



Faculty Publications

2005-12-11

Google and the Print Library Initiative: To Scan or Not to Scan

Peter A. Zuber
peter_zuber@byu.edu

Follow this and additional works at: <https://scholarsarchive.byu.edu/facpub>



Part of the [Library and Information Science Commons](#)

BYU ScholarsArchive Citation

Zuber, Peter A., "Google and the Print Library Initiative: To Scan or Not to Scan" (2005). *Faculty Publications*. 333.

<https://scholarsarchive.byu.edu/facpub/333>

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.

**Google and the Print Library Initiative:
To Scan or Not to Scan**

December 11, 2005

Peter A. Zuber

Introduction

On December 14, 2004, Google co-presidents Larry Page and Sergey Brin, Stanford Ph.D. students who founded the company in 1998, announced their intention to “organize the world’s information”¹ by effectively digitizing the content of the world’s books in a massive scanning project as part of the Google Print program. Enormous in scope, the initial content was made accessible through a cooperative agreement between Google and the libraries of Harvard, Stanford, the University of Michigan, and the University of Oxford in England, as well as The New York Public Library. The effort would include both public domain, out of copyright works, as well as all copyrighted works. Through harvesting efforts of its catalog records by the University of Michigan along with possible assistance of a digital registry through the OCLC,² the digitization process would be organized to avoid redundancy between collections.

Given the massive volume count in each collection, it is difficult to estimate how long the project will take to complete. Where Google has stated it anticipates completion in 6 years,³ it would take a Herculean effort to make that date. For example, the University of Michigan’s Library contains over 7 million volumes. Assuming a digitization rate of approximately 5000 volumes a year (19 volumes a day, 260 workdays a year), the effort would last 1400 years. To scan just the University of Michigan’s content within Google’s 6-year window; the rate would have to be 1,166,666 volumes a year, or 4,487 volumes a day.

Regardless of the practicality of the stated goal, the announcement nonetheless created an enormous stir among the literary industry ranging from authors to publishers to librarians. Already a strong presence on the web, Google has been of particular interest

Google and the Print Library Initiative

to those in the library industry, raising concerns over future roles that libraries and librarians would play in a Google-dominated, information-seeking world. Noted not only for its search capability, Google had previously demonstrated its desire to expand its role as an information provider with such tools as Google Earth (a remarkable 3-dimensional viewing tool of the planet), Google Scholar (a search tool designed to return only peer-reviewed scholarly content) and Google News and Images (capable of searching thousands of news stories and retrieving a myriad of images). Now the “Google Print Library Program” was striking, at least seemingly, at the core of a library’s offerings, its book collection.

Indeed, some of the verbiage used during and following the announcement didn’t promote anticipated excitement but, rather, concern among those in these industries. Larry Page voiced, “Even before we started Google, we dreamed of making the incredible breath of information that librarians so lovingly organize searchable online. Today we’re pleased to announce this program to digitize the collections of these amazing libraries so that every Google user can search them instantly.”⁴ Adding fuel to a reactionary viewpoint, John Wilkin, a University of Michigan librarian working with Google, was quoted as saying, “This is the day the world changes. It will be disruptive because some people will worry that this is the beginning of the end of libraries.”⁵ Barbara Quint reported, “when asked whether Google is building the library to replace all other libraries, Google representatives--after saluting the role of librarians--said they had ‘no such plans *at the moment.*’ ”⁶ These and other inflammatory comments, along with a careful reading and review of the Google proposal, has created a sense of perceived threat

Google and the Print Library Initiative

among some, while others have openly lauded the altruistic vision of this web giant.

Some of the more relevant issues brought to bear include the following:

- Copyright violation
- Reduction in library patronage
- Forces a paradigm shift in the concept of libraries
- A death knell to libraries

Considering each of these issues will provide some insight into the potential effects of this initiative on the library industry.

