Privacy Recommendations for the Google Book Search Settlement

July 27, 2009

Executive Summary

The settlement of the copyright infringement lawsuit against Google for its Book Search tool will, if approved, dramatically alter the way the public obtains and interacts with books. With its new offerings, Google will considerably increase public access to millions of books containing much of the world's written knowledge and ideas. Moreover, Google's powerful book search engine will transform how the public conducts research, interacts with written text, and shares information and ideas with others. The settlement deserves court approval because it will unquestionably provide a significant public benefit at a size and scale that is not otherwise likely to be replicated in the near term.

However, the settlement is not perfect and many parties are likely to provide recommendations as to how to address various perceived shortcomings. It will be important for the court to assess these claims. In particular, given the unique role that Google will assume upon approval of the settlement—that of a comprehensive library for research and browsing as well as a major bookstore—we believe that questions of reader privacy must be addressed. Libraries have a long history of protecting reader privacy and safeguarding the right to read anonymously. Indeed, patron circulation records are protected by law against undue disclosure in almost all states in the United States. But, because the settlement is focused specifically on resolving the copyright dispute between Google and the rightsholders, the agreement does not address the full range of reader privacy issues that will confront the new services.

Significant work remains to address that gap and ensure that historical reader privacy protections are not lost as library functions are centralized and moved online. At a minimum, before the settlement is approved, Google should issue a set of privacy commitments that explains both its general approach to protecting reader privacy and its process for addressing privacy in greater detail as Google Book Search moves forward. Since further detail regarding privacy matters may need to be fleshed out over time as the services are built, the court should monitor implementation of these privacy commitments as part of its ongoing supervision of the settlement. Critically, this structure—a set of evolving privacy commitments with court supervision—does not require an alteration of the current settlement.

This document identifies and analyzes the privacy risks posed by different aspects of the proposed services and makes specific policy recommendations for protecting reader privacy.

Introduction

In October 2008, Google, the Authors Guild, and the Association of American Publishers announced a sweeping settlement agreement in their class-action lawsuit concerning the Google Library Project, part of Google Book Search ("GBS").[1] GBS is an ambitious effort to enable full-text searches of a broad database of books—essentially, to do for books what Google's core search engine does for the Web. As originally conceived, Google would work with libraries to scan entire collections of books into a massive search database. Except for books for which the rightsholders opted out, this database would be available for searching, with search results displaying both the name of the book and a short snippet of text surrounding the text that matched the search query.

The authors and publishers had alleged that Google's scanning of copyrighted books and display of snippets amounted to copyright infringement, while Google maintained that its actions qualified as fair use. The tentative settlement—subject to a fairness hearing and judicial approval—would allow Google to offer greatly expanded access to scanned books under an ongoing arrangement to compensate rightsholders for the use of their books.

The array of services described in the settlement will greatly benefit the public by enabling unprecedented digital access to millions of books. Google stands to gain the right to offer new forms of access to the books it has scanned, including lengthy page previews of books returned in search results ("Previews"); sales of online access to books in their entirety ("Consumer Purchases"); subscriptions to the entire database of books for institutions ("Institutional Subscriptions"); and free access to the entire database via terminals in public libraries ("Public Access Service"). Many of these offerings are not possible today, nor would they necessarily have been developed if Google had litigated this case to the end and won an outright victory on fair-use grounds. In exchange, Google will provide ongoing compensation to rightsholders, in addition to a one-time payment for books already scanned. These payments to copyright holders will be coordinated and distributed by a newly established "Books Rights Registry."

