine companies are based primarily on the fair use doctrine. If a court concludes that Google’s scanning of millions of books is not a privileged fair use, then search engines’ scanning of millions of websites might not constitute fair use either, unless the court takes pains to distinguish one situation from the other. As discussed above, the owners contend that search engines have an “implied license” to scan works posted on the Internet. But the Ninth Circuit in *Kelly v. Arriba Soft* relied on fair use, not implied license.

The Impact on the Publishing Industry

The owners contend that if Google is permitted to assemble a search index of in-copyright books, it will have an unfair advantage over publishers that want to provide e-books. This is because the Library Project will lead consumers to perceive Google as the leading source for digital books.³⁹ This argument overlooks the fact that Google will be able to provide consumers with the full text of a book in its search index only with the permission of the copyright owner; fair use will not permit Google to make such a distribution without the owner’s authorization.

The owners similarly worry that the Library

Project will provide Google a bridgehead in the publishing industry, which it will be able to exploit with its enormous resources. Of course, more competition for publishers should benefit both authors and consumers. And if Google engages in anti-competitive conduct, the publishers can turn to the antitrust laws.

What do Authors Want?

The Authors Guild, which sued Google, represents only 8,000 authors. Thus, its positions do not necessarily reflect the views of the hundreds of thousands of authors whose books would be scanned under the Library Project. Most authors want their books to be found and read.⁴⁰ Moreover, authors are aware that an ever increasing percentage of students and businesses conduct research primarily, if not exclusively, online. Hence, if books cannot be searched online, many users will never locate them. The Library Project is predicated upon the assumption the authors generally want their books to be included in the search database so that readers can find them.

The Library Project is particularly important for authors of out-of-print books. While the publishers may participate in digitization projects such as the Partner Program with respect to in-print books, they have no incentive to devote any effort to the out-of-print books, which no longer have any economic value. But since the publishers of these out-of-print books may still hold the copyright, the authors of the books do not have the legal right to authorize Google to scan their books. This large class of authors probably is pleased that Google is providing users with a mechanism to find their abandoned books. Indeed, many of them might even be willing to pay Google to include their books in its search index, and are happy that Google is doing so free of charge. While the authors typically will receive no direct economic benefit from the rediscovery of their out-of-print works, it could enhance their reputations and disseminate their ideas. In any

event, if an owner does not want Google to scan her in-print or out-of-print book, Google will honor her request.

The Privatization of Knowledge

Some scholars have acknowledged that the Library Project can greatly assist research activities, but nonetheless voice concern that a corporate entity is assembling this vast search index rather than a public library.⁴¹ They feel that Google's ability to influence search results through its search algorithm will provide it with too much control over the access to knowledge. Additionally, they worry that Google will have an economic incentive not to respect the privacy of its users.

While in theory it might be preferable from a societal point of view for the Library Project to be conducted by libraries rather than a private corporation, libraries simply do not have the resources to do so. Thus, as practical matter, only a large search engine such as Google has both the resources and the incentive to perform this activity.

The Legality of the Library Copies

Google will provide each library participating in the Library Project with a digital copy of the books in its collection scanned by Google. The owners have not yet sued the libraries, nor have they expressed any intention to do so. In the event of litigation, the lawfulness of the library copies will turn on how the libraries are using them.⁴² A search index assembled by the libraries should receive even more favorable treatment than Google's given the libraries' non-commercial purpose. On the other hand, a court probably would find infringement if a library made the full text of in-copyright works available online to the general public.

The owners observe that because the University of Michigan is a state institution, the doctrine

of sovereign immunity will prevent the owners from suing the University if the University misuses its copies (*e.g.*, distributes them publicly as opposed to storing them in a restricted or dark archive). Although sovereign immunity would shelter the University from damages liability, the owners could still pursue injunctive relief against the University's officers and librarians. This would enable the owners to stop any misuse by the University of Michigan Library.

The Orphan Works Initiative

The Copyright Office has made recommendations to Congress on how to address the orphan works problem -- how to enable uses of works whose owners cannot be identified or located.⁴³ There are some similarities between orphan works and the Google Library Project, but there are significant differences as well. Certainly, many of the books Google seeks to include in its search index probably are orphan works. But Google's use of each these works is less extensive than the uses others hope to make of orphan works. As discussed above, Google will scan an entire work into its search index, but will make only snippets available to the public. In contrast, many of those who hope to use orphan works intend to make the entire work available to the public. For example, a library intending to digitize an archive of sound recordings of folk songs probably plans to make the sound recordings available on the Internet. This public distribution of entire orphan works will limit the availability of a fair use defense in many cases; hence, the user needs a new form of relief along the lines of what the Copyright Office is proposing. Conversely, Google can make a stronger fair use argument because it will display only snippets, and not entire works.

