The Google Library Project

Hal R. Varian*

January 2006
Revised: February 10, 2006

Abstract

This is an economic analysis of the Google Library project. I describe the project and outline why it is consistent with the legal doctrine of fair use. I go on to examine the transactions costs associated with opt-in and opt-out models for publisher participation. I conclude that the Google Library Project is legally sound and economically sensible. In particular, an opt-in model would incur very substantial transactions costs, making the entire undertaking problematic.

*Prepared for the AEI-Brookings discussion, "The Google Copyright Controversy: Implications of Digitizing the World's Libraries," February 24, 2006, Washington, DC. This document represents my personal views only and does not represent in any way the views of Google. I have consulted for Google since 2002 on various matters, but I have never had any direct involvement with the Google Book project. I have shown drafts of this document to members of that team in order to ensure that my description of their procedures is accurate but any remaining errors are, of course, solely my responsibility. I also wish to thank Pamela Samuelson and Stan Leibowitz, who offered helpful suggestions and references.

The Google Library Project is an attempt to organize the world's information about books available in libraries. It is essentially an online card catalog that allows for more efficient search than conventional card catalogs.

The Google Library Project is part of Google Book Search, which is available at `http://books.google.com`; the "About Google Book Search" link on that page contains the definitive description of the program. Note that Google Book Search is currently in beta, so various features of the program may change.

Google Book Search contains scanned images of books from two sources: the "Partner Program" and the "Library Project." In order to join the Partner Program, publishers send a message to Google asking that specific books be entered into a database of scanned images. Google then adds the scanned images to the Google Book Search database either by 1) scanning in a physical copy of the book provided by the publisher or a library or 2) using a PDF file provided by the publisher. Note that the Google Partner Program is an opt-in program: publishers have to specifically request that their content be added to the index.

The Google Library Project does not require the publisher to make a specific request to join the program, though publishers are free remove any of their books from the program at any time (assuming that they hold copyright to the work.). Google has made arrangements with several libraries such as Harvard University Library and the University of Michigan Library that enables them to scan in the works in their collections. This is done at Google's expense using special technology it developed for the project. If things go as planned, there will eventually be about 25 to 30 million books in the Google Library catalog.

It is important to distinguish the copyright status of the various types of content in the Google Library Project. The Partner Program generally consists of copyrighted material which is provided by the publisher. If the publisher provides physical books or scanned images, the works are typically "in print;" that is, available via normal retail channels. If the publisher simply provides permission to use books scanned in as part of the Library Project, the books may be in or out of print.

In contrast to the Partner Program, the Library Project contains books that are both in copyright and out of copyright and in print and out of print.

Users can search the contents in these two collections using the same interface: typing a query into a search box just as though they were searching the Web. Google searches through the text of books in its index using special

algorithms optimized for this purpose and finds books that are related to the query the user entered.

The search process is thus similar to using an online book catalog, but the Google search engine can actually look at the text inside the book, allowing it to give more accurate results than could be achieved from simple title and subject searches. Hence, the Google Library project allows users to find relevant material that would be almost impossible to find any other way.

The titles of relevant works are returned to the user in a ranked list, similar to the way web search results are returned. As in web search, book search results may contain both search results and sponsored links, which are ads triggered by the search terms. As with web search, Google receives no money from clicks on search results, but does receive money from ad clicks.

If the user clicks on one of the displayed search results, further information about the book is displayed. The information returned depends on which type of content is found.

If the work is out of copyright and therefore in the public domain, users can examine the book online, buy it from a used-book dealer or make whatever use of the contents that they wish, consistent with the public domain status of the work.

If the work is in copyright and permission has been received from the publisher for it to be in the Partner Program, links to several pieces of information are displayed. Currently, this information includes the front cover, the copyright page, the table of contents, the index and the back cover. The page also contains links to the publisher and to various book retailers where the user can purchase the book. Furthermore, the user can look at a few "sample pages" of the book so they can see how their search term appeared in the book itself.

If the link that the user clicks on refers to a book that was scanned in as part of the Library Project and a publisher has not provided specific permission to use the text in the Partner program, then he or she sees a screen like that depicted in Figure 1.

In this case, the results page displays a few snippets—two or three lines of text—from the book. It also displays links where one can buy used copies of the book or borrow them from libraries. Potentially links could be added allowing the user to purchase the book from the publisher or other suppliers via print on demand or via electronic books.

