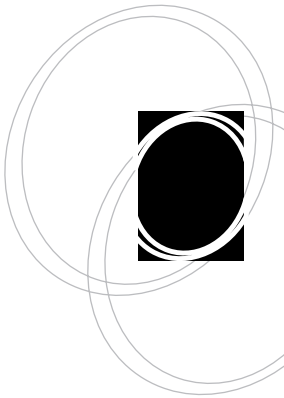




Project
MUSE[®]
Scholarly journals online



Fair Use or Exploitation? The Google Book Search Controversy

Corinna Baksik

abstract: The Google Book Search Library Project, in which millions of books from libraries will be scanned and made searchable on the Web, has led to controversy and legal action. Publishers are suing Google for copyright infringement, while Google claims their use falls under the fair use privilege of the *Copyright Act*. An overview of the library project is followed by an examination of the controversy and a look at the beta program in practice.

Introduction

Dear colleagues,

I would like to know your opinion about free access to full-text documents.

It's a good thing. Next question? Roy¹

Roy Tennant's succinct answer to a question posed on the Digital Libraries listserv represents what many feel about the increasingly common digitization of print materials. Likewise, Google, an innovator in information discovery, announced that it would begin scanning books in library collections "to make the world of books more discoverable."² Since then, there has been a barrage of good and bad press, numerous complaints, many glorifications, and two lawsuits. Although Google is not providing full-text documents for items still under copyright but is making them searchable and retrievable, there has been, nonetheless, a heated debate over the legality and ethics of what most would deem to be a "good thing."

About the Program

In support of Google's mission to "organize the world's information," the Google Book Search program is designed to digitize printed book content so that it may be searched and retrieved via Google's search engine.³ The program has two facets—one that involves publishers, which has not been controversial, and one that involves libraries, which has.

portal: Libraries and the Academy, Vol. 6, No. 4 (2006), pp. 399–415.

Copyright © 2006 by The Johns Hopkins University Press, Baltimore, MD 21218.

In the publisher program, publishers can either provide a hard copy of a book that Google scans at no charge, or they can supply books in a digital format. Authors can also supply books, as long as they are the copyright holders. Google then indexes the content and allows the full text to be searched online. Google advertises this program as “a free worldwide sales and marketing system.”⁴

The program has two facets—one that involves publishers, which has not been controversial, and one that involves libraries, which has.

The publisher program had been in existence for many months before Google announced its library program. Five libraries were asked to participate, and there are

no plans at this time to involve other collections.⁵ Libraries loan books to Google for scanning, and, in return, they receive a digital copy. Publishers expressed surprise at the project and concern over the lawfulness of Google scanning library books that were still in copyright, as well as Google supplying the library with digital versions of books.⁶

The University of Michigan has committed to allowing Google to scan its entire print and journal collection. In exchange, Michigan will receive image and OCR files. Of the five libraries in the project, Michigan is the only one to publicly disclose, under Michigan’s *Freedom of Information Act*, its contractual agreement with Google.⁷ Stanford University and Harvard University are participants as well, though neither library has committed to scanning its entire collection. The pilot projects at these universities included some in-copyright works, and the larger projects are initially focusing primarily on public domain materials.⁸ Both Oxford University and the New York Public Library have committed to only a subset of books, all in the public domain.⁹ In Oxford’s case, because of the United Kingdom’s copyright laws, these books are all published before 1885.¹⁰ Stanford, Harvard, and the University of Michigan have explicitly announced their intention to abide by copyright law.¹¹

It is estimated that Google is spending 200 million dollars on the library program. Despite a 26 billion dollar endowment, even Harvard University could not imagine funding such a project for its collection. Sidney Verba, director of the University Library, confessed, “I didn’t think it could be done by anyone, including Google.”¹²

Display of Search Results

The display of Google Book Search results varies depending on the source of the material and its possible copyright status. All search results display metadata and links to purchase the book. In cases of out-of-print materials, these links may connect to suppliers such as Alibris and Froogle. For books supplied by libraries, the metadata displayed is relatively sparse and appears to include only title, author, publisher, publication date, and pagination. Some books supplied through the publisher program appear to additionally include a general subject assignment, dimensions, format (for example, paperback), and ISBN. Metadata for publishers’ books are obtained from third party vendors.¹³

If books were submitted through the publisher program, users can see a full page, browse backward and forward two pages, but never see more than 20 percent of a book in any given month. Books supplied through the library program and published after



1922 will be treated as in copyright unless otherwise known. This cutoff date applies to users in the United States.¹⁴ For these books, users can only see three snippets per search, with each snippet not containing more than four or five lines of text. Since each snippet is a standard height, about 75 pixels, books with larger font reveal fewer lines of text. Books published before 1922 are treated as public domain material and can be browsed in their entirety.

In either copyright scenario (publisher permission versus library program), users must log in and are monitored so that viewing limits can be enforced. In addition, a set of pages in every book still under copyright is unavailable to all users unless explicitly permitted by the copyright owner.¹⁵ These pages do not display any text and are marked by a large yellow box reading “restricted page.”