Additionally, the relief the Copyright Office is proposing to Congress will not help Google. The Copyright Office's proposal limits the remedies available to a reappearing owner if the user made a good faith effort to locate the owner. Be-

cause of the scale of the Library Project, Google cannot attempt to locate the owners of the all the books it intends to include in its search index.

International Dimensions

Fair use under the U.S. Copyright Act is generally broader and more flexible than the copyright exceptions in other countries, including fair dealing in the U.K. Thus, the scanning of a library of books might not be permitted under the copyright laws of most other countries. However, copyright law is territorial; that is, one infringes the copyright laws of a particular country only with respect to acts of infringement that occurred in that country. Since Google will be scanning in-copyright books just in the United States, the only relevant law with respect to the scanning is U.S. copyright law. (Google will scan only public domain books at Oxford.)

Nonetheless, the search results will be viewable in other countries. This means that Google's distribution of a few sentences from a book to a user in another country must be analyzed under that country's copyright laws.⁴⁴ While the copyright laws of most countries might not be so generous as to allow the reproduction of an entire book, almost all copyright laws do permit short quotations. These exceptions for quotations should be sufficient to protect Google's transmission of Library Project search results to users.

Conclusion

Society would benefit significantly from a search index that includes the full text of a large percentage of all published books. Such a comprehensive index can be compiled only without the obtaining the permission of all the copyright owners; the transaction costs of obtaining all the permissions would be so large as to render the project an economic impossibility. At the same time, compiling such an index without obtaining the owners' permissions will not hurt the owners in any discernable way, provided that the search

results display only snippets of text. It will not diminish the market for the books, nor will it prevent licensed digitization projects that provide users with more text for a narrower range of books. Google further reduces the possibility of harm by permitting owners to opt-out of the Li-

brary Project altogether, or opt-in to the Partner Program. A court correctly applying the fair use doctrine as an equitable rule of reason should permit Google's Library Project to proceed.

REFERENCES

- American Geophysical Union v. Texaco Inc.*, 60 F.3d 913 (2nd Cir.), cert. denied, 516 U.S. 1005 (1995).
- Atari Games Corp. v. Nintendo*, 975 F.2d 832 (Fed. Cir. 1992).
- Basic Books, Inc. v. Kinko's Graphics Corp.*, 758 F. Supp. 1522 (S.D.N.Y. 1991).
- BMG Music v. Gonzalez*, 430 F.3d 888 (7th Cir. 2005).
- Campbell v. Acuff-Rose, Music, Inc.*, 510 U.S. 569 (1994).
- Field v. Google*, No. CV-S-04-0413-RCJ-LRL (D-NV Jan. 12, 2006).
- Kelly v. Arriba Soft*, 336 F.3d 811 (9th Cir. 2003).
- Matthew Bender & Co. v. Hyperlaw, Inc.*, 158 F.3d 693 (2nd Cir.1998).
- Princeton University Press v. Michigan Document Services, Inc.*, 99 F.3d 1381 (6th Cir. 1996).
- Sega Enterprises v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).
- Sony Computer Entertainment v. Connectix Corp.*, 203 F.3d 596 (9th Cir.), cert. denied, 531 U.S. 871 (2000).
- Stewart v. Abend*, 495 U.S. 207, 237 (1990).
- Thatcher, S. G. (2005). "Fair Use in Theory and Practice: Reflections On its History and the Google Case." NACUA conference on "The Wired University: Legal Issues at the Copyright, Computer Law, and Internet Intersection." Arlington, VA.
- Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151 (1975).
- UMG Recordings v. MP3.com*, 92 F. Supp. 2d 349 (S.D.N.Y. 2000).

