Policy & Internet

www.psocommons.org/policyandinternet

Vol. 2: Iss. 4, Article 3 (2010)

Anonymity in a World of Digital Books: Google Books, Privacy, and the Freedom to Read

Elisabeth A. Jones, ***The Information School, University of Washington***
Joseph W. Janes, ***The Information School, University of Washington***

Abstract

With its Books project, Google has made an unprecedented effort to aggregate a comprehensive public-access collection of the world's books. If successful, Google's collection would become the world's largest and most broadly accessible public book collection—indeed, project leaders have frequently spoken of their desire to create a "universal library" (Toobin 2007). Still, the Google "library" would differ from established contexts for the provision of free, public access to reading materials—like public libraries—along several policy-related dimensions, of which perhaps the most glaring is its treatment of reader privacy. This paper teases out the specific differences in reader privacy protections between the American public library and Google Books, and what those differences might mean for the values and goals that such contexts have historically embodied. Our analysis is structured by Helen Nissenbaum's "contextual integrity decision heuristic" (2009), which focuses on revealing changes in informational norms and transmission principles between prevailing and novel settings and practices. Based on this analysis, we recommend a two-pronged approach to alleviating the threats to reader privacy posed by Google Books: both data policy modifications within Google itself and inscription of privacy protections for online reading into federal or international law.

Keywords: book digitization, e-books, privacy, anonymity, Google Books, public libraries

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

Recommended Citation:
Jones, Elisabeth A. and Joseph W. Janes (2010) "Anonymity in a World of Digital Books: Google Books, Privacy, and the Freedom to Read," *Policy & Internet*: Vol. 2: Iss. 4, Article 3.
DOI: 10.2202/1944-2866.1072
Available at: http://www.psocommons.org/policyandinternet/vol2/iss4/art3

Introduction: The Promise and Peril of Google Book Search

Would you like to be able to search the full text of every book ever published, all at the same time? Would you like to download *David Copperfield,* peruse *Pride and Prejudice*, take *King Lear* for a spin on your Kindle? The dream of a comprehensive, free, universally accessible digital library of the world's books conjures up some truly fantastical possibilities. And through modern large-scale book scanning initiatives like the Million Book Project, the Open Content Alliance, and, most notably, Google Books, this dream is moving ever closer to reality. Even just the first five libraries to sign up with Google Books—Oxford, Harvard, Stanford, the University of Michigan, and the New York Public Library—have committed to contributing an estimated 10.5 million unique books (Lavoie, Connaway, and Dempsey 2005). And those five now represent a mere fraction of the whole: Google's library partners now number in the dozens, drawn from eight countries in Europe, North America, and Asia, and a recent estimate places the potential size of the system-wide collection at nearly 130 million volumes (Google, Inc. 2010a; Taycher 2010). In fact, thousands of works in the public domain—like that Dickens, that Austen, that Shakespeare—are already available on Google Books (and elsewhere), full text, free of charge, just waiting to reach new audiences. And should the proposed settlement in the copyright lawsuit against Google Books[1] be accepted, Google will be able to make the literature of the twentieth century findable—if not fully accessible—as well (Picker 2009; Samuelson 2010).

On the whole, this revolution in the accessibility of books seems likely to benefit readers, authors, and Internet users alike. It will improve individuals' ability to educate and entertain themselves; it will help books, and especially obscure books, to find new audiences; it will let book-based information compete more evenly with the vast amount of other, variably credible information currently found online; and it will go a considerable distance toward making access to high-quality, reliable information more egalitarian worldwide[2] (Center for Democracy and Technology 2009;

[1] We will not go into great detail about the structure and parameters of the proposed settlement in this case (*Authors Guild, Inc., Association of American Publishers, Inc., et al. v. Google, Inc*.). It has been discussed at great length elsewhere, and much, if not most, of that discussion is available through The Public Index (http://thepublicindex.org/), a website dedicated precisely to that purpose.

[2] Although differences in access to the Internet are not inconsequential, they are much less vast than existing differences in access to physical library collections, where

© 2010 Policy Studies Organization

Doctorow 2006; Toobin 2007). Still, as many have noted, the project is far from perfect. First, it is ultimately controlled by a single corporation, which under the proposed settlement would become responsible for providing or restricting access to particular works, devising algorithms to sort book searches and price e-book licenses, and ensuring user compliance with various copyright-related content restrictions (Kahle 2009; Samuelson 2009; 2010; Vaidhyanathan 2005). Also, the current state of the metadata on the site—titles, authors, dates, series designations, etc.—is abysmal, and likely will be for some time (Nunberg 2009a; 2009b). And of course, there are still those who believe, settlement or no, that Google is perpetrating a giant theft of content and value from authors and creators (LeGuin 2009; Writers' Representatives LLC and Epstein 2009). Most pertinent for this paper, however, are claims that the project poses serious risks to reader privacy, and that these risks, in turn, threaten intellectual freedom and free expression worldwide (Center for Democracy and Technology 2009; Electronic Privacy Information Center 2009; Privacy Authors and Publishers 2009; Vaidhyanathan 2005). In this paper, we will tease out the particulars of the privacy threat posed by Google Books using a theoretical lens not yet applied in this setting: Helen Nissenbaum's theory of privacy as "contextual integrity" (1998; 2004; 2009).

In particular, we contend that to the extent that Google Books seeks to provide free-of-charge reading material to the general public—and, in fact, has taken on the mantle of the universal library in its public statements—a strong historical comparison can be drawn to the established home of free, public reading in the United States—the public library system. This analogy, then, can be used not only to highlight the weaknesses currently present in Google Books, but to suggest useful ways of correcting them. Drawing on the framework provided by Nissenbaum's "contextual integrity decision heuristic," detailed in the next section, this paper addresses the following three questions:

(1) How do the norms of information flow within Google Books differ from those within the public library context?
(2) What moral or political factors are implicated in these changes in norms?
(3) How might these changes support or detract from the values, goals, and ends wrapped up in the provision of free, publicly accessible books, as established in public libraries?

---

barriers exist based not only on geography, but also, typically, on institutional affiliation.

Building upon this analysis, we then offer recommendations for the protection of reader privacy in Google Books and other similar initiatives: both adjustments to internal project policies and the introduction of supporting legal structures to make those policies less malleable and more enforceable.

Theoretical Framework: Privacy as Contextual Integrity

In recent years, myriad novel technological systems and practices—from ubiquitous surveillance to public records digitization to data mining—have radically altered the shape and direction(s) of information flow about individuals. Pieces of information that might once have seemed innocuous or even banal—photos from an office picnic, lists of friends on Facebook, real estate investment histories—can now easily slide between social contexts, merge together in unexpected ways, and reach audiences we never imagined or intended. And these slippages have serious implications for privacy. As philosopher James Rachels points out, "there is a close connection between our ability to control who has access to us and to information about us, and our ability to create and maintain different sorts of social relationships with different people," and privacy is thus valuable precisely because it allows us to "maintain the variety of social relationships with other people that we want to have" (1975, 326). Privacy protects our ability to maintain a different relationship with our supervisor at work than with our spouse, with our priest than with the clerk at the grocery store. Yet, as we lose control over how our information—whether "intimate" or seemingly innocuous—flows within and between social contexts, this valuable aspect of privacy is diminished. And such slippages can make it difficult, if not impossible, to maintain the contextual separations necessary for healthy human social life, at both the individual and societal levels.

Building upon this contention, then, Nissenbaum suggests that "a right to privacy is neither a right to secrecy nor a right to control but a right to *appropriate* flow of personal information," and offers the "framework of contextual integrity" to "[make] rigorous the notion of appropriateness" (2009, 127, emphasis in original). Contextual integrity, she suggests, is "preserved when informational norms are respected and violated when informational norms are breached" (2009, 140). For example, the norm in a Catholic confessional is for the priest to keep everything he is told in strictest confidence; were a priest to begin shouting everything learned in confession in the public square, that would breach existing informational norms, and thereby the contextual integrity, of his penitents. It is simply not

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

appropriate for a priest to shout confessions in the public square; even if those confessions are not especially salacious or intimate, such public retransmission violates the established expectations of the confessional space.

