

**Search 2.0:  
Web 2.0, Personal Information Flows, and the Drive for the Perfect Search Engine**

Michael Zimmer  
Fellow, Information Society Project  
Yale Law School

Presented at  
Association of Internet Researchers Conference  
Vancouver, Canada  
October 19, 2007

**Abstract**

The freedom to move through both physical and intellectual space resonates within many of the fundamental values and aspirations of American culture, including free and open inquiry, personal autonomy, and liberty. New information and communication technologies are frequently designed to foster increased mobility within these spheres in support of the preservation of these values. For example, Web search engines have emerged as a ubiquitous and vital tool for the successful navigation of the growing online informational sphere. As Google puts it, their goal is to “organize the world’s information and make it universally accessible and useful” and to create the “perfect search engine” that provides only intuitive, personalized, and relevant results. Meanwhile, new Web 2.0 infrastructures have emerged with the promise to empower creativity, to democratize media production, and to celebrate the individual while also relishing the power of collaboration and social networks.

The (inevitable) combining of Google’s suite of information-seeking products with Web 2.0 infrastructures – what I call Search 2.0 – intends to capture the best of both technical systems for the benefit of users. By capturing the information flowing across Web 2.0, search engines can better predict users’ needs and wants, and deliver more relevant and meaningful results. While intended to enhance intellectual mobility in the online sphere, this paper argues that the drive for Search 2.0 necessarily requires the widespread monitoring and aggregation of a users’ online personal and intellectual activities, bringing with it particular value externalities, such as the privacy of individuals’ online intellectual activities. These search-based infrastructures of dataveillance contribute to a rapidly emerging “soft cage” of everyday digital surveillance, where they, like other dataveillance technologies before them, contribute to the curtailing of individual freedom, affect users’ sense of self, and present issues of deep discrimination and social justice.

## Introduction

The rhetoric surrounding Web 2.0 infrastructures presents certain cultural claims about media, identity, and technology. It suggests that everyone can and should use new Internet technologies to organize and share information, to interact within communities, and to express oneself. It promises to empower creativity, to democratize media production, and to celebrate the individual while also relishing the power of collaboration and social networks. Websites such as Flickr, Wikipedia, del.icio.us, Facebook, and YouTube are all part of this second-generation Internet phenomenon, featuring user-generated content, opportunities for collaboration and harnessing collective intelligence, and relatively open platforms for anyone to participate, modify (mash-up) or share content (via RSS feeds, APIs, and the like).

The very notion of Web 2.0, of course, is not without controversy. Building from Tim O'Reilly's original conception of term (O'Reilly, 2005), many have hailed the emergence of Web 2.0 as the "new wisdom of the web" (Levy & Stone, 2006) and "a new cultural force based on mass collaboration" (Kelly, 2005), while others deride it as merely a hyped-up buzzword (Boutin & , 2006), "millennialist rhetoric" (Carr, 2006), and even an extension of Marxist ideology that is "inherently dangerous for the vitality of culture and the arts" (Keen, 2006).

This last notion, the relationship between Web 2.0 and Marxism, was suggested Andrew Keen, one of the loudest provocateurs of the Web 2.0 ideology (see Keen, 2007). Keen has received considerable criticism for making comparisons between the Web 2.0 meme and Marxism, but, between the vitriol, he does offer legitimate points about the utopianism and solipsism that seems to underlie much of the Web 2.0 discourse. In particular, he criticizes the fervent commitment to technological progress:

The ideology of the Web 2.0 movement was perfectly summarized at the Technology Education and Design (TED) show in Monterey, last year, when Kevin Kelly, Silicon Valley's über-idealist and author of the Web 1.0 Internet utopia Ten Rules for The New Economy, said:

"Imagine Mozart before the technology of the piano. Imagine Van Gogh before the technology of affordable oil paints. Imagine Hitchcock before the technology of film. We have a moral obligation to develop technology."

But where Kelly sees a moral obligation to develop technology, we should actually have – if we really care about Mozart, Van Gogh and Hitchcock – a moral obligation *to question the development of technology*. (Keen, 2006; emphasis added)

This moral obligation to question the development of technology compels Keen to identify (in his view) some of the unintended consequences of the emergence of Web 2.0 infrastructures, including the flattening of culture, the overabundance of amateur authors and producers, and narcissism run wild.

As I begin to study the Web 2.0 meme from the perspective of privacy and surveillance theory, a different set of unintended consequences emerges, including shifts in the flow of personal information across these open platforms that might threaten personal privacy in ways much more damaging than Keen's concern that content is now made and distributed by mere amateurs instead of honed professionals.

The focus of this paper is one particular unintended consequence of the increased flow of personal information across Web 2.0 infrastructures: the efforts by Web search engines to crawl and aggregate Web 2.0-related data flows in order to build profiles, predict intentions, and deliver personalized products and services. This drive for the perfect search engine through the capture of personal information flowing across the networks – what I call the quest for Search 2.0 – brings with it particular value externalities, such as the privacy of individuals' online intellectual activities. This paper argues that the externalities of Search 2.0 represent a new and powerful infrastructure of data surveillance – otherwise referred to as

“dataveillance” (Clarke, 1988) – for the aggregation of one’s online information-seeking activities, inflaming a growing environment of discipline and social control.

This paper is divided into four sections. The first section outlines how the Web search engine has become the center of gravity of many user’s information-seeking needs. The next section describes the quest for the “perfect search engine” – what I refer to as Search 2.0, with the requisite components of the “perfect reach” and the “perfect recall.” The third section reveals some of the externalities of Search 2.0, including its use as a tool for disciplinary power and panoptic sorting of users into economic categories. The final section outlines possible spaces for intervention, including the value-conscious design of future Search 2.0 platforms in order to mitigate its externalities.

