

Google and the Book Publishers: Testing the Limits of Fair Use in the Digital Environment

By Peter Givler

Editor's note: *We occasionally ask a non-lawyer to provide a perspective on a current issue in intellectual property law that has a significant impact on his or her business. Fair use is a notoriously difficult area of copyright law, and those difficulties are arguably more acute in the digital environment, in which the ability to make and distribute perfect copies on a scale never before possible poses an especially acute threat to copyright owners. At the same time, digital distribution also arguably presents previously unimagined opportunities to disseminate knowledge. We asked Peter Givler, Executive Director of the Association of American University Presses, to discuss the copyright issues raised by the Google Library Project, an incipient digitization project of gargantuan proportions that has alarmed many publishers, including the university presses represented by the AAUP and on whose behalf Mr. Givler has challenged Google to justify its plans under copyright law.*

The Association of American University Presses (AAUP) is a trade association for nonprofit scholarly publishers. About three-quarters of our 125 members are affiliated with research universities in the United States and Canada, and the rest are either the publishing arms of non-degree-granting organizations, like scholarly societies, or international members. In addition to their nonprofit status, AAUP members also have in common that their imprints are under the control of an editorial board, usually a committee of scholars on the faculty of the parent institution. Manuscripts being considered for publication by a press are sent out for peer review, and those reviews are then evaluated by the editorial board in determining whether to publish the work. This two-stage process of independent editorial review is one of the hallmarks of university press publishing and is intended to guarantee the intellectual merit (but, alas, not the sales) of the books that appear under the press's imprint.

Scholarly publishing means, almost by definition, publishing for small markets. Done within the financial context of a university, it usually means publishing on a shoestring as well: tight budgets, limited access to capital, low tolerance for aggressive pricing, thin margins. Many presses receive operating subsidies from their parent institution, but the amount varies considerably with the size of the press, with the smaller presses, for all the obvious reasons, much more heavily dependent on subsidies. Still, subsidies play a smaller role in a university press's finances than most people realize—on average, only eight to ten percent of their operating revenue.

So while university presses have an obvious and deep commitment to spreading knowledge, they have an equally strong commitment to copyright, the legal regime that permits them to recover from the marketplace almost all of the expenses of publishing. It was in that spirit that, writing on behalf of our members, I FedEx-ed a six-page letter to the Senior Intellectual Property Counsel at Google on May 20, 2005, asking a series of questions about the Google Library Project (described below), most of them going to Google's legal

justification for digitizing copyrighted material without permission.

To my astonishment, on the morning of May 23, the lead story in *Business Week Online* was "A Google Project Pains Publishers," with a link to the full text of my letter (now at http://www.aaupnet.org/aboutup/issues/0865_001.pdf). The next few days were my fifteen minutes of fame: I was interviewed by NPR (twice), the BBC World Service, the Canadian Broadcasting Corporation, *Forbes*, *The Chronicle of Higher Education*, *The New York Times*, and an AP reporter whose story ran in over 300 newspapers. It was nuts. Who knew the world was so interested in the byzantine complexities of fair use and library exemptions? And to cap it all off, that week Google's share price went up another \$10.

So why am I fussing? Google is everybody's favorite search engine, the stock market's darling, and a great American success story. What's not to like? For most of Google most of the time, not much. As we all know, there is a staggering amount of information available on the Internet, and web search engines, like Google, are what librarians call finding aids or resources that do not themselves provide you with the information you are looking for but point you toward places it may be located. You type a word or phrase into the search box, hit return, and you get a list of "hits" or links to the web pages where those words appear.

A search engine performs this useful task by creating a dynamic index of web pages, so when you type in a query, the whole web is not searched, only the index. The beauty of this process is that it is blisteringly fast; the drawback is that, like everything else a computer does, it is methodical to the point of witlessness. Search engines are indiscriminate and scoop up everything in their index that matches your search terms. Lists of search results, compiled in hundredths of a second, can easily include tens of thousands, even millions of links.