Google earns revenue through “contextually targeted ads,” though ads only display for books submitted as part of the publisher program and with publisher permission.¹⁶ These earnings are shared with publishers, who receive the majority of the resulting revenue. No ads are placed on books scanned from libraries.¹⁷

Opt Out

In the summer of 2005, after complaints from copyright holders that Google did not ask their permission to scan books from libraries, Google announced a new opt-out policy and indicated that it would pause library scanning until November to allow publishers to consider the new policy.¹⁸ The policy allows copyright holders to upload a list of titles that they do not want scanned as part of the library project. When Google encounters these titles in a library, they will not scan them.¹⁹

Opponents declared that “permitting publishers to ‘opt out’ is not an acceptable substitute for proper licensing in the first place.”²⁰ One author of over 900 books, Jacob Neusner, claimed that Google insisted he fill out a separate form for each of his books. His publisher, Rowman & Littlefield, supported him by insisting Google remove all of their books as well.²¹ One publisher points out a significant problem with this opt-out plan: “The absence of any knowledge about what the collections of the participating libraries contain, and Google—again until very recently—refused to provide that information, [posed] for us the daunting task of researching the rights for every title on our backlist not already licensed through the Publisher Program.”²²

Those publishers who do not oppose the library project can let Google know which books they would like added to the account they already have via the publisher program. When one of their books is scanned at a library, the same links that apply to books submitted via the publisher program will apply to the book obtained from the library. That is, users will see a publisher-chosen link to where the book can be purchased and targeted ads for which publishers will share in the revenue.²³

Lawsuits

Despite Google’s opinion that “any copyright holder can easily exclude their titles from Google Print—no lawsuit is required,”²⁴ lawsuits were filed. In September of 2005 the Author’s Guild filed a suit against Google claiming copyright infringement and finan-

cial damages.²⁵ In October, the Association for American Publishers (AAP) filed suit claiming the same.²⁶ Google maintains that their program conforms with the meaning of copyright law and that it enhances the value of each copyright.²⁷ In his article “Leggo my Ego,” Tim Wu eloquently captures the debate: “One side trumpets the culture of authorial exposure, the other urges the culture of authorial control.”²⁸ Lawrence Lessig admits that “it is not at all clear that Google and these libraries have the legal right to do what is proposed,” but on the other hand “[content owners] simply want to be paid for the innovations of someone else.”²⁹ Other copyright owners “find the Guild’s position to be exactly backward.”³⁰

Some of the arguments equate the project with theft, with Pat Schroeder, president of the American Association of Publishers (AAP) and former congresswoman, describing the project as “intellectual embezzlement.”³¹ The complaint filed by five publishers from the AAP cites “continuing, irreparable and imminent harm that Publishers are suffering, will continue to suffer and expect to suffer due to Google’s willful infringement.”³² The filing reminds the court that publishers rely, in order to make a profit, on “licensing fees received for granting permission to make copies.”³³ The Association of American University Presses (AAUP) called the project “an appropriation of property for commercial use.”³⁴

Irrespective of market effects, many publishers are outraged on principle. They consider it “unfair and arrogant and disrespectful” that Google is digitizing without permission.³⁵ The position statement from the Association of Learned and Professional Society Publishers (ALPSP) declares that they are opposed to the project because “the fact remains that copying on such a scale is in clear contravention of copyright law and is not covered by any exception in any relevant legislation.”³⁶ Likewise, the AAUP believes that Google does not have a right to digitize the material, and that act in itself is violation of copyright, whatever the financial considerations.³⁷

Confident that both actions, scanning the books and displaying snippets, fall within the scope of the fair use privilege, Google announced at the end of October that it would resume scanning at libraries.³⁸

Four Factors

Fair use is determined by considering four factors: the purpose and character of the use, the nature of the work, the amount used in relation to the whole, and the effect upon the potential market value of the copyrighted work.

Many of the objections to the library project focus on the idea that Google will make money from others’ work.

There is no simple formula, and courts weigh the factors involved to determine whether a particular use is fair.³⁹

Purpose and Character

Many of the objections to the library project focus on the idea that Google will make money from others’ work. “Google’s intent with these programs is to make even more money. It earns 99% of its considerable revenue from the sale of advertising on its search engine.



Can it be so greedy that it seeks to bolster its profits by freely exploiting the rights of publishers and authors?"⁴⁰

Although publishers "charge that Google is infringing copyright to 'further its own commercial purposes,'"⁴¹ "most of the cases in which courts have found unlicensed uses of copyrighted works to be fair have involved projects designed to make money, including some that actually have."⁴² Simply because a use is commercial does not mean it is not a fair use. Furthermore, "because Google does not profit from the sale of any books, it is not highly exploitative."⁴³

If the use of a work is sufficiently transformative, that will weigh in favor of fair use.⁴⁴ Google's use could be considered transformative because the functionality they provide is unavailable in the print version, and this expression does not supplant the original.⁴⁵ Libraries do not need permission to create a card catalog, "neither should Google or other search engines be required to when they create an improved digital equivalent."⁴⁶ The University of Michigan believes Google's use is *critically* transformative in the modern era.⁴⁷

Kelly v. Arriba Soft has been widely cited as a precedent, wherein a search engine's creation of thumbnails based on copyrighted images was deemed fair use because it was not highly exploitative.⁴⁸ The AAUP pointed out, though, that there is a countervailing decision in *Video Pipeline, Inc. v. Buena Vista Home Entertainment, Inc.*⁴⁹ In this case, a company compiled two-minute clips of movies in order to give customers an idea of what the movies were about. The court ruled against them because these clips could serve as substitutes for the original Disney copyrighted derivatives (the previews), and the transformative quantity was so small it did not weigh sufficiently for fair use.⁵⁰

Sanford Thatcher does not believe Google's use is transformative because it is not a creative use of the original content in a value-added process. It is merely a "purely duplicative process that is no more than a form of parasitical publishing."⁵¹

Nature of Work

The AAUP was the only group to point out that copyright laws vary based on the nature of the work in question. They asked of Google, "What is your argument to justify treating books of haiku, dictionaries, novels, collections of letters...as identical so that they can all fall under the same four factors analysis?"⁵²

A plaintiff in one of the lawsuits against Google is former Poet Laureate of the United States Daniel Hoffman.⁵³ One of his compilations was provided to Google through the publisher program; and because most poems are one page or less, a combination of searching within the book and browsing two pages back and forth for each hit makes it possible to view almost 30 poems in their entirety (the entire book is 248 pages).