As indicated above, if the copyright owner requests that even snippets from the book not be made available via the Google Library program, Google

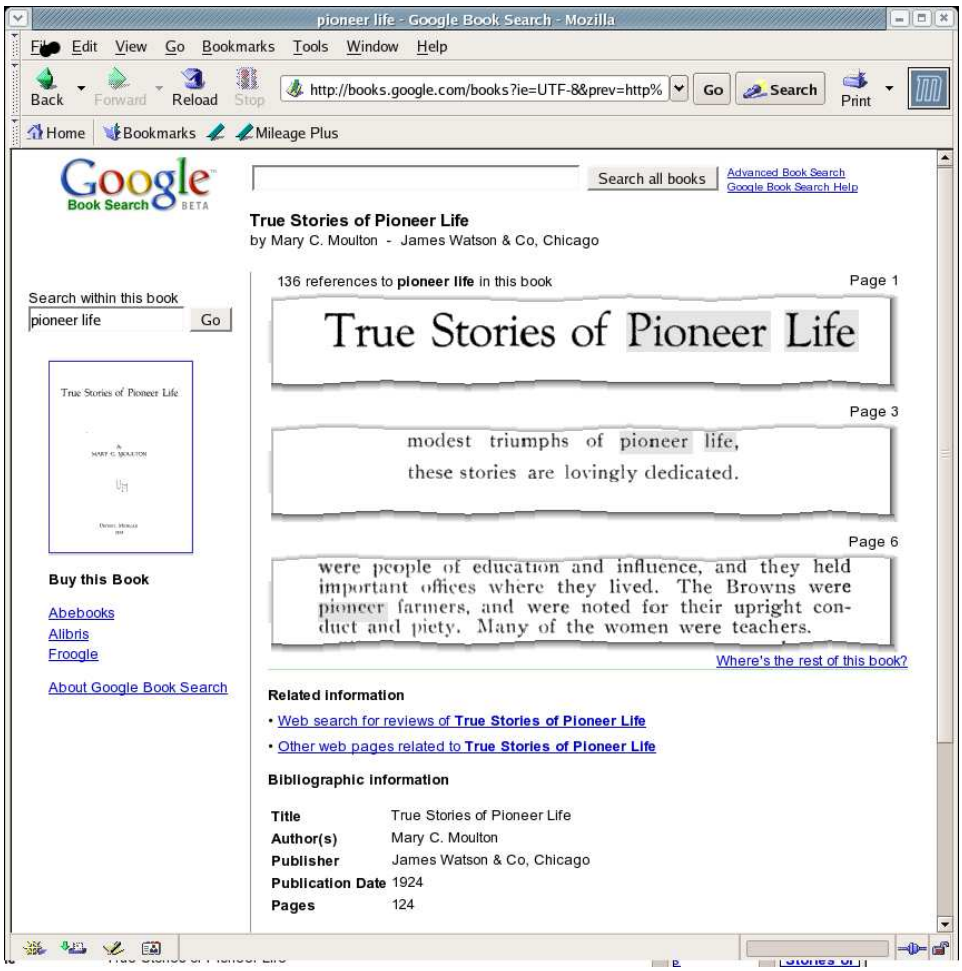

Figure 1: Screen shot from Google Library.

	Partner program	Library program
Content source	publishers	libraries
Publisher participation	opt-in	opt-out
Type of books	in print	in and out of print
Displayed results	sample pages	snippets
Links to	booksellers	booksellers and libraries

Table 1: Summary of Google Partner and Library Programs

will remove that material from the Library Project database.

Table 1 summarize the two programs along the dimensions we have discussed.

1 Nature of the controversy

As mentioned earlier, libraries generally contain three sorts of books: those that are out of copyright, those that are in copyright but out of print, and those that are in copyright and in print. Books that are out of copyright are in the public domain and there is no quarrel that they can be accessed and used by everyone. This category represents about 15 percent of the books in university libraries.

Books that are in copyright and in print are dealt with principally via the Google Partner program: if publishers choose to submit these books to the Google Partner program, users can gain limited access to parts of the book. These two categories of books are not particularly controversial.

The controversy involves books that are in copyright but have not been explicitly been submitted to the Google Partner program. Some of these books will be in print (i.e., available for purchase via normal retail channels) and some will the out of print. Out-of-print books will generally only be available to the public via libraries or used book sellers. Since this category represents about 65 percent of the books in university libraries, making such books searchable offers significant public benefits. Organizing such hard-to-find information and making it "universally accessible and useful" is central to Google's mission.

It is clear that this project will be beneficial to users as they can find useful material much more easily than they have been able to in the past.

It is also clear that this will be beneficial to libraries and other archives in advancing their mission to help users find the information they need. Used book stores, both online and off, would also benefit from this program, since it basically amounts to free advertising for their products.

Certainly, existing publishers benefit from the Google Partner program, as hundreds of them have chosen to submit their books to the program. Rights holders of the content in Google Library program could also potentially benefit from this material as they could collect royalties from print-on-demand, electronic books, and other services that could provide complete copies of these materials.