Copyright Violation

In the months following the announcement, publishers such as Random House and Houghton-Mifflin began to express concern about the prospect of scanning copyrighted work. The AAUP (the Association of American University Publishers) sent its now famous letter saying it felt “alarm and concern at a plan that appears to involve systematic infringement of copyright on a massive scale.”⁷ Since that time, Google has been filed with a lawsuit. The issue centers on copyright notices in almost every book with statements to the effect that any reproduction of this book without express consent of the publisher is prohibited. Where Google does not plan to provide the full text of copyright works online (“snippets” only, 3 page excerpts), it nonetheless states its intent to scan the entire work, providing copies to both Google and the library furnishing the volume. Google explains that full text versions are necessary to enable effective search capability. University of Michigan’s Wilkins states that their digital versions will be kept

Google and the Print Library Initiative

in a “dark” (non-visible) archive.⁸ Regardless of these statements depicting seemingly innocuous usage, the issue remains focused on the limited use the copyright statement allows. In its strictest sense, no one, not even a book owner, is allowed to copy or scan a book under copyright law. To this mixture of viewpoints is added a conflict between traditional print copyright law and common Internet practice, namely, the practice of web search engines scanning all online content to generate search databases without regard to infringing on the original content creator’s rights or copyright violation. This conflict creates a truly difficult landscape of precedent for a truly successful implementation. Google’s statement that non-conforming publishers can chose to “opt-out” of the program generated greater anti-Google sentiment among publishers as, according to Schroeder, executive of the Association of American Publishers, “Google’s procedure shifts the responsibility for preventing infringement to the copyright owner rather than the user, turning every principle of copyright law on its ear.”⁹ However the outcome, the role of several notable world libraries as supporters of Google’s efforts has the potential to be of harm to the library industry. Poor relations may be engendered between publisher and library depending on library association statements. In addition, there is a possibility that libraries, owners of the copyright works who allow their digitization, could be named in a lawsuit. The issue of what constitutes “fair use” will be critical in determining what and what cannot be done legally. Where the law seems murky in regards to copyright use, Google insists that “essentially creating a card catalog”¹⁰ is fair use. The fact that Google is a for-profit company makes that statement problematic.

Reduction in Library Patronage

Clearly a concern when the Internet began to become more widespread, the issue of reduced patronage with the ability to retrieve content on-line in seconds instead of driving to the nearest library continues today, invigorated most recently by the promise of all books available on-line. Much to the contrary, however, statistics show that the opposite proves true. Carol Brey-Casino, immediate past president of the American Library Association is quoted as saying, “We had this conversation when the Internet began to get popular, and what’s happened is that library visits have doubled in the last decade to 1.2 billion.”¹¹ As part of the explanation, one could consider the ease of retrieval and quick provision of limited information as potent advertising, whetting the appetite for the more satisfying act of reading the content in full at the library or after purchasing it locally. In fact, in one of the more humorous exchanges in the Google book battle, Allan Adler, vice president for legal affairs at the Association of American Publishers, spoke in response to Google’s general counsel, David Drummond’s contention that Google Print would increase book sales. Adler stated, “When people make inquiries using Google’s search engine and they come up with references to books, they are just as likely to come to this fine institution (referring to the public library) as they are to buy them.” To which Drummond replied, “Horrors.”¹²

Forces a Paradigm Shift in the Concept of Libraries

If Google were to successfully digitize all the world’s content and make it available on-line, the argument has been made that the traditional role of libraries would shift and that libraries would be forced to remake themselves in order to provide value. Indeed, even today, libraries are finding patrons are more search savvy, having run

Google and the Print Library Initiative

several Internet search attempts prior to coming into the library or asking for assistance. This has, to the benefit of the industry, drawn attention to the role of the librarian. The consulting firm Outsell speaking on the shift said, “It’s another shove to get librarians out from behind the stacks and harness their expertise, including subject-matter expertise, and to enhance users’ ability to find, use, and access information in any format. Getting out of the business of simply storing books should be a welcome goal.”¹³ An ongoing message coming from the industry is that each time something like this (i.e., internet, prior digitization efforts, etc.) happens, the role of librarians is better defined, improved and directed toward more of what they are meant to do, rather than what they have been doing. Marjorie Hiava of Access Innovations pointed out one practical benefit and potential shift in library emphasis, “It costs \$200 a square foot to maintain a library collection. If I had 132 miles of shelf space and someone offered to digitize half of it, I’d be real interested.”¹⁴ The potential to re-purpose funds and to focus more on special collections would further distinguish libraries and add enormous value.

