WHAT GOOGLE KNOWS: PRIVACY AND INTERNET SEARCH ENGINES

Omer Tene[*]

[*] Lecturer, College of Management School of Law, Israel. LL.M., J.S.D (New York University); LL.B., LL.M. (Tel Aviv University); MBA (INSEAD). All web sites cited were last visited September 2007.

"Don't be evil" Google motto[1]

Introduction

Search engines are the most important actors on the Internet today and Google is the undisputed king of search. Google dominates the Internet, guiding users to the information they seek through an ocean of unrelated data with astonishing precision and speed. It is a powerful tool, evoking ambivalent feelings.[2] On the one hand, we adore Google for its simple, modest-looking interface masking a hyper-complicated algorithm, which is the very essence of online ingenuity. We admire it for providing superb services at no (evident) cost, a practical miracle in today's market economy. On the other hand, we grow wary of Google's increasing clout as the ultimate arbiter of commercial success ("to exist is to be indexed by a search engine"[3]) and as a central database for users' personal information, not only logging their search queries but also storing their e-mail (Gmail), calendars (Calendar), photos (Picasa), videos (YouTube), blogs (Blogger), documents (Docs & Spreadsheets), social networks (Orkut), news feeds (Reader), credit card information (Checkout) – in short, their entire digital lives.

Google's access to and storage of vast amounts of personal data create a serious privacy problem, one that Princeton computer scientist Edward Felten recently called "perhaps the most difficult privacy [problem] in all of human history."[4] Every day, millions upon millions of users provide Google with unfettered access to their interests, needs, desires, fears, pleasures and intentions. Counter to conventional

[1] Google Code of Conduct, Preface (Jan. 30, 2007), available at http://investor.google.com/conduct.html.

[2] For notable works in the growing body of literature on "search engine law," *see* Urs Gasser, *Regulating Search Engines: Taking Stock and Looking Ahead*, 8 YALE J. L. & TECH. 201 (2006); James Grimmelmann, *The Structure of Search Engine Law*, New York Law School Public Law and Legal Theory Research Paper Series 06/07 No. 23 (2007), available at http://ssrn.com/abstract=979568; Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 8 YALE J. L. & TECH. 188 (2006); Heidi S. Padawer, *Google This: Search Engine Results Weave a Web for Trademark Infringement Actions on the Internet*, 81 WASH. U. L.Q. 1099 (2003); Lauren Troxclair, *Search Engines and Internet Advertisers: Just one Click Away from Trademark Infringement?*, 62 WASH. & LEE L. REV. 1365 (2005).

[3] Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16(3) INF. SOC. 169, 171 (2000).

[4] Economist Special Briefing, *Inside the Googleplex*, ECONOMIST, Aug. 30, 2007, available at http://www.economist.com/business/displaystory.cfm?story_id=9719610.

wisdom, this information is logged and maintained in a form which may facilitate the identification of specific users for various purposes, including not only their targeting with effective advertising but also prosecution by the government or pursuit by private litigants. As John Battelle memorably put it, "[l]ink by link, click by click, search is building possibly the most lasting, ponderous, and significant cultural artifact in the history of humankind: the Database of Intentions."[5] This "Database of Intentions" constitutes a honey pot for various actors, ranging from the NSA and FBI, which expend billions of dollars on online surveillance and cannot overlook Google's information treasure trove, to hackers and data thieves, who routinely overcome information security systems no matter how robust.

A leading advocate for human rights, Privacy International, recently ranked Google's privacy practices as the worst out of more than 20 leading Internet service providers, including Microsoft, Yahoo, Amazon and eBay.[6] Privacy International describes Google as "an endemic threat to privacy."[7] It criticizes Google's "aggressive use of invasive or potentially invasive technologies and techniques" and claims the company "fails to follow generally accepted privacy practices such as the OECD Privacy Guidelines and elements of EU data protection law."[8] EU data protection regulators have recently launched an investigation into Google's data retention and privacy practices,[9] which was quickly expanded to cover other search engines as well.[10] And the Electronic Privacy Information Center (EPIC), a leading privacy group, filed a complaint with the Federal Trade Commission, arguing Google's contemplated merger with long-time privacy nemesis Doubleclick must be blocked.[11]

[5] JOHN BATTELLE, THE SEARCH: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE 6 (Penguin Group 2005).

[6] *See* Gemma Simpson, *Google scores lowest in privacy rankings*, ZDNET, Jun. 12, 2007, available at http://news.zdnet.co.uk/internet/0,1000000097,39287492,00.htm.

