

Google Books and the Threat to Privacy from Aggregation of Search Data

Allan Verman Y. Ong*

I. INTRODUCTION.....	1001
II. THE PRIVACY POLICIES OF GOOGLE AND GOOGLE BOOKS.....	1002
A. Google Privacy Policy	
B. Google Books Privacy Policy	
III. WHERE PRIVACY CONCERNS STEM FROM.....	1007
IV. SOLVING THE PROBLEM.....	1008
V. CONCLUSION.....	1011

I. INTRODUCTION

The Slovak proverb “*povedz mi, čo čítas a ja ti poviem, kto si*”¹ captures the heart of the debate on privacy in the Google Books matter. Because Google

* '10 LL.M. cand., Columbia University Law School; '09 LL.M., Chinese Law, Civil and Commercial Law, Peking University Law School; '04 J.D., *with honors*, Ateneo de Manila University School of Law; The author was a Member of the Board of Editors and was Lead Editor for Vol. 48, Issue No. 1 of the *Ateneo Law Journal*. He is Han Depei Fellow in Private International Law of the Chinese Journal of International Law. His previous works published in the *Journal* are *Testing the Capacity of the Shareholders Meeting to Protect Minority Shareholder Rights under the Chinese Company Law*, 53 *ATENEO L.J.* 682 (2008); *Regulatory Transition in Employee Stock Options as Exempt Transactions from the Securities Regulation Code*, 52 *ATENEO L.J.* 153 (2007); *Foreign Investments Law: Studying the Administrative Treatment of Legislative Restrictions*, 51 *ATENEO L.J.* 314 (2006); co-authored *Examining the Standards for the Approval of Rehabilitation Plans in Corporate Rehabilitation Proceedings*, 50 *ATENEO L.J.* 878 (2006); *Infusing Principles of Corporate Governance in Corporate Law: A Proposed Hierarchy of Director Accountabilities to Various Stakeholders*, 50 *ATENEO L.J.* 480 (2005); *Upholding The Right to Freedom From Fear: An Examination of the Jurisprudential Response of Criminal Procedure and the Law on Public Officer to the Threat of Kidnapping For Ransom*, 48 *ATENEO L.J.* 1260 (2004); *The Commercial Ends That Motivate the Legal Profession: An Alternative Model to Understanding the Rules that Govern the Practice of Law*, 48 *ATENEO L.J.* 749 (2003); *The Implications of the Principles of Corporate Governance on the Doctrine of Piercing the Veil of Corporate Fiction: Rethinking MR Holdings v. Baja*, 48 *ATENEO L.J.* 209 (2003); *Judicial Policy in the Law on Public Offices*, 47 *ATENEO L.J.* 71 (2002).

Cite as 54 *ATENEO L.J.* 1001 (2010).

1. In English, the proverb states “tell me what you read and I will tell you who you are.”

Books is able to gather data — the books that individual site users are reading; how long each page is read; and the annotations made on the margins of the book — critics fear that this data can be used unscrupulously.

In some scenarios envisioned by critics of Google Books' policies on privacy, one's profile may be created based on the books that one searches for, reads, or annotates. This profile can be used for purposes including the sending of targeted ads that one would be uniquely susceptible to; the sharing of this information with government authorities as would lead to prosecution or persecution; and the sharing of this information with third party businesses as would affect one's employment or credit standing. Imagine when the due diligence of banks on potential borrowers can involve determining if such persons have read, or even searched for, books on defaulting on loans, or when insurance companies determining the insurable risk in life insurance can check if a person has accessed material on diabetes, cardiovascular disease, or other illnesses.

While the benefit of Google Books appears to be acknowledged as considerable, concerns on privacy have led to criticism. Some institutions including the Electronic Frontier Foundation have filed oppositions to the Google Books settlement on concerns of the potential violations of the user's privacy.