None of the documentation on the library project makes any distinction between factual or highly creative works, one of the factors considered in fair use analyses.

None of the documentation on the library project makes any distinction between factual or highly creative works, one of the factors considered in fair use analyses.

Amount

Google has been specific that at no time will a user be able to view more than 20 percent of a book provided through the publisher program, as specified in their contractual arrangement. Users will only see a few sentences of works obtained through the library program for works that *may* not be in the public domain. This falls well within fair use guidelines in terms of amount of work displayed. Many opponents focus on the fact that the entire work is digitized in the process of creating the derivative product—Google’s index.

Market

Supporters of the project argue that print books will not be supplanted, and “without question, the Print Library Project will increase the demand for some books.”⁵⁴ “When Cardinal Ratzinger became pope, millions of people who searched his name saw the Google Print listing for his book *In the Beginning*...in their results. Thousands of them looked at a page or two from the book; clicks on the title’s ‘Buy this Book’ links increased tenfold.”⁵⁵

When Amazon introduced their Search Inside the Book product, there was a 9 percent increase in sales for books that were enabled with full-text searching. “This is no surprise, considering the fact that ‘Search Inside’ shows shoppers many books that they would never have considered—or even known of—otherwise.”⁵⁶ In fact, “if the Amazon statistics are any guide, the market for books that are not even sold in the average bookstore is larger than the market for those that are.”⁵⁷ This suggests that discovering books through the Google Books Library Project would create an expanded market. Google and others firmly believe this and feel that this is what copyright law was designed to do.⁵⁸

Similar evidence comes from OCLC’s Open WorldCat program. When the first two million records from WorldCat were made available for open searching on the Web, statistics from search engines were in the millions per month, more than OCLC had ever seen. In addition, 8 percent of these clicked for library holding information, “at the rate of hundreds of thousands each month.”⁵⁹ OCLC reports that 83 percent of users who arrive at a “find in a library” link in a search engine end up in a library’s online catalog.⁶⁰ It is not a far leap to suggest that search engines with full-text indexes of books, as opposed to only metadata, would lead to more discovery and more purchases.

One publisher representative has confirmed that “publishers have been pleased with the increased hits although, as far as we are aware, actual sales have not increased dramatically.”⁶¹ Another publisher, however, feels that he “can predict with confidence that online-search capabilities will boost book sales: A university press that joins Google will find itself using ‘print-on-demand’ technology to fill orders from its backlist for that 1958 tome on the Maginot Line that it never dreamed would have a life in the 21st century.”⁶²

Book sales are only one factor in a market effect analysis of the project. The lawsuit filed by McGraw-Hill et al. claims that “Google’s continuing and future infringements are likely to usurp Publishers’ present and future business relationships and opportunities for the digital copying, archiving, search and public display of their works.”⁶³



Google's acts are preempting copyright holders from licensing their books to search engines for indexing. The fact, though, that most large copyright holders participate in Google's publisher program in the absence of any license fee, "negates the notion that a commercial harm is occurring."⁶⁴

For the majority of works in the library project, there is no market on which this project could have a negative effect. That is, for orphan works, "awarding the right to the copyright owner will not lead to fruitful negotiations or any negotiations at all. . . . No market can form around this right if it belongs to someone who is either dead and buried or who does not know or care that he has the right."⁶⁵

For the majority of works in the library project, there is no market on which this project could have a negative effect.

Analysis

"It's unlikely that either side will win this case on the strength of a fair use analysis regardless of how subtle, clever or even brilliant it may be."⁶⁶ This is because judges make their decision based on other criteria and then report that decision within the four-factor framework.⁶⁷

The question will come down to "whether the public service will outweigh the commercial exploitation."⁶⁸ The terrifying precedent for publishers is that "if Google can do this, anyone can do this."⁶⁹ Whether their intent or not, Google is taking up the cause of many groups whose "ability to communicate effectively is being restricted by an overly rigid approach to copyright compliance, and [the] public suffers as a result."⁷⁰ Jessica Litman notes that, historically, copyright owners have never had rights "as expansive as those that they have recently argued were their due."⁷¹ She suggests that reproduction is no longer an appropriate way to measure infringement because of the centrality of copying in digital technology.⁷²

Library Copies

Google has indicated that "each library will receive a digital copy of every book we scan at their specific library. This was a precondition of digitization, in order to help each library fulfill their mission to archive and preserve knowledge in all forms. Each library will treat their copies in accordance with copyright law."⁷³ The existence of the library copy is clearly not covered by the library exemption of the *Copyright Act*.⁷⁴ Libraries can only defend themselves by creating a transformative use of the work, "to build on and reconceive vital library services for the new millennium."⁷⁵

Publishers fear that "there is no control over how the library or others who gain access to this digital copy can use what Google has given them."⁷⁶ Both the University of Michigan and Stanford University have been explicit that their use will fall within fair use, and Harvard stated that "those second copies will be used only for archiving and preservation, in keeping with a research library's charter. . . . We think and hope it is legally the appropriate approach."⁷⁷ This is not reassuring to some publishers, who fear that a library's idea of fair use "nowadays may well mean deposit in an e-reserve system that functions as a coursepack-producing facility."⁷⁸

Litman's declaration that "copyright today is less about incentives or compensation than it is about control"⁷⁹ was confirmed by Sally Morris of ALPSP who stated: "We shall also be recommending (as suggested by Google) that [publishers] can protect both in- and out-of-copyright print and electronic works by placing them in the Google Print for Publishers program."⁸⁰ Perhaps this was a slip on her part, for why would publishers feel the need to "protect" out-of-copyright works? Denise Troll Covey discovered that there is at least one publisher who, if asked, will deny permission to digitize works even when they are no longer copyright protected.⁸¹