There is substantial evidence that programs such as the Google Publisher Partner program and Amazon's Look-Inside-the-Book program have helped book readers find the books they want, thereby benefiting publishers and authors. Jeff Bezos claims that participation in Amazon's Look-Inside-the-Book program increases sales by about 15 percent.

Who, then, could possibly complain about the Google Library Project?

In September 2004, the Authors Guild, which has about 8,000 authors as members, filed a complaint against Google claiming that the Library Project violated copyright.[1] The Association of American Publishers filed a similar lawsuit on October 19, 2005, on behalf of five of its members, The McGraw-Hill Companies, Pearson Education, Penguin Group (USA), Simon & Schuster and John Wiley & Sons.

At first glance, it is hard to understand their opposition to the Library Project. After all, any copyright holder can opt out of the Google Library project simply by sending Google a note indicating that they don't want their works to be listed. Furthermore, the rights holders to the most relevant category of works, books that are in copyright but out of print, are making no revenue from these works currently. The Google Library program at offers the opportunity to generate future revenue from these works via print-on-demand or other arrangements.

No sensible person would argue that having the works searchable with 2 or 3 line snippets shown to searchers would have any *negative* impact on the already minimal revenues from such works.[2]

[1] `http://www.boingboing.net/images/AuthorsGuildGoogleComplaint1.pdf`

[2] I say that rights holders are making minimal revenue from out-of-print books rather than no revenue, as they could potentially make small amounts from reprint rights via Copyright Clearance Center and other similar organizations. However, these amounts would typically be very small indeed.

On reflection, however, the motivations become clearer. From the viewpoint of a publisher, the only thing worse than having their works listed in the Google Library catalog is *not* having their works listed. Even if it is true that the copyright holders currently make minimal revenue from out-of-print books, they would certainly welcome the opportunity to charge someone, such as Google, who wanted to make organize that content. It's not that the publishers don't want to participate in the program—there is already simple way to opt out. What the publishers want is to be listed in the Google catalog and have Google compensate them for being listed.

Normally information aggregators such as Books in Print, or library cataloging services, do not pay publishers when they catalog or organize their books. In fact, publishers are typically more than willing to be included in such lists since it helps users find their works. This is, of course, part of the reason for the success of programs such as Google Partner program or the Amazon Look-Inside-the-Book program.

In this essay I will argue that 1) Google's Library Project fits comfortably with the copyright doctrine of "fair use;" 2) Google's opt-out model for participation is an appropriate procedure for minimizing transactions costs, 3) the program provides significant intellectual, cultural, and historical benefits to users at no cost the possibility of significant benefits to both authors and publishers.

2 Consulting works to create an index

The central complaint about the Google program is that is that Google intends to scan books into the program without the prior permission of the rights holders. As we have mentioned repeatedly, any rights holder can opt-out of the program simply by sending a request to Google.

This opt-out procedure is, in fact, the same as that used by Google and other search engines on the World Wide Web. A webmaster can put a file called `robots.txt` in the top-level directory of his site that requests that search engines do not index some subset of the material on the site. Google, and most other legitimate search engines, scrupulously honor such requests.

The system works well, in general. In fact, it is hard to imagine how an opt-in system would work for web search. Consider how difficult it would be to make bilateral arrangements with every search engine and every web site for permission to index the material. The opt-out procedure used on the web

is a sensible one since in minimizes transactions costs involved in indexing web sites.

We discuss transactions costs further below. Here we look at the question of whether scanning content in order to construct an index is permissible under US Copyright law; I argue that it is permissible under the doctrine of fair use.

3 Fair use

"Fair use" is a legal doctrine that is an exception to the charge of copyright infringement. It "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."[3] This is exactly the sort of case to which fair use was meant to apply.

It is worth stating section 107 of the U.S. Copyright Act of 1976 in its entirety as the interpretation of this section critical to understanding the legal background surrounding the Google Library program.

> Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified in that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include
>
> - The purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
> - The nature of the copyrighted work;
> - The amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
> - The effect of the use upon the potential market for or value of the copyrighted work.

[3] Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc., 109 F.3d 1394, 1399 (9th Cir. 1997)

Note that this provision specially states that reproduction for purposes of scholarship or research is not an infringement of copyright. This is noteworthy since the Google Library program is precisely a program to allow users to research topics of interest to them using the catalog generated by the Google technology.

Furthermore, the Act spells out four factors that are relevant to whether or not an activity counts as fair use. Let us consider these factors one by one.