## Search as the Center of Gravity

As the Internet has become increasingly important to modern citizens in their everyday lives (see Horrigan & Rainie, 2006), *web search engines* have emerged as today’s prevailing information interface for accessing the vast amount of information available on this global network. A web search engine is an online service that allows users to submit queries to locate Web content meeting specific criteria, typically in the form of keywords or phrases, and delivers a list of links to Web pages that match those criteria.<sup>1</sup> According to the Pew Internet & American Life Project, 84% of American adult Internet users have used a search engine to seek information online (Fallows, 2005, p. 1). On any given day, more than 60 million American adults send over 200 million information requests to web search engines, making searching the web the second most popular online activity (behind using e-mail) (Rainie, 2005). The first Web search engines were developed in the mid-1990s, often as not-for-profit research projects at university computer or information science departments.<sup>2</sup> Since those scholarly beginnings, Web search engines have grown to become a multi-billion dollar industry driven by large multi-national corporations like Google, Ask, Yahoo!, and Microsoft.

Originally designed to provide easy access to Internet websites, search engines now provide gateways to online images, news reports, Usenet archives, financial information, video files, e-mail and even one’s desktop files. Recently, search engine providers have started to digitize items in the “material” world, adding the contents of popular books, university libraries, maps, and satellite images to their growing, searchable indices. Reflecting on the rapid emergence of search-related applications, Silicon Valley venture capitalist Roger McNamee noted that “search is the new center of gravity for the computer industry” (McNamee, 2005). The same can be said more generally for the role of search engines as today’s dominant information interface: *Search engines have become the center of gravity for people’s everyday information-seeking activities.*

Consider, for example, the web search engine Google. Google has become the prevailing knowledge tool for searching and accessing virtually all information on the Web. Originating in 1996 as a Ph.D. research project by Larry Page and Sergey Brin at Stanford University (see Brin & Page, 1998; Page et al., 1998), Google’s Web search engine now dominates the market, processing almost 3.6 billion search queries in

---

<sup>1</sup> While many non-Web search engines exist, such as those for private computer networks, it has become common in popular discourse to use the non-qualified “search engine” to refer specifically to those search engines that index the World Wide Web. I will follow this practice and use the term “search engine” interchangeably with the more precise “Web search engine.”

<sup>2</sup> One of the very first search engines was WebCrawler, developed in 1994 at the University of Washington. Other search engines with roots in academia include Lycos (Carnegie Mellon), Excite (Stanford), Inktomi (Berkeley), Yahoo! (Stanford), and Google (Stanford).

February 2007, over half of all Web searches performed (Nielsen//NetRatings, 2007).<sup>3</sup> Google's mission, stated quite simply and innocuously, is to "organize the world's information and make it universally accessible and useful" (Google, 2005a). In pursuit of this goal, Google has developed dozens of search-related tools and services to help users organize and use information in multiple contexts, ranging from general information inquiries to academic research, news and political information, communication and social networking, personal data management, financial data management, shopping and product research, computer file management, and enhanced Internet browsing (see Table 1). Consequently, users increasingly search, find, and relate to information through Google's growing information infrastructure of search-related services and tools.<sup>4</sup> They also use these tools to communicate, navigate, shop, and organize their lives. By providing a medium for various social, intellectual, and commercial activities, "Planet Google" has become a large part of people's lives – their center of gravity – both online and off (Williams, 2006).

## **Search 2.0: The Quest for the Perfect Search Engine**

Since the first search engines started to provide a way of interfacing with the content on the Web, there has been a drive for the "perfect search engine," one that has indexed all available information and provides fast and relevant results (see Kushmerick, 1998; Andrews, 1999; Gussow, 1999; Mostafa, 2005). A perfect search engine would deliver intuitive results based on users' past searches and general browsing history (Pitkow et al., 2002; Teevan et al., 2005), knowing, for example, whether a search for the keywords "Paris Hilton" is meant to help a user locate a hotel in the French capital or the latest gossip about the celebrity debutante. Search engine companies have clear financial incentives for achieving the "perfect search": receiving personalized search results might contribute to a user's allegiance to a particular search engine service, increasing exposure to that site's advertising partners as well as improving chances the user would use fee-based services. Similarly, search engines can charge higher advertising rates when ads are accurately placed before the eyes of users with relevant needs and interests (Hansell, 2005).

Search engine companies recognized early on the importance of designing the perfect search engine, such as Google co-founder Sergey Brin's assertion in the company's very first press release that "a perfect search engine will process and understand all the information in the world... That is where Google is headed" (Google, 1999). When later asked what a perfect search engine would be like, Brin replied quite simply, "like the mind of God" (quoted in Ferguson, 2005, p. 40). Web journalist John Battelle summarizes how such an omniscient and omnipotent search engine might work:

Imagine the ability to ask any question and get not just an accurate answer, but your perfect answer – an answer that suits the context and intent of your question, an answer that is informed by who you are and why you might be asking. The engine providing this answer is capable of incorporating all the world's knowledge to the task at hand – be it captured in text, video, or

---

<sup>3</sup> Nielsen/NetRatings figures represent U.S. searches only, and include local searches, image searches, news searches, shopping searches and other types of search activity from Google's various services. If only web searches at [www.google.com](http://www.google.com) are considered, Google's share increases to 60% (Sullivan, 2006).

<sup>4</sup> Yahoo!, and to a lesser extent, Microsoft and AOL, also offers search-related tools beyond just locating relevant websites. Google, however, remains the clear market leader at 49.2% of all search activity, with Yahoo! following at 23.8% and MSN at 9.6%. Given their strong dominance of the overall marketplace, and recognition as the "gold standard" in search engine practices and innovation (Hellweg, 2002; Clark, 2006), Google will be the primary focus of this paper.

audio. It's capable of discerning between straightforward requests – who was the third president of the United States? – and more nuanced ones – under what circumstances did the third president of the United States forswear his views on slavery?

