

Copyright and the Digital Library

Jeff Erwin

LIBR 287

Prof. Kroski

San Jose State University

School of Library and Information Science

November 2, 2008

**Preface**

This paper concerns a continuing and developing story. After the Google Books settlement at the end of October, the number of articles and references to digital libraries has exploded; researching and remaining current on these analyses is a bit like swimming upstream. I will do my best.

**Introduction**

The future of libraries in the digital future is uncertain. Much of this uncertainty revolves around the library's ability to continue as a loaning institution as media becomes exclusively digital and users become accustomed to using computers and the internet to access the resources they need.

To remain relevant and recognizable, libraries need to continue to provide access to all types of media, not just those media that have been converted into a physical object (books, DVDs, videocassettes, tapes, DVD-ROMs and other texts). If one accepts that books themselves will become electronic documents, portable to a variety of forms, then the problem becomes: copyright.

Under current copyright law, American libraries have the right to dispose of their physical books as they see fit. This usually means that they can circulate and that this circulation is restricted only by the library's own policies. This is not a violation of copyright, despite the fact that the patron is not paying for the loan.

Digital texts have certain definitive advantages over printed works, advantages which will work to assure their importance and eventual primacy: firstly, they are profoundly more portable. A book is a container of size and heft that additionally requires implicit costs as a library item: it takes up space both for the customer and the library; it requires physical maintenance, and, most importantly, it requires physical delivery. The patron is in most cases required to obtain the book personally, and return it within a set period of time. These are transportation and infrastructure costs that subtly impact

the economic decision making which can either benefit or doom a library.

Secondly, the natural state of any digital media is its openness to manipulation, copying, alteration and editing. Implicit to the terms of copyright which assure libraries their rights is the assumption that the works are not easily duplicated, altered, re-edited or kept indefinitely. These considerations must be coded into the access points or the media itself if they are to restrict their use, and they remain generally both hackable and incomplete.

In reaction to this the copyright laws as they stand now place digital works in a different arena than physical texts. They require licensing specific to the item as they are all “copies.” A library with a digital copy of a text is in essence required to follow the copying provisions of copyright which require explicit consent from the rights holder.

According to Intellectual Property (a pseudonym):

“[L]ibraries probably can not [sic] make most of the non-digital works in their collections available electronically without owner authorization. Perhaps the owners of most works will be sufficiently motivated to make their works available electronically to keep this category of "digitally unavailable works" small. But wait a minute; even if owners rush to authorize digital conversion of everything they own, haven't the libraries already paid copyright owners once for many of these works (a price that includes a surcharge for all the copies that owners theoretically won't sell because potential buyers will read the library's copy instead of buying their own), and should they have to pay again because an activity (reading or browsing or access) that once did not involve a use of copyright (copying) now inherently does? Changes to the copyright law seem to confirm the copyright owners' position that the First Sale Doctrine should not apply to electronic works, but no one has addressed the profound implications of this change for

library operations. Perhaps it will be of no consequence ultimately once all works are acquired digitally to begin with under licenses that authorize distribution. For the moment, however, there are significant gaps in the legal authority to make works available electronically to the public and important questions about the costs.”

It is important to note that a truly digital library prices out commercial access to products. There is no incentive to buy a license for a digital work if the same work is available for the required length of time from a library, and the item is portable to the user’s own devices. Thus the economics of the current environment rule out the norms of library use. Not only that, the restrictions which balance the costs of the library and the incentives of the publishers to supply libraries, namely due dates and the costs, institutional (class based) and transportation are not naturally a component of the item: to make an item have a limited shelf life for the patron, or require a physical appearance at the library are costs that would be assumed by the library itself, reducing its usefulness to its community, to the advantage of the publisher alone.

Even from the viewpoint of classical economics, this is untenable. If an industry is supported through manipulation of the markets alone, and simply to advantage the industry against the general population, then this intervention acts to create unbalancing and unequal anti-competitive behavior.

### **Google Books and the Lawsuit**

From 2005 to very recently, Google Books was subjected to a lawsuit alleging infringement of copyright by the Authors Guild and the Association of American Publishers. Google Books operated by scanning books, both in and out of copyright. Books in copyright were restricted to either small fragments of browseable text or pages, or to publishing and physical data; books out of copyright were generally viewable as full text documents. The lawsuit argued that Google Books’s archiving of inaccessible copyrighted material was a violation of fair use; and that GB’s assertion that fragments of

text amounted to fair use was erroneous. However, the settlement reached created a new paradigm which favored Google. (Lardinois, 2008)

Under the settlement, Google Books would create a tiered system, allowing various levels of access to copyrighted materials based on payment. Under the new system, out of print, but still copyrighted works would become accessible—to those who can pay, and all works would be accessible to unspecified academic institutions and libraries, again with a pay schedule.