The settlement provisions establishing the operational framework for Google's expanded offerings are quite complex. Books will be included in Google's search database unless a rightsholder expressly objects. For books that are still in print, Google will be able display portions of actual book text on an opt-in basis only (i.e., only with the rightsholder's express agreement). For out-of-print books that are still covered by copyright, by contrast, Google will be able to display portions of text unless rightsholders come forward and identify themselves to opt out. Where such display is permitted, GBS users will have free access to expanded Previews—showing up to 20% of a book in most cases—with Google sharing ad revenues from each book's Preview

[1] Proposed settlement agreement in *The Authors Guild, Inc., et al., v. Google, Inc.*, No. 05 CV8136, http://www.googlebooksettlement.com/r/view_settlement_agreement ("Settlement").

pages with rightsholders via the Registry. New forms of full-text access (Consumer Purchases, Institutional Subscriptions) will generate additional revenue to be shared with the Registry and distributed according to usage statistics. In addition, the settlement provides that Google and the Registry can agree to implement new revenue models in the future, specifically listing PDF downloads, consumer subscriptions, print-on-demand, and automated summary creation as examples.

Managing and differentiating among these services will require extensive data collection on Google's part, including the collection of sensitive personal information about which books people are reading or are interested in reading. In addition, Google will need to share some information about usage of the services with the Registry so that it can administer payments to rightsholders.

The settlement, however, does little to describe what specific information will be collected, how it will be shared, and how it will be protected. This is understandable given the circumstances. First, the services have not been fully designed and implemented, so precise data flows are to some degree undefined. Second, the settlement is the result of a two-party negotiation aimed at resolving a copyright dispute; it is unsurprising that a detailed consideration of user privacy was not incorporated. Nonetheless, from a public interest standpoint, it means that privacy remains a significant open question.

Overall, CDT's view is that the services outlined in the settlement will considerably increase public access to millions of books containing much of the world's written knowledge and ideas. Moreover, Google's powerful book search engine and other dynamic tools will transform how the public conducts research, interacts with written text, and shares information and ideas with others. Because it will unquestionably provide a significant public benefit at a size and scale that could not otherwise be replicated in the near term, CDT believes the settlement as written deserves court approval.

At the same time, reader privacy must be addressed as functions traditionally performed by libraries migrate to the consolidated online environment of GBS. Our top-level recommendation, therefore, is that Google should issue a set of privacy commitments before the settlement's approval, explaining both its general approach for safeguarding reader privacy and its process for fleshing out full details as GBS moves forward.[2] Adherence to these commitments should be subject to ongoing court supervision over time as the services are designed and implemented. The incorporation of privacy promises need not and should not impact the structure of the agreement or in any way jeopardize approval of the settlement.

[2] Shortly before the publication of these recommendations, Google posted a "Privacy FAQ" list to the GBS blog. CDT is pleased that Google has made such a public statement—and indeed the statement speaks to several of our concerns—but we believe that this represents the beginning, not the end, of discussions of reader privacy, and that an explicit commitment subject to court oversight is nonetheless required. See "The Google Books Settlement and Privacy: Frequently Asked Questions," *Inside Google Books* blog, July 23, 2009, http://booksearch.blogspot.com/2009/07/google-books-settlement-and-privacy.html.

Below, CDT spells out the privacy issues in greater detail and offers concrete recommendations for specific safeguards and policies that should be developed.[3] In doing so, we recognize that Google by no means will be the sole provider of electronic access to books.[4] Many of the privacy recommendations that we make below could apply equally to other providers. However, the settlement will allow Google to provide a unique offering for at least the near future, especially with regard to in-copyright but out-of-print books. Only Google will have what amounts to a court-backed license to scan and provide access to in-copyright but out-of-print books except where rightsholders have come forward to opt out. Critically, this includes the books of rightsholders who do not register with the Registry at all. Potential competitors will lack the legal device of a class action settlement to secure rights from unregistered rightsholders and hence may have little practical ability to match Google's scope of coverage. The settlement therefore will put Google in a unique position to step into the role of a universal online library, at least for the near term. This in turn makes it incumbent upon Google to make a strong commitment to reader privacy. Indeed, we believe that the adoption of a robust privacy framework here will set a high standard for other providers as the online market for electronic books expands.