This perfect search also has perfect recall – it knows what you've seen, and can discern between a journey of discovery – where you want to find something new – and recovery – where you want to find something you've seen before. (Battelle, 2004)

To attain such an omnipresent and omniscient ideal, search engines must have both “perfect reach” in order to provide access to all available information on the Web and “perfect recall” in order to deliver personalized and relevant results that are informed by who the searcher is. And, as we will see, capitalizing on the flow of information inherent to Web 2.0 infrastructures makes the gravitational pull of the perfect search engine that much more powerful.

### Perfect Reach

To achieve the reach necessary for the realization of Search 2.0, web search engines amass enormous indices of the Web's content. Expanding beyond just HTML-based Web pages, search engine providers have indexed a wide variety of media found on the Web, including images, video files, PDFs and other computer documents. For example, Yahoo! claims to have indexed over 20 billion items, including over 19.2 billion Web documents, 1.6 billion images, and over 50 million audio and video files (Mayer, 2005). Google claims to have an index more than three times larger than that of any other search engine (Google, 2005c), and it is estimated that Google has indexed nearly 70% of the total World Wide Web (Sullivan, 2005). The increasing sophistication and reach of Web crawler and indexing technology provide search engine companies a powerful mapping of the web to fuel the drive for the perfect search (see Table 1). So powerful, that philosopher Lawrence Hinman (2005) has updated George Berkeley's eighteenth-century proclamation that “*esse est percipi*” (to exist is to be perceived) to the twenty-first-century equivalent “*esse est indicato in Google*”: to exist is to be indexed on Google.

Notwithstanding concerns over the increased corporatization of the user-generated content from Web 2.0 infrastructures (see, for example, Carr, 2006; Keen, 2006), search engine providers have capitalized on the growing Web 2.0 infrastructure to complement the reach of their searchable indices. Through the purchase of Web 2.0 properties like Flickr and del.icio.us, Yahoo!, for example, will be able to integrate user-generated photos and folksonomies of bookmarks into their web search engine results and other products and services.<sup>5</sup> Web 2.0 infrastructures are featured prominently in “Planet Google” as well: Blogger, Dodgeball, Orkut, Google Maps, Docs & Spreadsheets, and Calendar all share characteristics with archetypal Web 2.0 applications, enabling social networking, collaborate editing, blogging, and so on. Google's \$1.65 billion purchase of YouTube reveals their commitment to capitalizing on Web 2.0 platforms to expand their powerful reach, quickly integrating these user-generated videos into their own Google Video search results (Ackerman & Blitstein, 2006). In addition, their recent acquisition of Feedburner provides Google direct access to nearly a million feeds of Web 2.0-based traffic (Feedburner, 2007).

By capitalizing on the open flows of information provided by Web 2.0 platforms, the reach of Web search engines has increased far beyond the 1994 incarnation of “Jerry and David's Guide to the World Wide Web”<sup>6</sup> or the “Large-Scale Hypertextual Web Search Engine”<sup>7</sup> envisioned by Brin and Page a decade

---

<sup>5</sup> Flickr images have already been integrated with Yahoo! search results (Tickoo, 2007).

<sup>6</sup> The predecessor to what became Yahoo!.

<sup>7</sup> See (Brin & Page, 1998).

ago. The perfect reach of Search 2.0 extends far beyond the traditional Web pages created by organizations, corporations, or individuals, and includes the cornucopia of user-created content and personal information flows that make Web 2.0 so distinct.

### Perfect Recall

The more potent externality, however, is the impact of Search 2.0 on the tracking and collection of the bits of personal information left behind in one's "digital data trail" upon using various search – and Web 2.0 – related services. To achieve the "perfect recall" of Search 2.0, web search engines must be able to identify and understand each searcher's intellectual wants, needs and desires when they perform information seeking tasks online in order to deliver personalized and relevant results. The primary means for personalizing search results is to rely on a user's search habits and history (see, for example, Speretta, 2000; Pitkow et al., 2002; Teevan et al., 2005). To gather users' search histories, most web search engines maintain detailed server logs recording each web search request processed through their search engine, the pages viewed, and the results clicked (see, for example, Google, 2005b; IAC Search & Media, 2005; Yahoo!, 2006). Search engines also rely heavily on web cookies to help differentiate users and track activity from session to session, and increasingly push the creation of user accounts to help associate particular users with their online activity.

While public attention has recently focused on the industry practice of archiving users' web search queries in server logs (Hafner & Richtel, 2006; Hansell, 2006; Mintz, 2006), less attention has been paid to how search engine providers are able to monitor and aggregate activity across the myriad products and services – including Web 2.0 platforms – within these companies' larger web search information infrastructures. If we consider Google's wide array of products and services (Table 1), their server logs potentially contain much more than simply a user's Web search queries. Other searches logged by Google include those for images, news stories, videos, books, academic research, and blog posts, as well as links clicked and related usage statistics from within Google's News, Reader, Finance, Groups, and other services (see Table 2).

Inherent in Google's suite of products and services is the ability to collect and aggregate a wide array of personal and intellectual information about its users, extending beyond just what website they search for, but also including what news they read, what interests they have, the blogs they follow, the books they enjoy, and even the files on their desktop. By incorporating Web 2.0 services in "Planet Google," the server logs in Mountain View can also aggregate and cross-reference one's friends and interests in the Orkut social network, the special interest RSS feeds subscribed to in one's Google Reader account, or the key terms that trigger AdWords contextual ads in one's personal Blogger posts.