Superficially, this is a victory, and it is certainly not a defeat to the principle of digital access. But certain important caveats exist which disadvantage libraries. Lawrence Lessig (2008) comments:

“IMHO, this is a good deal that could be the basis for something really fantastic.

The Authors Guild and the American Association of Publishers have settled for terms that will assure greater access to these materials than would have been the case had Google prevailed. Under the agreement, 20% of any work not opting out will be available freely; full access can be purchased for a fee. That secures more access for this class of out-of-print but presumptively-under-copyright works than Google was initially proposing. And as this constitutes up to 75% of the books in the libraries to be scanned, that is hugely important and good...

“Here, too, however, there is an important challenge for Google. It has provided important value by making available works that have no rights attached to it. But it should do more to make available works that have some rights attached to it. Critical for evaluating whether the long term interest of Google is GOOD or GOOey, Google needs to build into its architecture assets that are licensed freely, or under noncommercial terms, to complement the assets that it claims are free for "noncommercial" download (namely, the public domain works it has). Acting to clearly

support the non-proprietary movement as well as the proprietary is an important way for it to show that it stands in the middle, and that it, with the AAP/Authors Guild, have now done some real good.”

The main problem with this structure is that Google has become a commercially dominant force in digital access to books. Google is easily in a position to dictate pricing and access based on profit or institutional needs at the expense of regular users, libraries or other subscribers to this service. It has not, as Lessig points out, made their archive Creative Commons licenseable, or otherwise integrate into the Open Source community.

The second problem is the obvious one that Libraries would be constrained to offer access to users within their physical surrounds, on computers they must maintain and make available, while the patrons cannot remove the titles to a more advantageous location, such as home, work, or school. In the library I work at alone, the 15 to 30 minute limit on computer use would essentially make this service useless to most consumers.

These issues will (like the main ones which affect digital copyright) become increasingly problematic as works become exclusively digital. Such a work would be unlendable by a library.

The Director of Harvard University’s Library, Robert Darnton, has criticized the settlement. As he points out:

“As we understand it, the settlement contains too many potential limitations on access to and use of the books by members of the higher education community and by patrons of public libraries... The settlement provides no assurance that the prices charged for access will be reasonable... [E]specially since the subscription services will have no real competitors.” (Albanese, 2008).

Wade Roush describes other issues:

“And there’s another provision of the settlement that spells out, to me, just how parsimonious the plaintiffs’ attitude really is. Under the agreement, the authors and publishers give Google permission to provide every public library in the United States with free access to the books database. That sounds great, on the surface. As Authors Guild president Roy Blount Jr. put it in a message to members about the settlement, “Readers wanting to view books online in their entirety for free need only reacquaint themselves with their participating local public library: every public library building is entitled to a free, view-only license to the collection.”

But the devil, again, is in the details. If you read the agreement, you’ll see that it restricts each public library to exactly *one* Google terminal. Tens of millions of books online—but at any given moment, no more than 16,543 people are allowed to read them without paying. (That’s how many public libraries and branches there are in the United States, according to the American Library Association—one for every 18,500 Americans.)” (Roush, 2008)

### **The future of digital works**

A digital work has one further advantage over a physical object: it massively reduces production costs. The author no longer requires a publisher; the publisher’s pricing becomes nearly all margin. It is here that we will see some of the most interesting consequences of digital work. At present these works are priced in relation to their printed counterparts. A book which costs \$25 to \$30 costs perhaps \$5 or \$6 in a digital form, or even less. But massive profits could be made if the work is inaccessible in any other form. A new Stephen King novel, for instance, might be priced, as a digital only product at a much

higher cost, and the public would be compelled to pay.

This situation, however, would exist mainly because the publishing business itself controls access to printing and distribution. These technologies and infrastructures are almost imaginary in a digital environment: it is simply economically a much better deal for the author to self-publish and charge a competitive price. The author as free agent receives all of the marginal profit and pays mainly in time spent composing the work.

What does this mean? It means that the direction of publishing will inevitably head towards (from either a publisher's or author's lead) purely digital data. The only constraint is the need to control copying and republication. If security measures can be created which overshadow the advantages of piracy for the general public, the work remains an income generator.

There are several ways in which Libraries are ill-equipped to handle this situation. Firstly, a digital work remains almost indefinitely in print, and rights owners are in no way economically predisposed to release their fees. Libraries would be depend on services like Google merely to sustain their collection. It would become overhead. Secondly, working with an Open Source model is incompatible with digital copyright, and as it stands, creative works in particular are only advantageous as OS if the creator derives some other tangible benefit from the act of releasing the data.

Secondly, a hypothetical digital-only novel of considerable public interest would be, under the Google Books paradigm, only accessible for limited stretches of time within the confines of the library itself or the patron must pay.