Purpose and character. The law has tended to allow broader use of copyrighted works for educational purposes than for commercial purposes. Google is, of course, a public company and does make a profit. However, note that in this case the economic nature of the activity is somewhat indirect. Google does not profit directly from use of its online catalog. It makes no profit by sending a user to library to access the book. It does not reproduce or sell the work itself. At most, Google receives payment when a user clicks on a sponsored link in the results list, which contains no copyrighted material. Sponsored links would typically be ads for related works, used book stores, document delivery services and the like. The particular sponsored links that are shown are based entirely on the user's query, *not* on the the contents of the copyrighted work.

Nature of the copyrighted work. The law has allowed broader use of materials containing factual material (i.e., non-fiction) as opposed to creative material (fiction). Most of the searches in Google Book Search are expected to be for factual research, as is currently the case for Google Web search. Furthermore most of the books in the Partner Program are non-fiction.

Portion used. Only a few lines (a "snippet") of a copyrighted work are displayed in order to show the user the context in which the query appears. For almost all books, the displayed snippet will be a tiny part of the entire work.

Effect on the market or value. Providing a catalog that enables users to find the works they want surely enhances the value of such works. Since publishers currently make essentially nothing from out-of-print works, the Google Library program can hardly have a negative impact

on that current market. Since it enables to find works that are in print, the project could almost certainly have a positive impact on book sales, even if the publisher did not opt-in to the Partner Program. Furthermore, it will certainly have a positive effect on the used-book market, and enable new markets, such as print-on-demand, which could potentially positively impact current copyright holders.

3.1 Precedent for online fair use

The most relevant precedent for the Google Library program is *Kelly v Arriba Soft*, 336 F.3d 811 (9^{th} Cir. 2003). This case has been examined by Band [2005, 2006] so we will limit ourselves to a brief summary of the arguments he advances.[4]

Arriba Soft (now named ditto.com) was an image search engine which crawled the web searching for images. Presumably some of the accompanying text for the images was also captured to assist in matching the queries to the images. The images it found were transformed to thumbnail images which were displayed in response to user queries. When the user clicked on a thumbnail image, she was sent to the original URL where the image was found.

In other words, Arriba Soft used procedures similar to those currently used by Google, Yahoo!, and other providers of web image search engines.

Kelly, a photographer, found that thumbnails of images from his web site were in the Arriba Soft database and sued for copyright infringement. The case was resolved in favor of Arriba Soft, with both the lower court and the Ninth Circuit finding that Arriba's reproduction of Kelly's photographs was fair use.

With respect to the first of the four factors, the Ninth Circuit noted that although Arriba was commercial, its use of the Kelley images was "more incidental and less exploitative in nature than more traditional types of commercial use." It specifically noted that Arriba did not try to profit by selling Kelly's images and so "the commercial nature of the use weighs only slightly against a finding of fair use."

The court also considered the "transformative" nature of the use. Did Arriba's use of the images merely supersede the use of the originals, or did

[4]See his publications in this area at `http://www.policybandwidth.com/publications.html`

it add further purpose or character? The court pointed out that Kelly's "images are artistic works intended to inform and engage the viewer in an aesthetic experience," Arriba's search technology "functions as a tool to help index and improve access to images on the Internet." The court emphasized that the thumbnail images "do not supplant the need for originals" and that they "benefit the public by enhancing information gathering techniques on the Internet."

One could hardly wish for a clearer statement. Certainly the two- or three-line snippets shown by Google Library do not "supplant the need for originals but clearly "benefit the public" by enabling them to find works that were hitherto difficult to locate.

In turning to the second fair use factor, the Court pointed out that "[w]orks that are creative in nature are closer to the core of intended copyright protection than are more fact-based works" and that "[p]ublished works are more likely to qualify as fair use because the first appearance of the artist's expression has already occurred." As Band [2005] points out, "The Print Library Project involves only published works. And while some of these works will be creative, the vast majority will be non-fiction." It appears from a casual examination of the sampled query stream in Google's lobbies, that the vast majority of queries to Google web search and book search are seeking answers to facts, not an aesthetic experience.

The third factor involves the amount of the work copied. The Court specifically indicated that "if the secondary user only copies as much as is necessary for his or her intended user, then this factor will not weigh against him or her." The Court went on to say

> 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.

Again, this is complete in accord with the Google Library program. Google must copy the entire work in order to provide appropriate search coverage for the end user. However, Google will only show a few lines of text

to the end user thereby ensuring that it does not reduce the value of the work in question.

This leads us to the fourth factor, the effect on the potential market or value of the copyrighted work. In *Kelly* the court pointed out that "A transformative work is less likely to have an adverse impact on the market of the original than a work that merely supersedes the copyrighted work" and that Arriba "search engine would guide users to Kelly's web site rather than away from it." Furthermore, the thumbnail images would not harm Kelly's ability to sell or license full size images.