The result is what John Battelle calls a "database of intentions":

This information represents, in aggregate form, a place holder for the intentions of humankind - a massive database of desires, needs, wants, and likes that can be discovered, subpoenaed, archived, tracked, and exploited to all sorts of ends. Such a beast has never before existed in the history of culture, but is almost guaranteed to grow exponentially from this day forward. This artifact can tell us extraordinary things about who we are and what we want as a culture. (Battelle, 2003)

While many of our day-to-day habits – such as using credit cards, ATMs, cell phones, or automated toll collection systems – leave countless "virtual footprints" of our activities, the panoptic gaze of Google's quest for Search 2.0 inevitably tracks our search histories, e-mails, blog posts or general browsing habits, providing "an excellent source of insight into what someone is *thinking*, not just what that person is doing" (Hinman, 2005, p. 23). By including Web 2.0 products and services within "Planet Google," and

linking them together with a common Google Account (see Weinberg, 2005; Google, 2006), the web search giant has gained the ability to track, capture, and aggregate a wealth of personal information – much stemming from the increased flow of personal information made available by growing use and reliance on Web 2.0-based applications – in support of their quest for Search 2.0.

### **Externalities of Search 2.0**

While it is easy to think of a search engine as a one-way information interface where you just provide a simple search term, and a Google gives you millions of pages of information in return, in practice there is an important *feedback loop*; the interface is actually two-way. In order to use many of Google's products and services, and in order for Google to achieve the perfect search, users must provide a large amount of potentially personal information to the search engine providers. To extend the metaphor, Google, the center of gravity of information seeking online, also acts as black hole, exerting its gravitational forces to pull as much information about its users *into* its databases as possible.

Thus, the emergence of Search 2.0 infrastructure represents a vital shift in the norms of information flow: previously, a person's information-seeking and intellectual activities were mostly offline, and distributed between sources like books, newspapers, letters, paper files, etc. While perhaps the librarian, or the bookseller, or the newsstand owner might know what you are reading at a particular time (if they happen to notice at all), there was no way for any one person to have access to *all* of your information-seeking activities. But with the drive towards Search 2.0, single entities, such as Google, have the means of monitoring, collecting and aggregating nearly all of one's online information-seeking and intellectual activities, extending beyond just what website they search for, but also including what news they read, what interests they have, the blogs they follow, the books they enjoy, the stocks in their portfolio, and perhaps even the URL every website they visit.

By monitoring and aggregating the results of every web search performed, every result clicked, every website bookmarked, or every page visited with the Toolbar, Google has created sophisticated infrastructure of dataveillance, capable of creating a robust digital dossier of each of the millions of users of their products. The full effects of this Search 2.0 infrastructure are difficult to predict, and we don't have enough time to discuss them in detail today, but we can identify a few that impact individual's ability to freely seek information online: the exercise of disciplinary power, panoptic sorting, and the general invisibility of both its gaze and its power.

### **Disciplinary Power**

Clive Norris warns of how such infrastructures could be used to “[render] visualization meaningful for the basis of disciplinary social control” (Norris, 2002, p. 251). Instances of how users of Google's infrastructure were made visible for the exercise of disciplinary power include a court ordering Google to provide the complete contents of a user's Gmail account, including e-mail messages he thought were deleted (McCullagh, 2006) and the introduction of evidence that a suspected murderer performed a Google search for the words “neck snap break” (Cohen, 2005). Yahoo! has been widely criticized for providing e-mail and other account data to Chinese officials, resulting in the jailing of dissidents within that country (Olesen, 2005; Schonfeld, 2006). While Google appears to recognize, at least partially, the disciplinary threat of storing such robust records of its users activities when it announced it would move user data collected from its Chinese site outside of the country in order to prevent China's government from being able to access the data without Google's consent (McMillan, 2006), the company recently agreed to comply with a Brazilian court order to release data on users of its Orkut social networking site to help Brazilian authorities investigate use of the site related to racism, pedophilia and homophobia (Downie, 2006). The possibility of search providers providing log data to government bodies for disciplinary action has reached new heights within the United States with the passage of the USA

PATRIOT Act, greatly expanding the ability of law enforcement to access such records, while restricting the source of the records from disclosing any such request has even been made (see Battelle, 2005, pp. 197-204).

### **Panoptic Sorting**

Search 2.0's infrastructure of dataveillance also spawns instances of "panoptic sorting" where users of search engines are identified, assessed and classified "to coordinate and control their access to the goods and services that define life in the modern capitalist economy" (Gandy, 1993, p. 15). Google, like most for-profit search engine providers, is financially motivated collect as much information as possible about each user: receiving personalized search results might contribute to a user's allegiance to a particular search engine service, increasing exposure to that site's advertising partners as well as improving chances the user would use fee-based services. Similarly, search engines can charge higher advertising rates when ads are accurately placed before the eyes of users with relevant needs and interests (Hansell, 2005). Through the panoptic gaze of its diverse suite of products – fueled by the growing Web 2.0 portion of their offerings – Google collects as much information as possible about an individual's behavior, and considers it to be potentially useful in the profiling and categorization of a user's potential economic value: recognizing that targeted advertising will be the "growth engine of Google for a very long time", Google CEO Eric Schmidt stressed the importance of collecting user information, acknowledging that "Google knows a lot about the person surfing, especially if they have used personal search or logged into a service such as Gmail" (Miller, 2006).

### **Invisibility and Allure of Search 2.0**

Perhaps the most potent externality of Search 2.0 stems from its relative invisibility, indispensability, and apparent inescapability. The majority of Web searchers are not aware that search engines have the ability to actively track users' search behavior (Fallows, 2005, p. 21; Kopytoff, 2006), and as Google continues to expand its information infrastructure, it becomes arduous for everyday users to recognize the data collection threats of these services, and easier to take the design of Google merely "at interface value" (Turkle, 1995, p. 103). Greg Elmer warns of the dangers of such an environment where the collection of personal information is a prerequisite of participation inevitably entrenches power in the hands of the technology designers:

Ultimately, what both requesting and requiring personal information highlight is the centrality of producing, updating, and deploying consumer *profiles* – simulations or pictures of consumer likes, dislikes, and behaviors that are automated within the process of consuming goods, services, or media and that increasingly anticipate our future needs and wants based on our aggregated past choices and behaviors. And although Foucault warns of the self-disciplinary model of punishment in panoptic surveillance, computer profiling, conversely, oscillates between seemingly rewarding participation and punishing attempts to elect not to divulge personal information. (Elmer, 2004, pp. 5-6)