[7] Privacy International, *A Race to the Bottom - Privacy Ranking of Internet Service Companies, A Consultation report*, Jun. 9, 2007, available at http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-553961.

[8] *Ibid.*

[9] Article 29 Working Party Letter to Mr. Peter Fleischer, Global Privacy Counsel, Google (May 16, 2007), available at http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_google_16_05_07_en.pdf;

[10] Article 29 Working Party Press Release (Jun. 21, 2007), available at http://ec.europa.eu/justice_home/fsj/privacy/news/docs/pr_21_06_07_en.pdf.

[11] In the Matter of Google and DoubleClick, Complaint and Request for Injunction, Request for Investigation and for Other Relief, before the Federal Trade Commission (Apr. 20, 2007), available at http://www.epic.org/privacy/ftc/google/epic_complaint.pdf.

How did Google evolve from being a benevolent giant seeking to "do no evil" into a privacy menace, an unruly private sector "big brother" reviled by human rights advocates worldwide?[12] Are the fears of Google's omniscient presence justified or overstated? What personal data should Google be allowed to retain and for how long? What rules should govern access to Google's database? What are the legal protections currently in place and are they sufficient to quell the emerging privacy crisis? These are the main issues addressed in this article.

Part I will untangle two distinct types of privacy problems raised by Internet search engines. First, there is the privacy of search targets, *i.e.*, the privacy rights of people you search for on Google, yielding increasingly detailed profiles ripe with personal information. The search target privacy problem is rooted in the ease of access to personal data, which might always have been publicly available, but were practically hard to reach. With Google, such data are but a mouse click away. Second, there is the problem of search engine users' privacy, which is raised by Google's meticulous collection of each user's search queries and their retention in search logs. After initially laying out these separate privacy issues, the article will proceed to focus on the question of user privacy.

Part II will analyze whether user search logs constitute personally identifiable information, which is subject to privacy protection. The data contained in user search logs are undoubtedly of a highly personal nature. They often include information about one's medical needs, sexual preferences, financial condition, political and religious beliefs. The more vexing question is whether these data may be linked to a specific individual, an identifiable person, thus rendering them personally identifiable information. The combination of users' IP addresses, persistent cookie files and personal details gleaned from registration forms, renders users' search logs personally identifiable. In addition, reporters have demonstrated the ability to link even fully anonymized search logs to specific individuals by a simple process of reverse engineering.

[12] *See*, Leaders, *Who's Afraid of Google*, ECONOMIST, Aug. 30, 2007, available at http://www.economist.com/opinion/displaystory.cfm?story_id=9725272.

Having established that user search logs constitute personally identifiable information, Part III proceeds to examine what such information is typically used for. Google utilizes search logs to improve its search service by tweaking its algorithm. In addition, it uses search logs to prevent online fraud and abuse, including the spread of viruses, spam and "black hat search engine optimization" techniques. Last but not least, Google analyzes search logs for revenue generating purposes, namely for targeting and maximizing the effectiveness of advertisements. User search logs are sought not only by Google itself but also by interested third parties. First and foremost is the government, which argues it must have Google's information to combat anything from terrorism and pedophilia to Internet porn. Private litigants too may try to subpoena Google search logs as evidence in copyright, defamation, employment, and family disputes. And of course there are hackers, data thieves and rogue employees, who will try to appropriate valuable personal information through illicit means.

Part IV uses Daniel Solove's comprehensive and thorough taxonomy of privacy[13] to classify the main privacy harms caused by Google's collection, retention and use of search logs. Google's activities raise the problem of *aggregation*, because intimate and comprehensive user profiles are assembled from bits of information revealed over time; *distortion*, because information in search logs may be highly misleading with potentially harsh results for users; *exclusion*, because search engine users are not granted access their files; *secondary use*, because Google uses data collected from users for one purpose (search) to different ends (commercial, security, law enforcement, litigation); and *breach of confidentiality*, because Google owes users a duty of confidentiality based on an implied term of contract or on the private nature of the information itself. Additional privacy problems, such as *disclosure*, *surveillance*, and *insecurity* are discussed briefly, as well as the *chilling effect* that Google's privacy practices could have for search and online activity generally.

[13] *See* Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477 (2006) [hereinafter Solove, Taxonomy]; for notable previous attempts to organize the field, *see* Ken Gormley, *One Hundred Years of Privacy*, 1992 WIS. L. REV. 1335; Jerry Kang, *Information Privacy in Cyberspace Transactions*, 50 STAN. L. REV. 1193 (1998). The best known taxonomy is of course Prosser's, William L. Prosser, *Privacy*, 48 CAL. L. REV. 383 (1960).