This fundamental concept—the informational norm—is central to understanding the theory of contextual integrity, and thus merits further explanation. Informational norms, Nissenbaum explains, are made up of four parameters: contexts, actors, attributes, and transmission principles (2009, 140–141). Of these parameters, the first three—contexts, actors, and attributes—are relatively straightforward:

- *Contexts* form the backdrop for informational norms: the social setting in which they hold true. Contexts and informational norms are co-constitutive: contexts simultaneously shape and are shaped by the informational norms they include (Nissenbaum 2009, 140–141).
- *Actors* relevant to informational norms can be divided into three types: senders, recipients, and subjects. Information *senders* and *recipients* form the two poles of information transmission—one sends, the other receives—and could be individuals, groups, or entities like organizations or committees. Information *subjects*, however, are generally individuals. They are the actors to whom the information refers: quite literally, the *subjects* of the information transferred. Information subjects and senders are frequently the same person: we often share information about ourselves. Still, Nissenbaum suggests that "it is crucial to identify the contextual roles of all three actors to the extent possible," even if in some cases the roles overlap in the same individual (Nissenbaum 2009, 141–142).
- *Attributes* refer to "the nature of the information in question: not only who it was about, and to whom and from whom it was shared, but what it was about" (Nissenbaum 2009, 143). Different attributes—different *types of information*—are appropriate in different contexts. For example, it is perfectly appropriate, and often necessary, for doctors to ask patients very detailed questions about their bodies. However, it would generally be inappropriate for one's supervisor at work to ask the same questions (Nissenbaum 2009, 143–144).

Nowhere in this definition is a clear line drawn between "private" or "intimate" contexts or information types and "public" or "innocuous" ones. The pivot point for contextual integrity is not the circumscription of private spheres, or the concealment of intimate facts, but the observance of context-relative informational norms: ensuring that *appropriate* types of information

http://www.psocommons.org/policyandinternet/vol2/iss4/art3
DOI: 10.2202/1944-2866.1072

about *appropriate* subjects are revealed by and to *appropriate* actors in *appropriate* ways. And it is these "appropriate ways," in turn, that form the fourth, and most complex, parameter of informational norms—transmission principles.

Transmission principles, Nissenbaum asserts, may be the "most distinguishing element of the framework of contextual integrity." They function to constrain "the flow (distribution, dissemination, transmission) of information from party to party in a context. The … terms and conditions under which such transfers ought (or ought not) to occur" (2009, 145). Confidentiality, reciprocity, desert, entitlement, need—these are just a few of the terms and conditions we commonly place on the sharing or withholding of information in everyday life (Nissenbaum 2009, 145). These principles might be codified or informal, firmly established or in ongoing flux. For example, there are numerous laws constraining the ways in which U.S. law enforcement officers can gather evidence and what they can do with that evidence once collected; this is a highly formal system of transmission principles with a strong grounding in historical and legal precedent (Nissenbaum 2009, 146–147). At the other end of the spectrum, however, there are also many situations where the principles are fuzzy or uncertain—for example, when and to whom HIV status should be disclosed for public health, or the criteria one might use to decide what information to share on a Facebook profile.

In order to guide determinations of whether or not a new system or process ought to be questioned on privacy grounds, Nissenbaum systematizes these concepts into an actionable set of discrete analytical steps, called the "contextual integrity decision heuristic" (2009, 148). This heuristic provides a framework for identifying breaches in contextual integrity, and for assessing the ethical and political ramifications of such breaches in light of the values, goals, and ends of the relevant context—and it is this structure that will guide our analysis of the privacy issues raised by Google Books.

The Contextual Integrity Decision Heuristic

The contextual integrity decision heuristic has nine steps, divisible into two basic phases. In the first phase, a comparison is drawn between the new practice and a preexisting context that served a parallel function or functions—as, for example, we have done in posing an analogy between Google Books and U.S. public libraries. Within this comparison, then, the relevant norms of information flow (including subjects, senders, recipients, and transmission principles) must be specified, and the entrenched

© 2010 Policy Studies Organization

informational norms of the old practice must be used to highlight where the new practice departs from those norms. At this point, then, a prima facie judgment can be made regarding whether or not contextual integrity has been violated, based on the extent of departure from established informational norms. If the new practice is found to violate contextual integrity, however, this does not necessarily indicate an a priori need to reject the practice: in some cases, the benefits of breaching contextual integrity in a particular context may outweigh the risks or harms, or the change might be value-neutral (Nissenbaum 2009, 182). The second phase of the decision heuristic weighs these risks and benefits on moral and teleological terms.

Specifically, the second phase prescribes two sets of assessments: first, identification of the moral and political factors—e.g., justice, fairness, equality, social hierarchy, democracy—affected by the practice in question, and second, clarification of the values, goals, and ends of the context. This second assessment allows the moral and political factors identified in the first to be evaluated in light of the context's overall reason for being. For example, in a hospital, the central goal is to promote health. Thus, if a new system interferes with that goal, it should most likely be rejected, even if it provides other moral or political benefits. Consider: it might serve the moral end of fairness if emergency rooms were first-come first-serve instead of triage-based, but that would likely imperil the emergency room's overall goal of promoting health, since urgent, life-threatening cases would be forced to wait behind those with more minor complaints.

In the remainder of this analysis, we will apply the contextual integrity decision heuristic to Google Books. Specifically, we compare Google Books to a well-established context of free, public reading—the American public library system—in terms of context-relative information norms, and then assess the moral and political implications of the divergences in informational norms between these contexts for their shared goal of democratized access to information.

Contextual Integrity and the Future of Anonymous Reading

The first step in conducting a contextual integrity analysis is to establish the prevailing context for the types of information behaviors supported by the novel systems or practices under discussion. For Google Books, we have already noted our sense that the public library seems to be the most relevant

prevailing context. Nonetheless, it is worth considering other settings that might be candidates. For example, why not bookstores, or for that matter simply reading in public? The act of reading or browsing in a public library is not that different from reading in a park or community center or shopping in a bookstore, and given that Google is likely to eventually start selling e-books, the bookstore provides parallels in other ways. Yet, we stand firm on our selection of the library, for two reasons. First, if the proposed settlement is approved, Google will in fact begin to offer library services—both access terminals in public libraries and institutional subscriptions for academic libraries. These services will almost certainly supplant existing library offerings, making Google all the more central to the information access landscape (American Library Association et al. 2009). And second, Google itself has most frequently held up the *library*, and not any of these other settings, as the ideal toward which it intends to strive. From the very beginning of the project, Google representatives have spoken of the potential of Google Books to "[democratise] access to human knowledge" (Redmer 2007, 1) and ultimately create "a universal digital library" (Toobin 2007). And this message has been even more forcefully echoed by representatives of the libraries being digitized, who typically depict Google Books as a natural extension of their longstanding commitments and missions (e.g., Coleman 2006; Milne 2006). Since a library is what the company and its partners purportedly wish to construct, then, it seems only fair to judge Google Books against the standards of that context. And libraries, as our analysis will indicate, provide a strong—and demanding—standard for comparison regarding reader privacy and the legal protection thereof. In the sections below, we outline the parameters of the informational norms present in the public libraries—context, actors, attributes, and transmission principles—and compare these norms to those emerging within Google Books.[3]

Prevailing Context: Anonymous Reading in Public Libraries

Public libraries are, by their very nature, public physical spaces (though with increasingly virtual, online presence); they generally present an open, inviting, relaxed, and informal setting for information searching, asking questions, browsing, quiet reading, community gatherings and meetings, and even rest and contemplation. Most library buildings of any size are

[3] It is also worth noting that many of the norms—and laws—protecting privacy in libraries have analogues in the bookselling context (American Library Association et al. 2005; Barringer 2001).

