

Copyright Reforms for the Digital Age: A Closer Look at Google

Daniel Holevoet and Sarah Price

When one of the authors discovered that her brother kept a personal blog, she was immediately curious about what he had written about her. Unwilling to pour through hundreds of entries, she used Google to search the domain of his blog for various terms, such as "sister," and used the pages Google cached to find out what exactly he had said. Needless to say, her brother probably did not anticipate this availability of information to his sister. It would seem that Google's services might actually be a detriment to copyright holders. Yet, in 2006, companies that Google had banned from its returns pages, such as KinderStart, filed a suit against Google for unfairly blocking them, because hits to their website dropped 70% when Google stopped showing information about their site.¹

The extent to which computers have changed everyday life is undeniable. It is now possible to obtain vast amounts of information more quickly and easily than ever before. With the conveniences brought by the digital revolution, however, complications far exceeding those of this example are inevitable. Not only does this new availability of information create novel situations that challenge former legal and social situations, but even the very nature of computing itself strains pre-existing legislation. The nature of governmental processes has meant that old legislation and concepts have been twisted to match a changing reality, and that new legislation is often reactionary and ill-advised.

This paper examines the strains placed on copyright law by the dynamic environment created by new technology. We first consider the history of copyright law from its origins to the present. Next, we consider Google as an example of a business, which has formed a robust economic model, based on opportunities presented by new technology, and how its business strategy and the limitations of existing copyright legislation challenge each other. Finally, we

¹ SiliconValley.com, "Lawsuit alleges Google improperly banishes Web sites." 03/17/2006. Available at: <http://www.siliconvalley.com/mld/siliconvalley/news/editorial/14125988.htm>

consider what interests should prevail in this battle and how the legal and technological environment should be modified in order to attain balance.

Part One: United States Copyright Law

Foundations of United States Copyright Law

It is reasonable that the government of a nation offer to its citizens some way to protect their works. In the United States, this protection takes the form of copyright laws, which secure a monopoly on a work to its owner for a limited time period. The power for the United States Congress to establish laws regarding copyright is found in Article I, Section 8, Clause 8 of the United States Constitution: “The Congress shall have Power [. . .] To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.” The meaning intended by the framers of the Constitution can be easily gathered from historical context, as the world’s first copyright legislation appeared 70 years earlier. The Statute of Anne, passed by the British Parliament in 1710, reads:

Whereas Printers, Booksellers, and other Persons, have of late frequently taken the Liberty of Printing, Reprinting, and Publishing, or causing to be Printed, Reprinted, and Published Books, and other Writings, without the Consent of the Authors or Proprietors of such Books and Writings, to their very great Detriment, and too often to the Ruin of them and their Families: For Preventing therefore such Practices for the future, and for the Encouragement of Learned Men to Compose and Write useful Books; May it please Your Majesty, that it may be Enacted . . . [that copyright owners] shall have the sole Right and Liberty of Printing such Book and Books...

Within a year of when the Constitution took effect, the United States Congress passed its own federal copyright legislation, the Copyright Act of 1790. It begins by describing itself as “an Act for the encouragement of learning, by securing copies of maps, Charts, And books, to the authors and proprietors of such copies, during the times therein mentioned.” It secures to copyright owners “the sole right and liberty of printing, reprinting, publishing and vending” their works for

a period of fourteen years, and provides relief to copyright owners against violators in the form of injunctions and monetary compensation. Both the Copyright Act of 1790 and the Statute of Anne required registration and the receipt of the work in an official depository for all copyright protections to apply.

Copyright law has changed in many substantial ways in the intervening time period, including the duration of copyrights, the kinds of materials that can be copyrighted, and what a copyright owner must do to ensure that his works are protected. Earlier Acts of Congress required rights to be contingent on proper adherence to formalities; after the 1976 Act, some of these formalities were relaxed, and US adoption of the standards of the Berne Convention in 1989 relaxed them further, though US Copyright law still requires formal registration for the right to sue for statutory damages. The most recent Berne Convention agreement basically dictates that member nations recognize the copyrights given to works produced in the other member nations, and requiring member nations to protect these works as they do their own.

Why Copyright?

The fact that the interests of authors and related parties are protected by *copyright* is an interesting point that can be easily overlooked. When copyright laws first originated, the best way for a creator to distribute his ideas and work, and still gain compensation, was to publish in a printed medium and charge consumers. Printing and publishing, however, were still relatively expensive. Copying required complicated machinery that was owned primarily by publishers, and not easily available to the average citizen. By preventing the unauthorized copying and publishing of his work, an author or his publisher was effectively blocking other publishers from reproducing the same work for their own personal gain. Laws of economics make it clear to us that these copies would dilute the market and decrease the profits available to the original author

and publisher. If the author were unable to receive compensation, he would be disincentivized from producing further works; assuming his works have value, his failure to produce more would cause a detriment to society. Copyright law is therefore much more complicated than it seems at first glance, and whose interest should be primary (the interest of the author or the interest of the public) is open to debate. In the world that gave birth to copyright law, however, both interests were often served by the same legislation.