Libraries have long understood that sustained and free access to popular materials is part of their rationale and a social and economic necessity. The purposes of libraries remains at odds with publishing per se, because the library takes over part of the potential customer base for a book. Thus

the publisher or creator would be benefited by increasing that market share, shaving off middle-income library users into the buyer pool. Without these very patrons, libraries would lack much of the political and social capital that keeps them open.

### **Practical Limitations of the Google Books and NetLibrary paradigm**

I recently used NetLibrary while researching a paper. The experience was telling, and frustrating. NetLibrary repeatedly froze up, exhibited consistent compatibility problems with FireFox, and did not permit me to skip ahead through the book. It needed the formal existence of a book in the San Jose Library catalog before it would permit its use, and it ultimately crashed FireFox before I could access the information I needed.

These are technical issues that surround the functioning of a restricted access database. Databases and wikis which are multi-server and hosted by innumerable computers simply do not have these issues. Tying down a system into a commercial site means the library is at the mercy of the site and its maintenance people. If the Library relies on it to supplement its holdings digitally, it can become near useless with no warning. This is akin to half the books in the library suddenly being checked out.

### **“Orphan Books” and Open Access**

Lawrence Lessig has written a considerable amount on the subject of digital copyright. He has proposed that Libraries and digital archives be allowed to gain inexpensive rights (he suggests \$1) to books which remain in copyright but are out of print (so-called “orphan books”); and that these funds would be used to compensate authors of these works. Authors would have the right to remove or place books on this registry.

Lessig argues, fairly strongly, that these orphan books, which make up by far the largest amount of all books ever printed (printing has exploded in the 20<sup>th</sup> century) could become the basis of an effective

digital library, one which would serve the public without infringement of copyright or straightforward challenge to the publishing industry:

“The idea is a simple one: Fifty years after a work has been published, the copyright owner must pay a \$1 maintenance [sic] fee. If the copyright owner pays the fee, then the copyright continues. If the owner fails to pay the fee, the work passes into the public domain. Based on historical precedent, we expect 98% of copyrighted works would pass into the public domain after just 50 years. They could keep Mickey for as long as Congress lets them. But we would get a public domain...

“The need for even this tiny compromise is becoming clearer each day. Stanford's library, for example, has announced a digitization project to digitize books. They have technology that can scan 1,000 pages an hour. They are chafing for the opportunity to scan books that are no longer commercially available, but that under current law remain under copyright. If this proposal passed, 98% of books just 50 years old could be scanned and posted for free on the Internet.” (Lessig, 2003)

To a certain degree, this approach resembles that of Google Books' settlement. However, Lessig's proposal, instead of centralizing access in the hands of a commercial interest, would have the funds and access controlled by a public or nonprofit agency. This would alleviate some of the concerns about proprietary control of these texts, but it is increasingly unlikely that this model would come about, unless, of course, Google could be persuaded to spin off their project under a non-profit business umbrella.

It exhibits, however, many of the same downsides. A library which contains out of print books, selected

premium titles and a great deal of out of copyright and Open Access works becomes, essentially, a glorified archive.

### **The Internet Archive**

Currently, a mirror of the Internet as a whole exists through Brewster Kahle's *Internet Archive* (<http://www.archive.org/index.php>). Some of the legal situation that faces Digital Libraries is reflected in the Internet Archive's policies:

“You agree to abide by all applicable laws and regulations, including intellectual property laws, in connection with your use of the Archive. In particular, you certify that your use of any part of the Archive's Collections will be noncommercial and will be limited to noninfringing or fair use under copyright law.”

“While we collect publicly available Internet documents, sometimes authors and publishers express a desire for their documents not to be included in the Collections (by tagging a file for robot exclusion or by contacting us or the original crawler group). If the author or publisher of some part of the Archive does not want his or her work in our Collections, then we may remove that portion of the Collections without notice.”

The Internet Archive innovates through placing the onus on copyright on the owner of the copyright, and upon the user. Essentially, if it was available publically on the Internet, it is assumed to be released, until the owner says otherwise. This model seems possibly applicable (assuming legislative change) to the vast number of “Orphan Books.”

### **Some radical suggestions**

It is my belief that the survival of the library as an institution relies on several key factors:

1. Its ability to be responsive and customer/patron-centered as well as community-oriented trumps

the essentially commercial interests of for profit information clearinghouses. This is accomplished through human interaction.

2. Libraries are free or have minimal costs. This makes them attractive to all sectors of the population, but most significantly, they can offer their services to people on the lower end of the economic spectrum. These individuals' access to information and assistance can impact their success in improving their financial, educational and personal situations.
3. Libraries endeavor to offer current and topical holdings to their patrons. Libraries alleviate the problem of access, but in modern American culture, newness is itself a strong indicator of value. This capability is strained by budgetary, space and processing costs.