Privacy Concerns Raised by the Settlement

Both as a legal and policy matter, readers have long enjoyed a high level of anonymity and privacy with respect to their reading habits. The First Amendment protects the right to receive information anonymously.[5] Accordingly, libraries have a longstanding commitment to intellectual freedom and patron privacy. The American Library Association *Code of Ethics* states, "We protect each library user's right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired, or transmitted."[6]

Furthermore, state laws protecting library patron records from public disclosure have for many years reinforced individuals' right to privacy with respect to what they read. Forty-eight states have enacted statutes that either expressly protect these records or

[3] Other parties have offered concerns and recommendations regarding various perceived shortcomings of the settlement. For example, a number of observers have argued that the settlement will leave both Google and the Book Rights Registry with dominant positions that merit close oversight from a competition perspective. Although these concerns merit consideration—indeed, the Department of Justice is conducting an inquiry into the matter—this paper focuses solely on privacy.

[4] Amazon, for example, recently announced an initiative to sell on-demand prints of public domain books from the University of Michigan's library. See Dave Gershmann, "University of Michigan, Amazon announce book-printing deal," *The Ann Arbor News*, July 21, 2009, http://www.mlive.com/news/ann-arbor/index.ssf/2009/07/university_of_michigan_amazon.html.

[5] See, e.g., Julie Cohen, "A Right to Read Anonymously: A Closer Look at 'Copyright Management' In Cyberspace," 28 Conn. L. Rev. 981 (1996).

[6] Code of Ethics of the American Library Association, Jan. 22, 2008, http://www.ala.org/ala/aboutala/offices/oif/statementspols/codeofethics/codeethics.cfm.

exempt them from public disclosure rules. In several states, violating a library privacy statute is a misdemeanor offense.[7] With respect to book purchases, courts have generally been reluctant to compel bookstore owners to reveal information about their customers' purchases.[8]

As more and more content—books and otherwise—migrates online, the goal of preserving individuals' privacy is by no means a challenge uniquely faced by Google. However, the GBS settlement represents a sea change with respect to the treatment of material that has historically been highly protected, and presents a unique opportunity to ensure that strong protections remain in place. Google is in many ways taking on the role of the public library as a gateway to information, only on a much larger and more comprehensive scale.[9] Providing such breadth of electronic access to so many published books will give Google an unparalleled view of people's reading and information-seeking habits. By hosting the scans and closely managing user access, Google will have the capability to collect data about individual users' searches, preview pages visited, books purchased, and perhaps even time spent reading particular pages. Whereas in the offline world such data collection is either impossible or widely distributed among libraries and bookstores, Google will hold a massive centralized repository of books and of information about how people access and read books online.

Furthermore, as noted above, Google is likely to be the only *comprehensive* source for digitized out-of-print books. Although Google's contracts with the Registry, rightsholders, and libraries are non-exclusive, Google alone will have the broad right to scan and display the works of any class members that do not register with the Registry. The ability to offer a comprehensive collection, together with the magnitude of the task of scanning and hosting millions of books, is likely to make Google the dominant provider of online universal library functions for at least the immediate future.

While the settlement agreement does not fully describe the types of data Google will collect, it does offer some indication. Detailed user information will be collected and used to differentiate among the services offered, to calculate payments to rightsholders, and to prevent unauthorized access to the scanned books. For the purposes of limiting Previews to 20% per book per user, Google will track page-by-page access using "IP address, cookies, and similar signals that may be available."[10] Google will manage and track individual consumer purchases using "account login or other equivalent method."[11] For controlling access to the Institutional Subscriptions, Google will authenticate users using "IP address authentication, user login, and/or leveraging

[7] See, e.g., Ariz. Rev. Stat. § 41-1354 (2008); Ark. Code. Ann. § 13-2-702; D.C. ST § 39-108; Fla. Stat. Ann. § 257.261; Mont. Stat. Ann. § 22-1-1111.

[8] See note 26 and accompanying text.

[9] Eric Schmidt, Google's CEO, compares the service to a library in a recent article, calling GBS "a really oh-my-God kind of change." David Carr, "How Good (or Not Evil) Is Google?" *New York Times*, June 21, 2009, http://www.nytimes.com/2009/06/22/business/media/22carr.html.