Early in its history, the United States government took a preliminary stance on this issue when the Supreme Court decided *Wheaton v. Peters*.² In a decision which denied relief to Wheaton, a former court reporter who sought damages from Peters, a subsequent court reporter who republished Wheaton's work, Justice McLean ruled that only federal copyright law applied and that the restrictions given by Congress must be strictly followed for copyright to be valid; however, he added that "no reporter has or can have any copyright in the written opinions delivered by this court; and that the judges thereof cannot confer on any reporter any such right." This sentiment seemed to promote the interests of the public above the interests of the individual author. A resulting theme of copyright law is the approach that "the original expression of factual matters is protected, but the facts themselves, once they are published, can be used by anybody."³ This attitude was further codified by the Supreme Court in its decisions in *Baker v. Selden*⁴ and *Feist v. Rural Telephone*.⁵ Feist was specifically chosen by the Court to counter the decisions of lower federal courts that rewarded "sweat of the brow," despite a lack of "original expression."

² 33 US (8 Pet.) 591 (1834)

³ Godwin, Mike. *Cyber Rights: Defending Free Speech in the Digital Age*. Cambridge: MIT Press, 2003. pg. 189

⁴ 101 US 99 (1879)

⁵ 499 US 340 (1991)

Fundamentals of Copyright

The debate over whose interests copyright law should serve must be kept in mind when examining United States copyright law, as accumulated in Title 17 of the United States Code.⁶ The specific kinds of works that can receive protection are given in Section 102 (a). Specifically, works must be “fixed in a tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device” and include (1) literary works, (2) musical works, (3) dramatic works, (4) pantomimes and choreographic works, (5) pictorial, graphic and sculptural works, (6) motion pictures and audiovisual works, (7) sound recordings, and (8) architectural works. Section 102 (b) relates the important principle that copyright protection can never extend to “any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work [of original authorship].” This statement encodes the distinction between fact and presentation: the ideas themselves cannot be copyrighted, but they can be embodied in protected works.

The owner of a copyrighted work is established by Section 201 of Title 17. Initial ownership is granted to the author of the work, unless the work was a work-for-hire as defined by Title 17, in which case the employer is considered to be the proper owner. Contributors to a collective work retain the copyright to their individual contribution while the compiler of the collective work retains the copyright to the collective work as an inseparable whole. Ownership may be transferred or bequeathed through legal methods, as if it were a piece of property. Ownership of the copyright itself, however, is quite different from ownership of a material in

⁶ A very navigable copy of Title 17 is provided by Cornell Law School at: http://www.law.cornell.edu/uscode/html/uscode17/usc_sup_01_17.html

which the copyrighted work is expressed. Section 202 specifically states this distinction, and adds that “transfer of ownership of any material object . . . does not of itself convey any rights in the copyrighted work embodied in the object; nor, in the absence of an agreement, does the transfer of ownership of a copyright or of any exclusive rights under a copyright convey property rights in any material object.”

Once copyright and copyright ownership have been established, the copyright owner is entitled to six exclusive rights by Section 106 of Title 17: (1) to reproduce the copyrighted work, (2) to prepare derivative works, (3) to distribute copies of the work to the public by sale or other transfer of ownership, (4) to perform the work publicly, (5) to display the work publicly, and (6) to perform the work publicly via digital audio transmission. The term “exclusive right” means that the copyright owner has the power to exclude for the duration of his copyright; no one but the copyright owner is permitted to engage in the actions given by Section 106 without express permission.

Traditional Limitations on Copyright

However, certain exceptions to these exclusive rights apply. Chapter 1 of Title 17 includes many limitations, from Section 108’s provisions about copying by libraries and archives to Section 121’s provisions about reformatting material for use by blind or disabled persons. The limitation most notable, however, and the limitation to which would-be infringers most often appeal, is the doctrine of fair use. Fair Use is described in Section 107 of Title 17. “The Fair Use of a copyrighted work, including such use by reproduction . . . for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright.” This language is protected from ambiguity by the description of specific factors, which must be applied to any consideration of

whether a Fair Use argument is applicable. These specific factors are “(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use on the potential market for or value of the copyrighted work.”

Digital Copyright and the Digital Millennium Copyright Act

Digital media differs drastically from analog media in several important regards. First, perfect copies of digital works can be created, whereas in analog media, flaws and loss of quality are generally introduced (tracing artwork, recording a radio broadcast on tape, photocopying a page out of a library book). Second, digital copies can be created at little no cost, with common hardware — the technological barrier to duplicate that was present when duplication required devices like the printing press is removed. Third, the use of digital media tends to *require* duplication in order to be utilized — transmission of a web page requires that page be duplicated over the wire, on intermediary servers, and on the destination computer in order for it to be viewed. These differences require new legislation to prevent all computer users and service providers from being held liable for copyright infringement during the normal use of digital media.