Part V will discuss a range of solutions to the search logs privacy problems, emphasizing existing shortcomings and proposing solutions thereto. First, it describes technological solutions, such as cookie blocking, proxy servers and anonymizing software, arguing that these tools, while useful, do not afford complete protection and are not readily available to the average user. Second, it addresses search engine privacy policies, which are self imposed, often opaque, contractual terms drafted by companies to protect their own interests as opposed to users' privacy. Moreover, user consent to such documents is implicit, uninformed, and partially coerced. Third, it presents Fourth Amendment constitutional doctrine, under which a person has no "reasonable expectation to privacy" in information she turns over to a third party. Consequently, U.S. constitutional privacy ends where EU privacy protection only begins. Fourth, it illustrates the Byzantine statutory scheme governing electronic communications stored by online service providers, which provides surprisingly weak protection to the contents of user communications. Fifth, it outlines the recent spate of national security inspired data retention legislation, which not only permits, but actually mandates the retention of users' search logs, further eroding users' privacy. Finally, it reintroduces the law of confidentiality and evidentiary privileges as a potentially effective solution to users' deficient privacy rights.

Part VI concludes, taking note of some of the issues worthy of future research.

Throughout this article I use Google as a proxy for the entire search engine industry. While Google dominates search, it is by no means the only actor in the field, and, setting aside the Privacy International report discussed above, nor is it worse than any of its major competitors.[14] I use Google for comfort of exposition and since, truth must be said, I would not think of using another search engine myself.

I. Two types of search engine privacy

[14] Some search engines do provide a greater degree of privacy, competing with Google, Yahoo and Microsoft on precisely this issue. *See*, *e.g.*, Jacqui Cheng, *Ask.com to offer anonymous search with AskEraser*, ARSTECHNICA, Jul. 20, 2007, available at http://arstechnica.com/news.ars/post/20070720-ask-com-to-offer-anonymous-search-with-askeraser.html. Yet the differences between the privacy practices of the major players are mundane and in some aspects Google has a better track record than the competition.

Search engine privacy comes in two flavors. On the one hand, there is the privacy interest of the search target. The power of search has significantly reduced the transaction costs of compiling digital dossiers profiling a person's activities. Before the advent of search engines, we enjoyed a degree of "practical obscurity," protecting our privacy interest in issues such as litigation, asset ownership, past employment and political opinion. Although such information has always been in the public sphere, it was protected *de facto* from all but skilled investigators or highly motivated researchers, due to the practical difficulty and costs involved in uncovering and compiling the data. Today such information has become available instantly and costlessly. On the other hand, there is the privacy interest of the person conducting the search ("user"). Search engines maintain comprehensive logs detailing users' search history. Such logs contain strikingly revealing records of user fears and aspirations, personal and professional data, financial condition, political affiliation, sexual orientation, health situation, and more. As such, they are highly personal and private in nature.

c) Search target privacy

Imagine a hypothetical (albeit daily) situation: you meet Sue, an old high school classmate, in an airport and out of curiosity enter her name in Google (that is, you "Google her"). You instantly obtain 25,000 search results, which are hyperlinks to the following documents: a court decision awarding Sue a restraining order against Bob, her second husband, an alcoholic who, the record shows, beat Sue and infected her with Hepatitis C; a public record listing Sue and her sister as the sole heirs of their grandfather, a well known industrialist; a copy of the Wichita Falls Queer Voice, referring to Sue as a major contributor to and frequent participant in gay and lesbian community events; a site featuring Sue as sixth grade teacher of the month in Wichita Falls Middle School; an online resume displaying Sue as former partner in a large New York law firm; and remarks (apparently) written and signed by Sue in an online forum titled "L. Ron Hubbard, my savior."

To be sure, most if not all of the information you find about Sue has always been in the public sphere. Yet much of it was traditionally protected by what Chris

Hoofnagle called "practical obscurity."[15] To access it, one would have to physically go to places, search through dusty file cabinets and incur potentially significant search costs.[16] And if finding the information used to be difficult, cross-referencing it to compile a personal profile was nearly impossible. Enter internet search engines with their nearly unrestricted capacity to store, organize, index, uncover and recover information. With convenient and costless search, anyone can be a private investigator, profiling targets at the click of a mouse.