[10] Settlement Attachment D (Security Standard) § 3.9.1

[11] Settlement Security Standard § 3.9.2

authentication systems already in place at an individual institution."[12] Google may offer a "Book Annotation" feature to allow readers to provide their own commentary and other content on individual pages of a digital book. For annotations made on purchased books, the settlement requires Google to have each reader "identify (*e.g.*, by name, login or user id) each individual with whom such Book Annotation will be shared" (for up to 25 individuals).[13] In addition, Google will monitor usage information to prevent advertising fraud and any pages users choose to print from GBS will contain an encrypted watermark containing session- and user-identifying information.

In short, the settlement gives Google the potential, and in some instances the need, to collect substantial quantities of sensitive reader information. Google will also need to share some usage data with the Registry. Specifically, Google will share sales and subscription usage data for calculating and distributing payment to rightsholders,[14] market research data concerning various Preview options,[15] and data pertaining to audits and security breaches.[16] Some collection and sharing is of course necessary to effectuate the settlement, but the settlement does not—and should not—require sharing anything other than aggregate data. However, the settlement does not contain a broad restriction on the sharing of user data. The agreement does state that Google cannot be forced to disclose "confidential or personally identifiable information except as compelled by law or valid legal process" in the case of a security breach, but it does not address voluntary disclosure by Google.[17] More generally, it also does not address Google's collection, use, retention, and sharing of user data outside the specific context of a security breach. More formal privacy safeguards would ensure that readers maintain the privacy they have traditionally enjoyed, preserving the right to read anonymously and allowing readers to feel free to access and read books of any sensitive sort.

The risk of not adopting explicit privacy protections is compounded by the fact that Google currently offers dozens of other services and software products on the Web and on end-user devices. In the absence of binding limits on what Google can do with the data it collects about readers through GBS, Google would remain free to combine that data with other data that Google collects, adding a rich and personal dimension to the profiles that Google already maintains about individuals' searching and Web surfing habits. Reading habits add an intimate element to profiles that may already be attractive for a variety of uses, from marketing to litigation.

Google's increasingly comprehensive stores of user data will likely be a tempting information source for government surveillance as well. Law enforcement has already shown considerable interest in search engine data and other kinds of records showing

[12] Security Standard § 3.9.3
[13] Settlement § 3.10(c)(ii)(5)(d)
[14] Settlement § 6.6(v); Settlement Attachment C (Plan of Allocation)
[15] Settlement § 4.3(e)(i)
[16] Settlement §§ 8.2, 8.3
[17] Settlement §§ 7.3(b), 8.6(a); see also 4.6(e) and 15.1 (establishing confidentiality requirements in the case of audit).

Internet usage.[18] However, statutory privacy protections have not kept pace with technological development.[19] Some kinds of data about Internet usage are provided very weak protection and there is no statutory standard at all for data collected by some new communications and information services. As Google's data collection capability grows, it is imperative that its thresholds for governmental disclosure are adequate to ensure user privacy and due process.

In short, Google is already faced with a multitude of privacy challenges in its existing services, but GBS significantly ups the ante. Given the settlement's sweeping potential impact on important policy matters, it is critical that strong privacy practices be put in place to safeguard the public interest.

Specific Recommendations

The magnitude of Google's undertaking and the slim odds that any other entity will soon compete with Google as a comprehensive digital library demand that the company craft and publish strong privacy protections to govern GBS. Because the settlement only addresses a handful of privacy issues and because Google has yet to design and implement many aspects of GBS, there is little public information about how Google intends to safeguard reader privacy once the services established by the settlement are launched.[20] This section describes in detail the kinds of protections that CDT would expect Google to commit to before launching the services described in the settlement.

We recognize that because the implementation of the GBS service is not yet fully conceptualized, it may not be possible for Google to commit to every privacy detail now. We therefore urge that Google set out, with as much specificity as possible, a baseline approach to safeguarding reader privacy that it can commit to now, as well as a process for articulating further detail once the settlement is approved and Google begins to design the implementation of GBS. That process, and the detailed privacy practices that emerge from it, should be subject to court oversight. CDT believes the end result should be a set of privacy protections that includes the following elements.