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

composed of various areas and sections, each devoted to a particular type of activity or resource topic. Because of this, the location in which a patron is observed—at the reference desk, say, or in the travel books section—can be taken as an indication of interest (which may occasionally be notable, as with an adult alone in the children's area). As public spaces, libraries engender an awareness of the openness and visibility of one's activity, but it would be unusual to have the feeling of constant surveillance by staff, other patrons, or any outside entity.

Information Subjects, Senders, and Receivers

Within the public library space there are many potential *subjects* of information transmission—that is, after all, a core purpose of the space. However, for the purposes of this analysis, we will focus on *readers*[4]: individuals who use public libraries to select, peruse, and check out books. Information can be gathered on readers in various ways: most directly, through circulation records, but also by noting which books are left out for re-shelving, or simply through visual observation of what readers do—what sections they spend time in, what books they pick up and/or read.

The *senders* in this case exist on two levels. The primary senders are the subjects themselves: simply by interacting with the library, readers transmit information about their reading preferences and behaviors. And while most of this information is ephemeral—unless you have a stalker, nobody is likely to follow you around writing down what books you glance at as you walk down the history aisle—other bits of information are more persistent. In particular, circulation records, which track the books readers check out, provide documentation of reader behaviors. And this documentation has secondary senders: namely, the libraries—or perhaps more precisely, the staff members responsible for tracking circulation.

The *receivers* in a library setting are again twofold. First, there are those in a position to directly observe reader behaviors—mainly library staff and other patrons. And second, there are those who are granted access to library circulation records—typically only internal library staff, but also potentially law enforcement, parents of readers who are minor children, or others with a demonstrable, reasonable, and legal right to see them.

[4] Throughout this paper, we use the terms "reader," "user," and "patron" interchangeably to describe individuals who utilize public libraries and/or Google Books to find and/or use reading materials.

Information Attributes

The information *attributes* in this setting—the types of information available—are more varied. Informally, readers' activities within and around the library can provide particular sorts of information to those able to observe them. For example, staff or other patrons might happen to notice the titles on a stack of books you have stacked next to you at a reading table; they might see if you have a friend or a child with you; they may note in passing various physically observable characteristics like your sex, age, skin color, height, or weight. Still, as previously noted, these types of information, in a library setting, are all usually ephemeral; it would be very odd for a stranger to record any of this information, or even to remember it for any substantial amount of time. Of course, that could change if you were observed doing something out of the ordinary, like doing cartwheels down the aisle or stocking up on bomb-making manuals—then your fellow patrons might be more likely to remember you, or even report you to library staff or beyond.

Most of the official, documented information about readers in the public library setting relates to keeping tabs on which patrons have what books—a process that falls into two stages: (1) issuing library cards and (2) maintaining records of book circulation based on those cards. Several types of information are typically required simply to apply for a library card.[5] First, a potential cardholder must usually demonstrate that they reside in the library's service area by showing an official document like a driver's license, utility bill, or signed lease. Additional information requested would likely include the patron's name, address, phone number, date of birth, and perhaps email address or preferred mode of contact. All these fall into the category of "personally identifiable information" as referred to in many statutes.

The second type of documentation, the circulation record, is created when a person borrows an item. At that time, their patron record, populated with the information enumerated above, is linked to a record for the item, which contains data such as the item's title, author, publisher, date of publication, subject headings, and call number, as well as somewhat more exotic information like the item's physical size or how much it cost. Some bibliographic records, particularly for newer works, also include chapter titles and notes on the contents of the item.

[5] Since public libraries operate independently, under local control, there is no single, uniform standard for these operations, and as such there will often be variation as to specifics and details, but these principles and attributes would be widely familiar.

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

The circulation record linking the patron record and the item record thus contains all of this information—personal and item-related—as well as the date (and probably the time) borrowed, any special conditions on the loan, and the due date. In the predigital era, these records were on paper or another tangible format: often a card was removed from a pocket in the book when borrowed, and stamped with the due date and, in some cases, also the name of the borrower. Such systems, however, have largely given way to digital, and now networked, systems in the last two decades. Due to this shift, circulation records are now subject to all the potential risks and safeguards applicable to any digital record maintained in a distributed, networked system.

Circulation records are maintained, of course, as inventory control—to ensure that the library knows who has what items, and that they are returned in a timely fashion for other readers to use. They are also used to generate reports, and to calculate fines for overdue materials or penalties for lost or damaged materials. Once an item is returned in good order, the circulation record may be deleted and purged from the system. Though maintaining them could have benefits for the library in, for example, better understanding reading tastes and interests, loan patterns, and so on, in the years after the passage of the USA PATRIOT Act, which eased restrictions on government snooping in library records, a number of libraries have decided to forgo those potential benefits to forestall having to provide "old" circulation records in response to a future request (Regan 2004).

Principles of Information Transmission

The transmission of official library records—those pieces of information the library collects about the reading habits of its patrons—is governed by strong ethical and legal commitments to *confidentiality*. For a librarian to transmit patron reading records to anyone without a demonstrable and well-justified need to see them would violate not only the ethics of the profession, but in most cases also state law.

The American Library Association (ALA) adopted the "Library Bill of Rights" in 1939 to codify the profession's stance on matters such as intellectual freedom, censorship, and access to information. They have also issued a number of interpretations, expounding on the Bill of Rights. The "Interpretation" on the issue of privacy states that in "physical or virtual" libraries, "the right to privacy is the right to open inquiry without having the subject of one's interest examined or scrutinized by others" (ALA Council 2002, 1). The ALA Code of Ethics, intended to guide librarians in their work (though not in a legally or professionally binding way as with attorneys),

adds that librarians “protect each library user’s right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted” (American Library Association 1997). And these ethics are not limited to librarians per se: the Interpretation cited above goes on to say that “[everyone] (paid or unpaid) who provides governance, administration, or service in libraries has a responsibility to maintain an environment respectful and protective of the privacy of all users” (ALA Council 2002, 2).

These ethical commitments to privacy within libraries and library systems, moreover, are reinforced by a nearly unanimous collection of U.S. state statutes. Though no federal regulation specifically addresses the issue of library patron privacy in the United States (Adams 2005, 48),[6] 48 states and the District of Columbia have enacted laws requiring that these records be kept confidential—and the two remaining states, Kentucky and Hawaii, have the attorney general’s opinions (which function essentially the same way that court opinions do) to the effect that library circulation records merit privacy protection.[7] In general, these state laws prohibit the release of library circulation data or records to any unauthorized third party, except in response to a subpoena or order from a court of competent jurisdiction, or for use by staff or administration for the purposes of administration, management, or operation of the library itself; a few even provide specific exemptions for public library circulation records with regard to Freedom of Information Act (FOIA) requests (Chmara 2009).

There is considerable variety in the terms and scope of these statutory protections, especially with regard to the conditions under which records can be released and the breadth of the information protected; still, a few common themes emerge. First, many states provide for library patrons to consent to the release of their own records, and several also allow parents or guardians to access the circulation records of their minor children; in many cases parents co-sign minors’ library card applications, and assume responsibilities for fines and replacement costs for lost or damaged items (Chmara 2009). Further, several states provide conditions under which library records can be released to law enforcement, including situations

[6] Though the USA PATRIOT Act’s requirement that individuals and institutions must turn over “tangible things and records” in terrorism investigations does apply to libraries, this provision of the law has been broadly criticized—in fact, some elements of this provision were ruled unconstitutional based on a lawsuit in which several plaintiffs were librarians (“Opinion Decision and Order” 2007).

[7] Many of these statutes were enacted in response to the FBI’s “Library Awareness Program,” which during the 1970s attempted to recruit librarians to “report on any ‘foreigners’ using America’s unclassified scientific libraries” (Adams 2005, 54).