This Essay surveys the questions on privacy that have arisen in relation to Google Books and raises some questions on the implications of privacy policies of Google Books. The first part summarizes the privacy policy of Google and Google Books. The second part surveys the challenges and perceived weaknesses of the privacy policies that have been enacted. The third part situates the debate on privacy on the issue of Google Books in the larger framework of the changing nature of consumer media. The last part concludes.

II. THE PRIVACY POLICIES OF GOOGLE AND GOOGLE BOOKS

In order to use the search function and other services of Google, one must first agree to the Terms of Service² imposed by Google. One accepts these terms either by (a) clicking to accept or agree to the terms, where this option is made available by Google in the user interface for any service, or (b) by actually using the services.³ Because users of Google voluntarily agree to these Terms of Service, this removes the relationship of Google and Google users from public law and makes the relationship essentially contractual in nature. It is therefore the Google Privacy Policy and the Google Books

2. Google Terms of Service, *available at* <http://www.google.com/accounts/TOS> (last accessed Feb. 10, 2010).

3. *See* Google Terms of Service, ¶ 2.2 (A) & (B).

Privacy Policy that will apply when one uses the services of Google, rather than the law of the jurisdiction where one is using Google services.

A. *Google Privacy Policy*

Under the Google Privacy Policy,⁴ information about users is obtained through two ways. First is through the use of “cookies.” Cookies are small files containing a string of characters.⁵ When one visits Google, Google sends one or more cookies to one’s computer or other device that uniquely identifies one’s browser.⁶ Google uses cookies “to improve the quality of our service, including for storing user preferences, improving search results and ad selection, and tracking user trends, such as how people search.”⁷ Google also uses cookies in its advertising services for its advertisers and publishers to serve and manage ads.⁸

Second is through the recording of log information. Google’s servers automatically record information that one’s browser sends whenever one visits a website.⁹ These server logs may include information such as one’s web request, internet protocol address, browser type, browser language, the date and time of one’s request, and one or more cookies that may uniquely identify one’s browser.¹⁰

It is interesting to note that Google is also able to receive information, and may therefore record such information, based on a user’s actual location.¹¹ When one uses Google’s location-enabled services such as Google Maps for mobile, Google may receive information about one’s actual location through GPS signals sent by one’s mobile device or information that can be used to approximate a location such as a cell ID.¹²

How does Google use this information that it collects? Google states that it only processes personal information for the following purposes:

- (a) Providing Google’s services, including the display of customized content and advertising;

4. Google Privacy Policy, *available at* <http://www.google.com/privacypolicy.html> (last accessed Feb. 10, 2010).

5. *Id.*

6. *Id.*

7. *Id.*

8. *Id.*

9. *Id.*

10. Google Privacy Policy, *supra* note 4.

11. *Id.*

12. *Id.*

- (b) Auditing, research and analysis in order to maintain, protect and improve its services;
- (c) Ensuring the technical functioning of its network;
- (d) Protecting the rights or property of Google or its users;
- (e) Developing new services; and
- (f) Other purposes described in the Google Privacy Policy¹³

Google states that it will not collect or use sensitive information for purposes other than those described in the Privacy Policy unless prior user consent is obtained.¹⁴

On the sharing of information with third parties, Google classifies information into personal information, that is, information that may enable the identification of a user, and aggregated non-personal information.¹⁵ The policy provides that Google shares personal information with other companies or individuals outside of Google in the following circumstances:

- (a) Google shares this information when it has the user's consent. Google states that it requires opt-in consent for the sharing of any sensitive personal information.¹⁶
- (b) Google provides such personal information to its subsidiaries, affiliated companies or other trusted businesses or persons for the purpose of processing personal information on its behalf. Google states that it requires these parties to agree to process such information based on our instructions and in compliance with the Google Privacy Policy and any other appropriate confidentiality and security measures.¹⁷
- (c) Further, Google will disclose personal information when it has a "good faith belief" that access, use, preservation or disclosure of such information is reasonably necessary to (a) satisfy any applicable law, regulation, legal process or enforceable governmental request, (b) enforce applicable terms of service, including investigation of potential violations thereof, (c) detect, prevent, or otherwise address fraud, security or technical issues, or