In 1998, the Digital Millennium Copyright Act was signed into law. The goal of the Act was to resolve two issues. First, it strove to add additional legal protection to works protected by a copy-protection mechanism, replacing the economic cost to copy present in analog media (like the printing press) with a legal one. Second, it attempted to remedy the issues caused by copying of protected works during normal transmission and use.

Title I of the DMCA addresses the World Intellectual Property Organization Treaty, of which the US is a signatory. Primarily of interest are the modifications to US copyright law designed to prohibit “circumvention of technical measures used by copyright owners to protect their works.” This has the result of limiting Fair Use, specifically because uses that would follow under the Fair Use exemption are not necessarily still protected if they are achieved by circumvention of a copy-protection mechanism (such as DeCSS, which breaks the protection mechanism on DVDs to allow them to be trans-coded into another format).⁷

In addition, Title I set up a prohibition on any tool that expressly circumvents copy-protection mechanisms — be it because this is its primary function, it has limited commercial purpose other than circumvention, or it is marketed as circumvention tool. In the case of DeCSS, the software has no purpose other than the breaking of the DVD protection mechanism, and is, therefore, illegal. This seemingly valuable addition to the legislation can actually be quite harmful: even after the copyright on a work has expired, the copyright protection mechanism is still inviolable by law. Assuming that technology will continue to change and certain current forms of data storage and digital rights management will become obsolete, the prohibition on circumventing copy protection mechanisms could lead to the loss of the information that was once protected.

Title I provides some exceptions to the limits on breaking protection mechanisms, but a Fair Use defense is not among them. The exemptions include: nonprofit libraries, archives and educational institutions, where these groups may break protection in order to decide whether or not to gain *authorized* access to the work; applications of reverse-engineering, where the engineer has acquired, lawfully, the right to use such software and is reverse-engineering only for the purpose of achieving interoperability; encryption research, where protection is broken in

⁷ Information about DeCSS can be found at: <http://www.lemuria.org/decss/decss.html>

order to identify flaws and vulnerabilities in the design of the mechanism; protection of minors, where protection may be circumvented to prevent minors from accessing information on the internet; protection of privacy, where the protection may be circumvented if the protection or the protected program may be collecting personal data about a natural person; and security testing, where protection may be circumvented in order to test the security of a computer or system.

Title II provides exceptions to copyright for parties that act as service providers in four areas of conduct: transitory communications, where the service provider is acting as a conduit between the request and the provider of the content; system caching, where the service provider keeps an unmodified copy of the content for locality purposes (such as quickly mirror the content to users) which acts in a way identical to it being otherwise stored on the host of origin; storage of information on systems or networks at the direction of users, where users lacking sufficient knowledge of the infringement, and without financial gain, host material, which is removed at the direct request of a copyright-holder; and information location tools, such as hyperlinks, where the provider does not have sufficient knowledge of the infringement, or does not benefit financially, and removes the offending location tool at the direct request of a copyright-holder.

Unfortunately, Title II does not rectify concerns that holding a copy of a website in RAM or on disk may be illegal. Even though these may be activities protected by Fair Use, without legislation that specifically deals with these scenarios, it will be up to the courts to keep Fair Use preserved.

Part Two: Google as an Internet Business

Fundamentals of Google's Economic Model

Google describes their business strategy in their March 2006 SEC filing. According to the corporation, their constituencies fall into three categories. First is the category of users.

Users obtain practical benefits from Google, and the service provided by Google is the end result. Second is the category of advertisers. Advertisers take advantage of Google's reputation and targeted advertising system to promote their own businesses, and the service provided by Google is an intermediate step connecting them with customers. Third is the category of advertising partners, who ask Google to display ads on their web sites. These advertising partners, known collectively as the Google Network, use Google to generate revenue for their sites.

The first group Google must satisfy is the group of end users, who rely upon the internet giant for information dissemination and organization. To satisfy users, Google seeks to provide services that allow users to find, create and organize information with greater speed, accuracy and efficiency. Google derives 1% of its revenue by providing search engine technology, based around its proven formulas, to customer web sites. Far better known, however, and a far greater boon to Google in the long run are the services it provides for free from its portal. Most notable is its web page search, followed by such services as Google Images, Google News, G-mail and many others.⁸

Google provides targeted advertising to advertisers, from which it receives 99% of its revenue. In order to effectively market itself as an advertising space, Google must ensure that it provides an ample viewing audience by providing effective services. Its web page search, for example, must quickly provide relevant hits, and users must be provided with an interface that allows them to quickly peruse the information and determine which web pages best suit their needs. Google tackles this problem by first approaching the issue of germane returns with its

⁸ At the time that this paper was written, all of these services could be found by navigating Google's site at: <http://www.google.com>

Page Rank system.⁹ The Page Rank system seeks to ensure that the returns are “important” (the more sites link to a given page, the more important it is) and “relevant” (Google claims to analyze the context of search terms, as well as the linking pages, to determine whether returns are good fits for the search criteria). Google then provides a usable interface by displaying the title and url of the relevant pages in its index, followed by an excerpt of text from the web page itself with the user’s search terms displayed in bold. Important to Google’s reputation is its commitment to “integrity” — no site can pay to have itself displayed higher in the list of returns.

