

## Google Book Search and the Future of Libraries

**Abstract:** Google Book Search (GBS) is a service provided by Google that allows for the searching of the full text of millions of books from some of the world's most renowned library collections. Two distinct view points are found amongst Information Management (IM) professionals regarding GBS: those who support the Google Books' Search project as an opportunity for libraries and alternately, those who believe GBS will ultimately lead to the demise of libraries and the way individuals retrieve and process information. Information professionals objectively weigh the pros and cons of the GBS Project: being aware of the privacy concerns of the project, the problematic metadata of the search engine and the problematic dependency which it may cause. Contrarily, the pleasing of patrons in libraries, and digitization happening through the project must not be overlooked. Library professionals must determine the best method of bringing GBS into library environments, and this can be done effectively by considering context, concerns and benefits.

### About the Author(s):

**Melanie Parlette**, MLIS Candidate (2010) holds a BA (Hons) in Political Studies and History from Trent University. Melanie currently works for Dalhousie's Killam Library as a Reference and Technical Services Intern and is completing her thesis which looks at the pleasure reading habits and experiences of undergraduate students. In addition, she is the blogger behind the [Dal Reads! Blog](#)

**Leanne Owen**, MLIS Candidate (2010) holds a BA (Hons) in Criminology and English from York University. She is working on an independent study course concerning how libraries serve inmates in North American prisons. She plans to pursue a career as a legal librarian or to work with children in public libraries.

**Jessica Babineau** holds a BA in Anthropology and Classical Studies from Mount Allison University. An MLIS Candidate (2010) in the School of Information Management at Dalhousie University, she is co-chair of the Dalhousie student chapter of the Canadian Library Association. Jessica is currently working for Dalhousie University's Killam Library as an intern, both with the Reference department and for the Map and Geospatial Information Collection.

The contents of this paper are based on the authors' research and papers written for INFO 5590: Information Management Systems, at Dalhousie University's School of Information Management.

## **An Introduction to Google Book Search**

Google Books can be understood under three main branches: the Google Books Partner program, the Library Project, and Google Book Search (GBS). The Google Books Partner program (Google, "Google Books Settlement Agreement", 2009) has partnered with more than 20,000 publishers and authors to give users access to their books through the Google and GBS. The Library Project (Google, "Google Books Settlement Agreement", 2009) involves a partnership between Google and libraries around the world to scan millions of books. The final branch, GBS, is a service provided by Google that allows the search for the full text of millions of books from some of the world's most renowned library collections.

Google began scanning the University of Michigan's library collection in July 2004 and since then has developed partnerships with twenty major libraries worldwide in hopes of reaching its goal of digitizing 15 million books by 2010 and 32 million books in ten years (Crow, 2007). The digitization process involves the scanning and converting of texts using optical character recognition, and then storing them in a digital database. Today, over seven million full-text books are searchable through Google Books. Google will not actually disclose the official number of titles in the database.

GBS is one of the most contentious topics surrounding the information management (IM) profession today. The project has the potential to drastically alter the way individuals read and access books. However, matters such as incorrect metadata, poor cataloguing, copyright issues, and the finance driven direction of the of the GBS system is a major concern for librarians. For others, both in and outside the IM profession, the GBS project is perceived as an extremely attractive opportunity since the initiative is a way of preserving and opening up collections when libraries could not digitize their own materials because of finances. Despite the great deal of support that exists in favour of the project, libraries are undeniably facing an uneasy position. Roush (2005) put it succinctly when he stated, "[l]etting a for-profit organization like Google mediate access to library books, after all, could either open up long-hidden reserves of human wisdom or constitute the first step toward the privatization of the

world's literary heritage" (p. 55). As Albanese (2009) explains, "libraries will be the key beneficiaries of the settlement, but they also recognize the enormous risk that the settlement could pose to access going forward" (p. 1).

Because GBS is still in its beginning stages, a great deal of uncertainty remains concerning the project and therefore a degree of caution on the part of IM professionals is expected. In order to understand the controversies surrounding the GBS project one must examine the issues which have led to the contrasting perspectives held by IM professionals in both public and academic library settings. The Canadian experience, while unique from the American in terms of privacy, holds many of the same concerns. Canada is included in the revised GBS settlement as were other countries that were deemed to hold a similar legal structure and practices. However, it remains to be determined if the Office of the Privacy Commissioner of Canada will specifically address any of the privacy concerns they may have with GBS.