---

[18] Jeremy Pelofsky and Michele Gershberg, "FBI Wants Internet Records Kept 2 Years: Source," *Boston Globe*, June 1, 2006, http://www.boston.com/news/nation/washington/articles/2006/06/01/fbi_wants_internet_records_kept_2_years_source/.

[19] Center for Democracy &Technology, "Digital Search and Seizure: Updating Privacy Protections to Keep Pace with Technology," 2006, http://www.cdt.org/publications/digital-search-and-seizure.pdf; Pelofsky and Gershberg, "FBI Wants Internet Records Kept 2 years: Source"; Eric Lichtblau and James Risen, "Officials Say U.S. Wiretaps Exceeded Law," *New York Times*, April 15, 2009, http://www.nytimes.com/2009/04/16/us/16nsa.html.

[20] Google has recently made some preliminary statements addressing the settlement and reader privacy. See note 2.

Notice

The settlement agreement currently contains no provision requiring Google to notify readers about the data it collects in connection with GBS. We believe that Google should clearly and prominently disclose the following:

- What information Google collects in connection with GBS, including information that can be used to identify individual readers (IP addresses, cookie information, and account information, for example);
- What information Google collects about individuals' use of GBS (search terms, book selections, page selections, or length of stay on a particular page, for example);
- The purpose for which this information is collected;
- How long each type of data is retained;
- What technical mechanisms Google uses to track readers on the GBS site;
- How readers can exercise choice about having their data collected and used in connection with GBS; and
- How reader data is safeguarded against theft or misappropriation.

In light of the special sensitivity of readership information and library browsing, a link to this notice should be displayed more prominently than the usual privacy notice associated with other Google services. The notice language itself should be provided in a dedicated GBS privacy policy (similar to the dedicated privacy policies that Google already provides for many of its other services[21]), as well as any other location within Google's collection of Web sites where readers might reasonably look for such information.

Collection Limitation

The terms of the settlement indicate that Google will be collecting reader data for a variety of purposes, including differentiating among its services, calculating payments to rightsholders, and preventing unauthorized access to scanned books. The settlement does not say whether Google may collect details about how readers interact with books, but there are some indications that Google can already track which pages of a book users view and how long each page is viewed in the current incarnation of GBS.[22]

Google's potential technical capability to intimately track reader behavior should not trump individuals' long-standing ability to read books anonymously. Thus, CDT believes that Google should commit to collecting only the data necessary to provide the services laid out in the settlement. For example, Google generally does not need to

[21] Google Privacy Center, http://www.google.com/privacy.html.

[22] See Motoko Rich, "Google Hopes To Open a Trove of Little-Seen Books," *New York Times*, January 4, 2009, http://www.nytimes.com/2009/01/05/technology/internet/05google.html (revealing that a Google employee knew how many book pages a particular user had viewed and for how long).

collect details about how readers are consuming its digital books—which pages they view, how often they view them, how long they view them, and so forth—and thus the default should be that it will not do so. Some exceptions might be where necessary to do certain usage accounting required for Previews, or for other purposes with the full understanding and express consent of readers.

Google should also limit the data it collects when patrons of libraries and other institutions access the service. The settlement contains a detailed "Security Standard" primarily aimed at ensuring the security of the digitized copies of books that Google will come to possess. The Standard has specific provisions about authenticating users accessing GBS from institutions; it requires that "Google shall use commercially reasonable efforts to authenticate individual End Users for access to Books in an Institutional Subscription" by using techniques that "may include IP address authentication, user login, and/or leveraging authentication systems in place at an individual institution."[23] Google, however, should have no need to know the identity of any individual user of an Institutional Subscription, only that such a user is authorized under the subscription. Institutions should therefore be responsible for authenticating their own end users without sharing authentication credentials or other personal information with Google.

Use Limitation

The settlement agreement does not currently restrict Google from using data collected in connection with GBS, other than to prohibit the unauthorized dissemination of copyrighted works.[24] Given the potential sensitivity of information surrounding reading habits, Google should refrain from using information collected through GBS for purposes other than to provide and secure the GBS service. By default, information collected through GBS should not be used in connection with any other Google services or combined with data from other Google services, including advertising services provided outside the GBS site. Google could offer users the option of having their GBS data used for other purposes, as long as those choices were offered with appropriate user control and consent mechanisms (see the User Control recommendation, below).