A Death Knell to Libraries

The growth of Google’s popular search engine along with innovative content tools such as Earth, Scholar, News and Images may have been considered somewhat redundant, especially when compared to current library content or Internet offerings. However, the Google Print Library initiative seems to be striking a blow at the library’s most prized content, its book collection. Long considered difficult to provide online, several digitization efforts have previously addressed this goal and currently provide full text content, such as Project Gutenberg, the Million Book Project, the Library of Congress, netLibrary, ebrary, not to mention numerous digitization projects at many

Google and the Print Library Initiative

academic and other libraries. However, all of these efforts are small in comparison to what Google envisions. Where most of these mentioned efforts number their content in the tens of thousands of books, Google sees their collection being in the millions of volumes. Naturally, the question is raised: does this signal an end to traditional brick and mortar libraries? Is the information age leaving the past content provider behind for the new electronic landscape? How will the industry be able to withstand the challenge of instant access, scholarly content, wide distribution of content and the potential of future, full text resources of all known books? These questions can create highly speculative scenarios of libraries (and consequently, librarians) being under the proverbial gun. Clearly, an acceptance of what is happening and an awareness of how to use it is considered more helpful than not. On the willingness to assist students who wish to use the Google search interface, some librarians may be concerned about the message they are sending. Quite the opposite, industry analysts suggest, it presents perfect opportunities to educate and apply information-science principles to the use of such search tools and their advantages and limitations.

There is also a lesson to be learned from the Internet “dot com” business boom a few years ago. The need for personal assistance, hands on evaluation, local and dependable support and long-term business presence made the exorbitant visions of pure “dot com” business models appear silly after a few years. In that same sense, libraries have nothing to fear from Google or any other book digitization effort. Our own fears and lack of knowledge may be what sets limits on the future. Mary Case, library director at the University of Illinois at Chicago, spoke to this fear: “If we dig in our heels, we’ll just look stupid. It’s coming. We must use it.”¹⁵ Michael Gorman, current president of ALA

Google and the Print Library Initiative

says it is “premature to prepare to mourn the death of libraries and the death of the book... This latest version of Google hype will no doubt join taking personal commuter helicopters to work and carrying the Library of Congress in a briefcase on microfilm as ‘back to the future’ failures, for the simple reason that they were solutions in search of a problem.”¹⁶ Where the notion of personal helicopter implementation for commuter traffic is appealing on many levels, it still remains unattainable for practical reasons. Gorman, it would appear, feels that the same grand vision of scanning all the world’s knowledge may follow the same flight path of the commuter helicopter; having nothing favorable to say about Google, he goes so far as calling the co-presidents (we assume) as “the boogie-woogie-Google boys.”¹⁷ Indeed, others have seen this attempt or effort as not the end but the beginning of an opportunity to enhance librarianship. Wilkins has said “...this is something we have to do to revitalize the profession and make it more meaningful.”¹⁸

Summary

One issue raised is whether the Google Print Library initiative is being portrayed as another “Napster,” an ideal that generated enormous attention only to fail in the courts. The analogy is certainly useful, for in its efforts to devise a new business model for the music industry, Napster fell short and very hard, seeing its concept, and more importantly, its value transferred to another business concern (read iTunes). This second adopter then created a model that was satisfactory to all concerned. The message seems clear and instructive, especially to the library industry. The issue of digitizing books and content does not seem as much at question as the when and how much will be done? The effort began years ago when the technology became available; it is the scope of Google’s vision that has generated concern of late.