The first two points serve as positive touchstones for the Library; the third is an important factor in their relevance. All three are threatened by digital access.

The first issue, however, is mainly a technical one. Online, real time reference is actually already available. Better implementation of online presences (from being able to speak to a librarian or chat with one as one looks at the catalog) and a fuller embrace of contemporaneous digital communications means that savvy libraries can continue to be the human face of information.

Crucial to the library's mission of free access is being able to negotiate licenses which can be applied across the board, or better yet, specialized and publisher-donated free access (if only available on site).

This brings us to the third point, which is the main focus of this paper. Thus far, various proposals for bridging the differences between the library's free use of physical texts and those of digital texts have avoided or sidestepped the issue of uneven access to newer works. These would be neither "orphans" nor covered by any sort of blanket license.

At this moment, it is possible to download books online via Peer to Peer networks; and if books do

migrate to a digital format in the next few decades, then this sort of access will be just as prevalent as the gray and not so gray areas of music and film downloading.

The obvious question is: how is society, and by extension, the library, to deal with these sorts of developments? As it is, if one borrows a piece of digital media from a library, you are simply on your honor not to duplicate it without permission.

To answer this, we must consider what it means to live in a society where the legal restrictions of copyright are enforceable only upon large institutions and citizens apparently selected arbitrarily by the copyright holders.

The Google settlement clearly is ineffective in serving libraries and their users. Restricting libraries to one (!) terminal with unfettered access to Google Books is useless.

There may be a partial loophole in the Copyright law. Under Sections 107 and 108 of the law, libraries and other institutions may make copies of works without permission if the copying is necessary for interlibrary loan, research purposes and for archival reasons. (Intellectual Property, 2003) However, in terms of digital works, the copying must be done by the library and thus can't be an automatic process initiated by the patron, and the work must be owned by the library. In other words, we should expect that interlibrary loans will be conducted with copy protected digital copies of books and other media in the near future (assuming the library is able to create or obtain a digital counterpart to their physical holdings).

As I see it, there is but one solution that supports libraries and their patrons in an effective way: extending the First Sale Doctrine to allow the "lending" of copies of the licensed product with robust copy protection or in a read only format. But—you may ask, these measures are useless against a skilled hacker... well, photocopying a whole book or stealing it are loopholes that are exploited by a

few library users; but only by a rare few. Most users will have little reason to reduplicate their items; and one should expect that films and other media will be available in pirated versions in the greyer areas of the Internet. The practical limits of hard drives will keep patrons from downloading a whole collection.

Does this mean that libraries will be forced to bear the brunt of fees? Perhaps. Residency and a library card would still be required; and portability from the computer to another device would be restricted. Thus there will be numerous customers for whom a digital copy obtained directly from the publisher or bookstore would be superior.

One thing is certain, however. Whether libraries play a role in this or not, the truth is, digital texts of popular interest are almost certain to become pirated. Publishers of bestsellers can expect that their position will resemble that of music publishers today: fighting a rear-guard and unpopular battle.

Bibliography

- Albanese, Andrew (2008). "Harvard Slams Google Settlement; Others React with Caution." *Library Journal*, 30 Oct 2008. Retrieved on 2 Nov 2008 from <http://www.libraryjournal.com/article/CA6610115.html?desc=topstory>.
- Intellectual Property (pseud.) (2003). "Copyright in the Library – The Digital Library." In *Crash Course on Copyright*, retrieved on 2 Nov 2008 from <http://www.utsystem.edu/ogc/intellectualproperty/1-diglib.htm>.
- Lardinois, Frederic (2008). "The End of Snippet View: Google settles Lawsuit with Book Publishers." *New York Times*, 28 Oct 2008.
- Lessig, Lawrence (2003). "We Need Your Help." *Lessig Blog*, May 2003. Retrieved on 25 Oct 2008 from [http://lessig.org/blog/2003/05/we\\_need\\_your\\_help.html](http://lessig.org/blog/2003/05/we_need_your_help.html).
- Lessig, Lawrence (2008). "On the Google Book Search agreement." *Lessig Blog*, 29 Oct 2008. Retrieved on 29 Oct 2008 from [http://lessig.org/blog/2008/10/on\\_the\\_google\\_book\\_search\\_agre.html](http://lessig.org/blog/2008/10/on_the_google_book_search_agre.html).
- Roush, Wade (2008). "In Google Book Search Settlement, Readers Lose." *XEconomy*, 31 Oct 2008. Retrieved on 2 Nov 2008 from <http://www.xconomy.com/national/2008/10/31/in-google-book-search-settlement-readers-lose/>.