The Book Annotation feature may create especially sensitive user records because it involves users generating their own content and identifying other users with whom to share it. Google should not assume that because this content is actively generated by users it deserves lower protection than any other data generated by using GBS. Accordingly, Google should limit its use of the annotations themselves and the data about which users are sharing annotations strictly to providing the Book Annotations feature. If Google wants to use this data for other purposes, such as targeted advertising or integration with other Google services, it should only do so with full user control and consent.

[23] Settlement Security Standard § 3.9.3

[24] Settlement § 2.2

User Access

Google should provide users with the ability to access the information that has been collected in connection with their user accounts, including purchase histories, annotations and records of annotation sharing, and, to the extent they are tied to user accounts, search histories. Should Google decide to maintain more detailed data about how individual readers make use of Consumer Purchases (for example, which pages they have viewed, how long they viewed each page, or which pages they printed), that data should be made accessible to readers themselves. Google should also provide readers with the option of deleting this usage data or disassociating it from their user accounts.

User Control

Should Google seek to use GBS data in connection with its other services, it should obtain affirmative consent from readers before doing so. This permission should be easy to change at any time, and Google should be required to honor readers' choices persistently. Readers who have accounts with Google may optionally be provided a way to tie their preferences to their accounts so that they persist across different computers or Web browsers. Readers should not be prevented from using GBS or other Google services should they choose not to have their GBS data associated with those services.

Readers who use the Consumer Purchase function of GBS should have the ability to delete individual purchases and/or their entire purchase histories at any time should they decide that they no longer need online access to those purchased books. In addition, readers making use of the Book Annotation feature should have the ability to delete their annotations and the record of the users with whom they shared their annotations.

Disclosure Limitation

The settlement agreement is silent with respect to general law enforcement and civil litigant access to the data that Google collects in connection with GBS. Yet it is almost certain that at some point in the not-too-distant future a governmental entity or private litigant will seek disclosure from Google of information that could be used to identify a user or to associate a user with access to particular books.

The burden in the first instance will fall upon Google to resist any request that is not supported by proper legal process and to ensure that the request is not overbroad. This is something that Google routinely does already when receiving law enforcement and civil litigant requests for email on its Gmail service, for example. In the case of Gmail, the standards for access are fairly clear—if not always adequate[25]—under existing law.

In the case of GBS, however, given the unique comprehensiveness of the services and the special sensitivity associated with reading, Google has an obligation to state, in advance, what kinds of process it will comply with and what kinds it will resist. At a

[25] See CDT, "Digital Search and Seizure," note 19, above.

minimum, Google should state publicly that, except in cases of emergency or situations in which Google determines that it has little chance of prevailing, it will take reasonable steps in response to government requests to insist that the government obtain a court order or warrant issued upon probable cause to compel disclosure of information that could be used to identify a user or to associate a user with access to particular books.[26] "Reasonable steps" would include, at a minimum, filing a motion to quash any process less than a search warrant and conscientiously litigating the matter through the appropriate court of first instance. Google should also commit that, unless otherwise prohibited from doing so, in the case of any demands by a governmental entity, it will provide prompt notice to the user sufficient to allow that person to oppose the disclosure request.[27]

With respect to access by civil litigants, Google should likewise commit that, in all circumstances, it will resist demands for disclosure and will provide prompt notice to the user sufficient to allow that person to oppose the disclosure request. Google should commit that it will take the position in response to civil litigation requests that it will not, unless otherwise required by law, disclose any information about GBS users to a third party in a civil or administrative action absent a court order issued following a judicial determination that the party seeking the information has a compelling interest in the information, that the information cannot be obtained by less intrusive means, that the case has prima facie validity, that the user has been given the opportunity to object, where allowable by law, and that the First Amendment right to anonymously speak and receive information has been appropriately considered.[28]

Because it is difficult to foresee exactly what types of requests will be made, and because experience with those requests might reveal the need for adjustments to these disclosure standards, Google should commit to making available to the public certain details about the compulsory disclosure of GBS information. Specifically, it should make public the number of requests by government and civil litigants for GBS usage or user-identifying data it has received, the types of information sought, the types of legal action underlying the requests, Google's response to each request, and the types of information, if any, that were in fact disclosed.