Orphans

Barbara Quint states, "Experts estimate that 80 percent of books in copyright fall into the 'orphan' category."⁸² If this estimate is even near accurate, then obeying the publishers' request to ask for permission before digitizing would mean that most printed works from 1923 to the introduction of born-digital materials would never be digitized because permission could never be obtained. Congress is making efforts to resolve the problem of orphan works, however, not only is congressional action "a wild card,"⁸³ but

Congress [also] lacks the interest, expertise, and institutional memory to represent the public on this particular project, and has found significant political benefits in deferring to the interests the legislation affects. Thus, what Congress has done more often than not is delegate the job of coming up with legislation to interested private parties, which is how the statute got so long and convoluted in the first place.⁸⁴

As for the holdings of the five participating libraries, estimates suggest that more than 80 percent of the collections are still in copyright.⁸⁵ Using the estimate cited earlier that 80 percent of books in copyright are orphans, the large majority of copyrighted

works in these libraries would never be digitized if the policy were to avoid liability. Google is taking on a "copyright system [that] thus denies public access to these orphan works, without creating any countervailing benefit either to authors or the public at large."⁸⁶

Carnegie Mellon's recent study on obtaining permission to digitize and fully display books showed that "ultimately, 21 percent of the publishers,

accounting for 19 percent of the titles in the sample, could not be located," and another third did not respond.⁸⁷ In their random sample feasibility study, "a crude, retrospective speculation about the transaction cost...is roughly \$200 per title for which permission was granted."⁸⁸ Permission was granted for approximately 25 percent of the sample.⁸⁹ Lavoie calculates that 80 percent of the 10.5 million unique books in the combined collections of the five libraries participating in Google's project are still in copyright.⁹⁰ If, as in Carnegie Mellon's project, permission could be obtained for one-quarter of these books, and the rough transaction cost is \$200 per title, then the Google project would equate to a 420 million dollar effort to determine the copyright status and obtain permission from

Using the estimate cited earlier that 80 percent of books in copyright are orphans, the large majority of copyrighted works in these libraries would never be digitized if the policy were to avoid liability.



copyright owners. Extrapolating further, if permission for all works could be obtained, the transaction cost would be 1.7 billion dollars before digitization even began.

Furthermore, “the law does not protect an individual who conducts a good faith search for a copyright holder, but cannot find him. If the individual uses the work, and the copyright holder resurfaces, the user is subject to the full panoply of penalties the copyright law provides.”⁹¹ It is not shocking, therefore, that many online digital library collections focus on the nineteenth century and earlier.

Discrepancy, Hypocrisy, and Hyperbole

On The Publisher’s Side

There has been significant misinformation published regarding the Google Book Search Program, and even the lawsuit supported by the AAP contains, to the laywoman’s eye, discrepancies. “There is no principled distinction between the Google Print Program for Publishers and the Google Library Project, with respect to...the public accessibility and display of the copied works. The only distinction is Google’s decision not to seek permission for books included.”⁹² The appendix to the complaint, though, contains screen shots from Google’s domain showing precisely what the distinction is between “public accessibility and display” for a book submitted by a publisher and a book from the library program.⁹³ The images in the complaint explicitly reveal that a book in the publisher program will display “a full page and a few pages on either side,” whereas the library program books reveal “a few short sentences of text around your search term.”⁹⁴

As for publishers’ exception to Google’s opt-out policy, letters sent to 209 publishers as part of the Carnegie Mellon study included the statement: “If we do not receive a response from you within 60 days of mailing this letter, we will assume that you have granted permission to digitize the book and offer it free to read by anyone on the Internet.”⁹⁵ Although Carnegie Mellon had no intention of digitizing books without permission, only one publisher out of 209 commented on this approach.⁹⁶

As late as a year after the launch of the Google Book Search beta, Pat Schroeder of the AAP was still getting facts wrong regarding the library project. “[Google] plans to scan the entire contents of the Stanford, Harvard and University of Michigan libraries and make what it calls ‘snippets’ of the works available online, for free.”⁹⁷ As previously indicated, only the University of Michigan has committed their entire collection to the project. She also stated that “the creators and owners of these copyrighted works will not be compensated, nor has Google defined what a ‘snippet’ is: A paragraph? A page? A chapter? A whole book? Meanwhile Google will gain a huge new revenue stream by selling ad space on library search results.”⁹⁸ The University of Michigan has documents dating back to 2004 that clearly define what a snippet is, and Google has specified that there will be no ads next to search results from library books.⁹⁹

Many publishers have expressed praise for the approach taken by the Open Content Alliance (OCA), which seeks permission before digitizing books. Yahoo, an OCA participant, will perform the full-text searching and indexing. “Yahoo could make money on the project indirectly if the digitized books can attract people to Yahoo’s site.”¹⁰⁰ In the OCA project, there is no arrangement for publishers to receive any revenue from the use of their books, whereas in the Google program, the arrangement is for publishers

to receive revenues from contextually targeted ads placed with their permission. Why do opponents decry Google's ads, which will benefit copyright owners as well, but not ads from other search engines from which they will not share the profit?

On Google's Side

If Google is intent on information discovery, why do they not attribute the library or libraries from which the book was scanned? Many libraries have no holdings in OCLC for older books, which means that the "find in a library" link will not lead users to accurate

If Google is intent on information discovery, why do they not attribute the library or libraries from which the book was scanned?

holdings. For some books, all a user will see is "from the Google Books Library Project." For example, the book *Rules for a Dictionary Catalogue* by Charles Ammi Cutter can be viewed in its entirety, but there is neither a link to locate it in a library nor any indication of where the book was obtained. Browsing to page two, though, reveals a stamp indicating

that the book belongs to the New York Public Library through the Astor, Lenox, and Tilden Foundations. For many of the library books that are fully browsable, it is eventually possible to discover what library provided the print copy through attributions in bookplates, stamps, and embossments. If the library book is in copyright and displayed in snippet form, then this front and back matter will not be available.