© 2010 Policy Studies Organization

where there is imminent danger of physical harm (Illinois), to protect public safety (Texas), in response to a court order (Iowa, Nevada, Utah), or to investigate crimes which took place within the library itself (Louisiana, Wisconsin). Tennessee also allows the release of patron records to collection agencies in cases where the library needs to seek reimbursement for lost, stolen, or overdue materials (Chmara 2009).

Though most of these statutes were written primarily to cover circulation records, some have been interpreted more broadly, to cover the use of various library resources and services. Arkansas and New York, for example, provide explicit protection for database searches, interlibrary loan, reference inquiries, photocopies, and the use of reserve or audiovisual materials, and North Carolina includes protection for any evidence of having used the library. All of this leaves the status of library Internet usage (including pages viewed, searches conducted, email, and, it might be noted, the use of Google Books on library terminals) blurry. This is an area where legislation has yet to catch up with the advance of technology, despite the longstanding presence of Internet access in virtually all public libraries in the United States (Bertot, McClure, and Jaeger 2008).

In most cases, the penalties or sanctions for violations of library privacy statutes are minor, and are civil rather than criminal in nature. Still, a handful of states do make unauthorized release a criminal act, prosecutable as a misdemeanor or petty offense (Chmara 2009). Regardless of the penalties, however, it seems reasonable to surmise that the mere fact of such laws' existence may act as a deterrent to dispersal of these records: most people, after all, tend to follow laws in any case, and in this case in particular, the laws merely reinforce the already-strong norm of confidentiality within the community—librarians—that they govern.

Informational Norms: Public Libraries versus Google Books

Shifting the context of free, public reading from public libraries to Google Books alters the informational norms relevant to that activity in several significant ways. Though certain elements—the information subjects and senders, many of the information types being transmitted—remain the same, others—the context, information receivers, and transmission principles—differ radically.

The primary information subjects and senders in Google Books, as in the public library, are the readers themselves. In both contexts, it is the reader's behavior that causes information to be transmitted. In the library, readers' browsing and borrowing habits telegraph information about them with varying degrees of formality and permanence; on Google Books,

readers transmit information about themselves by searching for particular terms, browsing through particular authors or subjects, clicking on particular results, or spending time on, copying, or printing out particular pages. The information attributes here are also quite similar to those in public libraries: in both cases, information is revealed about what a given reader is either reading or thinking about reading. There is a difference in degree, however: the information that readers transmit on Google Books is considerably more detailed—potentially at the level of pages or even words—and much more closely tracked than that which they might transmit in public libraries.

More different between public libraries and Google Books are the characteristics of the underlying contexts themselves. Google Books is a fully digital, Internet-based context. It entirely lacks physicality, and thus any physical cues. In a library, there is a reasonable chance that if someone is watching you, you will be able to see (or otherwise sense) them watching, and perhaps even watch them back; on Google Books there is no such reciprocity. Someone may be tracking you (and likely is), but you would have no way of directly sensing that this is the case. You cannot see who, if anyone, is watching you—much less watch what they are doing in return.

Further, because Google Books might be used anywhere one might use the Internet, the relative "publicness" of the context—and the expectations and behaviors that follow from that—becomes less clear. In a public library, while users may not always be—or expect to be—observed, there is always an awareness of that possibility. And for most people, this possibility of observation is enough to deter them from behaving in particular ways—for better or worse. Such social norms might keep certain individuals from watching pornography on library computers, but they might also keep readers from picking up or admitting interest in books on controversial or potentially embarrassing topics, like sexually transmitted diseases or locally unpopular political perspectives. Google Books, on the other hand, might be used all alone in one's home, one's office, or a desert island. There is no sense inherent to the use of Google Books that one must behave in particular ways or according to particular rules; it can thus *feel* entirely private, anonymous, and unobserved, which might allow individuals a greater sense of freedom than they might feel in a more decisively public, observable setting.[8]

Still, however anonymous and private Google Books might feel to users, this feeling is misleading, at least in its current state. In fact, virtually

[8] In this sense, Google Books is identical to the Internet in general: though there is an increasing awareness of the ubiquity of online tracking, many users continue to use the Internet as though nobody is watching (Fox 2000, 1–2; Madden and Smith 2010, 6).

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

all behavior on Google Books is tracked—though not necessarily tied to users' identities—and the receiver of all of that tracking information is Google. This centralization marks another departure in informational norms between public libraries to Google Books. In public libraries, most information transmission is diffuse—one person might see you walk down the sociology aisle, another might notice you sitting at a table reading Hemingway, yet another might glance over the titles of the DVDs you're returning—but as previously noted, short of having a stalker, it is unlikely that any one person will gather up all this information. And the more formalized information transfers in libraries—library card signups, circulation records—go only as far as the library computer systems and the occasional librarian, with various internal rules and external regulations preventing them from reaching further. On Google Books, however, all of this information—your browsing behavior, your reading selections, your user profile, your printing or purchase records, etc.—goes to Google, and whether to share that information with secondary recipients—law enforcement, advertisers, business partners, your contact list—is entirely up to Google, based exclusively on internal privacy policies, which are continually subject to change.

Finally, and most importantly, the principles governing transmission of information about users' reading behaviors in Google Books differ markedly from those in public libraries. One of these altered principles—*reciprocity*—has already been noted: in public libraries, if someone is watching you, you can watch them right back. In Google Books, by contrast, you might not even know that Google is tracking your every click—and you certainly can't track them in return. However, the most pressing shift in transmission principles relates to libraries' strong commitment to patron *confidentiality*. As discussed above, in public libraries, the primary recipient of information about readers—the librarian[9]—is bound by both professional ethics and state regulations to protect the confidentiality of patron reading patterns. The strength of this commitment has been expressed in a huge variety of forms and venues—from self-proclaimed "Radical Militant Librarians" sporting *t*-shirts declaring "*Scimus quae legis, et non dicimus*"[10] to the four librarians who went to federal court to challenge provisions of the USA PATRIOT Act that they felt unlawfully abridged reader privacy (American Library Association 2007; Instant Attitudes T-Shirts). Given the strength of these ideals and the expressed passion behind them, one suspects

[9] Here we use "librarian" as shorthand for library staff in general. Not all library staff are librarians, but they are all subject to the same legal and ethical constraints with regard to patron reading records.

[10] "We know what you read, and we're not saying."

DOI: 10.2202/1944-2866.1072

that even if no laws prohibiting the release of library records existed, librarians would go to the mat protecting them anyway. Still, the laws do add force and substance to these principles, and reinforce librarians' practical ability to uphold their ethical standard of confidentiality.

For readers on Google Books, the maintenance of confidentiality is much less certain. Google does have a privacy policy; in fact, it has many. In addition to its general policy, many of its specific products and services, including Books, have their own. The Google Books privacy policy explicitly states that Google will collect several types of information, including "the query term or page request (which may include specific pages within a book you are 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," and reserves the company's right to aggregate usage data from Google Books with other data linked to users' Google Accounts—so your book purchase history or personalized reading lists may be combined with your usage data from Google Search, Gmail, Google Reader, Google Maps, Picasa, or any of the company's myriad other services (Google, Inc. 2009a).

Further, the general Google privacy policy lists three conditions under which any or all of this data might be shared with third parties: (1) if Google has your direct consent; (2) for the purpose of processing the information on Google's behalf; or (3) if the company 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 (d) protect against harm to the rights, property or safety of Google, its users or the public as required or permitted by law (Google, Inc. 2009b).

Though the first and second conditions listed strongly resemble existing stipulations placed on release of library records, the third is considerably broader: in addition to compliance with legal warrants, this policy grants Google the right to share user information with third parties based on a quite vague, and apparently internally determined, standard of "good faith"

© 2010 Policy Studies Organization

suspicion.[11] And although the Books policy does also acknowledge the existence of "special books laws" in some jurisdictions, and maintains that where such laws apply, the company will attempt to use them to protect the privacy of Google Books's readers (Google, Inc. 2009a), Google's willingness to act on such pledges remains untested.