By satisfying users, Google increases the volume of traffic to their site as well as their reputation, leading to their second target constituency: advertisers. Google employs an auction-based advertising program called Google AdWords, which allows advertisers to select the keywords that will trigger the display of their ads.¹⁰ By contracting with Google, advertisers ensure that their ads are displayed in a popular location, and are specifically displayed to users who read or search for content related to the advertiser’s business. Furthermore, most advertisers only pay when a customer clicks on an ad.

Google’s first advertising program was introduced in the first quarter of 2000. It allowed advertisers to place ads on Google’s sites; these ads were specifically triggered by users’ search queries. Advertisers paid for the number of times their ads appeared. The AdWords program, an online self-service program, appeared in the fourth quarter of 2000, and by 2002 the program was offered exclusively on a pay-per-click scale. In 2005, Google began offering a pay-per-impression pricing structure for use with Google Network sites, as well as a Publication Ads Program through which Google distributes advertisers’ ads in Google Network magazines.

⁹ Google describes its Page Rank system at: <http://www.google.com/technology>

¹⁰ Google’s AdWords and AdSense services can be found at: <http://www.google.com/ads/>

The Google Network brings us to Google's third constituency, which consists of web sites that wish to earn revenue through Google's AdSense program.¹¹ AdSense includes the two branches of "search" and "content." The "search" branch, released in 2002, works with sites that have licensed Google search software for their own web sites, and displays ads on search results pages. The "content" branch displays ads on the web sites targeted to the content found within the pages.

AdSense works because the majority of the revenues generated by advertising on Google Network pages returns to the Google Network members, meaning that web sites can easily obtain advertisement revenue without needing to find advertisers. Because the ads displayed are selected to match the content of the sites, the number of clicks increases and therefore the site gains more revenue. Finally, members of the Google Network are given control to specifically filter or block advertisements of their choice, such as those of their competitors, or those that would be inappropriate for their site.

Thus, Google has developed a coherent business model that works by appealing to different levels of constituents. The level of satisfaction attained by each has a ripple effect that increases the satisfaction levels of the others: the Google Network members receive revenue through the AdSense program, the advertisers receive effective advertising because their ads are targeted and most pay only for clicks, and Google promotes itself and gains additional name recognition and validity by mediating between advertisers and paying customers.

A Hidden Constituency

It is true that all of the parties Google describes in its SEC filing are valid constituencies whose interests it must serve. These are not, however, the only interest groups to whom Google

¹¹ See previous footnote.

must pay attention. Google's business strategy is dependent at its roots upon providing useful services, principally the service of organizing information. It does not own much of the information it organizes — on the contrary, Google provides tools that allow users to gain access to copyrighted works, and profits financially from the use of these tools. The current legal environment leaves Google open to lawsuits filed by the original copyright owners, around whom Google must tiptoe to ensure the questionable legality of its actions. We would argue, for example, that Google's search engine forms a fact-based index, and that it has as much right to profit from ads placed in the margins of search result pages as a phone book has the right to profit from ads taken out in its pages; yet some of the copyright owners of the pages Google presents would argue that the situation is far less clear than it seems. Many of the services Google provides, however, must be examined carefully in light of the restrictions given by Title 17 and the DMCA. Determining the legality and whose interests are more fundamental is often a difficult matter, suggesting the need for changes in the legal and technological world.

Legal Issues Presented by Google

Text Excerpts and Keywords

Users familiar with Google's search results page will certainly be aware of the short excerpt of text that accompanies most results of a query. Within the box are highlighted phrases from the search, with a short, surrounding string to hopefully provide context to the user, who then decides whether or not the link is relevant. Google's text excerpts immediately sound like they could be covered under Fair Use — they use only the smallest portion of the original work to demonstrate relevancy in the search results, and would seem to have little to no negative economic impact on the author of the work (since it is unlikely a text excerpt could or would be used as an adequate replacement for the author's original). However, the issue that arises is the

fact that at the same time that Google displays these excerpts, it also displays ads, which are context-sensitive, and makes a profit.