Google has indicated that ads will not appear alongside books scanned from libraries and that they will share the revenue from ads that appear alongside books from the publisher program. Who earns the money from ads on the main search results page, where books from multiple publishers and libraries are listed in short form? Here Google's assurances are somewhat misleading, as ads do appear alongside library books if they appear on the main results page. Ads do not appear only when the "search within this book" function is used on a library book.

The founders of Google themselves have questioned the intentions of "advertising funded search engines" because they are "inherently biased towards the advertisers and away from the needs of the consumers. ... The issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm."¹⁰¹

On Both Sides

Publishers sign an agreement with Google allowing them to digitize "all text, images, photographs, illustrations and all material and artwork on the book jackets and covers."¹⁰² Do publishers always own the copyright for this material? Covey estimated that 11 percent of books in her study had complicated third-party copyright ownership due to charts, illustrations, and so on.¹⁰³ One book in the Google Book Search beta, *Way of the Peaceful Warrior*, lists the author as the copyright owner, contains credits for a cover artist and a designer, and indicates the book was published in a joint venture between two companies. Is it clear, even to them, who has the authorization to permit digitiza-



tion? In the Tasini case, the *New York Times* ended up at the Supreme Court before it was determined that they did not always have authorization to license their content.¹⁰⁴ What are the chances book publishers would not at times make the same mistake? There is evidence that for some books precautions have been taken not to display images, particularly with children's books in which users will see a white page reading only "copyrighted image."

Beta in Practice

Although Google Book Search is a very exciting tool even in its beta version, there are some areas that do not function well. The most notable and significant problem is that numerous non-copyrighted books are restricted. For example, the *WIPO Copyright Treaties Implementation Act* by the House Committee on the Judiciary, Subcommittee on Courts and Intellectual Property is displayed in snippet form. This is the case for most works published by Congress after 1922, even though works by the U.S. Government are not eligible for U.S. copyright protection.¹⁰⁵

A less obvious problem comes from occasional errors in book metadata. A search for the title *Statistics on Vegetable [sic] Oils and Allied Products* reveals snippets that show that *vegetable* is spelled correctly on the cover and title page, but the bibliographic information displayed for the book lists the title with the misspelling. A search for *vegetable* within the book yields no results. According to WorldCat, the New York Public Library is the only one of the five Google libraries with this book, but *vegetable* does not appear anywhere in the bibliographic description in CATNYP (the NYPL online catalog). A search for this title in the catalogs of the other four libraries reveals that this error stemmed from Stanford's bibliographic record. Google's emphasis, though, is on resource *discovery* not resource description; and, because the text is searchable, the item can be retrieved with either spelling.

Google has stated that each book from the library project will be displayed with a "find this book in a library" link. It is extremely easy to find examples for which this is not the case, namely for books without ISBNs. Google has acknowledged this problem, and while they work on fixing it, at least one user has distributed his own solution in the form of a script for the Firefox Web browser.¹⁰⁶ It is unfortunate that Google will not display the "find this book in a library" link for books provided via the publisher program, but luckily the hacker community has provided an alternative.

As for targeted ads, they are as non-intrusive as ads on Google's Web search engine, and they appear to have improved somewhat over time in terms of their targeted nature. For example, in November 2005, an ad for laser hair removal was being shown alongside Sidney Verba's *Participation and Political Equality*. By April 2006, the ads included links to a Web directory, free phone ring tones, and a site to buy college essays.

There are some books with partial or bad images, such as *Plaidoyes de Mre Loys Servin* or *Oxford Book of English Verse, 1250–1900*, though each page includes a link beneath it allowing users to report a problem. For other books, the text is too small to be read, and the image cannot be magnified. In some cases, it is not clear what version of a book was scanned. *The Three English Brothers: Sir Thomas Sherley His Travels, with His Three Yeares Imprisonment...* is listed in Google Book Search with a publication date of 1607.

Google, in fact, did not digitize a book from 1607, which a user can discover by linking the call number that appears on one of the images with a book held in the Harvard College Library. An inspection of the physical book reveals that it is a photostat copy. Coincidentally, this book is part of the *Early English Books Online* collection in both image and full-text form, which could allow for a comparison of Google's full-text searching although this Old English example might be an unfair basis for evaluation.

In addition, book navigation leaves something to be desired. For the book discussed above, page one is displayed by default, but the numbering is incorrect, and it is necessary to scroll backward from what is marked as page cxv to get to the title page. The only other way to start at the title page is to search within the book for the title and click on the first result. On a positive note, search terms are highlighted in yellow in the text, and a number of books have hyperlinked tables of contents.

The most disheartening aspect of Google Book Search is its failure to make use of the bibliographic metadata available for library books. It was shown earlier that Google uses title data from library catalogs—why stop there? They label the library project as an “enhanced card catalog of the world's books,”¹⁰⁷ but it makes no use of the power and value of controlled vocabularies, descriptive standards, and other tools of the library profession. What is particularly exciting is the potential future in which the retrieval capabilities and relevancy ranking of search engines in a full-text environment are combined with traditional methods of information organization.

Conclusion

“One thing is clear in looking at the Google library digitization; it is a project that excites people with its possibilities, a result that seems to fit in well with the spirit of innovation that the copyright laws are meant to protect, and not something the law or its application should obstruct.”¹⁰⁸ Unfortunately, there are those who feel that fair use is an “obsolete privilege” and a “doctrine that has outlived much of its usefulness.”¹⁰⁹ They argue that modern technology has minimized the transaction costs of licensing works and enabled digital rights management (DRM), so there is no longer a defense for using a work without the explicit permission of the copyright owner.¹¹⁰

An argument for DRM in theory is quite different from DRM in practice. Sony BMG's efforts in DRM have resulted in a class action lawsuit against them. The technology they used on CDs monitors users' listening and installs “undisclosed and in some cases hidden files on users' computers that can expose users to malicious attacks by third parties, all without appropriate notice and consent from purchasers.”¹¹¹ This is an example of a copyright owner whose efforts at control have infringed on users' fundamental right to privacy.