Additionally, as has been noted elsewhere (e.g., Center for Democracy and Technology 2009, 7; Grimmelmann 2009, 16; Privacy Authors and Publishers 2009, 21), these privacy policies are all non-binding, and thus fully dependent on the goodwill of a large, public corporation—not a reassuring prospect.[12] Where in public libraries a highly developed system of ethical principles and legal protections prevents the vast majority of reader information from going beyond its primary recipient (the librarian or library system), this is simply not true for Google Books, both because their existing privacy policies already allow for a broad range of sharing and secondary use and because even the protections that they *do* offer lack enforceable stability.

Prima Facie Assessment: Google Books Breaches Contextual Integrity

As should be clear from the discussion above, we would assert that Google Books breaches the informational norms surrounding free, public reading as established in the U.S. public library context, by (1) increasing the kind and amount of information about readers being formally tracked, (2) shifting the direction of primary information flow from libraries and librarians to Google itself, (3) expanding the set of potential directions for secondary information flow, and (4) altering the transmission principles of the setting to exclude reciprocity and diminish confidentiality. In breaching these norms, in turn, Google Books represents a prima facie violation of contextual integrity—and thus privacy—for readers.

[11] One wonders, as well, how far a liberal interpretation of protecting Google's "rights, property, or safety" might extend—does that include not using Google services to create things that compete with—or perhaps even criticize—Google?

[12] And this particular corporation has recently demonstrated an increasing tendency to alter or even reverse its stances on significant policy issues to suit the business prerogatives of the moment (e.g., Crovitz 2010; Price 2010).

DOI: 10.2202/1944-2866.1072

Evaluating the Breach

It would be overly conservative, however, to conclude that all violations of informational norms are negativeper se: sometimes norms can—and should—change, or must be overridden by other, more important values.[13] As Nissenbaum suggests, "if a way can be found to demonstrate the moral superiority of new practices, this presumption [of objectionability] could be overcome and what was recognized as a prima facie violation may be accepted as morally legitimate" (2009, 164). In order to determine whether or not this is the case for Google Books, we must follow the contextual integrity decision heuristic through its second phase: assessment of the moral and political issues affected by the shift, and how those moral and political issues factor into the ends, goals, and values of free, public reading.

Moral and Political Considerations

In considering how to judge the alterations Google Books will make in context-relative informational norms, we must first identify the fundamental values at stake in the context in question. For free, public reading, we suggest that the most relevant moral tenets are (1) freedom, (2) autonomy, and (3) justice, which in turn have implications for the future of certain political values—especially intellectual freedom and freedom of expression.

The first two of these values, freedom and autonomy, are strongly intertwined; indeed, certain kinds of freedom are often cited as fundamental prerequisites for the condition of autonomy. For example, Mendus suggests that autonomy requires "that the agent possess not only the standard negative liberties (freedom from coercion, constraint, and threat of punishment), but also freedom from the suffocating constraint of social mores and customs" (1986, 108). Both freedom from external oppression and freedom from subtler internalized coercion thus factor in to individuals' ability to conduct themselves as autonomous, self-actualizing beings. And protecting reader privacy, we contend, supports both of these types of freedom, and thus autonomy, by reducing the impact of such internal and external constraints on the breadth of readers' inquiry.

[13] The classic example here would be slavery; other examples might include the disenfranchisement of women/non-whites, the colonial concept of "white man's burden," etc. Less extreme examples occur frequently throughout the legal system whenever an existing precedent is found to be unjust or no longer in keeping with societal values—for example, *Lawrence v. Texas* (2003), which ruled state sodomy laws unconstitutional throughout the United States, or more recently *Kalman v. Cortes* (2010), which overturned a Pennsylvania law against blasphemy.

© 2010 Policy Studies Organization

Inhibitions on inquiry may flow from a number of sources: some topics are themselves broadly controversial or taboo (e.g., terrorism, erotica); some reading materials might lead others to draw conclusions about the reader (e.g., self-help books, pregnancy or child-rearing reference materials, get-out-of debt guides); some readers may simply have tastes or preferences they would rather not broadcast (e.g., the varsity football player reading *Sweet Valley High*, or the high-powered ad exec reading Harlequin romances). It is not that any of these readers have "something to hide" per se; it is simply the case that sharing information about ourselves, no matter how innocuous, opens us to judgment—and in those judgments lie the power for others to coercively shape our behavior, thus diminishing our freedom and autonomy. By shielding readers from this kind of coercion, privacy protects their freedom to seek whatever information they need or desire, as well as the underlying autonomy of thought that allows for the development of those needs and desires in the first place.[14]

Moreover, within the U.S. legal system, freedom of inquiry has long been recognized as a fundamental element of free expression—itself perhaps the most strongly held belief enshrined in the Bill of Rights of the U.S. Constitution. As the Electronic Privacy Information Center points out,

> An American right to send and receive information anonymously is as old as the country itself; even the Federalist Papers, a collection of 85 essays written to support ratification of the Constitution, were all signed pseudonymously by the Founding Fathers (2009, 15).

And this belief underscores the laws protecting library records confidentiality: if readers cannot be sure that their reading records will be kept confidential, the argument goes,

> they may be unwilling to ask questions, perform a search, read a book on the premises, or check out a book on a controversial

[14] An offshoot of this argument has been offered in the "Privacy Authors and Publishers" comments on the proposed settlement. The essential issue raised by these authors and publishers—each of whom have published books that particular segments of society have found objectionable—is that if readers' ability to read potentially controversial books is chilled, then (a) their book sales will suffer, and (b) as readers, they will lose some of their own freedom to read (Privacy Authors and Publishers 2009). Thus, in a certain way, interfering with reader autonomy may also interfere with these authors' ability to express themselves as freely as they would like (i.e., by expressing potentially objectionable ideas), because it deprives them of an audience.

DOI: 10.2202/1944-2866.1072

> subject for fear of judgment by the community they live in or society at large, or for fear of retribution by the government (Bowers 2006, 377).

These kinds of chilling effects on inquiry, in turn, reduce not only the diversity of viewpoints available on various issues, but the depth and complexity of opinion and critique that individuals can develop on those issues: it reduces their ability to participate in the political process as autonomous agents.

This last point—that speech and inquiry are inextricably linked, and that this has implications for the autonomous formation of opinions—is one that has been made forcefully elsewhere by Julie Cohen, who writes that,

> Thoughts and opinions, which are the predicates to speech, cannot arise in a vacuum. Whatever their content, they are responses formed to things heard or read. It is this iterative process of "speech formation"—which determines, ultimately, both the content of one's speech and the particular viewpoint one espouses—that the First Amendment should shield from scrutiny (1996, 1008).

And so, for the most part, it has. Indeed, Cohen goes on to assert that:

> The freedom to read anonymously is just as much a part of our tradition, and the choice of reading materials just as expressive of identity, as the decision to use or withhold one's name [e.g. from a political pamphlet]. Indeed, based purely on tradition, the freedom to read anonymously may be even more fundamental than the freedom to engage in anonymous political speech. Anonymous advocacy has always been controversial. Anonymous reading, in contrast, is something that is taken for granted (1996, 1014).

The autonomy to freely cast about for ideas to inform our opinions and statements, without worrying about who might be looking over our shoulder, is a basic assumption underlying free expression. If we self-censor our inquiry, we will have less to say—the very definition of a chilling effect on speech.

The moral value of reader privacy, beyond its benefits to individuals, can also be seen in the support it provides for the perpetuation of justice within social systems. Justice, broadly speaking, describes the fair

© 2010 Policy Studies Organization

distribution of rights and privileges among actors within a social context, based on such criteria as equality, desert, or harmony (Lamont and Favor 2007). Reader privacy, we submit, protects this value by maintaining a more balanced distribution of power between readers and the various social actors who might be interested in what they read. As Solove suggests, whenever institutions—governmental or otherwise—gather information about individuals, there is a risk of creating an unjust power imbalance between the information gatherer and the information subject. And this risk, he argues, grows along two dimensions relevant in this context: (1) decreasing transparency and (2) increasing potential for secondary use.