Google and the Print Library Initiative

Libraries are allowed to store and loan content and materials for public use partly because material is paid for by the library at higher than usual rates, because someone who reads an author is more likely to purchase in the future, because marketing has shown library use as a viable advertising mechanism, and because, quite simply, of the difficulty experienced when reproducing printed content. This is one of many reasons the five libraries that work with Google are involved; they get free digital versions of their holdings. However, once digitized, a book becomes a different type of commodity as far as the publishers are concerned. A digital book can be “reprinted” or copied in a matter of seconds at almost no cost and sent to many users all over the world. This alone will continue to feed the battle between publishers and Google. But regardless of outcome, it is reasonable to expect (and responsible to prepare for) digitization of more and more book content.

The method and timeliness of delivery makes digitization a very important consideration regarding future library content, management, and function. If the model continues to be one where copyrighted works are shown as snippets only, it stands to reason library patronage would most likely increase. If this began to show an impact on the publisher’s bottom line, libraries may find themselves paying more and more for collection development. If full text is granted on a cost per page or other monetary basis, it will be an interesting test of everyman’s native preference of paying for on screen or paying for printed form. Regardless, the library has shown in its past a strong resilience to similar challenges and, in fact, as mentioned previously has grown from the opportunity. Library redirects like OpenWorldCat, the use of library web pages for collection content, and most importantly, the ability to improve the quality of patron

Google and the Print Library Initiative

questions are all examples of effective librarianship in the age of the Internet. Stephen Abram, in his recent presentation to library professionals on the Google initiative said, “Google and other search engines answer Who, What, When, and Where questions. Librarians answer How and Why questions. That is our expertise, that is where we excel.”¹⁹ Clearly, the How and Why questions will always remain. Perhaps the future is best described by Mary Ellen Bates of Bates Information Services in the title of her recent work, “You still Google? That is so last week.”²⁰

REFERENCES

- 1) Google, Inc., Mountain View, CA. (2004, December 14). *Google Checks Out Library Books*. Retrieved December 5, 2005 from http://www.google.com/intl/en/press/pressrel/print_library.html
- 2) Quint, B. (2004, December). Google's Library Project: Questions, Questions, Questions. *Information Today*. Retrieved December 9, 2005 from <http://www.infotoday.com/newsbreaks/nb041227-2.shtml>
- 3) Ibid.
- 4) Google, Inc. *Google Checks out Library Books*.
- 5) Quint, B. *Google's Library Project: Questions, Questions, Questions*.
- 6) Ibid.
- 7) B. Helm. (2005, May). A Google Project Pains Publishers. *Business Week*. Retrieved December 5, 2005 from <http://www.businessweek.com>
- 8) Helm, B. (2005, June). A New Page in Google's Books Fight. *Business Week*. Retrieved December 5, 2005 from <http://www.businessweek.com>
- 9) Association of American Publishers, New York. (2005, August). *Google Library Project Raises Serious Questions For Publishers and Authors*. Retrieved December 5, 2005 from <http://www.publishers.org/press/releases.cfm?PressReleaseArticleID=274>
- 10) Helm, B. (2005, September). For Google, Another Stormy Chapter. *Business Week*. Retrieved December 5, 2005 from <http://www.businessweek.com>
- 11) Quint, B. *Google's Library Project: Questions, Questions, Questions*.

Google and the Print Library Initiative

- 12) Wyatt, E. (2005, November). Googling Literature: The Debate Goes Public. *The New York Times*. Retrieved December 5, 2005 from <http://www.nytimes.com>
- 13) Quint, B. *Google's Library Project: Questions, Questions, Questions*.
- 14) Ibid.
- 15) Ibid.
- 16) Ibid.
- 17) Ibid.
- 18) Ibid.
- 19) Abram, S. (2005, November). *Google Inc. and Libraries. Our Common Future*. Retrieved November 17, 2005 from <http://dynix.com/institute>
- 20) Bates, M. E. (2005). *You Still Google? That Is So Last Week*. Retrieved November 11, 2005 from <http://www.batesinfo.com/index.html>