The primary question is whether or not Google's model of profiting from the display of other's copyrighted works invalidates its claim that the textual excerpts are Fair Use; however, a further complication lies in the fact that AdWords is designed to tether advertisements to specific keywords. This means that a user searching for a particular company will receive a search results page that will in all likelihood include the URL and some text from the particular company's web site, yet, if a competing company has requested association between its Google AdWords advertisement and the trademarked name of the first company, then the user will see advertisements on the search results page for the competitor company.

Needless to say, this process has caused a controversy over whether its steps are legal. Previously, a district court judge decided in a case brought by Geico that the sale of trademarked names as AdWords keywords was legal, so long as the resulting ads did not contain the trademark, as ads containing the trademark were likely to cause confusion. The matter, however, remains unsettled, as Check 'n Go has sued Google over the same issue.¹² Assuming that the ads that appear on the results page are search and content based, just like the ads appearing on the Google Network, then Google profits from the triggering of ads by the text excerpts it displays.¹³ Under the current legal system, there are therefore questions over the legality of Google's actions.

¹² OUT-LAW.COM, "Google sued for selling Check 'n Go keyword." 02/09/2006. Available at: http://www.theregister.co.uk/2006/02/09/google_checkngo/

¹³ We carefully examined the information at <http://www.google.com/ads> and were unable to find any information to the contrary.

Google Images

Google’s image search behaves similarly to its text search with one major distinction — instead of being presented with a small excerpt of the resulting text on a search hit, the user is presented with a small thumbnail that is meant to be representative of the original work. As is the case with Google’s textual results, it would seem fair to assume that the relatively limited display of content from the host of origin would allow Google to be protected by Fair Use. But, just as is the case with textual search, there is the wrinkle provided by Google’s AdWords advertisements, which are contextually displayed along with the thumbnails.

The original precedent for the legality of thumbnail display by search engines was established with *Kelly v. Arriba Soft*¹⁴, in which the court decided that thumbnails were protected under Fair Use, essentially due to the fact that the thumbnails generated by Arriba Soft’s search engine were a transformative use of the work that served a different function from Kelly’s originals. The court held that the purpose of the thumbnails was to “improv[e] access to information on the internet” as opposed to Kelly’s intent to use the images for aesthetic or illustrative purposes.

This makes it sound like Google is acting legally with regards to their Images service; however, the court held a different opinion in *Perfect 10 v. Google Inc.*¹⁵ While not in direct contrast to the court’s earlier holding, it was found that Google’s image thumbnails were not exempt under Fair Use due to the fact that Google profits from their service, and that Perfect 10, in partnership with another company, already offered thumbnails of their images for sale.

¹⁴ 67 U.S.P.Q.2D (BNA) 1297 (2003)

¹⁵ 2006 U.S. Dist. LEXIS 6664

Caching

One of the most useful features of Google's search engine is the caching ability, which is unique among search engines. As Google trawls the web it keeps a local copy of the plaintext of the pages it searches, which it makes publicly accessible to users who turn up the page in a search. The page remains unmodified on Google's servers, and links to external media such as image and video files are unchanged, still referring to the host of origin. However, requests for the cached copy of the page will pull HTML and text directly from Google's servers rather than from the original host. The primary purpose of this feature is to allow users to view pages that are no longer online (whether this is due to technical issues at the host of origin, or because the page has been removed).

In the case where pages have been removed, Google's cache acts very similarly to the archiving site Internet Archive, which hosts the popular Wayback Machine. The Wayback Machine takes snapshots of pages on the Internet (or complete websites) and preserves them for later reference. These archived pages can be of major benefit to web users. Specifically, the Wayback Machine is of use to intellectual property attorneys because it allows the investigation of historical copyright and IP violations even if they have been later corrected. One specific example of such a use is the case of Healthcare Advocate, Inc., which used archived copies of its website to defend itself from a case of trademark infringement. Healthcare Advocate's defense firm, Harding Earley Follmer & Frailey, located old versions of the plaintiff's site in the Wayback Machine to demonstrate prior use of their client's trademark, which was under dispute from a similarly named company, Healthcare Advocates.¹⁶

¹⁶ Flynn v. Healthcare Advocate, Inc. 2004 U.S. Dist. LEXIS 293; 2004 U.S. Dist. LEXIS 12536; 2005 U.S. Dist. LEXIS 1704

However, most likely because the Wayback Machine was responsible for Healthcare Advocate's successful defense of their trademark, the original plaintiff (Healthcare Advocates) went on to sue both the Wayback Machine and Harding et al. for damages resulting from copyright infringement and violation of the DMCA and Computer Fraud and Abuse Act.¹⁷ According to the suit, the Internet Archive was accused of breach of contract, negligence and other charges because it failed (in a small set of circumstances) to follow the robots.txt file on the Healthcare Advocates' site. As a result of this, the suit claims Harding et al. was in violation of the DMCA by following links to material behind a copy-protection mechanism (in this case, robots.txt). The firm was in further violation because of copies of archived documents created and transmitted as a result of their defense of Healthcare Advocates in prior litigation.

