

**Thoughts on Privacy and the Google Book Settlement:
What's At Stake, Why We Need to Advocate, and What We Can Do**
Rough draft of remarks by: Michael Zimmer

The Google Books Settlement and the Future of Information Access
UC-Berkeley School of Information
August 29, 2009

For those not familiar with my work, I am largely concerned with the ethical dimensions of technology, especially as they impact intellectual freedom, intellectual privacy, and access to knowledge, and the intersections between these three essential human values. Particularly, I have written and blogged extensively on how our intellectual privacy is increasingly threatened as more and more of our information-seeking activities move online and become consolidated in a single source, like Google.

I like to think of this in terms of *spheres of mobility*, where individuals have historically enjoyed the ability to engage in social, cultural, and intellectual activities free from answerability and oversight. These spheres might be geographic (escaping the wild frontier, taking a road trip towards self discovery), intellectual (the protected sphere of the library or the relative anonymity of reading a newspaper in a cafe), or digital (the sometimes-utopian promise of cyberspace, where ideas are sought & exchanged effortlessly).

Within these spheres, individuals have historically enjoyed the presumption of liberty and autonomy, of un-answerability, self-determination, and self-definition. Individuals create, discover, and enjoy spaces for personal growth, exploration, and escape within these spheres of mobility. Whether on the highways, in the library, or on

the Internet, spheres of mobility provide the means to break down barriers, expand our horizons, offer new insights, and lead us into new directions.

Web search engines, as the center of gravity of information-seeking activities, represent the latest addition to these spheres of mobility, providing an interface to new worlds of information, new spaces for communication, and new means of experiencing the world. In many ways, the physical, intellectual, and digital mobilities of the past converge in Web search engines, providing new means of physical escape, of intellectual exploration, and digital freedoms.

Given the centrality of search engines in our information-seeking lives, my concern focuses on the emergence of a large-scale infrastructure to track user search activity through cookies, user accounts, and co-mingled services. To make a long story short, the infrastructures of dataveillance constructed by search engines threaten our ability to navigate, to inquire, and to explore anonymously and free from oversight, constraining our spheres of mobility, and limiting our intellectual privacy and intellectual freedom.

These are the reasons why we were so upset about the AOL search data release. And why are so concerned over data retention policies for search engine logs. Why we pushed for short-term expiration dates in search tracking cookies. Why we argue that one's IP address might be "personally identifiable information." This is why we push companies like Google – who I do believe have the users interest in mind – to be even more proactive, more transparent, and more, for lack of a better term, ethical, when it comes to users' intellectual privacy.

So, what of the proposed Google Book Settlement? The speakers before me have highlighted the key issues and concerns regarding the possible impact of GBS on intellectual privacy, and it would be hard for me to improve on their remarks. Suffice it to say that given these concerns, GBS represents a very potent threat to our spheres of mobility.

Building from what has already been said today, I do wish to highlight two key factors that make the privacy concerns of the proposed GBS even more problematic than search query logs in general.

First, as Angela has explained to us, there are particular expectations of privacy when it comes to seeking information in a library setting. The context of the library brings with it specific norms of information flow regarding patron activity, including a professional commitment to patron privacy. But the insertion of Google into this context raises an alarm.

Certainly, searching for information on the World Wide Web was a new experience for most everyone. It was a new frontier of information-seeking, which developed its own business model, its own technical infrastructure, and its own technical standards and best practices, ones that rely heavily on the tracking and capturing of user data. The Web isn't the library, and many simply accept that in order to benefit from what companies like Google have created, we need to acquiesce to this kind of tracking of our searches. I don't agree, but that's a debate for another time.

However, with the proposed GBS settlement, we are talking about a more direct transfer of library practices to a Web-based infrastructure powered by Google. It will be reasonable to expect the same informational norms that exist in the library settings –

limited tracking, short-term data retention, possibilities of anonymous browsing – to translate into the proposed digital system for browsing books. If these are the expected norms of information flow, we must ensure they are respected by any system designed for browsing digitized books. In short, the system must be conceived as an extension of the library – with its informational norms in tow – and not an extension of Google, burdened with its standard practice of gathering user information.

Second, there will be limited competition in the proposed digital book-browsing universe. With web search engines, we saw evidence of how a competitive marketplace can have positive impacts with regard to user privacy. In recent years we've witnessed how Google, Yahoo and Microsoft tightened their data-retention practices in reaction to public (and government) pressure, with each not wanting to be the one with the longest retention period. To try to gain a competitive advantage, Ask.com created AskEraser that effectively removes your search activity from their logs. And the new Microsoft-Yahoo deal holds promise of bringing new competitive pressures to search privacy practices.