---

[26] The probable cause standard was drawn from the decision of the Colorado Supreme Court in *Tattered Cover, Inc. v. City of Thornton*, 44 P.3d 1044 (2002). In that case, the court required not only a warrant, but also additional protections. It is important to note that the decision in *Tattered Cover* was based on a provision in the Colorado constitution protecting freedom of speech. See Colorado Constitution, Article 2, §10; see also California Constitution, Article 1, § 2(a).

[27] At the federal level, the Video Privacy Protection Act, which safeguards the privacy of videocassette and other similar audiovisual records, expressly requires notice prior to any disclosure of personally identifiable information to a law enforcement agency pursuant to a court order (although apparently not when a warrant is used). 18 U.S.C. § 2710(b)(2)(B), (b)(3). Similarly, the Cable TV Privacy Act requires prior notice and an opportunity to object before a governmental entity may obtain personally identifiable information regarding a cable subscriber. 47 U.S.C. §551.

[28] The compelling interest test is drawn directly from *In re Grand Jury Subpoena to Kramerbooks & Afterwords Inc.*, 26 Med. L. Rptr. 1599 (D.D.C. 1998), and *Tattered Cover*. The least intrusive means standard is drawn from *Tattered Cover*.

With regard to the Registry, although the settlement agreement requires Google to disclose data for the purposes of reporting on tests of the service,[29] calculating payments,[30] sharing audit results,[31] and disclosing security breaches,[32] no provision of the settlement would require Google to disclose anything but aggregate data to the Registry. To effectuate the services described, the Registry need only know that a particular book was read or accessed, not which individual user was responsible.

As noted above, the settlement stipulates that Google will not be required, absent valid legal process, to disclose to the Registry "confidential or personally identifiable information" in the event of a security breach.[33] However, the settlement contains no broad prohibition to guarantee that Google will not voluntarily share personal information with the Registry. Google should therefore expressly commit to disclosing only aggregate data to the Registry, without any information that identifies or might be used to identify an individual (including user account information, IP addresses, and cookie identifiers). The Registry should have no reason to request such data, but in any event Google should not provide it.

The settlement also requires that any book pages printed by users include a visible watermark containing "encrypted session identifying information . . . which could be used to identify the authorized user that printed the material."[34] However, there is no requirement that Google identify individual users to the Registry or rightsholders. Should a rightsholder wish to pursue legal action against a user of the GBS service whose identity is not known, the appropriate mechanism to attempt to identify the user is—as in the case of any third party civil litigant—a John Doe lawsuit leading to a court-approved subpoena,[35] not a direct request to Google.

Data Retention Limitation

Operating the GBS service in compliance with the settlement agreement will likely require Google to retain data in identifiable form (or in association with a reader identifier like an IP address or cookie ID) for several purposes: to secure the GBS service and defend against breaches,[36] to protect against advertising fraud,[37] to test different Preview displays for individual books,[38] to limit Preview page views per individual,[39]

---

[29] Settlement § 4.3(e)(i)
[30] Settlement § 6.6(v); Settlement Attachment C (Plan of Allocation)
[31] Settlement § 8.2
[32] Settlement § 8.3
[33] Settlement § 7.3(b), 8.6(a); see note 18 and accompanying text. The settlement contains other narrow confidentiality provisions. See Settlement § 4.6(e) (binding third-party auditors to confidentiality) and 7.2(b)(ii)(2) (protecting against the disclosure of the identity of any blind user authorized to use special tools at participating libraries).
[34] Settlement § 4.1(d)
[35] See, e.g., *Dendrite Int'l v. Doe*, 775 A.2d 756 (N.J. App. Div. 2001) (setting out leading standard for piercing anonymity).
[36] Settlement § 8.3
[37] Settlement § 4.6(b)
[38] Settlement § 4.3(e)
[39] Settlement Security Standard § 3.9.1

and to facilitate the purchase and consumption of books through the Consumer Purchase feature.[40] However, the settlement agreement does not impose any restrictions on how long Google may retain data it collects in connection with GBS for these purposes.