Google has also been the target of a lawsuit relating to Internet caching, and recently the case of *Field v. Google*¹⁸ was decided in Google's favor, by Google's appeal to the protections of Fair Use. Google's Fair Use appeal consisted of several arguments that have a direct consequence on both the Internet Archive case as well as other cases where cached content is at issue. First, Google argued that the purpose of its cached copy was fundamentally different than the purpose of Field's original works: Google offered a cached page in order to "improv[e] access to information on the internet" while Field intended his works as art. This implies that a digital copy of a work that is used in a different manner than the way that the copyright owner originally intended could perhaps be considered legal.

Second, Google argued that the intention of the author was to widely disseminate his works, by offering them to the public, for free, on the internet; so, therefore, Google was

¹⁷ At the time that this paper was written, we could not provide a citation for this case because it had yet to be decided. A copy of the complaint, however, could be found at:

http://www.geocities.com/bledrydudenet/Healthcare_Advocates_v._Harding_Complaint_FINAL.pdf

¹⁸ 412 F. Supp. 2d 1106 (2006)

assisting the goal of the author by providing a cached copy of the page. One interpretation of this is that by publishing a page on the internet, the author is creating an implicit license to copy for search engines that have the intention of helping to disseminate the content of the pages, especially those sites which also include a robots.txt that does not prohibit archival.

Third, Google argued that the amount of content provided by the web cache is just sufficient to allow the copy to be useful to the user. The court case cites the fact that highlighted search terms would not be useful without the surrounding text also being present, as well as the fact that the Supreme Court has upheld uses of entire copyrighted works when the original work is already available for free (as it is here).

Fourth and finally, Google argued that the case should be decided in its favor partly on good faith demonstrated by its caching system: Google will not cache pages the owners specify not to cache, and Google provides on its website clear instructions on how to prevent caching.

Unfortunately, while *Fields v. Google* seems to put to rest, or at least partially alleviate, some of the tensions caused by digital copyright, it does not address one key issue of the Internet Archive case — the DMCA. At least part of the burden placed on Google's Fair Use claim is the reliance on adherence to robots.txt, which the Wayback Machine did not follow. If the robots.txt file is considered to be a significant enough form of digital protection, Harding et al. are still liable under the DMCA for accessing the archived copies of its client's site.

Caching raises questions outside of the United States as well. In 2005, Canada's Parliament proposed Bill C-60, an amendment to Canada's Copyright Act, which attempts to give copyright holders more recourse in the digital age. One clause, however, reads: "...the owner of copyright in a work or other subject-matter is not entitled to any remedy other than an injunction against a provider of information location tools who infringes that copyright by

making or caching a reproduction of the work or other subject matter." This clause suggests that the very act of caching is illegal. Should the bill pass, there will be serious implications for companies such as Google.

Google News and RSS News Feeds

Google News provides a single page portal to news from a variety of media outlets. Again, following the trend of its other services, news entries are provided as simply an article title, a short excerpt, and a list of links to read more. The links themselves are to actual providers of content, and the relationship between Google and the news item ends at the textual excerpt — Google does not cache the articles, nor do they advertise on the News site.

For once, one of Google's services seems relatively benign, considering that the hot-button issue of their profits is not a factor in this instance. However, this has not prevented Google for being the target of copyright infringement suits (or the threat of such suits), such as from the French news agency AFP.¹⁹ As a result of the AFP complaint, Google removed all references to AFP articles and content.

Google Books

Google Books and Google's Library Project aims to digitize entire libraries of books, beginning with those of the University of Michigan, Harvard and Stanford, as well as public domain books in the New York Public Library and Oxford University library.²⁰ Books still protected under copyright are available to search, but only the book's identifying information and small excerpts are made available to users who find the book via searches. Books no longer under copyright are presented in their entirety.

¹⁹ Sandoval, Greg. "Newspapers want search engines to pay." CNET News.com: 01/31/2006. Available at: http://news.com.com/Newspapers+want+search+engines+to+pay/2100-1025_3-6033574.html

²⁰ Wyatt, Edward. "Authors sue Google over library." The New York Times: 09/21/2005. Available at: <http://www.iht.com/articles/2005/09/21/business/google.php>

Google behaves very carefully with books that are still under copyright, requiring users to log in to view copyrighted works (in order to make sure that the works are not viewed in their entirety by one user). Also, Google makes sure to offer links to purchase the books (from which it does not profit), but shares the profits derived from its AdWords ads with the authors and publishers. In addition, Google held back from scanning copyrighted books, delaying their project for three months in order to give publishers and authors ample time to opt-out.²¹ Current copyright legislation, however, is not framed as an opt-out situation, leaving Google's legal standing somewhat shaky.