Yet, with the proposed Google Books Search settlement, there will be little-to-no competitive pressure to design innovative privacy-protecting measures. Google will not be competing with a rival service whose privacy practices might be more attractive to users. For those who argue we don't need privacy regulation because the marketplace works, the Google Book Settlement provides no such market for users to select a service provider based on their privacy preferences, leaving us at the mercy of the practices of a single provider of digital book services.

So, where do we go from here. We can look at what Google is promising, we can look at their history, and we can look at ways to design GBS in ethically-conscious ways that empower users.

First, Google has responded to many of the concerns expressed here today. They note that Google Account will *not* be required for general browsing, they won't sell user data, and that the general principles of Google's privacy policy will apply to book search data. Google also notes that the settlement agreement is intended to address the complicated IPR issues, and not user privacy, thus the absence of any mention of privacy protections or practices.

So, we are asked to trust Google. I want to trust Google – I've had the pleasure of meeting many Googlers, and they're all good people trying to do the right thing. Yet, Google's track record when it comes to protecting user privacy is mixed, and causes me some pause.

Google says they are dedicated to "always giving users clear information about privacy." Yet, it took a surprising amount of pressure and arm twisting to get Google to simply place a 7-letter hyperlink to their privacy policy on their homepage. Their level of resistance to this minor request doesn't instill a lot of confidence about this "dedication".

Google also states they are "thinking hard about how best to build privacy protections into the products authorized under the settlement". I know that Google is filled with really smart people capable of "thinking hard," but the problem is more about the direction of that thinking.

Consider Street View. Google thought hard about privacy for that product, and I know they included legal counsel on the product design team before that product went to

market. “Thinking hard” about privacy in Street View led Google to the conclusion that “The imagery is not different from anything each of us can photograph themselves” and that “The United States has “a long tradition of saying that it is legal and appropriate to take pictures from public spaces and publish them”

These statements are the result of legal “hard thinking”, and the privacy protections within Street View were limited, because US law didn’t require them.

Compare this to the European and Canadian versions of Street View. Here, “thinking hard” about privacy led Google to the conclusion that privacy must be protected by blurring faces....because the law required it.

So, it matters who is “thinking hard”. If it is a lawyer, we’ll get a legal solution to the privacy concerns (do only what the law requires). If it is an engineer, we’ll get a technical solution to privacy concerns (make the data secure). These are certainly necessary perspectives, but not sufficient. If we have an ethicist doing the “hard thinking” at Google, we would have had a very different release of Street View in the US...perhaps no release at all.

It becomes imperative, then, that the “hard thinking” about privacy and GBS comes from a wide range of disciplines and perspectives, ranging from librarians to engineers to ethicists. I again call on Google to add ethicists to their design teams; I’m happy to volunteer, and know there are people much smarter than I ready to help out as well.

Google also reminds us that only after the settlement is approved – *if* it is approved – will they begin to actually design the proposed system and fully contemplate

what kind of privacy protections will be embedded within it. Again, they promise that “that whatever we ultimately build will protect readers' privacy rights, upholding the standards set long ago by booksellers and by...libraries”.

Again, I want to trust Google, and I want to help them with these design decisions so they can succeed with their venture and respect user privacy at the same time. I have 3 quick suggestions for Google to design privacy into the system, and one for the community of coders/hackers out there.

First, Google should build anonymization (proxies, TOR) into the system, and do some “hard thinking” to find a way to accomplish this while also respecting the permission limits proposed by the settlement.

Second, Google should apply the same privacy-protecting measures it employs with other “sensitive” user information – in Google Health, for example – with the Book Search product. This would include SSL encryption of searches to prevent “man-in-the-middle” attacks, encryption of log files held by Google, and so on. If users’ health data deserves special protections, so do their book searches that might very well be related to their health

Third, whatever special steps Google takes to protect the privacy of users Book search activities, they must – MUST – recognize that the same concerns and principles apply to general search activity, and Google’s overall privacy practices should be changed as well. This will be a hard sell for many, but one I intend to close.

And finally, I call on the coders and hackers in the room to help empower individual users. Some of you might be familiar with the TrackMeNot Firefox extension that sends ghost queries to the major search engines in an effort to pollute your “data

cloud” within their transaction logs. I’ve been chatting with Helen Nissenbaum and Daniel Howe, the two primary designers of TMN, and we agree that a TMN for Books might be in order. So, despite what steps Google takes to protect one’s privacy, TMN can be used to further obfuscate my books search history, and at least providing me some comfort that my real book search activity can’t be easily profiled or mined for reasons unknown to me.

In closing, I am hopeful about the levels of access to knowledge the Google Book Search product promises. Yet, as this panel has revealed, there are significant implications on privacy and intellectual freedom that demand serious attention by scholars, librarians, advocates, as well as Google. That said, I remain hopeful that we can find a way to “think hard” about these issues and arrive at an ethically-driven solution to provide online access to digital books that preserves the norms of information flow within the library context.

Thank you, and I look forward to our discussion.