13. *Id.*

14. *Id.*

15. Privacy FAQ — Google Privacy Center, *available at* http://www.google.com/privacy_faq.html#toc-terms-aggregated-info (last accessed Feb. 10, 2010). Google defines aggregated non-personal information as "information that is recorded about users and collected into groups so that it no longer reflects or references an individually identifiable user." *Id.*

16. Google Privacy Policy, *supra* note 4.

17. *Id.*

(d) protect against harm to the rights, property or safety of Google, its users or the public as required or permitted by law.¹⁸

For example, Google states that aggregated non-personal information such as the number of users who searched for a particular term or how many users clicked on a particular advertisement may be shared with third parties.¹⁹

B. Google Books Privacy Policy

The Google Books Privacy Policy²⁰ was created in addition to the main Google Privacy Policy. Google describes the relationship between the two statements:

The main Google Privacy Policy describes how we treat personal information when you use Google's products and services, including Google Books. This additional Policy for Google Books does three things: (1) it highlights key provisions of the main Google Privacy Policy in the context of the Google Books service, (2) it describes privacy practices specific to the Google Books service, and (3) it describes planned privacy practices for services proposed in the Google Books legal settlement, which is currently awaiting court approval.²¹

Thus, it provides that Google will not share users' personal information with third parties, "except in the narrow circumstances described in the Privacy Policy," such as emergencies or in response to a valid legal process.²²

What worries most critics is that when one uses Google Books, Google receives log information similar to what Google receives when one uses the Google Web Search. The log information includes: the query term or page request (which may include specific pages within a book one is browsing), internet protocol address, browser type, browser language, the date and time of your request, and one or more cookies that may uniquely identify your browser.

Google states that where law enforcement or civil litigants ask Google for information about what books an individual user has looked at, Google (a) is committed to notifying the affected user if it receives such a request that may lead to disclosure of their information, and (b) if Google is permitted to do so by law and if it has an effective way to contact the user, Google will seek to do so in time for the user to challenge the request.²³

18. *Id.*

19. *Id.*

20. Google Books Privacy Policy, available at <http://books.google.com/googlebooks/privacy.html> (last accessed Feb. 10, 2010).

21. *Id.*

22. *Id.*

23. *Id.*

Apart from the search function currently available in Google Books, the settlement agreement will potentially authorize additional services for users in the United States, some of which will be for a fee.²⁴ This fee will be shared with the authors of the books.²⁵ To enable the processing of the information on usage and fees, the Book Rights Registry created under the settlement will receive aggregate non-personally identifiable information about Google Books usage. Google states that the Registry will not have access to individual user information unless it goes through proper legal processes or in other narrow circumstances set out in the Privacy Policy.²⁶ Google will not require users to create Google accounts, or in any way register their identity with Google, in order to use the following planned services:

- (a) Free online viewing of pages from books covered under the settlement;
- (b) Use of the Institutional Subscription. Schools or other institutions that sign up for subscriptions will be able to authenticate users based on the user's or the institution's IP address, or using other technologies that allow Google to confirm that a user is part of a subscribing institution without knowing who that user is;
- (c) Use of Public Access Service terminals in public libraries. Users will not need to register with Google to use these terminals. Google will receive IP address and cookie information that may identify internet connection or browser, but not the actual user.²⁷

Google addresses some additional privacy issues in its answers to a list of frequently asked questions,²⁸ but do not add anything to the policies already set forth in the Google Books Privacy Policy. However, it clarifies why the privacy policy is not among the matters included in the settlement with the copyright holders. The settlement was a negotiation between the plaintiffs in the lawsuits and Google and it settles the copyright claims that were raised.²⁹ If approved by the court, the settlement will authorize new services to be offered, but the privacy policy for these new services is not yet specified. Google suggests that this is a good thing, both for users and for privacy, first,

24. For a complete discussion, see FAQs — Google Book Settlement, *available at* <http://www.googlebooksettlement.com/help/bin/answer.py?hl=en&answer=118704> (last accessed Feb. 10, 2010).