Like in other cases where Google is providing excerpts of copyright-protected material, without having first obtained express permission, they have found legal trouble. With Google Library, three authors filed an injunction against Google for copyright infringement caused by the project, each author claiming copyright of one of the books in the University of Michigan library. The authors did not know whether or not their books had yet been copied, but held that the project was in violation of the Copyright Act because Google did not ask explicit prior permission to use the works.

With each of Google's services, and similarly with services like those provided by Google, problems stem from the fact that opt-out implicit licensing is not a legal standard, and that Google or others may not be protected from copyright violations simply by providing ways to opt-out. Secondly, services are also hindered by the fact that they are commercial enterprises, which, while at the same time are showing excerpts of third-party content, may also be showing ads and earning revenue. Third, there is another variable, the DMCA, which may possibly be

²¹ Band, Jonathan. "The Google Print Library Project: A Copyright Analysis." Available at: <http://www.policybandwidth.com/doc/googleprint.pdf>

used against services such as these if it is ruled that mechanisms such as robots.txt are suitably strong enough to be protective of copyright-protected material. Finally, differences in copyright law between nations complicates the system even further: what may be legal in one country (such as caching in the United States) may be illegal in another country (such as caching in Canada).

A legal climate that provides so many roadblocks to what seems to be a fair dissemination of information is in conflict with the original goals of copyright legislation. It is critical to protect the rights of authors and publishers, but these rights are protected in order that the general population may also receive some benefit, despite recent trends that attempt to find real property rights in intellectual property. The ideal legal environment should be one of balance, and not one of intellectual protectionism. In order for Google and similar services to be confident that they can provide helpful services and still have the financial incentive to provide such services, the problems given above must be addressed.

Part Three: Legal and Technical Solutions

In *Kelly v. Arriba Soft*, the court held that:

The Copyright Act was intended to promote creativity, thereby benefiting the artist and the public alike. To preserve the potential future use of artistic works for purposes of teaching, research, criticism, and news reporting, Congress created the fair use exception. Arriba's use of Kelly's images promotes the goals of the Copyright Act and the fair use exception. The thumbnails do not stifle artistic creativity because they are not used for illustrative or artistic purposes and therefore do not supplant the need for the originals. In addition, they benefit the public by enhancing information-gathering techniques on the internet.

It would seem that the court would promote the idea of dissemination of information on the internet, specifically information that is already provided, for free, to the public. Therefore, we propose to amend the Copyright Act in the spirit of dissemination of free information.

Proposals

First, we propose to modify the Copyright Act in this spirit — the purpose of copyright should be for the promotion of sciences, useful arts, and the dissemination of public information. This is accomplished by making the publishing of freely accessible media online carry an implicit license granting permission to copy, pursuant to the requirement that it is for dissemination of free information (as in *Kelly v. Arriba Soft*).

Second, in the interest of preserving the rights of diligent copyright owners, we allow for the possibility of opting-out of this implicit licensing, using a technical solution similar to the rights granted by robots.txt but with additional benefits and legal protections. This technical solution, either implemented as an extension to robots.txt or as another separate solution, provides lists of either Uniform Resource Identifiers (such as URLs) or directories that should be omitted from the implicit license, or that should be provided under a modified license.

Third, we propose an amendment to the DMCA that allows circumvention of copy-protection mechanisms in instances that would otherwise fall under the provisions of Fair Use, such as DeCSS. Furthermore, breaking copy-protection mechanisms should only be illegal for works that are under copyright and not within the public domain.

Finally, we would move to have these suggestions implemented by the Berne Convention so that they would apply near-equally for a large portion of internet-using companies and individuals.

Implementation: Why it Works...

The modified license proposal would work hand-in-hand with the proposed changes to Title 17 and represents the major modification to how the explicit licensing would work. Copyright holders would be able to specify the duration their work may be held by another party,

to prevent an eternal hoarding of information, especially information that may be updated and outdated quickly. The holders would also be able to dictate terms of use, similarly to how Creative Commons licenses operate.²² With the ability to set terms, holders would be able to specify whether or not the content could be used in for-profit enterprise, only in non-profit use, or to distribute the content but not create derivative copies. The holders would also be able to specify different rights for different parties; for example, giving the Internet Archive the right to cache a page indefinitely, but giving Google the right to cache a page for only several days at a time.²³ Corporations would be bound by law to respect the desires of the copyright-holder as expressed in this license file, so long as the terms set by the copyright-holder do not violate pre-existing standards of Fair Use. In this way, a balance is achieved between the interests of the copyright-holders and the interests of those who wish to use the material.