## **The Strengths and Weaknesses of Google Books Search**

Currently the GBS project is centered on two distinct view points: those who support the project as an opportunity for libraries and IM professionals and alternately, those who believe GBS will ultimately lead to the demise of libraries and the way individuals retrieve and process information. Potential solutions may be identified by examining the strengths and weaknesses of the project.

### **Privacy Concerns**

A major criticism of the GBS project focuses on privacy. This issue has recently been propelled to the forefront of the project, with the assistance of the Electronic Frontier Foundation, the ACLU of Northern California, and the Samuelson Clinic at the University of California at Berkeley, through the filing of an objection to the Google Books settlement on September 8, 2009 (Electronic Frontier Foundation, "National Coalition," 2009). Google currently provides an overarching privacy policy along with a supplemental document explaining issues specific to Google Books. These documents may continue to be modified until a final settlement is reached. On November 19, 2009, preliminary approval of the Amended Settlement was achieved. A Fairness Hearing for this Amended Settlement has not yet been scheduled (Google Book Settlement, 2009).

Because Google notoriously monitors and tracks the reading habits of users, there are concerns that the confidentiality of the user could become compromised. It is common knowledge that Google collects and retains large amounts of personal information on each individual that uses its services. Anyone with a Google toolbar on their personal computer is effectively sending Google a record of every website they visit regardless of whether they use the toolbar (Waller, 2009). GBS will be no exception. However, some librarians regard this as an opportunity rather than an invasion of privacy because it enables personalized searches.

This essentially means that the search engine “remembers” a user’s previous Internet activities, and as a result, Google is able to provide more relevant searches in theory. As a result, Einsenberg (2008) explains how “some librarians have welcomed the prospect of such personalized search as a way of improving search for the user” (as cited in Waller, 2009, p. 6).

Lesk & MacKle-Mason (2009) explain that most users will be tracked at some point when they use a Kindle or Google Book Search. By tracking their users, they might become aware of this through the advertising that they receive or readers advisory. While organizations such as the Electronic Frontier Foundation suggest that users might be discouraged if they thought someone could track what they read, Lesk & MacKle-Mason (2009) contradict this through studies conducted by the Electronic Privacy Information Center. These studies demonstrate while individuals might assert that they are highly concerned about their privacy on the Internet, they ultimately do little to tackle the issue.

The American Library Association (ALA) highlights that, in terms of privacy (Terry, 2009), they are concerned about the confidentiality of users of the Google Books service. The ability to track and monitor reader habits is of concern. The policies of libraries may become subject to the Google approach to managing privacy and other personal information. Libraries are under constant pressure to increase the availability of technology. They may not fully understand the ramifications of what they are implementing (Fifarek, 2002). The type of policy that a library chooses to implement regarding privacy might not solely be the choice of the library. The ability of libraries to protect the privacy of their patrons would be subject to the policies and procedures of Google. Quint (2009) assures readers that the Privacy Policy of Google is one of the best in the business, comparing it to that of Amazon and Microsoft. Quint also states that nothing short of court orders would be likely to encourage Google to change its policies about privacy.

Dan Clancy, the Engineering Director of Google Books, responds to questions from a variety of stakeholders through the Google blog on public policy. In a July 23, 2009 post, he explains that at this point it is difficult, without an approved settlement from the courts, to determine a detailed policy on privacy. An updated post by Jane Horvath on September 3, 2009 following the settlement highlights the release of an official Google Books Privacy Policy. They also share that they are engaging with the Federal Trade Commission on their Privacy Policy. The frequently asked questions section on privacy provides further information on the policy, highlighting that they will not be providing data about individual users unless specifically asked to under certain circumstances. This highlights that Google will not sell the data of its users. It further explains that in the case of academic institutions Google will authenticate the subscription through the IP address and uses the example of Shibboleth, a way of securely identifying members of community over the Internet. The post further emphasizes that in the case of public libraries and public access terminals, no information will be gathered that could

identify a user, even in cases where they are required to log on to a computer.