Enter Google, the first to market with a search engine that returned lists of search results organized in a useful way: they are ranked based on the frequency

with which other people who have used the same search terms have “voted” for their relevance by clicking on them. Google calls this, unsurprisingly, ranking by relevance, and while googling is not efficient for a specialist—a lawyer, for example, does not need Google to tell her where to find 17 U.S.C. § 108(a)(2)—it usually works pretty well if you do not know where to start. Google searches still may return a staggeringly high number of links, but usually the one you are looking for is in the first five; rarely do you have to look beyond the first ten. So the Google search engine has proved to be a very useful and hugely popular tool for locating information on the web.

While there is a great deal of information on the web, there is also a great deal of information that is not, namely, the contents of all the millions of books that are only available publicly in traditional, ink-on-paper form. Several years ago Google began to make plans to index this material as well through Google Print, which, as Google puts it, “makes offline material searchable.” Google Print is made up of two quite different programs: Google Print for Publishers, sometimes known as the Google Publisher Program, and Google Print for Libraries, sometimes referred to as the Google Library Project.

In Google Print for Publishers, Google has been negotiating agreements with publishers for the last year or so under which Google gets permission to add the contents of published books to a Google database, either by scanning the books to create a digital copy or by using a digital file supplied by the publisher, so that they can be indexed by the Google search engine. This allows links to the contents of these books to be included in Google search results. A user clicking on a such a link would be shown the page in the book where the search terms appeared, have limited ability to browse further in the book, and be offered the opportunity to “Buy this book” by clicking on a link to the publisher’s website and/or to online vendors, like Barnes & Noble or Amazon. So far, so good.

Google Print for Libraries is similar in that Google is digitizing books and adding them to a Google database so that they can be indexed and references to them included in Google search results. However, the similarity ends there. In the Library Project, Google entered into agreements, not with publishers, but with five libraries: the New York Public Library, the Bodleian Library at Oxford University, and the university libraries at Harvard, Stanford, and Michigan. The New York Public and the Oxford libraries are only supplying Google with books in the public domain for scanning. However, Harvard, Michigan, and possibly Stanford are supplying Google with a mix that includes books under copyright as well, and that is where the problems begin.

As part of their agreements with the libraries, Google not only is making digital copies of works protected by copyright, but also is giving each participating library a digital copy of all the books, including those under copyright, that Google has digitized from the library’s collection. How the libraries intend to use these copies is not clear, at least in part because their contracts with Google are apparently protected by non-disclosure agreements. However, the University of Michigan is a public institution subject to the state’s sunshine laws, and in response to a FOI request, the Michigan contract was recently posted on the university’s website (<http://www.lib.umich.edu/mdp/>).

Paragraphs 4.4.1 and 4.4.2 of the agreement state that “U of M shall have the right to use the U of M Digital Copy” in “services” provided on the U of M website, and cooperatively with “partner research libraries such as the institutions in the Digital Library Foundation.” Those paragraphs then go on to stipulate restrictions on use of the U of M digital copy: technological measures will be implemented to restrict automated access, access will be restricted to people “having a need” for it, reasonable efforts will be made to prevent third parties from downloading or redistributing it, and so on.

As I read these restrictions, though, there is nothing that would prevent an authorized user of the U of M website—presumably registered students and faculty—from accessing the U of M digital copy for use in teaching, study or research, downloading portions of it, and printing them out, as long as the portions are not “substantial,” which is undefined, nor the downloading “systematic,” and the portions were not “redistributed,” not used for a “commercial purpose,” and not “disseminated to the public at large.” In other words, there appears to be little in the agreement that would restrict use of the U of M digital copy by students and faculty at the U of M for many arguably educational purposes, including electronic reserves.

It appears that the agreement extends the same privilege to the U of M’s “partner research libraries.” It provides that before “making any such distribution, U of M shall enter into a written agreement with the partner research library” that will “contain limitations on the partner research library’s use of the materials that correspond to and are at least as restrictive as the limitations placed on U of M’s use.” Who are the partners contemplated under the agreement? The Digital Library Federation (DLF) has 34 members. They comprise twenty-eight research universities and the Bibliotheca Alexandrina, the British Library, the Council on Library and Information Resources, the Library of Congress, the National Archives and Records Administration, and the New York Public Library (<http://www.diglib.org/about.htm>). Note, though, that the agreement says “with partner institutions *such as*” (emphasis added)

members of the DLF. As I read the agreement, U of M's partners could just as easily include any or all of the 124 members of the Association of Research Libraries.