The Google Library Project—Band

NOTES

1. Earlier versions of this article appeared in *E-Commerce Law & Policy* and a briefing paper prepared for the Office of Information Technology Policy of the American Library Association.
2. Many articles incorrectly suggest that users can access the full text of in-copyright works. Google's supporters discuss the enormous social value of a digital index of the world's books, while Google's opponents stress Google's use of copyrighted material without permission.
3. Many of the arguments recounted here emerged in various public debates concerning the Library Project, including debates in which the author participated. See, e.g., "Gutenberg meets Google: The Debate About Google Print," pff.org/issues-pubs/pops/pop13.1googletranscript.pdf.
4. Displays of the different treatments can be found at <http://books.google.com/googlebooks/library.html>.
5. As discussed later in the article, only the agreement with the University of Michigan has been made public.
6. Because the author, the publisher, or a third party can own the copyright in a work, this paper will refer to "owners."
7. Google initially required owners to state under penalty of perjury that they owned the copyright in the books they wished to opt-out. Google relaxed this requirement after the owners complained that they felt uncomfortable making assertions of ownership "under penalty of perjury" because of the complexities of copyright law. See Sanford G. Thatcher, *Fair Use in Theory and Practice: Reflections on its History and the Google Case* (pp. 11-12).
8. The contract was disclosed as required under the Michigan Freedom of Information Act.
9. See "Cooperative Agreement Between Google Inc. and the Regents of the University of Michigan," section 4.4.1.
10. UM Library/Google Digitization Partnership FAQ, August 2005.
11. HarperCollins recently announced that it intends to scan 20,000 books on its backlist and make the digital text available on its server for search engines to index. It will offer this service to search engines free of charge. The technological feasibility of this distributed indexing has not yet been proven.
12. The copyright issues relating to the copies Google makes for the participating libraries are discussed later in the article.
13. Association of American Publishers Press Release, "Google Library Project Raises Serious Questions for Publishers and Authors," August 12, 2005.
14. In its answer to the Authors Guild lawsuit, Google raised numerous other defenses, including merger doctrine, scenes a faire, failure to comply with copyright registration formalities, lack of suitability for class action treatments, and the plaintiffs' lack of standing. In the Publishers' suit, Google raised many of these defenses, as well as license to scan and the publishers' lack of ownership of electronic rights.
15. *Kelly*, 336 F.3d at 818.
16. *Kelly*, 336 F.3d at 818-20.
17. *Kelly*, 336 F.3d at 820.
18. *Id.*
19. *Id.*
20. *Kelly*, 336 F.3d at 821.
21. *Id.*
22. See also *Field v. Google*, No. CV-S-04-0413-RCJ-LRL (D-NV Jan. 12, 2006). Blake Field brought a copyright infringement lawsuit against Google after the search engine automatically copied and cached 51 stories he posted on his website. Google argued that its Google Cache feature, which allows Google users to link to an archival copy of websites indexed by Google, does not violate copyright law. The court granted summary judgment in favor of Google on five independent bases:
 - (a) Serving a webpage from the Google Cache does not constitute direct infringement, because it results from automated, non-volitional activity initiated by users;