Google Books also makes public their Privacy Policy. This document refers to an overarching document of Google Privacy but highlights some aspects that are specifically of interest to those who use Google Books. This highlights the provisions of the policy including when user information may be provided (Google cites such circumstances as “emergencies or in response to valid legal process”) (“Google Books Privacy Policy, 2009). In addition, Google (2009) dictates that certain information such as log information such as IP address and date and time of request along with cookies that might identify the browser one is using.

Google Books products such as the My Library, allow for the creation of a public profile that lists a user’s favourite books. Features such as the IP address of a users and cookies are used to abide by the contracts that Google shares with those who hold the rights to the books they are making available. Schools and other organizations that acquire an institutional subscription will not have to require users to create Google accounts; they will instead be able to be authenticated through other methods. It is necessary to have a Google account to purchase books. Google highlights access to their Privacy Center and provides further opportunities for engagement through other avenues such as Twitter.

To support the provision of privacy to library patrons, IM professionals must determine their next steps and re-evaluate the GBS project and how it meets their individual and collective needs on a continual basis. This may be determined by outlining a review process for the privacy policy of the library and the education of users utilizing GBS. GBS as a library tool may evolve as Google develops the service. The privacy needs and considerations of users are yet to be fully addressed in a complete document, as the settlement is not yet final. Therefore, libraries and users must be informed of the transitional nature of the service. Education, policies, and ‘procedures must be consistently evaluated to ensure currency and to provide users with consistent service which provides them the utmost privacy.

## **Academic Institutions and Google Books**

According to Grogg and Ashmore (2007) the libraries participating in Google’s digitization project represent an effort to achieve two key goals: to preserve materials for future generations and to provide increased access and functionality for the current generation. Google believes these goals to be achievable pending a favourable ending to the settlement. Albanese (2008) outlines the six central benefits that academic institutions will gain from GBS: 1) free preview and the ability to find a book at a local library or to purchase it, 2) creation of a database of public domain works which scholars can use to conduct research, 3) institutional subscriptions providing access to in-copyright, out-of-print books, 4) accommodated services

for persons with print disabilities, 5) digital copies of works digitized by Google provided to its partner libraries for long-term preservation periods, and 6) free full-text access at public libraries at a single designated terminal (p.1). However, due to the scale and scope of the GBS project, IM professionals and scholars are only just beginning to consider some of the negative developments this project may bring to their institutions. These issues range from technical and organizational problems to the development of information scaffolds for users.

Joint (2009) comments on GBS in academic institutions from a technical perspective, claiming that Google's "[...] ultimate goal is to work with publishers and libraries to create a comprehensive, searchable, virtual card catalogue of all books in all languages that helps users discover new books and publishers discover new readers", yet their site is riddled with metadata errors (p.337). These technical issues create a catalogue of information that is not properly searchable. GBS is developed to be a primarily implicit search-oriented system, thus the explicit elements of searching that are available through OPAC metadata are not available in GBS.

The issue of metadata in GBS is recurrent in literature. One strong voice is that of Nunberg (2009), who discusses how the GBS metadata are problematic to users, the high level of errors can be seen by searching for books, for example "Barack Obama", where many books dated back to before 1900. While Nunberg's piece is focused on the needs of scholars, it can be noted that these issues regarding the organizational aspects of GBS are also a concern for librarians.

Moving beyond the technical and organizational aspects of GBS, there are a number of cultural biases surrounding the project. Jeanneney (2007) and a number of other European politicians and intellectuals are criticizing the project on "language-imperialism" grounds, arguing that because nearly all of the books being scanned are in written in English, it will result in disproportionate representation of natural languages in the digital world. For example, German, Russian, and French are prominent languages in scholarship and critics like Jeanneney fear the disproportionate online emphasis on English could shape access to historical scholarship, and, ultimately, the growth and direction of future scholarship (2007). Google has attempted to shape perceptions of language bias in the selection of works by expanding its partnerships to non-English speaking libraries in Switzerland, Japan, Belgium and Spain as well as a German State Library and the National Library of Catalonia (Waller, 2009). Despite Google's attempts, criticism has yet to subside. Jeanneney (2007) continues to profess that GBS has the potential to adulterate the world's cultural heritage as a result of its bias in favour of English publications. If users see GBS as a one-stop resource for academic research, they may be developing special and temporal biases in their research work. Jeanneney (2007) points out the commercial nature of Google's undertaking, along with the

issues and conflicts of interest inherent to such a revenue-focused approach to highly valued international cultural works. Furthermore, he questions the notion of a private corporation, like Google, leading an effort that he believes should be left in the hands of the public at large.