What about Google's use of the files? Google says that a user in Google Library who finds material he or she is looking for in a book in the public domain will be able to browse the book at will. If the book is under copyright, however, he or she will be shown only bibliographic information about the book and "snippets" of text: the search term and a couple of lines of text before and after it, with a three-snippet limit. For simplicity's sake let us assume there is a simple and infallible bright-line test for telling what is in the public domain, and that there are no legal questions about either Google's or the libraries' use of public domain material. If all the Google Library project were doing was making it easier for people to find information from texts in the public domain, and helping libraries provide access to them for their patrons, we all would be cheering them on. But would we be cheering any more loudly than we already do for Project Gutenberg, Ask Thomas, Find-Law, or any of the hundreds of other websites that already provide straightforward access to classic texts and useful information in the public domain? Maybe a little more loudly because Google Library would offer one-stop shopping, but my hunch is not by much.

The attraction of Google Library—or in the current jargon, the source of its ability to attract eyeballs—is the promise that it will bring to your computer screen some form of instant access to every book you would find in the combined collections of five world-class libraries. What would happen if you limited that access only to works in the public domain? You still would have a resource that would be of great interest to scholars and people who use government information. It would be of some use to students. For everybody else, though, using Google Library would be like looking things up in an encyclopedia that was last updated in 1923, and how attractive is that?

So here is the basic problem: Google, a company whose current market capitalization is over \$80 billion and growing, plans to further expand its business by making digital copies of copyrighted works in the collections of at least two major university libraries, distributing copies to those libraries, and displaying portions of those copies, all without permission of the copyright owners. On the libraries' side, they are turning copies of copyrighted books over to Google for digitization and receiving digital copies of the books in return. At least one of them apparently contemplates some form of further distribution, both to users of its own website and to other institutions—and all of it, again, without permission of the publishers.

What are Google's possible legal defenses? Section 108 of the Copyright Act does offer libraries some

exemptions from the exclusive rights of copyright owners, but those exemptions are very specific and limited, and they do not apply here. Both the language of the statute and its legislative history make clear that section 108 neither permits libraries to subcontract copying to a commercial entity, nor to engage in the wholesale and indiscriminate copying of copyrighted works in their collections. Rather, section 108 is designed to permit, for example, the making of a copy when a book is damaged or lost.

Google has taken the position that its copying is fair use, and as precedent it cites a Ninth Circuit case, *Kelly v. ArribaSoft*, in which the court found that ArribaSoft's copying of photographic images from a photographers' website in order to create an index using low-resolution "thumbnails" of the original images was a fair use. That case, however, is very different than the Google Library Project in several respects, including that ArribaSoft was indexing material that already was available on the web. No precedent authorizes the systematic copying of entire library collections contemplated by Google. The fair use argument made for the Library Project is tantamount to arguing that every work ever published that is still under copyright is subject to the same fair use claim.

The next to last paragraph of my May 20 letter states:

Google Print for Libraries has wonderful potential, but that potential can only be realized if the program itself respects the rights of copyright owners and the underlying purpose of copyright law. It cannot legitimately claim to advance the public interest by increasing access to published information if, in the process of doing so, it jeopardizes the just rewards of authors and the economic health of those non-profit publishers, like the members of AAUP, who publish the most thoroughly vetted and highest quality information in the first place.

The letter concludes with an offer to meet with Google representatives to discuss our concerns and attempt to find a way to resolve them.

We hope that Google can be persuaded to proceed with the Library Project in a manner that respects the rights of copyright owners to authorize the copying and distribution of digital copies of their works. Notwithstanding Google's contention that it is providing a public service and that it will stimulate, rather than harm, book sales, AAUP and its member presses are deeply concerned by what appears to be a large-scale usurpation of the prerogatives of copyright ownership.