This blurring of punishments and rewards – subtle requests and not so subtle commands for personal information – reoccurs in Google's information interface where the default settings and arrangement of services make the collection of personal information automatic and difficult to resist. Give the rising ubiquity of Web 2.0 services – and Google's attempts to bring such services into their own product suites – many are willing to join "Planet Google" with only scant hesitation: "I don't know if I want all my personal information saved on this massive server in Mountain View, but it is so much of an improvement on how life was before, I can't help it" (Williams, 2006). Google's infrastructure of dataveillance places its users under an almost invisible gaze, resulting in a kind of anticipatory conformity, whereby the divulgence of personal information become both routinized and internalized.

## Conclusion

In conclusion, by amassing a tantalizing collection of, admittedly, innovative and useful tools – including many from the so-called Web 2.0 revolution – coupled with requiring the divulgence of personal information as a precondition for using many of their new information and communication technologies, Google has constructed an information-seeking environment whereby which individuals are continuously integrated into a larger infrastructure of dataveillance. This un-ending quest to achieve Search 2.0 has resulted in the emergence of a robust infrastructure of dataveillance that can quickly be internalized and become the basis of disciplinary social control. Roger Clark provides a prescient warning about the effects of dataveillance on the individual:

[The] real impact of dataveillance is the reduction in the meaningfulness of individual actions, and hence in self-reliance and self-responsibility. Although this may be efficient and even fair, it involves a change in mankind's image of itself, and risks sullen acceptance by the masses and stultification of the independent spirit needed to meet the challenges of the future. ...In general, mass dataveillance tends to subvert individualism and the meaningfulness of human decisions and actions. (Clarke, 1988, p. 508)

Thus a kind of Faustian bargain emerges: Search 2.0 promises breadth, depth, efficiency, and relevancy, but enables the widespread collection of personal and intellectual information in the name of its perfect recall. If left unchecked, potential cost of this bargain is nothing less than the “individualism and the meaningfulness of human decisions and actions.”

What options exist for renegotiating our Faustian bargain with Search 2.0? One avenue for changing the terms of the Faustian bargain is to enact laws to regulate the capture and use of personal information by Web search engines. A recent gathering of leading legal scholars and industry lawyers to discuss the possibility of regulating search engines revealed, however, that viable and constitutional solutions are difficult to conceive, let alone agree upon.<sup>8</sup> Alternatively, the search engine industry could self-regulate, creating strict policies regarding the capture, aggregation, and use of personal data via their services. But as Chris Hoofnagle reminds us, “We now have ten years of experience with privacy self-regulation online, and the evidence points to a sustained failure of business to provide reasonable privacy protections” (2005, p. 1). Given search engine companies’ economic interests in capturing user information for powering Search 2.0, relying solely on self-regulation will likely be unsatisfying.

A third option is to affect the design of the technology itself. As Larry Lessig notes, “how a system is designed will affect the freedoms and control the system enables” (2001, p. 35), I argue that technological design is one of the *critical junctures* for society to re-negotiate its Faustian bargain with Search 2.0 in order to preserve a sense of “individualism and the meaningfulness of human decisions and actions.”<sup>9</sup> Potential design variables include whether default settings for new products or services automatically enroll users in data-collecting processes – or whether the process can be turned off. Or the extent to which

---

<sup>8</sup> See “Regulating Search: A Symposium on Search Engines, Law, and Public Policy” held in December 2005 at Yale Law School (<http://islandia.law.yale.edu/isp/regulatingsearch.html>).

<sup>9</sup> Various pragmatic frameworks have recently emerged to broaden the criteria for judging the quality of technological systems to include the advancement of ethical and human values, and to proactively influence the design of technologies to account for such values during the conception and design process. These include Design for Values (Camp, n.d.), Values at Play (Flanagan et al., in press, 2005), and Value Sensitive Design (Friedman, 1999; Friedman et al., 2002).

different products should be interconnected: For example, if a user signs up to use Gmail, should the Personalized Search automatically be activated? Should the user automatically be logged in to other services? Ideally, new tools can be developed to give users access and control over the personal information collected: In the spirit of the Code of Fair Information Practices, a Google Data Privacy Center should be built to allow users to view all their personal data collected, make changes and deletions, restrict how it is used, and so on. Through such an intervention in the design of Search 2.0, there is hope that our Faustian bargain can be re-negotiated to counter its externalities.