Brewster Kahle has argued that copyright in its current form is unconstitutional, but so far these arguments have not succeeded in court.¹¹² As long as they continue to fail, even a noble visionary like Mr. Kahle will be forced to have a free library of open-access works in which most content was published before most of its users' parents were born.

In its “marvelous, if imperfect” way, Google has used the legal system in such a clever manner that no one is really getting hurt, and everyone is better off.¹¹³ As for au-



thors, "obscurity is a far greater threat... than copyright infringement, or even outright piracy."¹¹⁴ As for users and libraries, "[Google's] model is more likely to help more people find library resources and publishers' works than anything else on the horizon."¹¹⁵

Michael Gorman, president of the American Library Association, has been one of the few critics from the library community. He fears that the project will only increase the number of undergraduates who think all their research can be done in Google.¹¹⁶ If Google continues to digitize library holdings and enhances Google Scholar, then it may be that undergraduates *can* do all their research in Google, at least as a means of discovery. Would that be that bad? A generation of students who have the ability to perform full-text searching of centuries of material from around the world will have the potential to discover information that earlier generations never knew existed.

As far as Google is concerned, they "look forward to the day that the program's opponents marvel at the fact that they actually tried to stop an innovation that, by making books as easy to find as web pages, brought their works to the attention of a vast new global audience."¹¹⁷

Corinna Baksik is systems librarian, Harvard University Library, Cambridge, MA; she may be contacted via e-mail at: corinna_baksik@harvard.edu.

Notes

1. Roy Tennant, "Access to Documents," International Federation of Library Associations and Institutions, Digital Libraries Research Mailing List, archive, <http://infoserv.inist.fr/wwsympa.fcgi/arc/diglib> (accessed July 1, 2006).
2. Joseph O'Sullivan and Adam M. Smith, "All Booked Up," *Google Blog* (December 14, 2004), <http://googleblog.blogspot.com/2004/12/all-booked-up.html> (accessed July 1, 2006).
3. Google, "About Google Book Search," Google Book Search beta, <http://print.google.com/intl/en/googleprint/about.html> (accessed July 1, 2006).
4. Google, "Author Resources" Google Book Search beta, <http://print.google.com/googleprint/author.html> (accessed July 1, 2006); "Partner Program—An Online Book Marketing and Sales Program," Google Book Search beta, <http://books.google.com/googlebooks/publisher.html> (accessed July 1, 2006).
5. Google, "Help," Google Book Search beta, <http://print.google.com/googleprint/help.html> (accessed July 1, 2006).
6. Peter Givler to A. Macgillivray, senior intellectual property and product counsel, Google, Inc., 20 May 2005, http://aaupnet.org/aboutup/issues/0865_001.pdf (accessed July 1, 2006).
7. University of Michigan Library, "Cooperative Agreement," <http://www.lib.umich.edu/mdp/um-google-cooperative-agreement.pdf> (accessed July 1, 2006).
8. Stanford University Libraries and Academic Information Resources, "Stanford University and Google Mass Digitization Project: FAQ," SULAIR, http://www-sul.stanford.edu/about_sulair/special_projects/google_sulair_project_faq.html (accessed July 1, 2006); Harvard University Library, "Harvard Google Project," Harvard University Library, http://hul.harvard.edu/news/2004_1214_faq.html (accessed July 1, 2006); Information on Harvard's pilot can be found on a Web page archived at <http://web.archive.org/web/20050310090049/hul.harvard.edu/publications/041213faq.html> (accessed July 26, 2006).
9. New York Public Library, "NYPL Partners with Google to Make Books Available Online," New York Public Library (December 14, 2004), <http://www.nypl.org/press/2004/google.cfm> (accessed July 22, 2006); Reg Carr, "Oxford-Google Digitization Agreement," Bodleian Library, University of Oxford, <http://www.bodleian.ox.ac.uk/google/> (accessed July 1, 2006).