The Google Library Project—Band

- (b) Field's conduct (posting an "allow all" robot.txt header and then intentionally failing to set a "no archive" metatag) indicated that he impliedly licensed search engines to serve his archived web page;
 - (c) Fields is estopped from asserting a copyright claim because he induced Google to infringe by using software code that invited Google to cache and serve his website;
 - (d) The Google Cache is a fair use; and
 - (e) The Google Cache qualifies for the Digital Millennium Copyright Act's section 512(b) caching "safe harbor" for online service providers.
23. In *Field v. Google*, the court dismissed the argument that Google was a commercial entity by stressing that there was no evidence that Google profited from its use of Field's stories. The court observed that his works were among the billions of works in Google's database. In the Library Project cases, Google will be able to make the same argument with respect to any one owner.
24. Google Blog post, "Making Books Easier to Find," August 11, 2005. This tool includes not only digital copies of the books, but also an index of all the words in the books, and sophisticated software that enables the user to search the index and access search results.
25. In *Field v. Google*, the court considered an additional factor: "whether an alleged infringer has acted in good faith." Google's allowing owners to opt-out, its refusal to display any snippets for certain reference works, and its willingness to upgrade any book into the revenue sharing Partner Program give Google strong evidence that it is acting in good faith.
26. Additionally, in *Field v. Google*, the court found Google's presentation of caches of the full text of Field's stories to be a fair use.
27. In *Field v. Google*, Google raised implied license as a defense. But Google's implied license argument in *Field* does not support the owners' attempt to distinguish *Kelly* on the basis of the unique characteristics of spidering the Web. In *Field*, the court treated implied license and fair use as distinct defenses. Thus, the absence of an implied license for the scanning in the Library Project does not weaken Google's fair use defense based on *Kelly*. Moreover, Field used a software header that specifically invited Google's spider to crawl his website. There is no evidence that Kelly made a similar invitation to Arriba Soft.
28. *Kelly*, 336 F.3d at 818, citing *Campbell*, 510 U.S. at 579.
29. *Kelly*, 336 F.3d at 820, citing *Campbell*, 510 U.S. at 586-87.
30. *Kelly*, 336 F.3d at 821, citing *Campbell*, 510 U.S. at 591.
31. *Kelly*, 336 F.3d at 817.
32. *Stewart v. Abend*, 495 U.S. 207, 237 (1990).
33. *Twentieth Century Music Corp. v. Aiken*, 422 U.S. 151, 156 (1975).
34. See Georgia Harper, "Google This: The Bottom Line," [utsystem.edu/ogc/intellectual property/googlethis.htm](http://utsystem.edu/ogc/intellectual%20property/googlethis.htm).
35. Testimony of Paul Aiken on Behalf of the Authors Guild, House Committee on Energy and Commerce, Subcommittee on Commerce, Trade, and Consumer Protection, Hearing on "Fair Use: Its Effects on Commerce and Industry," November 16, 2005, p. 11: "And a negotiated license could pave the way for a real online library – something far beyond the excerpts Google intends to offer through its Google Library program" (Aiken Testimony).
36. The court in *Field v. Google* found that "there is no evidence before the Court of any market for licensing search engines the right to allow access to Web pages through "Cached" links, or evidence that one is likely to develop."
- The owners could argue that the Library Project might deprive them of the promotional value of their works, e.g., steering traffic away from their websites were they to offer search capability. See *Video Pipeline, Inc., v. Buena Vista, Inc.*, 342 F.3d 191 (3rd Cir. 2003), *cert. denied*, 540 U.S. 1178 (2004). Interpreting the fourth fair use factor to incorporate promotional value of this sort significantly limits the utility of the fair use privilege because every work theoretically has some promotional value. Additionally, if a particular owner believes that a search index of the works it owns does have promotional value, it can simply opt-out of the Library Project. In contrast, Video Pipeline did not permit Disney to opt-out of its service displaying film trailers.
37. Aiken Testimony, p. 8.
38. Aiken Testimony, pp. 8-9.
39. Aiken Testimony, pp. 10-11.

The Google Library Project—Band

40. See Tim O'Reilly, "Search and Rescue," *The New York Times*, September 28, 2005.
 41. See Siva Vaidhynathan, "A Risky Gamble With Google," *The Chronicle of Higher Education*, December 2, 2005.
 42. Some argue that while fair use might permit a library to make a digital copy of a work for archival purposes, it would not permit Google, a commercial entity, to make the archival copy for the library. This argument derives from cases such as *Princeton University Press v. Michigan Document Services, Inc.*, 99 F.3d 1381 (6th Cir. 1996)(en banc) and *Basic Books, Inc. v. Kinko's Graphics Corp.*, 758 F. Supp. 1522 (S.D.N.Y. 1991), where the courts ruled that commercial copy centers could not claim that their photocopying of coursepacks constituted a non-commercial use. In these cases, however, the copy centers often made hundreds of copies of a single work in a manner designed to supplant the copyright holder's rights. Here, by contrast, Google will be making just one copy of any given work for a library. If a library is permitted under Section 108(b) to make three copies of an unpublished letter for preservation purposes, surely it can retain a preservation specialist to make those copies for it.

In any event, as a practical matter, if the library copy became a focus of the litigation, it is unlikely that a court would find the library copy to be infringing, but Google's index copy to be a fair use. Ultimately, the library copy is ancillary to the index copy; Google makes the library copy as consideration for obtaining access to the book for the purpose of making the index copy.
 43. United States Copyright Office, *Report On Orphan Works*, January 2006.
 44. Google arguably is causing a copy of the sentences to be made in the random access memory of the user's computer. Thus, the lawfulness of this copy must be examined under the copyright laws of the user's country.
-

Jonathan Band represents Internet companies and library associations on intellectual property and Internet policy matters in Washington, D.C. He does not represent the parties in the litigation discussed in the article. He received his B.A. from Harvard College and his J.D. from Yale Law School, and is an adjunct professor at the Georgetown University Law Center.