Information gathered about reading habits within systems like Google Books, as on the broader Internet, is generally non-transparent to the user. Indeed, it can be very difficult for Internet users to even determine what kinds of information are gathered about them, who is doing the gathering, how broadly they are sharing that information, or how long they intend to keep it.[15] This simple fact—that online information gathering is non-transparent (and non-reciprocal, as noted earlier)—creates a power imbalance between Internet companies and Internet users. The companies know everything about us; we know next to nothing about them—and worse, we know very little about what they might be doing with their vast stores of our personal information.

Second, and perhaps more acute, is the power imbalance engendered by secondary information transmissions—when an information gatherer decides, perhaps even in contravention of its own stated policies, to share information with a third party, such as a government agency or business partner.[16] As Solove argues,

> there is a social value in ensuring that companies adhere to established limits on the way they use personal information. Otherwise, any stated limits become meaningless, and companies have discretion to boundlessly use data. Such a state of affairs can leave nearly all consumers in a powerless position.

---

[15] In some cases, these kinds of parameters are included in a site's privacy policy; however, most Internet users seem to either not read or not understand such privacy policies. For example, in a 2008 survey of Californians conducted by researchers at the University of California, Berkeley, over half the respondents misinterpreted the protections that privacy policies offered (Hoofnagle and King 2008).

[16] Solove cites the example of airlines sharing passenger information with federal agencies after the 9/11 attacks, for the purpose of studying airline security, in direct conflict with the airlines' prior confidentiality statements (Solove 2007, 769).

DOI: 10.2202/1944-2866.1072

> The harm, then, is less one to particular individuals than it is a structural harm (2007, 770).

If we cannot trust institutions to keep their promises to individuals with regard to privacy and confidentiality, this can lead to the breakdown of the layer of trust necessary for social order. In the context of reading, strong—and legally enforceable—privacy safeguards thus help to balance the scales: whatever information libraries gather about their patrons, they are both ethically and legally prohibited from sharing that information any further. Not only *would* they not say what you read; they *could* not if they wanted to—and the power imbalance between information gatherer and information subject in that context is thereby greatly reduced.

Impact on Contextual Values, Ends, Purposes, and Goals

The fundamental goal of the American public library has for more than a century been to support the freedom of inquiry, and thereby the freedom of expression, necessary to the functioning of a free society. Indeed, for more than 150 years, the rationale for public libraries has remained essentially identical to its description in the preamble to the first Massachusetts public library law, which proclaimed that:

> a universal diffusion of knowledge among the people must be highly conducive to the preservation of their freedom, a greater equalization of social advantages, their industrial success, and their physical, intellectual and moral advancement and elevation: and … there is no way in which this can be done so effectively, conveniently and economically as by the formation of Public Libraries (Ditzion 1947, 18–19).

By stripping away many of the traditional safeguards on reader privacy—whether legal, ethical, or situational—shifting free-of-charge, publicly available reading from libraries to Google Books complicates the capacity of the context to support truly unfettered inquiry and knowledge diffusion. For all the reasons already noted—controversial interests, the ability of reading material to reveal other things about the reader, or pure embarrassment—a lack of privacy with regard to the selection of reading materials can significantly chill individuals' desire and/or ability to explore as broadly as they might wish. In this way, Google Books's breach of contextual integrity threatens not only to violate privacy, but also to obstruct the fundamental goals of its context—the provision of publicly available reading as

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

historically offered in American public libraries. For this reason, we would have to conclude that in its present state, the contextual integrity decision heuristic recommends against a wholesale shift of free, public reading from libraries to Google Books: not only do its alterations to established informational norms imperil freedom, autonomy, and justice, they also interfere directly with the historic purpose of providing cost-free reading material to the public in the first place—the encouragement of an expansive, uninhibited quest for knowledge among as broad a segment of the population as possible.

Recommendations

So what is to be done? As indicated at the outset of this analysis, we believe that Google Books has amazing positive potential—and that potential is worth fighting for. Given this, however, what structures might be put in place to help Google's "universal library" perpetuate the social benefits ascribed to the institution it takes as its model? Based on the analysis above, we propose a twofold approach. First, we suggest that there are a variety of institutional and technological modifications that Google could make to better safeguard reader privacy, including reduction of the scope and quantity of data it gathers about readers and augmentation of the limits it places on the sharing of that data. Second, given that corporate privacy policies, however beneficent-seeming, are neither stable nor independently enforceable, we suggest that some form of public regulation is needed to protect the privacy of reading online—analogous to current state library privacy laws, but likely at the federal or international level.

Institutional Recommendations

Many of the internal adjustments that Google itself could make have been suggested elsewhere, especially in the responses to the dispute over the proposed settlement agreement in the Google Books copyright lawsuit ("The Public Index: Amended Settlement and Responses"). Most of these responses suggest that privacy protections ought to have been written into the settlement itself (e.g., American Library Association et al. 2009; Electronic Privacy Information Center 2009; Grimmelmann 2009; Privacy Authors and Publishers 2009). However, we would join the Center for Democracy and Technology in noting that "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"—it simply may not be the best or most appropriate venue for such a discussion (Center for Democracy and Technology 2009, 3). Still, while the settlement remains under discussion, it would certainly be worth considering writing into it some version of the recommendations that follow.

Whether or not privacy ultimately comes to be directly addressed in the settlement, however, we would suggest that Google take steps on its own to improve the reader privacy protections in its Books product. In particular, we contend that the departures from established informational norms described above, especially as they relate to tracking and secondary use, suggest two sets of ways in which Google might reduce the threat that its "universal library" currently poses to user privacy.

Recommendation 1: Google Should Reduce the Type and Amount of Information it Tracks and Stores About Readers. Especially within the last decade, libraries have gone to great lengths to reduce the quantity and identifiability of data they collect about their patrons, despite the fact that doing so may often reduce their ability to perform various valuable internal functions (Regan 2004). Though we recognize that Google also derives enormous value from the data that it is able to gather about its users, we suggest that it should make a similar concession to the interests of online readers. For Google Books, the company's financial interest in user data should be carefully weighed against the ethical and political risks that such data gathering entails for those using the product, and also against the broader social implications of those risks. The company should ask itself what data it actually needs about its users, how identifiable that data has to be, and how long they absolutely must retain that data—and beyond what is absolutely necessary for Google Books to function both technically and legally, they should refrain from collecting data about online readers (however innocuous that data may seem). And for the data that *is* deemed necessary, the company should make it perfectly clear to users what is being tracked and why, and should set strict and transparent limits on the duration of its retention—perhaps 90 days, as is the industry standard (Center for Democracy and Technology 2009, 13).

Recommendation 2: Google Should Increase the Stringency of its Limitations on Secondary Data Use. Currently, Google's privacy policy—and its additional policy for Google Books—outlines a wide variety of uses Google might make of user data, as well as a relatively broad set of circumstances under which the company reserves the right to share user data with third parties. And, of course, these policies are subject to change at any time. All of this is par for the course for online privacy policies: one would

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

expect to see similar provisions on websites run by Amazon, eBay, or any other major Internet company. However, we suggest that the privacy policy for Google Books should be grounded on a different precedent: the analogous policies stated by public libraries. The Seattle Public Library's privacy policy, for example, states that,

> Staff members and volunteers shall protect information about Library borrowers, their requests for information and materials, the online sites and resources they access, and their loan transactions, and s*hall not transmit such information to individuals or to any private or public agency without an order from a court of competent jurisdiction, or as otherwise required by law* (2002, emphasis added).

Such language is fairly standard across U.S. public libraries: the standard for deciding whether to share patron information is nothing short of a court order. Especially given Google's expressed hope that its Books project will assume many public library functions—and the fact that some of those functions are even written into the proposed settlement—we would recommend that they also consider adopting the restrictions those institutions have placed on their willingness and ability to share user data as well.