10. Bodleian Library, "Oxford University—Google Digitization Programme: FAQs," Bodleian Library, University of Oxford, <http://www.bodleian.ox.ac.uk/google/> (accessed July 1, 2006).
11. Katie Hafner, "At Harvard, a Man, a Plan and a Scanner," *New York Times*, November 21, 2005; University of Michigan Library, "Cooperative Agreement," University of Michigan Library, <http://www.lib.umich.edu/mdp/um-google-cooperative-agreement.pdf> (accessed July 1, 2006); and Stanford University Libraries, "Stanford and Google Book Search Statement of Support and Participation, December 2005," SULAIR, http://library.stanford.edu/about_sulair/news_and_events/stanford_google_project.html (accessed July 1, 2006).
12. Hafner; President and Fellows of Harvard College, "Harvard's Endowment Funds," The Harvard Guide: Finances, <http://www.hno.harvard.edu/guide/finance/> (accessed July 1, 2006).
13. Google, "Where does metadata come from?" Google Book Search beta, <http://print.google.com/support/publisher/bin/answer.py?answer=20025&topic=348> (accessed July 1, 2006).
14. Google, "Your Content is Protected," Google Book Search, http://books.google.com/services/print_tour/print4.html (accessed July 19, 2006); Google, "Google Books Partner Program Standard Terms and Conditions," Google Book Search beta (November 21, 2005), <https://books.google.com/partner/terms> (accessed July 19, 2006); and Google, "Library Project—An enhanced card catalog of the world's books," Google Book Search beta, <http://print.google.com/googleprint/common.html> (accessed July 19, 2006).
15. Google, "Help: Frequently Asked Questions," Google Book Search beta, <http://books.google.com/googleprint/help.html> (accessed July 19, 2006).
16. Ibid.
17. Eric Schmidt, "Books of Revelation," *Wall Street Journal*, October 18, 2005.
18. Jeffrey R. Young, "Google Adjusts its Book-Scanning Project in Response to Complaints from Publishers," *Chronicle of Higher Education*, September 2, 2005.
19. Google, "Information For Publishers and Authors about the Library Project," Google Book Search beta, http://print.google.com/googleprint/publisher_library.html (accessed July 1, 2006).
20. Sally Morris, "Google Print for Libraries—ALPSP Position Statement" (July 2005), <http://www.alpsp.org/2005pdfs/Googlestatement.pdf> (accessed July 1, 2006).
21. Vincent Kiernan, "Prolific Author Tells Google to Remove His Books from its Library-Scanning Project," *Chronicle of Higher Education*, October 21, 2005.
22. Sanford G. Thatcher, "Fair Use in Theory and Practice: Reflections on its History and the Google Case," Penn State: psupress.org, http://www.psupress.org/news/news_NACUA.html (accessed July 1, 2006).
23. Google, "Information For Publishers and Authors about the Library Project."
24. Schmidt.
25. "Authors' Group Sues Google Over Library-Scanning Project," *Chronicle of Higher Education*, September 30, 2005.
26. *McGraw-Hill Companies, Inc., Pearson Education, Inc., Penguin Group (USA), Inc., Simon & Schuster, Inc., and John Wiley & Sons, Inc. v. Google, Inc.*, No. 05 Civ. 8881 (S. D. N. Y. Oct. 19, 2005), <http://publishers.org/press/pdf/40%20McGraw-Hill%20v.%20Google.pdf> (accessed July 1, 2005).
27. Schmidt.
28. Tim Wu, "Leggo My Ego," *Slate* (October 17, 2005), <http://www.slate.com/id/2128094/> (accessed July 1, 2006).
29. Lawrence Lessig, "Let a Thousand Googles Bloom," *Los Angeles Times*, January 12, 2004, <http://www.freepress.net/news/print.php?id=6159> (accessed July 1, 2006); Lessig, "Google Sued," Lessig Blog (September 22, 2005), <http://www.lessig.org/blog/archives/003140.shtml> (accessed July 1, 2006).
30. Tim O'Reilly, "Search and Rescue," *New York Times*, sec. A, September 28, 2005, late edition, final.



31. Pat Schroeder, "Lack of Sense and Sensibility," *USA Today*, November 8, 2005.
32. McGraw-Hill et al.
33. *Ibid.*
34. AAUP Board of Directors, "Google's New Opt-out Policy," Association of American University Presses (August 19, 2005), <http://aaupnet.org/aboutup/issues/81905.html> (accessed November 20, 2005).
35. Kiernan.
36. Sally Morris, "ALPSP Position Statement on Google Print for Libraries" (August 24, 2005), <http://www.alpsp.org/2005pdfs/GooglePrintLibraries.pdf> (accessed July 14, 2006).
37. "A University Press Leader Has Questions for Google," *Chronicle of Higher Education*, June 3, 2005.
38. Jonathan Band, "The Google Print Library Project: A Copyright Analysis" (August 2005), <http://www.policybandwidth.com/doc/googleprint.pdf> (accessed July 1, 2006); Adam M. Smith, "Discovering Hard-to-Find Books," Google Blog (October 31, 2005), <http://googleblog.blogspot.com/2005/10/discovering-hard-to-find-books.html> (accessed July 1, 2006).
39. Carrie Russell, *Complete Copyright: An Everyday Guide for Librarians* (Chicago: American Library Association, 2004), 19–33.
40. Schroeder, "Google Cannot Rewrite U.S. Copyright Laws," *Wall Street Journal*, October 20, 2005.
41. Scott Carlson, "Publishers Sue Google to Prevent Scanning of Copyrighted Works," *Chronicle of Higher Education*, October 28, 2005.
42. Subcommittee on Commerce, Trade, and Consumer Protection of the House Committee on Energy and Commerce, *Fair Use: Its Effects on Consumers and Industry, Hearing*, 109th Cong., 1st sess., 2005, 16–22.
43. Band.
44. Russell, 19–33, 218–9.
45. Elisabeth Hanratty, "Google Library: Beyond Fair Use?," *Duke Law & Technology Review Rev. 0010* (2005), <http://www.law.duke.edu/journals/dltr/articles/2005dltr0010.html> (accessed July 1, 2005).
46. Electronic Frontier Foundation, "Google's Card Catalog Should Be Left Open," Electronic Frontier Foundation: Defending Freedom in the Digital World (September 21, 2005), http://www.eff.org/news/archives/2005_09.php#003994 (accessed July 1, 2006). Whether Google Book Search is an "improved digital equivalent" of a card catalog is open for debate.
47. University of Michigan Library, "U-M University Library—Google Digitization Project: A Brief Overview" (December 13, 2004), <http://www.lib.umich.edu/mdp/overview.pdf> (accessed July 1, 2006).
48. Band.
49. Givler.
50. "Clip Previews of Movies Did Not Qualify as Fair Use," *Computer & Internet Lawyer* 20, 11 (November 2003): 27.
51. Thatcher.
52. Givler.
53. "Authors' Group Sues."
54. Band.
55. Schmidt.
56. Michael A. Banks, "Amazon.com Opens the Books," *Online* 28, 2 (March/April 2004): 33.
57. Chris Anderson, "The Long Tail," *Wired Magazine* 12, 10 (October 2004), <http://www.wired.com/wired/archive/12.10/tail.html> (accessed July 1, 2006).
58. Susan Wojcicki, "Google Print and the Authors' Guild," Google Blog (September 20, 2005), <http://googleblog.blogspot.com/2005/09/google-print-and-authors-guild.html> (accessed July 1, 2006); Electronic Frontier Foundation.