GBS alters the information found by users twofold. Users flock to Google, and GBS, as it facilitates rapid and effortless searching for complex questions, but this may be resulting in a problematic dependency. GBS creates an elevated bar for information which Google deems as sufficient (Barbazon, 2007), yet there is a wealth of information beneath that is being ignored. It is ignored not only because GBS is not all encompassing, but because it is not well organized. This results in a loss of knowledge and the inability of users to think critically and holistically. For example, assume that GBS has been integrated into an academic library system without further effort from librarians or staff to proactively introduce it to users. This small change in the dynamic of the library system will have serious repercussions: users will develop a great dependency on this one element of the library system, creating a vast shift in usage dynamics. While change is expected, this dramatic modification in the structure of usage of the library resources can be expected in light of the discussion of Brabazon (2007), meaning that users will develop a greater dependency on GBS. This change should not be welcomed in terms of academic study; instead users should be aware of the multitude of resources available, rather than a select few. Through this discussion, it is clear that the integration of GBS into an academic library system by taking a piecemeal approach would be ineffective in creating an effective learning process for users. Waldman (2007) stated that "Thinking systems include humans, and humans always learn. They may learn the wrong things or lessons that the organization prefers that they should not learn, but [that] humans by their nature learn" (p. 273).

According to Grafton (2009), the problems identified with GBS create a danger for libraries. The change in user research strategies can leave the physical space of the library forgotten, and while this is a problem for libraries, it is also a problem for users, whether they are aware of it or not. Libraries must consider a means of creating an environment that is both digital and physical in order to meet the wants and needs of users. Thus, the previous issues are no reason for libraries to disregard GBS as a library and/or scholarly resource and tool. When considering implementing GBS in an academic setting, the future of the library must be kept in mind. IM professionals and users should be aware that Google is not all encompassing and does not reach the extent of information and organization that a library's collection is able to carry, while also considering the metadata and organizational issues that need to be resolved with the GBS project.

## **The Potential Benefits for Libraries and IM Professionals**

GBS poses a great deal of potential benefits for libraries and IM professionals. MacColl (2006) asserts that librarians should recognize GBS as an opportunity since it seems apparent that Google is doing libraries a major service by digitizing books that are out of copyright and need

to be preserved (p.1). Digitizing the collections of some of the world's most renowned libraries also means giving patrons what they want. According to Grafton (2009) university libraries find their clientele are happy to use virtual books explaining that for students today, "the primary source of information on life, the universe, and everything is the Web, and they normally seek information not by making a research plan but by entering words in a search engine—usually a non-specialist one like Google" (p. 94). Grafton (2009) believes that rather than competing, librarians should assist patrons in conducting more effective Google searches. For example, in Australia, a number of academic libraries run 'GoogleSmart' classes to help students use the search engine effectively. Similarly, a number of librarians advocate welcoming Google into the reference interview, not only to take advantage of Google's search function but also to capitalize on the relationship that patrons already have with Google (Grafton, 2009). Evidently Google has a lot to offer libraries and librarians need to be realistic—patrons will want to use GBS and in order to stay current, librarians need to find a way to use this to their advantage. In addition to pleasing patrons, librarians will enjoy some savings in purchasing prices since paying for one all-encompassing subscription to GBS will save on binding, shelving and preservation costs (Grafton, 2009). Furthermore, having digitized items also means that libraries will eventually no longer need to house large collections of physical books, freeing up physical space.

## **Debunking the Longstanding Fear of Libraries' Demise**

The literature demonstrates that there is a longstanding fear that Google will eventually replace libraries. Roush (2005) discusses these fears, stating that when people are able to retrieve books online from their homes, offices, or dorms, "libraries could turn into lonely caverns inhabited mainly by preservationists. Checking out a library book could become as anachronistic as using a pay phone..." (p.56). Similarly, Eterović and Šrndić (2009) point out that many users find traditional libraries to be outdated while the modern web-based libraries like GBS become more popular, however, they also believe that if libraries want to avoid this fate, they need to "evolve and enhance their services in order to fulfill the requests of modern users that are used to the comfort of computer-based information access" (p.43). Roush (2005) also reassures librarians that "the flood of new digital material will make job classifying, cataloguing, and guiding readers to the right texts" (p.56) will keep them busier than ever.