Google should therefore commit to retain data in identifiable form or in association with a reader identifier only as long as is necessary for the purpose for which the data was collected, and in any event no longer than 90 days. There are few purposes that will require a longer retention period on a consistent basis and 90 days has become the industry-leading standard for retention of identifiable data in the Web search industry.[41] Google should also delete all reader identifiers collected in connection with Consumer Purchases at the request of the reader. These requirements will allow Google to administer the GBS service in compliance with the settlement while reducing the privacy risks associated with storing GBS data in identifiable form.

Security and Compliance

The settlement's Security Standard outlines a comprehensive set of security and compliance requirements that Google must implement to protect digitized files, but no equivalent set of requirements to protect data about readers and their use of GBS. To the extent applicable (some requirements might be irrelevant to securing reader information, such as the requirement to watermark digital images served to users), we believe that Google should apply the same security standard to the data that it collects in connection with GBS.

Next Steps

As a class-action settlement, the agreement between Google and the authors and publishers is subject to court approval. The judge presiding over the case will hold a fairness hearing on October 7, 2009 to determine whether to approve or reject the settlement and the certification of the class.

Given the sweeping impact of the agreement and its potential impact on reader privacy, it is critical that Google make a public commitment to protecting reader privacy. In the long term, CDT would expect Google to commit to the set of protections outlined in the previous section or something substantially similar, and to incorporate those commitments into Google's privacy policy. However, given that the GBS services will likely not be designed or implemented prior to the settlement's approval, it may not be feasible for Google to commit to every privacy detail. Under these circumstances, we believe Google should commit to a baseline set of privacy protections and submit that

[40] Settlement §§ 1.32, 4.2

[41] See Jessica Guynn, "Yahoo to Purge User Data after 90 Days," *Los Angeles Times*, December 18, 2008, http://articles.latimes.com/2008/dec/18/business/fi-yahoo18; Anne Toth, "Your Data Goes Incognito," Yahoo! corporate blog, December 17, 2008, http://ycorpblog.com/2008/12/17/your-data-goes-incognito/.

commitment to the court prior to the settlement's approval, providing as much detail as possible. Importantly, such a commitment would not require the settlement or the services it describes to be altered. Making this commitment in support of the values of privacy and intellectual freedom would only increase GBS's value to readers and should be acceptable to all parties.

Recent statements indicate that Google is indeed thinking about reader privacy in relation to GBS.[42] However, while these statements or even a more detailed public commitment from Google would be valuable, they would still be subject to change at any time. It is therefore incumbent upon the court to ensure that Google's commitments are binding and adequate by exercising ongoing supervision over Google's privacy practices.[43] The court should use Google's initial set of commitments as a baseline for oversight and continue to monitor Google's evolving privacy practices and commitments as the GBS services are designed and implemented. Indeed, given that the level of detail with which Google will be able to describe its practices is likely to increase over time, ongoing oversight seems highly suitable.

It is important to keep in mind that neither party to this case was negotiating on behalf of the reading public, leaving the task of addressing the settlement's public policy implications to the court. It is critical for the court to protect the longstanding tradition of reader privacy even as GBS brings sweeping change to the way we find and read books.

CENTER FOR DEMOCRACY & TECHNOLOGY

FOR MORE INFORMATION

Please contact: Andrew McDiarmid, (202) 637-9800, x305, andrew@cdt.org;
David Sohn, x317, dsohn@cdt.org;
Alissa Cooper, x110, acooper@cdt.org.

[42] See note 4.

[43] The court could also exercise oversight by making adoption of privacy protections a condition of approval of the settlement, or by proposing amendments to the settlement agreement. However, these options are more likely to result in delays or threaten the tentative settlement.