The DMCA as modified should still stand as a piece of legislation that discourages piracy. Our proposed changes would not condone the breaking of copy-protection mechanisms on copyrighted works; rather, they would apply the same standards to digital works as already

²² Information about Creative Commons can be found at: <http://creativecommons.org/about/licenses/>

²³ A sample document could be named "license.xml" and might look something like this:

```
<DomainEntry>
  <Requestor>Google.com</Requestor>
  <Right>NonProfitDisplay</Right>
  <Right duration="5 days">PublicCache</Right>
</DomainEntry>
```

```
<DomainEntry>
  <Requestor>Alexa.com</Requestor>
  <Right>NonProfitDisplay</Right>
  <Right duration="indefinite">PublicCache</Right>
</DomainEntry>
```

```
<DomainEntry>
  <Requestor>Others</Requestor>
  <Right cache="5 days">PrivateCache</Right>
</DomainEntry>
```

This document provides domain-level rights management and offers the ability to confer specific rights based on the domain of the requestor. The requestor could be IP-based (perhaps with a range of acceptable IPs) or based on the DNS entry of the requesting IP.

exist for analog works. This change would be congruent with pre-internet legislation, and just as with analog works, violations would still be punishable by law. Therefore, the needs of those who would wish to manipulate copy-protected digital works under Fair Use, and of those who would wish to manipulate copy-protected digital works that are no longer under copyright, would be balanced with the needs of copyright holders who wished to protect their works with the use of digital rights management.

The Berne Convention already requires that member organizations treat the copyrighted works of other member organizations with the same respect that they treat their own copyrighted works. A philosophy such as this is even more critical in the digital age, which has removed many international barriers. The United States may pass whatever legislation it chooses, but without the cooperation of international allies, its copyrighted works may be subjected to different standards in other countries. Even more troubling is the stipulation that copyright violations be dealt with in the country in which the infraction occurred. The nature of the internet has blurred the concept of location, as has already been seen in countless jurisdiction cases just within the United States. Does a violation occur at the location of the infringer? At the location of the server with which he infringes? At the location of the server from which he made the copies? At the location of the corporate headquarters of the copyright owner? If the Berne Convention were persuaded to adopt more unilateral standards, the problems arising from the potential application of different laws depending on what definition of “location” was used would be minimized.

These proposals, for the reasons indicated above, should mitigate the legal problems highlighted by the specific examples we have given regarding corporations such as Google,

particularly in the re-articulation of Fair Use and in giving copyright-holders a clear method for expressing their wishes.

...Without Causing More Problems

Courts have already shared opinions that the dissemination of free information is a valid cause, as in *Fields v. Google Inc.*, so our first recommendation serves to strengthen this sentiment, following the standard set by *Kelly v. Arriba Soft* and partially reversing the ruling of *Perfect 10 v. Google Inc.* The Supreme Court has held that the fact that a piece of information is used commercially does not prohibit that piece of information from being used under Fair Use. And, specifically in *Kelly*, the court held that “The more transformative the new work, the less important the other factors, including commercialism, become.” In the case of search engine technology, the use *is* transformative and only acts to further aid the original intent of the author by extending the reach of the work (since it has been provided free, online, in order to be found by a search engine). Our proposals are therefore consistent with how judicial decisions have previously interpreted existing legislation, and therefore consistent with the existing mentality regarding copyright. Our proposals serve to clarify this mentality rather than confuse it.

Our modifications to the DMCA do not undermine any current copyright protections — the rights of copyright holders do not change, and previous illegal uses of copyright-protected material remain illegal. Our provision changes the protections to the containers of protected works, allowing them to be circumvented within the bounds of Fair Use. One objection to modifications of the DMCA could be that allowing some circumvention of copy-protection mechanisms would encourage the development of tools that would also be used for illegal purposes. However, the existence of the DMCA does not currently thwart the development of such software and merely defines it as illegal — such software would continue to exist both

within the United States and internationally, even without the modifications to the DMCA, and it is unlikely that further illegal uses would result.

As far as the Berne Convention proposal, it is difficult for us to find an objection to its implementation. We do recognize that there could be considerable resistance to the adoption of such a proposal by other nations, which may not come into the Berne Convention with the same mentality and legislative background that the United States would. However, were the proposal to overcome the barriers to its adoption, we do not see any problems arising except for inconsistencies between the Berne Convention and the individual laws of its members. The international environment in which the internet has grown would be best served by a reconciliation between the proposal and the laws of Berne Convention members — as previously mentioned, the more nations that follow the same rules, the more confusion will be minimized.

The internet is a powerful medium for the transmission of works of art, science, literature, news, and opinion. It has the potential to be an educational equalizer, provided that the information contained inside its digital walls remains accessible. It also has the potential to destroy the foundations of our incentive-based copyright system. But, with an appropriate balance of the rights of the public and the rights of copyright-holders, we can both encourage the development and fair treatment of new works while at the same time providing a near infinite stream of information to the knowledge-thirsty public.