25. *Id.*

26. *Id.*

27. Google Books Privacy Policy, *supra* note 20.

28. The Google Books settlement and privacy: frequently asked questions, *available at* <http://booksearch.blogspot.com/2009/07/google-books-settlement-and-privacy.html> (last accessed Feb. 10, 2010).

29. *Id.*

because the product has not yet been designed and developed thereby making detailed privacy policy drafting almost impossible, and second, that it is important to engage in discussions with the broader community and in particular with institutions such as libraries about the appropriate privacy policy for these services.³⁰

III. WHERE PRIVACY CONCERNS STEM FROM

It may be argued that Google is a private company providing a service, and that users can simply not avail of its service for reasons such as dissatisfaction with the Google Privacy Policy. Another argument is that Google is simply using the information to provide advertisements, a business model that is necessary to provide a free and useful service to individuals. What is so wrong about providing information in exchange for a free service? Lastly, the appeal to privacy always attaches itself to a suspicion of wrongdoing, that if one were otherwise not doing anything “wrong” then one could do anything openly.

First, concerns over privacy in the Google Books use stems from the ability of Google to amass information about its users, since its “current practices show it is capable of compiling ‘dossiers’ that reveal [peoples’] lives in intimate detail.”³¹ Online stores such as Amazon are likewise able to amass information about its users. But what makes Google Books different is that it is run by Google, which is a widely-used search engine. Under the Google Privacy Policy, Google states that it processes personal information that it is able to acquire using cookies and log information to provide and develop services. Thus, Google is able to link all the information that it has about individual users and create a profile of a person. This creates a culture of pervasive surveillance that permits every user of intellectual property that is capable of being “consumed” online to be logged and charged, and consequently studied as to how best to control the most important resource in an internet society: people’s attention.³²

Second, this ability to amass information about individuals used to be a capacity enjoyed solely by governments. But as commercial enterprises acquire this capacity, controls over the manner by which this information

30. See Google Book Settlement, *available at* <http://www.googlebooksettlement.com/> (last accessed Feb. 10, 2010).

31. Google Book Search Settlement and Reader Privacy, *available at* <http://www.eff.org/issues/privacy/google-book-search-settlement> (last accessed Feb. 10, 2010).

32. Eben Moglen, Anarchism Triumphant: Free Software and the Death of Copyright, First Monday, *available at* http://emoglen.law.columbia.edu/my_pubs/anarchism.html (last accessed Feb. 10, 2010).

can be used demands regulation, since traditional privacy laws have been limited to the regulation of government intrusion.

Third, while the business model of gathering information in order to provide targeted advertisement to segments of a market has existed in other forms, the Google model is different because of its capacity to aggregate personal information, thus:

Given the long and troubling history of government and third party efforts to compel libraries and booksellers to turn over records about readers, it is essential that Google Books incorporate strong privacy protections ... [w]ithout these, Google Books could become a one-stop shop for government and civil litigant fishing expeditions.³³

Fourth, the suspicious view surrounding some claims to privacy — that one would not need privacy if one were doing nothing wrong — assumes a well-functioning law enforcement agency. The world of the internet requires a wider view — as the policemen that stronger privacy protection will “[disappoint] are not just our good policemen but also other people’s quite evil policemen who commit crimes too. And when you remove from them the possibility of committing those crimes, you are preventing a lot of evil.”³⁴ Further, it is possible to imagine the change of circumstances where even persons who have done no wrong need to protect their privacy. Any cultural, religious, or racial group that may in the future face persecution may be identified and rooted out with, among others, their choice of literature.