There is little doubt that the same would be true of Google's Library project. The entire purpose of the project is to "guide users to" works that are currently only available only in libraries or used book stores.

As Band [2005] puts it, "It is hard to imagine how the Library project could actually harm the market for certain books, given the limited amount of text a user will be able to see ... any displacement of sale is unlikely." Indeed as we have argued, catalogs, indexes and other guides have the capability to enhance the market for a work. Even Pat Schroeder, AAP president has observed that the "Google Print Library could help many authors get more exposure and maybe even sell more books ..."

Band [2005] also makes the interesting observation that

> Publishers might argue that the Library project restricts their ability to license their works to search engine providers. The existence of the Print Publisher Program however, undermines this argument. By participating in the Print Publisher Program publishers receive revenue streams not available to them under the Library Project. And Google presumably prefers for publishers to participate in the Publisher Program; Google saves the cost of digitizing the content if publishers provide Google with the books in digital format.

It seems clear that the Google Library program satisfies the criteria set forth for fair use in the Ninth Circuit's analysis of *Kelly v. Arriba Soft*.

4 Copying is incidental

It is important to understand that copying the work is itself incidental to the purpose of the Google Library project. The copied material is only valuable

to the extent that it is converted by Optical Character Recognition to ASCII text that can then be used as part of a search algorithm. The snippets that are shown to the end user are only valuable in that they aid the user in determining whether or not this particular book is relevant to his or her query.

Though Google's computers have access to the full text of the book, the only parts of the book that are ever displayed to users are the tiny snippets of materials that contain their search term.

In trying to determine whether the copying of the book is fair use it is worthwhile considering hypothetical variations on Google's model and ask whether a critic would deem them permissible use.

- Is it the scanning of the work or the display of snippets that is objectionable?
- If it is the display that is objectionable, what would be appropriate to display? The title page? The copyright page? The table of contents? Fewer, or shorter, snippets?
- If it is the scanning that is objectionable, would it be so if the scanned image was discarded and only list of words in the book were retained?
- What if only a fraction of the list of words in the book were retained?

One could contemplate many more variations. Presumably no one would have objections to material extracted from works that contained the same information as current library catalogs. But what if additional information were extracted, to enable better search, but the only information displayed was similar to that by current catalogs? This is essentially what Google is doing, as the snippets are very minimal indeed.

Presumably no one would object to human readers skimming the book and assigning subject classifications—indeed, that is basically to what the Library of Congress, Amazon, and other metadata providers do already. If a human is allowed to do this, why shouldn't a computer be allowed to do it? The end result is the same: to enable users to find books that are relevant to their needs.

Although we will eventually discover what, exactly, the publishers are objecting to, it is reasonable to suppose that it is not the display of snippets per se, but rather the scanning of the book and the conversion via OCR to

searchable text. But this is certainly a transformative action, and thus falls under the precedent of the Arriba Soft case.

The typical representation of textual material for search is a so-called "inverted index" which is a list of words contained in the texts in the collection, along with their locations in the database. It is this index which is consulted by a search engine, not the work itself. The index is constructed solely for the purpose of search.

In this sense it is similar to a library card catalog—just better, as it allows for a more accurate searching. Consider, for example, searching for works on "public goods." With a conventional library catalog search such as `http://melvyl.cdlib.org` you could search for "public goods" in the titles of books, or search for the Library of Congress Classifications such as *Public goods — International cooperation*, *Public goods — Mathematical models* and so on.

Or you could use Google Book Search which would search all the text in the book, giving appropriate weighting to various parts of the work in able to return more relevant results.[5] The functionality is essentially the same, though one would expect that full text search via Google Book Search would be much more effective.

5 Legal cases involving incidental copying

There are some interesting legal precedents for copying works in order to produce new products and services. For example, consider *Sega v Accolade*.[6] In this case, Accolade copied the machine code of Sega's game machines from the device and disassembled it to produce human-readable source code. The purpose of this exercise was to reverse engineer the Sega system in order to make Accolade's games compatible with that system. Accolade argued that copying the copyrighted machine code was fair use. Initially, the district court rejected this fair use defense.

However, the Ninth Circuit reversed this decision, finding that Accolade's copying of the code was in fact fair use. With respect to the first statutory factor, the court found that "Accolade copied Sega's code for a legitimate,

[5] I urge the reader to try these searches at both `books.google.com` and in a conventional library catalog such as `melvyl.cdlib.org`.

[6] Sega Enterprises Ltd. v. Accolade Inc., U.S. Court of Appeals, Ninth Circuit, October 20, 1992, 977 F.2d 1510, 24 USPQ2d 1561.

essentially non-exploitative purpose, and that the commercial aspect of its use can best be described as of minimal significance." They also advanced an argument that Accolade's actions served the public interest by increasing the number of video games available for the Sega console.