Legal Recommendations

If Google were to make the institutional adjustments suggested above, it would go a considerable distance toward diminishing the extent to which Google Books might violate reader privacy based on a contextual integrity analysis. Yet, neither the opportunity nor the responsibility to protect reader privacy stops with Google. Such protections must also have the force of law. For this reason, we recommend the adoption of legal structures that will provide online readers with protections analogous to those they currently enjoy offline, in libraries. A legal framework for the protection of reader privacy online would most likely have to be constructed at the federal level—if not the international level—due to the geographic dispersal of the readers in question: as previously noted, Google Books readers could be anywhere there is a live Internet connection.[17] Still, the principles could

[17] Though the actual availability of books will vary according to the copyright laws of the countries of use, Google Books will at minimum be offering full-text access to public domain works (as locally defined), and full-text search of all books not explicitly opted-

DOI: 10.2202/1944-2866.1072

remain much as they already stand in U.S. state library laws: protection of individuals' right to read without fear of oversight or retransmission, as a necessary safeguard to those individuals' rights to free inquiry and freedom of expression.

To be clear, our call for regulation here does not emerge from any specific feeling of distrust for Google (or any passionate love for regulation). Trust Google or not, such sentiments are irrelevant to the need to provide a stable, enforceable foundation upon which to ground the basic rights of readers. Consider, again, the American public library. The level of trust it inspires with regard to privacy is virtually unparalleled, on the basis of the profession's ethical principles alone: if the average person could trust anyone with their information, it would be a librarian. Yet, 48 states and the District of Columbia still passed laws prohibiting the release of library patron records. This was not a statement of skepticism with regard to library ethics, but rather an affirmation that the protections offered by those ethics *deserve the force of law*.

Such reification of institutional privacy protections into legal ones, moreover, helps readers and information providers alike. It helps readers in all the ways noted throughout this paper: by providing them with an enforceable assurance that their choice of reading material will remain confidential; by helping them to maintain a greater freedom in their interests and choices with regard to informational exploration; by facilitating a more equitable balance of power between readers and information providers. But perhaps more intriguing, to the extent that an information provider has the desire to conform with particular ethical principles, it can help that information provider as well. Laws protecting library circulation records not only protect the patrons to whom those records refer, after all, but also the ability of librarians to withstand various kinds of pressure to release those records to third parties. In the case of Google Books, the company has stated its desire to protect reader privacy. A legal structure protecting such privacy online would reinforce their ability to do so, no matter who—repressive governments, potential business partners, warrantless federal agents—is doing the asking.

---

out by their copyright holders (with varying levels of text preview available, including no preview) (Google, Inc. 2010b).

© 2010 Policy Studies Organization

Conclusion

The Google Books project has the capacity to change the shape and nature of reading worldwide for years to come. Along with other, similar projects, it promises to place hundreds of millions of books into the hands (or at least, onto the screens) of readers who might otherwise never have dreamed of such access to knowledge. But this amazing promise comes with important responsibilities, for both Google and its surrounding regulatory environment. In this paper, we have argued, based on Nissenbaum's contextual integrity decision heuristic, that in its current state, Google Books represents a significant threat to reader privacy, and that this threat implicates readers' freedom and autonomy, as well as the maintenance of a just balance of power between readers and those who would monitor their reading. And these implications, in turn, imperil the fundamental goal of providing free, publicly available reading materials, as expressed over the long history of American public libraries: support for free inquiry as an essential element of citizenship in a free society.

Still, we remain hopeful. Indeed, particularly if some version of these recommendations were followed, Google Books could even become a better servant to the fundamental end of the public library than the public library is itself, in terms of both scope and—surprise!—confidentiality. First, the goal of diffusion of knowledge is one in which size matters—and the Google Books collection will ultimately offer access to more books than any public library could even dream of owning, in a virtual space accessible in any place and time one can go online. And second, as suggested earlier, the contexts in which one might use Google Books are as broad as the contexts in which one might use the Internet itself, and those contexts include ones of near-absolute privacy, like dark closets, locked offices, and desert islands. Where in a public library, even with all its rules and protections, there is nothing to prevent another patron—who might well be a friend, neighbor, or family member—from simply *noticing* what you are reading (and judging you for it, or even telling others), on Google Books, nobody you know can directly observe you; you can search and browse and peruse as you will, without fear of judgment or approbation. Given stricter institutional and legal limits on what Google would or could do with your data, it could thus approach the ideal of completely confidential inquiry, while simultaneously providing access to a wider universe of books than possible ever before. The potential contribution of Google Books to knowledge and freedom worldwide is truly inspiring—but both the company and its surrounding regulatory institutions have a great deal of work to do if they hope to bring that potential to fruition.

References

Adams, H.R. 2005. *Privacy in the 21st Century: Issues for Public, School, and Academic Libraries*. Westport, CT: Libraries Unlimited.

ALA Council. 2002 "Privacy: An Interpretation of the *Library Bill of Rights*." Chicago: ALA. http://www.ala.org/ala/issuesadvocacy/intfreedom/librarybill/interpretations/privacyinterpretation.pdf (accessed August 13, 2010).

American Library Association. 1997. *American Library Association Code of Ethics*. http://www.ala.org/ala/aboutala/offices/oif/statementspols/codeofethics/codeethics.cfm (accessed April 6, 2009).

American Library Association. 2007. "USA Patriot Act: Doe v. Gonzales." http://www.ala.org/ala/aboutala/offices/oif/ifissues/usactlibrarians.cfm (accessed August 11, 2010).

American Library Association, American Booksellers Association, Association of American Publishers, and Pen American Center. 2005. "Campaign for Reader Privacy." http://www.readerprivacy.org/index.jsp (accessed October 19, 2010).

American Library Association, Association of College and Research Libraries, and Association of Research Libraries. 2009. "Library Association Comments on the Proposed Settlement." *The Authors Guild, Inc., Association of American Publishers, Inc.*, et al. *vs. Google, Inc.* Case No. 05 CV 8136-DC. http://thepublicindex.org/docs/letters/acrl_ala_arl.pdf (accessed July 10, 2010).

Barringer, F. 2001. "Using Books as Evidence against Their Readers." *New York Times*, April 8. http://www.nytimes.com/2001/04/08/weekinreview/ideas-trends-using-books-as-evidence-against-their-readers.html (accessed October 19, 2010).

Bertot, J.C., C.R. McClure, and P.T. Jaeger. 2008. "Public Libraries and the Internet 2007: Issues, Implications, and Expectations." *Library & Information Science Research* 30 (3): 175–184.

Bowers, S.L. 2006. "Privacy and Library Records." *The Journal of Academic Librarianship* 32 (4): 377–383.

Center for Democracy and Technology. 2009. "Privacy Recommendations for the Google Book Search Settlement." Washington, DC (July 27). http://cdt.org/copyright/20090727_GoogleRecs.pdf.

© 2010 Policy Studies Organization
Published by Berkeley Electronic Press

Chmara, T. 2009. *Privacy and Confidentiality Issues : A Guide for Libraries and Their Lawyers*. Chicago: American Library Association.

Cohen, J. 1996. "A Right to Read Anonymously: A Closer Look at 'Copyright Management' in Cyberspace." *Connecticut Law Review* 28: 981. http://ssrn.com/abstract=17990 (accessed October 19, 2010).

Coleman, M.S. 2006. "Google, the Khmer Rouge, and the Public Good." Address to the Professional/Scholarly Publishing Division of the Association of American Publishers (February 6). http://www.umich.edu/pres/speech/archive/060206google.php (accessed April 12, 2006).

Crovitz, L.G. 2010. "Google's China Search Result." *Wall Street Journal*, July 12, A13.

Ditzion, S. 1947. *Arsenals of a Democratic Culture*. Chicago: ALA.