IV. SOLVING THE PROBLEM

In a letter³⁵ to Google, Inc. Chairman and Chief Executive Officer Eric Schmidt, the Electronic Frontier Foundation, and the Samuelson Law, Technology & Public Policy Clinic, urge that Google Books should, at the minimum, include the following privacy protections in its privacy policy, to wit:

- (1) Protection Against Disclosure: Readers should be able to use Google books without worrying that the government or a third party is

33. Letter to Eric Schmidt, Chairman and Chief Executive Officer, Google Inc., from Cindy Cohn, Legal Director, Electronic Frontier Foundation, Nicole A. Ozer, Technology and Civil Liberties Director, ACLU of Northern California, and Jennifer Lynch, Lecturer in Residence & Supervising Attorney, Samuelson Law, Technology & Public Policy Clinic, *available at* http://www.aclunc.org/issues/technology/asset_upload_file173_8515.pdf (last accessed Feb. 10, 2010).

34. Eben Moglen, *So Much for Savages: Navajo 1, Government 0 in Final Moments of Play*, *available at* http://emoglen.law.columbia.edu/my_pubs/yu-encrypt.html (last accessed Feb. 10, 2010).

35. Schmidt Letter, *supra* note 33.

reading over their shoulder. Google needs to promise that it will protect reader records by responding only to properly-issued warrants from law enforcement and court orders from third parties. It also must promise that it will let readers know if anyone has demanded access to information about them.

- (2) Limited Tracking: Just as readers can anonymously browse books in a library or bookstore, they should also be able to search, browse, and preview Google books without being forced to register or provide any personal information to Google. And for any of its Google Book Search services, Google must not keep logging information longer than 30 days. Google should also not link any information it collects about reader use of Google Book Search to that reader's usage of any other Google services without specific, affirmative consent.
- (3) User Control: Readers should have complete control of their purchases and purchasing data. Readers should be able to delete their records and have extensive permissions controls for their "bookshelves" or any other reading displays to prevent others from seeing their reading activities. Readers should be able to "give" books to anyone, including to themselves, without tracking. Google also should not reveal any information about Google book use to credit card processors or any other third parties.
- (4) User Transparency: Readers should know what information is being collected and maintained about them and when and why reader information has been disclosed. Google needs to develop a robust, enforceable privacy policy and publish the number and type of demands for reader information that are received on an annual basis.³⁶

How does the Google Books Privacy Policy measure up to this list of proposed measures? First, on the disclosure of information, the position taken by Google is that it will follow the laws that apply to it. Google states that it may share users' personal information with third parties in the circumstances described in the Privacy Policy, such as emergencies or in response to valid legal process.³⁷ Critics have complained that this does not go far enough. In saying that it will follow laws that apply to it, Google is not committing too much because these laws would bind Google regardless of whether or not Google also wrote about them in its privacy policy.³⁸ As for disclosure to third parties, Google has to commit to the highest legal standards for the disclosure of information — for instance, complying with a letter of request from a prosecutor would be "in response to valid legal

36. *Id.*

37. See Google Books Privacy Policy, *supra* note 20.

38. See generally Cindy Cohn, Google Book Privacy Policy: Good Start, Much More Needed, Electronic Frontier Foundation, available at <http://www EFF.org/deeplinks/2009/09/google-book-privacy-policy-good-start-more-needed> (last accessed Feb. 10, 2010).

process,” as provided in the Google Books Privacy Policy, but would not pass the threshold of the highest legal standard, which is a warrant.³⁹

Second, as to the tracking of user information, Google commits neither to co-mingle the Google Books data with web search and other data Google collects on users, nor use book data as part of their behavioral advertising campaigns.⁴⁰ In a response to queries from the United States Federal Trade Commission (FTC), Google stated that it does not currently engage in behavioral advertising with respect to Google Books, and that Google will adhere to the FTC Self-Regulatory Principles for Online Behavior Advertising (the FTC Principles)⁴¹ should Google decide to do so in the future. The FTC has urged Google to focus on appropriately limiting secondary uses of data collected through Google Books, including uses that would be contrary to reasonable consumer expectations. The FTC also asked that Google continue engaging in dialogues with the FTC regarding Google’s products and services and how to most effectively protect consumers’ privacy interests.⁴²

Google will eventually co-mingle personal information that it gathers from users from Google Books with its other services. When it does, the pressure from FTC to comply with the FTC Principles and to engage in dialogue would be a good start to safeguarding consumer interest.