Generally, access to information is a good thing, of course. We all benefit from finding the best consumer goods at rock bottom prices. We greatly value the increased access to information for research, education, business and pleasure. Indeed, search engines create enormous social benefits. Yet this comes at a great cost to search targets' privacy.

First, search engines facilitate the *aggregation* of search targets' personal data from a large number of disparate sources. As Solove recently explained, "[a] piece of information here or there is not very telling. But when combined together, bits and pieces of data begin to form a portrait of a person. The whole becomes greater than the parts."[17] Hence, even if one could find out offline that Sue is (apparently) rich or gay, the aggregation of online information about Sue, telling us she is a wealthy, unhappily married, scientologist school teacher, formerly practicing law in New York and infected with Hepatitis C, constitutes a privacy problem.

Second, search engines greatly increase *access* to search targets' personal data. As Justice Stevens holds in *Department of Justice v. Reporters Committee for Freedom of the Press*,[18] "there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single

[15] Chris Hoofnagle, *Search Engines and Individual Rights*, Pre-Conference Paper, "Regulating Search Conference", Yale Law School, Nov. 28, 2005, available at http://islandia.law.yale.edu/isp/search_papers/hoofnagle.pdf.

[16] As Battelle notes, "regardless of your prurient desire to know whether your new coworker has a messy divorce or a DUI in his otherwise well appointed closet, most of us will not spend an afternoon down in the basement of our county courthouse to find out." Battelle, *supra* note 5, at p. 191.

[17] Solove, Taxonomy, *supra* note 13, at p. 507.

[18] 489 U.S. 749 (1989).

clearinghouse of information."[19] To use the previous example, access to the Wichita Falls Queer Voice requires finding an outlet carrying the journal, approaching the vendor to ask for it (possibly blushing), and obtaining a great deal of irrelevant information (assuming one is interested only in Sue), whereas online access is swift, apparently anonymous and precise.

Third, personal data accessed through search engines might be *inaccurate* or misleading. This may be the result of an innocent error or intentional interference. Perhaps the Hepatitis-infected Sue is another Sue altogether, yet her fate becomes intertwined with that of your former classmate due to an innocent typo. Alternatively, the Scientologist forum entry may have been written by a malicious colleague, trying to tarnish Sue's credentials as an educator. In either case, Sue has few practical options available to her to prevent the wrong impression reverberating through cyberspace.

Fourth, personal data indexed by search engines may subject search targets to a risk of tangible harm. Consider the Amy Boyer "cyberstalking" case.[20] Liam Youens, a former classmate of Ms. Boyer, who had been obsessed with her since high school, obtained her personal data, including home and work address, from Docusearch.com, a self proclaimed "premier provider of on-line investigative solutions." Mr. Youens used the information to locate Ms. Boyer at her workplace, murder her, and commit suicide.[21] In another case, David Mullins, a U.S. government employee, argued that he had been unlawfully dismissed due to a Google search by a supervisor, which revealed that he had been discharged from the Air Force.[22] Another example is whosarat.com,[23] a web site devoted to exposing the identities of witnesses cooperating with the government. The site posts police and FBI informants' names and mug shots, along with court documents detailing what they have agreed to do in exchange for

[19] *Ibid*, at p. 764.

[20] Remsberg v. DocuSearch, Inc., 816 A.2d 1001 (N.H. 2003). Docusearch.com is an on-line search agency that requires a fee for its services.

[21] *See* Herman T. Tavani & Frances S. Grodzinsky, *Cyberstalking, Personal Privacy, and Moral Responsibility*, 4 ETHICS & INFO. TECH. 123 (2002).

[22] Mullins v. Department of Commerce, 2007 WL 1302152 (Fed.Cir. 2007).

[23] Available at www.whosarat.com.

lenient sentences.[24] Clearly, the aggregation of such information and ease of online access thereto places informants at a grave risk of harm.

Search target privacy is a vexing issue,[25] yet this article focuses on the privacy of search engine *users*, the issue I turn to next.

d) Search user privacy

In August 2005, as part of its longstanding effort to enforce the Child Online Protection Act ("COPA"),[26] the U.S. government issued a subpoena to AOL, Google, Microsoft and Yahoo, requesting the addresses of all web sites indexed by the search engines as well as every search term entered by search engine users during a period of two months. The government was seeking to refute the assertion that filtering devices may work as well as or better than criminal prosecutions in achieving the COPA's aims of keeping pornographic materials away from children. The government wanted to prove its point by showing what the average Internet user is searching for, surmising that many of the searches lead to material "harmful to minors." The government was not interested in *who* was doing the searching and did not request information that could link the searches back to individual users.[27]

Of the four companies approached, only Google objected to the government subpoena, claiming that the request for information threatened its trade secrets and image as a protector of user privacy. In January 2006, following negotiations with

[24] Adam Liptak, *Web Sites Listing Informants Concern Justice Department,* NY TIMES, May 22, 2007.