With respect to the fourth factor, the court found that Accolade's production of Sega-compatible games had only a minor effect on the market for Sega games and so found this factor weighed in favor of Accolade. The court also emphasized that "The fact that an entire work was copied does not, however, preclude a finding of fair use," citing previous case law to this effect.

Another relevant case is *New York Times v Roxbury Data Interface*, U.S. District Court for New Jersey, 1977. In this case, Roxbury Data hired individuals to read through the *New York Times Index* (copyrighted and published by the *New York Times*) and copy all personal names that appeared therein. These names then served as the basis for a Roxbury Data publication, *Personal Name Index to the New York Times Index*.

The court found that this activity was fair use, noting the "public interest in the dissemination of information." Of particular interest was the court's emphasis on the fourth fair use factor. They noted that "Since the defendant's index carries citations only to the New York Times Index, defendants' index is useless unless its user has access to the Times Index, from which he will be directed to the articles appearing in the New York Times." The court also emphasized that "defendant's index is not another version of the plaintiff's index, but a work with a different function and form."

Both of these cases are relevant to the Google Library project. We have already argued that the project will have no negative impact on the market for books; if anything it's impact will be positive. Furthermore, the Google index is clearly a completely different offering than the works themselves; it is, in fact, an aid to finding the original sources, as was the case for the*Personal Name Index.*

6 Opt-out versus opt-in

The other controversial part of the Google Library project is that publisher need to explicitly opt books out of the program rather than opt them in.

Indeed, another project, the Open Content Alliance (OCA) (`http://opencontentalliance.org`), which uses an opt-in model, has been endorsed

by several leading publishers. Pat Schroeder, president for the Association of American Publishers, called described the OCA's approach as "very encouraging." Sally Morris, chief executive for the Association of Learned and Professional Society Publishers, said she hopes Google follows the alliance's example. "The OCA's model of allowing rights holders to control which of their works are opened up ... and where they are hosted may encourage others to do so."[7]

However, I will argue that the opt-in approach is unlikely to be successful in creating a reasonably complete index of searchable book content due to the high transactions costs inherent in this model.

7 Transactions costs

There is a large literature in economics on transactions costs, which are defined as the economic costs necessary to actually complete a transaction. I will argue that the transactions cost are substantially higher in an opt-in model as compared to an opt-out model. Let us begin with an elementary descriptions of how transactions costs affect economic efficiency.

Suppose that Google values having a work available to be searched at 15 cents, above and beyond the direct costs of scanning, indexing and hosting the work. Suppose further that owner of the work would be willing to have it listed if he were paid 10 cents. In this case, the value of listing exceeds the cost so, by the normal economic calculus, the work should be listed. The value created by the deal is $15 - 10 = 5$ cents.

Suppose now that explicit agreement by the copyright owner is required to make the work available and that there is a 20 cent transactions cost to finding the owners and reaching an agreement. If Google has to pay those costs, the total cost to Google is $10 + 20$ which exceeds the total benefits of 15 so the work will not be made available. If the seller were required to pay the transactions costs, it would receive 15 cents from Google, but have to pay the 20 cent transactions costs and the assumed 10 cent costs of making its work available. Either way, the net value of the deal would be -15.

Large transactions costs are deal killers—they prevent mutually profitable

[7]The use of the term "opened up" is a good example of how the Google Library Program has been misrepresented. The works haven't been "opened up," they have been made searchable. And any rights holder is free to opt-out of the Google Library program, so they maintain complete control over "where their works are hosted."

transactions from occurring. However, transactions costs can be reduced by an appropriate assignment of property rights.

In an opt-in model, Google and rights holders would face two sorts of transactions costs: search costs—the costs of the rights holder and Google finding each other—and negotiation costs—the costs of reaching an agreement about the terms of the deal.

7.1 Search costs

The first problem Google would face would be identifying the rights holder. How, exactly, would someone know if they held copyright to a work?

You might think that this would be easy. But copyrights can now last 70 years after the death of an author. If an author wrote a book when he was 20 and died when he was 80, the book could be in copyright for 130 years. A lot can happen in that time.

Often publishers hold copyrights. But what if the publisher went out of business in 1933? What if the publisher merged with or was acquired by another firm? What if the publisher changed its name or address?

What if the rights were transferred to someone else? It is commonplace for book contracts to contain an out-of-print clause that transfers the copyright to the author when a book goes out of print. When the author dies, the copyright becomes part of his or her estate. Given the 70-years-from-death term of copyright, it could easily happen that the current rights holders are two or three generations removed from the author.