## References

- Ackerman, E., & Blitstein, R. (2006, October 9). Google buys YouTube for \$1.65 billion. *San Jose Mercury News*.
- Andrews, P. (1999, February 7). The search for the perfect search engine. *The Seattle Times*, p. E1.
- Battelle, J. (2003, November 13). The database of intentions. *Searchblog*. Retrieved May 16, 2006, from <http://battellemedia.com/archives/000063.php>
- Battelle, J. (2004, September 8). Perfect search. *Searchblog*. Retrieved May 16, 2006, from <http://battellemedia.com/archives/000878.php>
- Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York: Portfolio.
- Boutin, P. (2006, March 29). Web 2.0: The new internet "boom" doesn't live up to its name. *Slate*. Retrieved January 12, 2007, from <http://www.slate.com/id/2138951/>
- Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *WWW7 / Computer Networks*, 30(1-7), 107-117.
- Camp, L. J. (n.d.). Design for values, design for trust. Retrieved September 20, 2006, from <http://www.ljean.com/design.html>
- Carr, N. (2006, October 3). The amorality of Web 2.0. *Rough Type*. Retrieved January 14, 2007, from [http://www.roughtype.com/archives/2005/10/the\\_amorality\\_o.php](http://www.roughtype.com/archives/2005/10/the_amorality_o.php)
- Clark, H. (2006, August 23). Innovation: A waste of money? *Forbes.com*. Retrieved August 20, 2006, from [http://www.forbes.com/leadership/2006/08/23/leadership-innovation-requiredreading-cx\\_hc\\_0823moore.html](http://www.forbes.com/leadership/2006/08/23/leadership-innovation-requiredreading-cx_hc_0823moore.html)
- Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, 37(5), 498-512.
- Cohen, A. (2005, November 28). What Google Should roll out next: A privacy upgrade. *The New York Times*, p. A18.
- Downie, A. (2006, September 8). Google carves a middle path on privacy. *The Christian Science Monitor*, p. 1.
- Elmer, G. (2004). *Profiling machines: Mapping the personal information economy*. Cambridge, MA: MIT Press.
- Fallows, D. (2005). Search engine users: Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve. *Pew Internet & American Life Project*. Retrieved October 15, 2005, from [http://www.pewinternet.org/pdfs/PIP\\_Searchengine\\_users.pdf](http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf)
- Feedburner. (2007). About Feedburner. Retrieved July 23, 2007, from <http://www.feedburner.com/fb/a/about;jsessionid=C1462F0EACD520D8C83B5EB0BA2FB701.ap2>
- Ferguson, C. (2005). That's next for Google? *Technology Review*, 108(1), 38-46.
- Flanagan, M., Howe, D., & Nissenbaum, H. (2005). Values at play: Design tradeoffs in socially-oriented game design. *Conference on Human Factors in Computing Systems*, 751-760.
- Flanagan, M., Howe, D., & Nissenbaum, H. (in press). Values in design: Theory and practice. In J. van den Hoven, & J. Weckert (Eds.), *Information technology and moral philosophy*. Cambridge University Press.

- Friedman, B. (1999). Value-sensitive design: A research agenda for information technology. National Science Foundation, Contract No: SBR-9729633). Arlington, VA.
- Friedman, B., Kahn, P., & Borning, A. (2002). Value sensitive design: Theory and methods. (Technical Report 02-12-01). Seattle, WA.
- Gandy, O. (1993). *The panoptic sort: A political economy of personal information*. Boulder, CO: Westview.
- Google. (1999, June 7). Google Receives \$25 million in equity funding [press release]. *Google Press Center*. Retrieved August 18, 2006, from <http://www.google.com/press/pressrel/pressrelease1.html>
- Google. (2005a). Company overview. Retrieved May 3, 2006, from <http://www.google.com/corporate/index.html>
- Google. (2005b). Google Privacy faq. Retrieved May 3, 2006, from [http://www.google.com/privacy\\_faq.html](http://www.google.com/privacy_faq.html)
- Google. (2005c). Sizing up search engines. Retrieved December 1, 2006, from <http://www.google.com/help/indexsize.html>
- Google. (2006). Blogger help: Will my use of blogger be associated with my use of other Google services? Retrieved August 20, 2006, from <http://help.blogger.com/bin/answer.py?answer=42601&topic=8939>
- Gussow, D. (1999, October 4). In search of. *St. Petersburg Times*, p. 13.
- Hafner, K., & Richtel, M. (2006, January 20). Google resists u.s. Subpoena of search data. *The New York Times*, pp. A1, C4.
- Hansell, S. (2005, September 26). Microsoft plans to sell search ads of its own. *The New York Times*, pp. C1, C8.
- Hansell, S. (2006, August 8). AOL removes search data on vast group of Web users. *The New York Times*, p. C4.
- Hellweg, E. (2002, April 22). Google's need for speed. *CNN/Money*. Retrieved August 20, 2006, from <http://money.cnn.com/2002/04/22/technology/techinvestor/hellweg/index.htm>
- Hinman, L. (2005). Esse est indicato in Google: Ethical and political issues in search engines. *International Review of Information Ethics*, 3, 19-25.
- Hoofnagle, C. (2005, March 4). Privacy self regulation: A decade of disappointment. *Electronic Privacy Information Center*. Retrieved April 18, 2007, from <http://www.epic.org/reports/decadedisappoint.html>
- Horrigan, J., & Rainie, L. (2006, April 19). The internet's growing role in life's major moments. *Pew Internet & American Life Project*. Retrieved May 26, 2006, from [http://www.pewinternet.org/PPF/r/181/report\\_display.asp](http://www.pewinternet.org/PPF/r/181/report_display.asp)
- IAC Search & Media. (2005, July 13). Privacy policy for ask.com. Retrieved January 6, 2007, from <http://sp.ask.com/en/docs/about/privacy.shtml>
- Keen, A. (2006, February 15). Web 2.0: The second generation of the internet has arrived. It's worse than you think. *The Daily Standard*. Retrieved January 14, 2007, from <http://www.weeklystandard.com/Content/Public/Articles/000/000/006/714fjczq.asp>
- Keen, A. (2007). *The cult of the amateur: How today's internet is killing our culture*. New York: Doubleday/Currency.
- Kelly, K. (2005, August). We are the Web. *Wired.com*. Retrieved July 2, 2007, from [http://www.wired.com/wired/archive/13.08/tech\\_pr.html](http://www.wired.com/wired/archive/13.08/tech_pr.html)
- Kopytoff, V. (2006, January 24). Most Web users say Google should keep data private. *San Francisco Chronicle*, p. C3.
- Kushmerick, N. (1998, February 23). The search engineers. *The Irish Times*, p. 10.
- Lessig, L. (2001). *The future of ideas: The fate of the commons in a connected world*. New York: Random House.
- Levy, S., & Stone, B. (2006, April 3). The new wisdom of the Web. *Newsweek*. Retrieved July 2, 2007, from <http://www.msnbc.msn.com/id/12015774/site/newsweek/page/0/>