[25] *See*, *e.g.*, Herman T. Tavani, *Search Engines, Personal Information and the Problem of Privacy in Public*, 3 INT'L REV. INFO. ETHICS 39 (2005).

[26] Pub. L. No. 105-277, 112 Stat. 2681 (1998) (codified as 47 U.S.C. § 231 (2000)). The law, intended to protect children from access to online pornography (not to confuse with child pornography), has repeatedly been challenged by the ACLU and struck down by the Supreme Court. *See* Reno v. ACLU, 521 U.S. 844 (1997) (invalidating COPA's predecessor, the Communications Decency Act of 1996, Pub. L. No. 104-104, 110 Stat. 133); ACLU v. Ashcroft, 322 F.3d 240 (3d Cir. 2003), aff'd, 124 S. Ct. 2783 (2004) (invalidating COPA).

[27] *See*, Gonzales v. Google, Trial Motion, Memorandum and Affidavit, Reply Memorandum in Support of the Motion to Compel Compliance With Subpoena Duces Tecum, 2006 WL 733758 ((Feb. 24, 2006), stating: "the government has not asked Google to produce any information that could identify the users of its search engines, or the computers from which any search terms have been entered. Instead, the government has asked for the production only of the actual text of a sample of queries entered on to the Google search engine, without any additional information identifying the source of that text."

Google, the government significantly scaled-down its request to a random sampling of one million URLs[28] in Google's indexed database together with all queries that have been entered on the search engine during a one-week period; and later to only 50,000 URLs and 5,000 entries from Google's query log.[29] Despite these modifications, Google maintained its objection to the Government's request. A United States District Court finally ruled that the government was entitled to compel Google to provide a sample of URLs, but that Google would not have to disclose any of its users' search queries.[30]

Most people who followed the story asked themselves not whether the government subpoena complied with the Federal Rules of Civil Procedure, but rather: "what? Google keeps a record of all of my online searches?" Surprisingly for users not rehearsed on Google's intricate privacy policy, the answer is simply "yes." Google records all search queries linked to a specific Internet Protocol (IP) address.[31] In its privacy policy, Google states: "our servers automatically record information that your browser sends whenever you visit a web site. These server logs may include information such as your web request, 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."[32] In addition, Google records the hyperlinks users click after obtaining their search results.[33]

A user's search history contains highly revealing, sensitive personal data. We use search engines to explore job opportunities, financial investments, consumer goods, sexual interests, travel plans, friends and acquaintances, matchmaking services, political issues, religious beliefs, medical conditions, and more. One's search history

[28] A URL, or "Uniform Resource Locator," is the global address of documents and other resources on the World Wide Web. *See* URL, WEBOPEDIA, available at http://www.webopedia.com/TERM/U/URL.html.

[29] *See* subpoena at http://i.i.com.com/cnwk.1d/pdf/ne/2006/google-doj/notice.of.stark.declaration.pdf.

[30] Gonzales v. Google, Inc., 234 F.R.D. 674 (N.D.Cal. 2006).

[31] For example, if a user enters a search for "kill neighbor" and "dispose of body," the URL for Google's reply, which will be logged by the search engine, is: http://www.google.com/search?hl=en&q=kill+neighbor+dispose+of+body.

[32] Google Privacy Policy, available at http://www.google.com/intl/en/privacypolicy.html#information. Also *see* Google Privacy FAQ, available at http://www.google.com/intl/en/privacy_faq.html, at section 4: "What are server logs? Like most web sites, our servers automatically record the page requests made when users visit our sites. These 'server logs' typically include your web request, 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."

[33] Google Privacy FAQ, *ibid*, at section 5.

eerily resembles a metaphorical X-ray photo of one's thoughts, beliefs, fears and hopes. Data contained in user search logs may be far more embarrassing and privacy intrusive than that the contents of e-mail correspondence or telephone calls. Consider the scrutiny you give to an e-mail message prior to clicking "send," compared to the utter carelessness before Googling a search query. Imagine an online dossier of yourself, residing on the servers of a multinational company, laden with terms such as "Britney nude," "growing marijuana," "impotence pills," "job search," "genital warts," "prozac side effects," "married gay men," etc.