In 2005, the US Copyright Office sought comments about "orphan works," which are copyrighted works whose owners are difficult or even impossible to locate. They received over 850 written comments and concluded that "...there is good evidence that the orphan works problem is real and warrants attention, and none of the commenters made any serious arguments questioning that conclusion." (Register of Copyright [2006], page 2.)

Since copyright registration and renewals are not required for a great many works there is no easy way to find current owners. George [2002], Pritcher [2000], and Lesk [1997] document the difficulties of finding legitimate owners of works, as do many of the submissions to the orphan works project. For example, Carnegie-Mellon submission indicates that in their attempt to locate publishers to seek permission to digitize their works, 22 percent could not be found.

Even if the copyright holder could be identified at low cost, there is still the question of whether the copyright holder has the right to list the book in Google Library program. The reason is that the author has a contract with the publisher that may modify the standard copyright. For example, I hold the copyright on some of the books I have written , but I also have a contract with the publisher that limits what I can do—e.g., I cannot publish a "substantially competing work."

Book contract are not, of course, public information. I would venture to say that book contracts from 50 year-old works would be very difficult for either the author or the publisher to locate, even if both still existed. The question of who, exactly has the right to determine if a work should be listed in the Google Library Program is not at all simple.

Furthermore, even if owners were relatively certain about whether or not they owned a given work, and whether they had the contractual right to approve listing, it could easily not be worth the rights holder's time (in expected value) to opt-in to the Google Library program. It is true that an author might be enthused about seeing his work listed in the Google catalog. But would the author's descendants have such a degree of enthusiasm for their great-uncle's out-of-print book?

Imagine receiving a letter that told you that you had inherited the copyright on great-uncle Fred's autobiography. If you signed and returned the enclosed legal document, the book would be added to the Google Library index. What would the response rate be? The response rate would probably be about the same as to those letters telling you have won the Nigerian lottery.

And of course, rights holders would be perfectly rational not to respond to such requests since most works would be rarely consulted and thus have very low expected value. Despite this, having a complete listing could well be socially valuable since *some* of those works could well become valuable—it just might be hard to tell which ones they would be. For example, when Cardinal Ratzinger became Pope there was a significant increase in searches about him and his written work. His book *In the Beginning* happened to be listed in Google Partner program received a substantial increase in searches.

7.2 Negotiation costs

Suppose that Google somehow finds the legitimate rights holder to a work, and the rights holder knows that he holds the rights to the work. Consider

the problem of negotiating a payment.

First, consider the scale of the problem. There over 25 million of books in the libraries Google is working with and at least hundreds of thousands of possible rights holders.

Economists usually assume that if there is a bargain to be had, the players can find it. But in this case, with many parties all seeking to maximize their own utility, and highly ambiguous property rights, bargaining can easily break down. Liebbap and Wiggins [1984] describe the illustrative case of oil field unitization. In states where unanimity was required to coordinate oil-field extraction rates, many fewer agreements were reached and extraction rates were very inefficient. In states where a two-thirds majority was required, more agreements were reached, but extraction policies still weren't as efficient as those states where only a majority vote is required.

Bargaining is costly, particularly when there are many participants. The AAP suit involves only 5 publishers out of the hundreds or thousands of rights holders who have affected materials. The Author's Guild has 8,000 members, but I would guess that they have not been polled about their attitudes towards this case.

Furthermore there is highly imperfect information about the value to Google and to rights holders about being included in the Google Book Search program. Google is taking on the costs of the digitization but is not certain about what the revenues will be.

In such a circumstance, it makes much more sense to bargain ex post rather than ex ante. If the Library Search program somehow turns out to be profitable, rights holders could simply threaten to take their books out of the program unless they were appropriately compensated. Certainly it would be much easier to reach a deal once it was the parties have some idea about the actual benefits and costs faced by the different parties.

Given the intensely competitive nature of online search, if a particular corpus of material such as digitized books became profitable, one could imagine search engines bidding for exclusive access. I hasten to add that I don't see this as a likely outcome, simply because I am doubtful that digitization projects will lead anyone to substantial profits. However if by some chance it did, publishers would likely have the opportunity to capture some of that profit via agreements with competing search engines.[8]

[8]I owe this point to Stan Leibowitz.

7.3 Hold up

There are even more subtle issues involving the interaction of search costs and negotiation. The fact that the search costs have to be borne before the negotiation costs allows for a particular form of transactions costs known as "hold up." (Williamson [1985])

Suppose that Google values inclusion of a work at 15 cents and a rights perceives a 2 cent cost of inclusion. Suppose the transactions costs to Google of locating the rights holder is 10 cents. Further assume that rights holder knows Google's reserve price with certainty, perhaps because of other experience. It would seem that there is no problem in the transaction taking place since the benefits of 15 cents exceed the total costs of $10 + 2 = 12$ cents.