- Mayer, T. (2005, August 8). Our blog is growing up – and so has our index. *Yahoo! Search Blog*. Retrieved November 25, 2006, from <http://www.ysearchblog.com/archives/000172.html>
- McCullagh, D. (2006, March 17). Police blotter: Judge orders Gmail disclosure. *News.com*. Retrieved June 20, 2006, from [http://news.com.com/Police%20blotter%20Judge%20orders%20Gmail%20disclosure/2100-1047\\_3-6050295.html](http://news.com.com/Police%20blotter%20Judge%20orders%20Gmail%20disclosure/2100-1047_3-6050295.html)
- McMillan, R. (2006, March 1). Google moving search records out of China. *InfoWorld*. Retrieved August 20, 2006, from [http://www.infoworld.com/article/06/03/01/75996\\_030106HNgooglechina\\_1.html](http://www.infoworld.com/article/06/03/01/75996_030106HNgooglechina_1.html)
- McNamee, R. (2005, February 7). Google's desktop search. *The New Normal*. Retrieved May 2, 2006, from [http://thenewnormal.com/index.php/newnormal/googles\\_desktop\\_search/](http://thenewnormal.com/index.php/newnormal/googles_desktop_search/)
- Miller, M. (2006, March 17). Google's Schmidt clears the air. *PCMag.com*. Retrieved March 17, 2006, from <http://www.pcmag.com/article2/0,1895,1939257,00.asp>
- Mintz, H. (2006, January 16). Feds after Google data: Records sought in u.s. Quest to revive porn law. *San Jose Mercury News*. Retrieved January 19, 2006, from <http://www.siliconvalley.com/mld/siliconvalley/13657386.htm>
- Mostafa, J. (2005, January 24). Seeking better Web searches. *Scientific American.com*. Retrieved January 30, 2005, from [http://www.sciam.com/print\\_version.cfm?articleID=0006304A-37F4-11E8-B7F483414B7F0000](http://www.sciam.com/print_version.cfm?articleID=0006304A-37F4-11E8-B7F483414B7F0000)
- Nielsen//NetRatings. (2007, March 20). Nielsen//Netratings announces February u.s. Search share rankings. Retrieved March 27, 2007, from [http://www.nielsen-netratings.com/pr/pr\\_070320.pdf](http://www.nielsen-netratings.com/pr/pr_070320.pdf)
- Norris, C. (2002). From personal to digital: Cctv, the panopticon, and the technological mediation of suspicion and social control. In D. Lyon (Ed.), *Surveillance as social sorting*. (pp. 249-281). London: Routledge.
- O'Reilly, T. (2005, September 30). What is Web 2.0? Retrieved June 13, 2007, from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- Olesen, A. (2005, September 6). Rights group says Yahoo helped China jail journalist. *USA Today*. Retrieved March 13, 2007, from [http://www.usatoday.com/tech/news/2005-09-06-yahoo-china-journalist\\_x.htm](http://www.usatoday.com/tech/news/2005-09-06-yahoo-china-journalist_x.htm)
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the Web. Technical report. Stanford University, Stanford, CA.
- Pitkow, J., Schütze, H., Cass, T., Turnbull, D., Edmonds, A., & Adar, E. (2002). Personalized search. *Communications of the ACM*, 45(9), 50-55.
- Rainie, L. (November 2005). Search engine use shoots up in the past year and edges towards e-mail as the primary internet application. *Pew Internet and American Life Project*.
- Schonfeld, E. (2006, February 8). Analysis: Yahoo's China problem. *CNN/Money*. Retrieved March 13, 2007, from [http://money.cnn.com/2006/02/08/technology/yahoo\\_china\\_b20/](http://money.cnn.com/2006/02/08/technology/yahoo_china_b20/)
- Speretta, M. (2000). *Personalizing search based on user search histories*. University of Kansas.
- Sullivan, D. (2005, May 17). New estimate puts Web size at 11.5 billion pages & compares search engine coverage. *SearchEngineWatch*. Retrieved January 4, 2007, from <http://blog.searchenginewatch.com/blog/050517-075657>
- Sullivan, D. (2006, August 23). Hitwise search engine ratings. *SearchEngineWatch*. Retrieved August 23, 2006, from <http://searchenginewatch.com/showPage.html?page=3099931>
- Teevan, J., Dumais, S. T., & Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, 449-456.
- Tickoo, U. (2007, June 26). Flickr-izing image search. *Yahoo! Search Blog*. Retrieved July 23, 2007, from <http://www.ysearchblog.com/archives/000466.html>
- Turkle, S. (1995). *Life on the screen: Identity in the age of the internet*. New York: Simon & Schuster.
- Weinberg, N. (2005, September 11). Google unifyinig accounts. *Inside Google*. Retrieved August 20, 2006, from <http://google.blognewschannel.com/index.php/archives/2005/09/21/google-unifying-logins/>

Williams, A. (2006, October 15). Planet Google Wants you. *The New York Times*, p. 9.1.  
Yahoo!. (2006, November 11). Yahoo! Privacy policy. Retrieved January 6, 2007, from  
<http://info.yahoo.com/privacy/us/yahoo/details.html>