IM professionals need to remind themselves that the GBS project could not exist without libraries. Librarians are the ones providing Google with the books they are scanning; they are the institutions that house millions of out-of-print materials, often with public funding, long after publishers and authors have walked away from them (Oder, 2009). Once the digitization project is complete, libraries will play a major role in the future of GBS since the likely demand among academic libraries for an institutional subscription is high. Finally, Oder (2009) predicts, "faculty and students doing serious research will insist on the ability to search and read the full



text of out-of-print books. This means that libraries are likely to be among the fee-paying users of the services enabled by the settlement.

## **Conclusion**

Although there is no standardized checklist which can be produced, the following recommendations can support the appropriate development for integration of GBS into an IM setting. The needs of users and the education that will be necessary for all those that may participate and engage with the service must consistently be addressed through careful analysis of the situation at hand. Libraries must carefully consider the context of their own staff and users as it is necessary to understand the concerns of their users and staffs to best understand how to address them. IM professionals must determine the best method of bringing GBS to their library and users, and must determine how they wish to educate staff and users on Google Books and surrounding privacy issues. Appropriate policies and procedures must be developed and continuously evaluated and revised so that current and accurate information is disseminated. Consistently evaluating the service and the method of implementation and education also will ensure that users are educated and able to use the service effectively and to their benefit.

Finally, while GBS has been considered the one primary resource for digitization and a source not to be ignored by IM professionals and libraries, Google is not the only corporation with ambitious digitization plans. This means that those libraries reluctant to develop partnerships with Google will also have the possibility of considering other digitization services. GBS has simply been singled out because it is the most well-known, the most talked about in the literature and media, due to the enormity of the project's reach and ambition (Waller, 2009). Competitors like Amazon, who intend to offer every book ever printed in every language on the Kindle Reader, and Yahoo, who is hosting the Open Content Alliance which was formed by Microsoft and the Alfred P. Sloan Foundation in 2005, may be viable alternatives to GBS. Harper Collins also announced they would scan twenty thousand of their books and will provide free access to these materials through all search engines. Finally, on a much larger scale, the Million Book Project funded by the Governments of China and India in conjunction with the Library of Alexandria (Anon, 2008).

In summation, "The stakes are high, both for Google and for the library community—and the technologies and business agreements framed now could determine how people use libraries for decades to come" (Roush, 2005, p. 56). The relationship between IM professionals and GBS is an evolving process, and because GBS is still in the early stages and the project has suffered setbacks as a result of the ongoing legal proceedings, there is still a great deal of uncertainty concerning the project. Overall it seems that while eagerly awaiting these final legal proceedings, information professionals should continue to objectively weigh the pros and cons of the GBS Project, recognizing the benefits it can potentially have on the future of libraries

and the IM profession while also being careful not to lose sight of the longstanding values and traditions of libraries that need to be upheld.

## References

Albanese, A. (2008, November 11). One for all? As Google deal is evaluated, critics question single library terminal. *Library Journal*. Retrieved from <http://www.libraryjournal.com/article/CA6613723.html>

Albanese, A. (2009, February 17). Library leaders meet to discuss Google book search deal. *Library Journal*. Retrieved from <http://www.libraryjournal.com/article/CA6638100.html>

Anon. (2008, 13 October). Search engines as publishers. *E-world*. Retrieved from Businessline.

Brabazon, T. (2007). *The university of Google: Education in the (post)information age*. Great Britain: Ashgate Publishing.