Third, on user control of information that will be provided to Google, Google Books Privacy Policy currently commits to allow users substantial control. The My Library feature of Google Books allows users to maintain a public online list of your favorite books. Google states that this information may be reviewed and deleted any time.⁴³ Further, when the option to purchase access to books is made available, Google plans to build protections to limit the information (such as book titles) available to credit card companies about book purchases, and to enable users to delete or disassociate the titles of books purchased from your Google Account.⁴⁴

39. *Id.*

40. Richard Koman, Does Google Book Privacy Policy Go Far Enough?, *available at* <http://government.zdnet.com/?p=5363> (last accessed Feb. 10, 2010).

41. United States Federal Trade Commission Staff Report: Self-Regulatory Principles for Online Behavioral Advertising, *available at* <http://www.ftc.gov/os/2009/02/P085400behavadreport.pdf> (last accessed Feb. 10, 2010).

42. Letter to Jane Horvath, Global Privacy Counsel, Google, Inc., from David C. Vladeck, Director, Bureau of Consumer Protection, *available at* <http://www.ftc.gov/os/closings/090903horvathletter.pdf> (last accessed Feb. 10, 2010).

43. Google Books Privacy Policy, *supra* note 20.

44. *Id.*

On the last point of user transparency, what makes the privacy policies deficient is precisely that they are merely policy. “Those policies can be changed at any time and may be unenforceable by readers whose privacy has been violated.”⁴⁵ So much so that the Electronic Privacy Information Center (EPIC) has filed a Motion to Intervene in the settlement of the Google Books cases,⁴⁶ where it argues that the settlement fails to protect readers’ privacy interests.⁴⁷ EPIC argues that routine privacy safeguards are missing and should be included in the settlement agreement.⁴⁸ This argument is compelling — as the Google Privacy Policy is mere policy, Google is free to modify it in accordance with its business needs. “Even if Google would write it in blood, there is still the obvious problem that when the government comes knocking the policy doesn’t mean anything.”⁴⁹

V. CONCLUSION

The privacy concerns over the use of information by Google Books does not detract from the benefits that may be brought about by Google Books as a search tool for books and as a means by which out-of-print books may become available. Users, however, who avail of Google Books services will be exposing information about themselves to a commercial enterprise. As a commercial enterprise, Google will wish to capitalize on this aggregated information. This Essay describes the privacy issues that may arise with Google holding so much information about its users. In a world where commercial enterprises will become custodians of information, guidelines need to be provided on how the information will be handled. Should the issue be resolved satisfactorily, the treatment of the issue of privacy in the case of Google Books can be a model on how information will be treated by other commercial enterprises. But if unguided by consumer advocacy, adequate disclosure, and perhaps government regulation, the dangers are grave and manifold.

45. Cohn, *supra* note 38.

46. Electronic Privacy Information Center’s Motion to Intervene in *The Authors Guild, Inc., v. Google, Inc.*, Case No. 05 CV 8136 (D.C.) (S.D.N.Y. Sep. 4, 2009), available at http://epic.org/privacy/googlebooks/EPIC_Brief-GBS.pdf (last accessed Feb. 10, 2010).

47. *Id.* at 12.

48. *Id.* at 21.

49. Ryan Singel, *Privacy Group Asks to Join Google Book Lawsuit As Deadline Approaches*, available at <http://www.wired.com/epicenter/2009/09/privacy-group-asks-to-join-google-book-lawsuit-as-deadlineapproaches> (last accessed Feb. 10, 2010).