However, it is important to observe that the search costs have to be incurred *before* any licensing fee is negotiated. Once the 10 cent search costs have been incurred by Google they are sunk costs. When Google locates the rights holder and seeks permission to publish, the rights holder doesn't care about the costs it took to find him—he would want to set his price at the maximum he can get, 15 cents. If the rights holder can commit to that price, Google would be (just) willing to pay it.

But now roll the game back one stage to the question of whether Google should search for the rights holder. Since the search costs plus the price is greater than the value, the answer is "no." The opportunistic behavior on the part of the rights holder has eliminated the deal.

Admittedly this case is a bit artificial. But the fact that the search costs have to be incurred before negotiation can begin certainly affects the willingness to search, since the payoff if the search is successful is quite uncertain.

All of these arguments illustrate that transactions costs can be prohibitive in an opt-in model. All of these transactions costs go away, or at least are much reduced, in an opt-out model. With the opt-out model, no ex-ante negotiations are required—rights holders can unilaterally decide not to participate in the program simply be sending a message to Google. The entire process can be automated, making it far less costly than any solution requiring negotiation.

As Allen Adler, Vice President for Legal and Government Affairs, Association of American Publishers, points out, most works published in the last 35 works are subject to the ISBN system where the publisher is identified via a numerical code.[9] Hence any publisher can uniquely identify the books

[9]Singleton et al. [2006], page 9.

that it does not wish to be part of the program (assuming, of course, that they have retained copyright and have no other contractual impediments to doing so.)

Compare the transactions costs of a simple email message from a publisher to Google with a list of ISBN numbers of books it wishes removed from the program to the costs of Google locating and negotiating with thousands of publishers and authors. It is not hard to see that the opt-out model makes much more sense than opt-in.

Finally, as with any economic analysis, we have to ask how the publishers and authors would change their behavior if the Google Library Project proceeds as planned. Would they publish fewer books? Would the books cost more? Would the quality of the books published be lower? So far as I can tell, no one has argued that any of these dire events would occur as a result of the Google Library project. What is clear is that users would be able to find books that are relevant to their interests much more easily than is currently the case, benefiting not only publishers and authors, but society at large.

8 Conclusion

Google's mission is to organize the world's information and make it universally accessible and useful. This includes the information the mission locked up in out-of-print and out-of-copyright books in library. The model for the Google Library Program is essentially the same as the model for Google Web Search or Google Images: scan the relevant material, index it, and match it up with users queries, giving the users enough additional information to evaluate whether the material is worth looking at. Any rights holder can opt out of these programs, simply by sending Google a message.

Alternatives to this model, such as prior opt-in and negotiation, would be immensely costly and lead to minimal participation in the program leading to a collection that would be far inferior to what Google has proposed. Google's model will allow people to make much more effective use of material currently available only in libraries, making it "more accessible and useful" to us and to all future generations.

References

Jonathan Band. The Google Print Library Project: A copyright analysis. *E-Commerce Law & Policy*, August 2005. URL `http://www.policybandwidth.com/doc/googleprint.pdf`.

Jonathan Band. Copyright owners v. the Google Library Project. *17 Entertainment Law Review*, 21, January 2006. URL `http://www.policybandwidth.com/doc/googleprint-EntLaw.pdf`.

Carole A. George. Exploring the feasibility of seeking copyright permission. Technical report, Carnegie-Mellon Libraries, 2002. URL `http://www.library.cmu/edu/Libraries/FeasibilityStudy.ppt`.

Michael Lesk. How much information is there in the world? Technical report, Lesk.com, 1997. URL `http://www.lesk.com/mlesk/ksg97/ksg.html`.

Gary D. Liebbap and Steven N. Wiggins. Oil field unitization: Contractual failures in the presence of imperfect information. *American Economic Review*, 74(1), 1984.

Lynn Pritcher. Ad*Access: Seeking copyright permissions for a digital age. *D-Lib Magazine*, 2000. URL `http://www.dlib.org/dlib/frebruary00/pritcher/02pritcher.html`.

Register of Copyright. *Report on Orphan Works*. Library of Congress, 2006. `http://www.copyright.gov/orphan`.

Solveig Singleton, Allan Adler, Jonathan Band, Alan Davidson, and Keith Kupfershmid. *Gutenberg Meets Google: The debate about Google Print*. Progress and Freedom Foundation, 2006. `http://www.pff.org/news/news/2006/020106googletranscript.html`.

Oliver Williamson. *The Economic Institutions of Capitalism*. Free Press, New York, 1985.