Doctorow, C. 2006. "Why Publishing Should Send Fruit-Baskets to Google." *Boing boing: a directory of wonderful things*, February 14. http://www.boingboing.net/2006/02/14/why_publishing_shoul.html (accessed February 14, 2006).

Electronic Privacy Information Center. 2009. "EPIC's Motion to Intervene." *The Authors Guild, Inc., Association of American Publishers, Inc.*, et al. *vs. Google, Inc.* Case No. 05 CV 8136-DC. http://thepublicindex.org/docs/letters/epic.pdf (accessed July 10, 2010).

Fox, S. 2000. "Trust and Privacy Online." Washington, DC: Pew Internet & American Life Project (August 20). http://www.pewinternet.org/Reports/2000/Trust-and-Privacy-Online.aspx (accessed August 14, 2010).

Google, Inc. 2009a. "Google Books Privacy Policy." http://books.google.com/googlebooks/privacy.html (accessed February 25, 2010).

Google, Inc. 2009b. "Privacy Policy." http://www.google.com/privacypolicy.html (accessed February 25, 2010).

Google, Inc. 2010a. "Google Books Library Partners." http://books.google.com/googlebooks/partners.html (accessed August 9, 2010).

Google, Inc. 2010b. "Perspectives: Facts & Fiction." http://books.google.com/googlebooks/facts.html (accessed October 19, 2010).

Grimmelmann, J. 2009. "How to Fix the Google Book Search Settlement." *Journal of Internet Law* 12 (10): 1–20. http://works.bepress.com/james_grimmelmann/23/.

Hoofnagle, C.J., and J. King. 2008. "Research Report: What Californians Understand About Privacy Online." http://ssrn.com/abstract=1262130 (accessed August 13, 2010).

Instant Attitudes T-Shirts. "We Know What You Read, and We're Not Saying." http://www.instantattitudes.com/shirts/t052.html (accessed August 11, 2010).

Kahle, B. 2009. "A Book Grab by Google." *Washington Post*, May 19. http://www.washingtonpost.com/wp-dyn/content/article/2009/05/18/AR2009051802637.html (accessed May 19, 2009).

Lamont, J., and C. Favor. 2007. "Distributive Justice." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. http://plato.stanford.edu/entries/justice-distributive/ (accessed August 13, 2010).

Lavoie, B., L.S. Connaway, and L. Dempsey. 2005. "Anatomy of Aggregate Collections: The Example of Google Print for Libraries." *D-Lib Magazine* 11, no. 9 (September). http://www.dlib.org/dlib/september05/lavoie/09lavoie.html (accessed February 8, 2007).

LeGuin, U.K. 2009. "Petition Concerning the Google Book Settlement." *The Authors Guild, Inc., Association of American Publishers, Inc.*, et al. *vs. Google, Inc.* Case No. 05 CV 8136-DC. http://thepublicindex.org/docs/amended_settlement/LeGuin.pdf (accessed July 28, 2010).

Madden, M., and A. Smith. 2010. *Reputation Management and Social Media*. Washington, DC: Pew Internet & American Life Project (May 26). http://www.pewinternet.org/Reports/2010/Reputation-Management.aspx (accessed August 14, 2010).

Mendus, S. 1986. "Liberty and Autonomy." *Proceedings of the Aristotelian Society* 87: 107–120. http://www.jstor.org/stable/4545058.

Milne, R. 2006. "Upholding Bodley's Vision: The Google Mass Digitisation Project at Oxford." *Art Libraries Journal* 31 (3): 5–10.

Nissenbaum, H. 1998. "Protecting Privacy in an Information Age: The Problem of Privacy in Public." *Law and Philosophy* 17 (5/6) (November): 559–596.

Nissenbaum, H. 2004. "Privacy as Contextual Integrity." *Washington Law Review* 79 (February 17): 119–158.

© 2010 Policy Studies Organization

Nissenbaum, H. 2009. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Palo Alto, CA: Stanford University Press.

Nunberg, G. 2009a. “Google Books: A Metadata Train Wreck.” *Language Log*, August 29. http://languagelog.ldc.upenn.edu/nll/?p=1701 (accessed September 5, 2009).

Nunberg, G. 2009b. “Google’s Book Search: A Disaster for Scholars.” *Chronicle of Higher Education* (31 August). http://chronicle.com/article/Googles-Book-Search-A/48245/?sid=at&utm_source=at&utm_medium=en (accessed May 26, 2010).

“Opinion Decision and Order”. 2007. *Doe v. Gonzales*. 500 F. Supp. 2d 379. http://scholar.google.com/scholar_case?case=7873627742897079382&hl=en&as_sdt=100000000000002&as_vis=1 (accessed August 14, 2010).

Picker, R.C. 2009. “The Google Book Search Settlement: A New Orphan-Works Monopoly?” *Journal of Competition Law and Economics* 5 (3): 383–409. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1387582 (accessed July 31, 2010).

Price, E. 2010. “Net Neutrality Advocates Blast Google, Verizon Plan.” *PC World*, August 10. http://www.pcworld.com/article/202964/net_neutrality_advocates_blast_google_verizon_plan.html?tk=hp_blg (accessed August 11, 2010).

Privacy Authors and Publishers. 2009. “Privacy Authors and Publishers’ Objection to Proposed Settlement.” *The Authors Guild, Inc., Association of American Publishers, Inc.*, et al. *vs. Google, Inc.* Case No. 05 CV 8136-DC. http://thepublicindex.org/objections/privacy_authors.pdf (accessed July 10, 2010).

Rachels, J. 1975. “Why Privacy Is Important.” *Philosophy and Public Affairs* 4 (4) (Summer): 323–333.

Redmer, J. 2007. *Google Book Search: The Story.* London: Google. http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.co.uk/en/uk/press/files/book-search-en.pdf (accessed October 4, 2009).

Regan, P.M. 2004. “Old Issues, New Context: Privacy, Information Collection, and Homeland Security.” *Government Information Quarterly* 21 (4): 481–497.

Samuelson, P. 2009. "Legally Speaking: The Dead Souls of the Google Book Search Settlement." *Communications of the ACM* 52 (7) (July): 28–30.

Samuelson, P. 2010. "Google Book Search and the Future of Books in Cyberspace." *Minnesota Law Review* (forthcoming). http://ssrn.com/abstract=1535067.

Seattle Public Library. 2002. "Confidentiality of Borrower Information." http://www.spl.org/default.asp?pageID=about_policies_confidentiality (accessed August 15, 2010).

Solove, D.J. 2007. "'I've Got Nothing to Hide' and Other Misunderstandings of Privacy." *San Diego Law Review* 44: 745–772. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=998565 (accessed August 11, 2010).

Taycher, L. 2010. "Books of the World, Stand up and Be Counted! All 129,864,880 of You." *Inside Google Books*, August 5. http://booksearch.blogspot.com/2010/08/books-of-world-stand-up-and-be-counted.html (accessed August 13, 2010).

"The Public Index: Amended Settlement and Responses." New York Law School. http://thepublicindex.org/documents/amended_settlement (accessed August 16, 2010).

Toobin, J. 2007. "Google's Moon Shot: The Quest for the Universal Library." *The New Yorker*, February 5. http://www.newyorker.com/reporting/2007/02/05/070205fa_fact_toobin (accessed June 7, 2008).

Vaidhyanathan, S. 2005. "A Risky Gamble with Google." *Chronicle of Higher Education*, December 2. http://www.nyu.edu/classes/siva/archives/002445.html (accessed January 10, 2006).

Writers' Representatives LLC, and R.A. Epstein. 2009. "Objections to Class Action Settlement and Notice of Intent to Appear." *The Authors Guild, Inc., Association of American Publishers, Inc.*, et al. *vs. Google, Inc.* Case No. 05 CV 8136-DC. http://thepublicindex.org/docs/amended_settlement/Writers_Representatives_Objection.pdf (accessed July 28, 2010).

© 2010 Policy Studies Organization