Clancy, D. (2009, July 23). The Google Books settlement and privacy. Message posted to <http://googlepublicpolicy.blogspot.com>

Crow, D. (2007, March 17). Web giants scramble to spread the printed word on the internet. *The Business*. Retrieved from: <http://proquest.umi.com/pqdlink?did=1243160791&sid=1&Fmt=3&clientId=15814&RQT=309&VName=PQD>

Electronic Frontier Foundation. (2009, July). *Google Book search settlement and reader privacy*. Retrieved from <http://www.eff.org/issues/privacy/google-book-search-settlement>

Eterović, T. & Šrندیć, N. (2009). Introducing the unified e-book format and a hybrid library 2.0 application models based on it. *Преглед НЦД*, 14, 43-52. Retrieved from: <http://www.ncd.matf.bg.ac.yu/casopis/14/NCD14043.pdf>

Fifarek, A. (2002). Technology and privacy in the academic library. *Online Information Review*, 26(6), 366-374. doi: [10.1108/14684520210452691](https://doi.org/10.1108/14684520210452691)

- Google. (2009). *Google Books Settlement Agreement*. Retrieved from <http://books.google.com/googlebooks/agreement/>
- Google. (2009, September 3). *Google Books Privacy Policy*. Retrieved from <http://books.google.com/googlebooks/privacy.html>
- Google Book Settlement. (2009). FAQs. Retrieved from <http://www.googlebooksettlement.com/help/bin/answer.py?answer=118704&hl=en>
- Horvath, J. (2009, September 3). An update on Google Books and privacy. Message posted to <http://googlepublicpolicy.blogspot.com>
- Jeanneney, J. (2007). *Google and the myth of universal knowledge: A view from Europe*. Chicago: The University of Chicago Press.
- Joint, N. (2009). The Google book settlement and academic libraries. *Library review*, 58(5), 333-340. Doi: 10.1108/00242530910961756
- Lesk, M., & MacKle-Mason, J. (2009). Reading over your shoulder. *IEEE Security and Privacy* 7(3), 78-81. doi: 10.1109/MSP.2009.74
- Grafton, A. (2009). Apocalypse in the stacks? The research library in the age of Google. *Daedalus*, 138(1), 87-98. Retrieved from: <http://proquest.umi.com/-pqdweb?did=0000001636209031&Fmt=3&clientId=43168&RQT=309&VName=PQD>
- Grogg, J.E., & Ashmore, B. (2007). Google book search libraries and their digital copies. *Searcher* 15(4), 18-27. Retrieved from: <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=24865819&site=ehost-live>
- MacColl, J. (2006). "Google challenges for academic libraries". *Ariadne* 46. Retrieved from <http://www.ariadne.ac.uk/issue46/maccoll/intro.html>
- Nunberg, G. (2009, 31 August). Google's book search: A disaster for scholars. *The Chronicle*

of higher education. Retrieved from <http://chronicle.com/article/Googles-Book-Search-A/48245/>.

Oder, N. (2009, August 20). PW survey: Librarians on the fence regarding Google settlement. *Library Journal*. Retrieved from <http://www.libraryjournal.com/article/CA6678927.html>

Oshiro, D. (2009, September 8). Will privacy concerns kill Google Books? *ReadWriteWeb*. Retrieved from [http://www.readriteweb.com/archives/will\\_privacy\\_concerns\\_kill\\_google\\_books\\_settlement.php](http://www.readriteweb.com/archives/will_privacy_concerns_kill_google_books_settlement.php)

Quint, B. (2009). An unapologetic apologia for Google Book Search. *Information Today*, 26(9), 7-8. Retrieved from: <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=44408397&site=ehost-live>

Roush, W. (2005). The infinite library: does Google's plan to digitize millions of print books spell the death of libraries; or their rebirth? *Technology Review*, 108 (5), 54-59. Retrieved from <http://www.technologyreview.com>.

Terry, J. (2009, May). Google book search. *College & Research Libraries News*, 70(5), 306. Retrieved from: <http://vnweb.hwwilsonweb.com/hww/jumpstart.jhtml?recid=0bc05f7a67b1790e29dfa231af182b44149634f778f729a78901dc3f99068b06007fe38bfd1e434c&fmt=H>

Vladeck, D.C. (2009, September 2). *Horvath letter*. Retrieved from <http://www.ftc.gov/os/closings/090903horvathletter.pdf>

Waldman, J. (2007). "Thinking systems need systems thinking". *Systems research and behavioural science*, 24(3), 271-284. Doi: 10.1002/sres.828

Waller, V. (2009). "The relationship between public libraries and Google: Too much information." *First Monday* 14 (9), 1-9. Retrieved from: <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2477/2279>