59. Barbara Quint "Fame or Fortune? The Appeal of Numbers," *Information Today* 22, 5 (May 2005):7.
60. OCLC, "Quick Facts About the Open WorldCat Program," OCLC, <http://www.oclc.org/worldcat/open/facts/default.htm> (accessed July 1, 2006).
61. Morris.
62. Michael Jensen, "Presses Have Little to Fear from Google," *Chronicle of Higher Education*, July 8, 2005.
63. McGraw-Hill et al.
64. Hanratty.
65. Georgia Harper, "Google This," Copyright Crash Course, <http://www.utsystem.edu/ogc/intellectualproperty/googlethis.htm> (accessed July 1, 2006).
66. Ibid.
67. Ibid.
68. Hanratty.
69. Carlson.
70. Association of Independent Video and Filmmakers, Independent Feature Project, International Documentary Association, National Alliance for Media Arts and Culture, and Women in Film and Video, Washington, D.C. Chapter, "Documentary Filmmakers Statement of Best Practices of Fair Use," <http://www.centerforsocialmedia.org/rock/backgrounddocs/bestpractices.pdf> (accessed July 1, 2006).
71. Jessica Litman, *Digital Copyright* (Amherst, New York: Prometheus Books, 2001), 114.
72. Ibid., 171–86.
73. Google, "Information For Publishers and Authors about the Library Project."
74. Hanratty.
75. University of Michigan Library, "U-M University Library—Google Digitization Project."
76. Schroeder, "Google Cannot Rewrite U.S. Copyright Laws."
77. University of Michigan Library, "Cooperative Agreement"; Stanford University Libraries, "Stanford and Google Book Search Statement"; and Hafner.
78. Thatcher.
79. Litman, 80.
80. Morris.
81. Denise Troll Covey, "Acquiring Copyright Permission to Digitize and Provide Open Access to Books," Digital Library Federation, <http://www.diglib.org/pubs/trollcovey0509/> (accessed July 1, 2006).
82. Barbara Quint, "Books Online: The Fee Versus Free Battle Begins," NewsBreaks & The Weekly News Digest, Information Today, Inc., <http://www.infotoday.com/newsbreaks/nb051121-2.shtml> (accessed July 1, 2006).
83. Harper.
84. Litman, 74.
85. Brian Lavoie, Lynn Silipigni Connaway, and Lorcan Dempsey, "Anatomy of Aggregate Collections: The Example of Google Print for Libraries," *D-Lib Magazine* 11, 9 (September 2005), <http://www.dlib.org/dlib/september05/lavoie/09lavoie.html> (accessed July 1, 2006).
86. Stanford Law School Center for Internet and Society, "Kahle v. Ashcroft Case Page" (April 12, 2005), Stanford Law School Center for Internet and Society, http://cyberlaw.stanford.edu/about/cases/kahle_v_ashcroft.shtml (accessed July 1, 2006).
87. Covey.
88. Ibid.
89. Ibid.
90. Lavoie, Connaway, and Dempsey.
91. Subcommittee on Commerce, *Fair Use*, 36–43.
92. McGraw-Hill et al.
93. Ibid.



94. Ibid.
95. Covey.
96. Ibid.
97. Pat Schroeder and Bob Barr, "Reining in Google," *Washington Times*, November 3, 2005, <http://www.washtimes.com/commentary/20051102-093349-7482r.htm> (accessed July 1, 2006).
98. Ibid.
99. University of Michigan Library, "U-M University Library—Google Digitization Project"; Schmidt.
100. Scott Carlson and Jeffrey R. Young, "Yahoo Works with Academic Libraries on a New Project to Digitize Books," *Chronicle of Higher Education*, October 14, 2005.
101. Sergey Brin and Lawrence Page, "The Anatomy of a Large-scale Hypertextual Web Search Engine" (January 24, 2000), <http://www-db.stanford.edu/pub/papers/google.pdf> (accessed July 1, 2006).
102. Google, "Google Books Partner Program Standard Terms and Conditions."
103. Covey.
104. Cary Kenney, "Newspaper Publishers in the Post-Tasini Era," *Searcher* 10, 3 (March 2002), <http://www.infotoday.com/searcher/mar02/kenney.htm> (accessed July 22, 2006).
105. U.S. Copyright Office, "Copyright Basics: Circular 1," <http://www.copyright.gov/circls/circ1.html> (accessed July 1, 2006).
106. Ben Bunnell, "A Real Find," Google Blog (March 8, 2006), <http://googleblog.blogspot.com/2006/03/real-find.html> (accessed May 1, 2006).
107. Google, "Google Books Library Project—An Enhanced Card Catalog of the World's Books," Google Book Search beta, <http://books.google.com/googlebooks/library.html> (accessed July 1, 2006).
108. Hanratty.
109. Subcommittee on Commerce, *Fair Use*, 43–55.
110. Ibid.
111. Electronic Frontier Foundation, "Sony BMG Litigation Info," Electronic Frontier Foundation: Defending Freedom in the Digital World, <http://www.eff.org/IP/DRM/Sony-BMG/> (accessed July 1, 2006).
112. Stanford Law School Center for Internet and Society.
113. Jensen.
114. O'Reilly.
115. Jensen.
116. "One College Librarian Worries About 'Atomizing' Books," *Chronicle of Higher Education* (June 3, 2005).
117. David Drummond, "Why We Believe in Google Print," Google Blog (October 19, 2005), <http://googleblog.blogspot.com/2005/10/why-we-believe-in-google-print.html> (accessed July 1, 2006).