**Table 1: Google Suite of Products and Services**

Product	Description	Notes
<i>General Information Inquiries</i>		
Web search	- Query-based website searches	
Personalized Homepage	- Customized Google start page with content-specific modules	- Use in conjunction with Google Account is encouraged
Alerts	- E-mail alerts of new Google results for specific search terms	
Image Search	- Query based search for website images	
Video	- Query based search for videos hosted by Google	- Google Video Player available for download
Book Search	- Full text searches of books scanned into Google's servers	- Google Account required in order to limit the number of pages a particular user can view
<i>Academic Research</i>		
Scholar	- Full text searches of scholarly books and journals	
<i>News and Political Information</i>		
News	- Full text search of recent news articles	- With a Google Account, users can create customized keyword-based news sections
Reader	- Web-based news feed reader	- Google Account required
Blog Search	- Full text search of blog content	
<i>Communication and Social Networking</i>		
Gmail	- Free Web based e-mail service with contextual advertising	- Creation of Gmail account automatically results in activation of Google Account - Logging into Gmail also logs user into their Google Account
Groups	- Free Web based discussion forums	- Includes complete Usenet archives dating back to 1981 - Google Account required for creation of new Group;
Talk	- Web-based instant messaging and voice calling service	- Google Account and Gmail e-mail address required
Blogger	- Web-based blog publishing platform	- Google Account required
Orkut	- Web-based social networking service	- Invitation-only - Google Account required
Dodgeball	- Location-based social networking service for cellphones	
<i>Personal Data Management</i>		
Calendar	- Web-based time-management tool	
<i>Financial Data Management</i>		
Finance	- Portal providing news and financial information about stocks, mutual funds; Ability to track one's financial portfolio	- Google Account required for posting to discussion board

(Table continues)

**Table 1:** Google Suite of Products and Services (continued)

Product	Description	Notes
<i>Consumer Activities</i>		
Catalog Search	- Full text search of scanned product catalogs	
Froogle	- Full text search of online retailers	- Google Account required for shipping lists
Local / Maps	- Location specific Web searching; digital mapping	
<i>Computer File Management</i>		
Desktop Search	- Keyword based searching of computer files - Ability to search files on remote computer	
<i>Internet Browsing</i>		
Bookmarks	- Online storage of website bookmarks	- Google Account required
Notebook	- Browser tool for saving notes while visiting websites	- Google Account required
Toolbar	- Browser tool providing access to various Google products without visiting Google websites	- Some features require Google Account
Web Accelerator	- Software to speed up page load times for faster Web browsing	

**Table 2:** Personal Information Collected by Google's Suite of Products

Product	Information Collected	Notes
<i>General Information Inquiries</i>		
Web search	- Web search queries - Results clicked	- Search for own name, address, social security number, etc is common
Personalized Homepage	- News preferences - Special interests - Zip code	
Alerts	- News preferences - Special interests - E-mail address	- Alerts for a user's own name (vanity search) are common
Image Search	- Search queries - Results clicked	
Video	- Search queries - Videos watched/downloaded - Credit card information for purchased videos - E-mail details for shared videos	- Google Video Player contains additional DRM technology to monitor off-site video usage
Book Search	- Search queries - Results clicked - Pages read - Bookseller pages viewed	
<i>Academic Research</i>		
Scholar	- Search queries - Results clicked - Home library (Optional)	
<i>News and Political Information</i>		
News	- News search queries - Results clicked	
Reader	- Feed subscriptions - Usage statistics	
Blog Search	- Search queries - Results clicked	
<i>Communication and Social Networking</i>		
Gmail	- Text of email messages - E-mail searches performed - Email address or cellphone number (used for account creation)	
Groups	- Search queries - User interests - Usage statistics - Profile information	- Users are encouraged to create detailed profiles, including name, location, industry, homepage, etc
Talk	- Contact list - Chat messages - Usage statistics	

(Table continues)

**Table 2:** Personal Information Collected by Google's Suite of Products (continued)

Product	Information Collected	Notes
<i>Communication and Social Networking</i>		
Blogger	<ul style="list-style-type: none"> <li>- Weblog posts and comments</li> <li>- Profile information</li> <li>- Usage statistics</li> </ul>	- Users are encouraged to create detailed profiles, including name, location, gender, birthday, etc
Orkut	<ul style="list-style-type: none"> <li>- Profile information</li> <li>- Usage statistics</li> </ul>	- Users are encouraged to create detailed profiles, including name, location, gender, birthday, etc
Dodgeball	<ul style="list-style-type: none"> <li>- Profile information</li> <li>- E-mail address</li> <li>- Location</li> <li>- Mobile phone information</li> <li>- Text messages sent</li> </ul>	- User location when messages sent are tracked by Google
<i>Personal Data Management</i>		
Calendar	<ul style="list-style-type: none"> <li>- Profile information</li> <li>- Events</li> <li>- Usage statistics</li> </ul>	
<i>Financial Data Management</i>		
Finance	<ul style="list-style-type: none"> <li>- Financial quotes</li> <li>- Discussion group activity</li> <li>- Portfolio (optional)</li> <li>- Profile information</li> </ul>	- Names and e-mails are displayed with discussion posts
<i>Consumer Activities</i>		
Catalog Search	<ul style="list-style-type: none"> <li>- Product search queries</li> <li>- Results clicked</li> </ul>	
Froogle	<ul style="list-style-type: none"> <li>- Product search queries</li> <li>- Results clicked</li> <li>- Sites visited</li> <li>- Shopping list</li> </ul>	
Local / Maps	<ul style="list-style-type: none"> <li>- Search queries</li> <li>- Results clicked</li> <li>- Home/default location</li> </ul>	- Search queries might include geographic-specific information
<i>Computer File Management</i>		
Desktop Search	<ul style="list-style-type: none"> <li>- Search queries</li> <li>- Computer file index (Optional)</li> </ul>	<ul style="list-style-type: none"> <li>- Search queries visible to Google under certain circumstances</li> <li>- Desktop file index is stored on Google's services if using Search Across Computers</li> </ul>
<i>Internet Browsing</i>		
Bookmarks	<ul style="list-style-type: none"> <li>- Favorite websites</li> <li>- When visited</li> </ul>	
Notebook	<ul style="list-style-type: none"> <li>- Notes and clippings</li> <li>- Sites annotated</li> </ul>	
Toolbar	<ul style="list-style-type: none"> <li>- Search queries</li> <li>- Websites visited</li> </ul>	- Use of some advanced features routes all browsing traffic through Google servers
Web Accelerator	<ul style="list-style-type: none"> <li>- Websites visited</li> </ul>	- All browsing traffic is routed through Google servers