A surprising peek into precisely such digital dossiers was provided courtesy of AOL in August 2006. AOL posted on a newly established web site, research.aol.com, a list of 20 million search queries entered by 658,000 users over a period of three-months. It hurried to take the data offline amid a maelstrom of public criticism concerning its privacy implications. Yet much of the information had already been downloaded, reposted and made searchable at a number of third party web sites. The privacy debacle ended when AOL issued a formal apology and dismissed its chief technology officer.

The detailed search records revealed by AOL underscore how much users unintentionally reveal about themselves when they use search engines. Consider some of the search queries entered by user 1515830:
chai tea calories
calories in bananas
aftermath of incest
how to tell your family you're a victim of incest
surgical help for depression
oakland raiders comforter set
can you adopt after a suicide attempt
who is not allowed to adopt
i hate men
medication to enhance female desire
jobs in denver colorado
teaching positions in denver colorado
how long will the swelling last after my tummy tuck

divorce laws in ohio
free remote keyloggers
baked macaroni and cheese with sour cream
how to deal with anger
teaching jobs with the denver school system
marriage counseling tips
anti psychotic drugs

Queries entered by users such as number 17556639 appear to manifest criminal intent and may consequently be used at trial as evidence of wrongdoing:[34]

how to kill your wife
pictures of dead people
photo of dead people
car crash photo

Similarly, consider the searches of user 336865:

sexy pregnant ladies naked
child rape stories
tamagotchi town.com
preteen sex stories
illegal child porn
incest stories
illegel anime porn

Other queries, such as those entered by user 100906, are less ominous but no less revealing:

cinncinati bell iwireless
addicted to love
women who love to much
learning to be single
should you call your ex

[34] *See, e.g.*, U.S. v. Schuster, 467 F.3d 614 (7th Cir. 2006); also see Harriet Ryan, *Florida man convicted of killing his wife during faked mugging, now faces death*, COURT TV NEWS, June 26, 2006, available at http://www.courttv.com/trials/barber/062406_verdict_ctv.html.

when your ex goes out of his way to run into u
slim upper thighs
prophet mohamed life teaching
missed period or light spotting
birthcontrol for morning after pill
l&n federal credit union
hes just not that into u
i dont have a career
should i get back with my divorced husband
questions about the bible
do i quailfy for food stamps in kentucky

And while the AOL query data were anonymized and users assigned random serial numbers, the New York Times demonstrated how by simple reverse engineering, the identity of anonymous users becomes easy to discern.[35]

II. User search logs – personally identifiable information?

Privacy concerns relate to personally identifiable information ("personal data"),[36] that is, information which can be used to uniquely identify, contact, or locate a specific individual person. Federal privacy legislation protects personal data in a number of contexts, such as health information,[37] financial data,[38] or credit reports.[39] Similarly, the European data protection framework applies to "personal data," defined as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by

[35] Michael Barbaro & Tom Zeller, *A Face Is Exposed for AOL Searcher No. 4417749*, NY TIMES, Aug. 9, 2006.

[36] In this article I use the term "personal data," which is the European term for personally identifiable information.

[37] See Health Insurance Portability and Accountability Act ("HIPAA") of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996) (codified as amended in various sections of 42 U.S.C. (2000)).

[38] See Gramm-Leach-Bliley Financial Modernization Act of 1999, Pub. L. No. 106-102, 113 Stat. 1338 (codified in scattered sections of 12 U.S.C. (2000) and 15 U.S.C. (2000)).

[39] See Fair Credit Reporting Act of 1970, Pub. L. No. 91-508, 84 Stat. 1114 (codified at 15 U.S.C. §§ 1681-1681t).

reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."[40]

Information that cannot be linked to an individual person is not problematic from a privacy standpoint. Imagine we have highly revealing data about AOL user 100906, but we do not know, nor can we find out, who the user *is*. Or consider I tell you that X is a heroin-addicted, schizophrenic Satan worshipper, who earns $10,000 a month, half of which is spent on diet pills. Absent any indication as to the identity of X, this information is meaningless from a privacy perspective.

Do users' search logs constitute "personal data"? Can the data in search logs be traced to specific individuals? I show that they do, and therefore raise serious privacy problems. First, as noted above, search engines log a user's queries under such user's IP address. An IP address is a unique string of numbers assigned to a user's computer by her Internet Service Provider (ISP) in order to communicate with her computer on the network.[41] Simply put, it is the cyberspace equivalent of a real space street address or phone number. An IP address may be dynamic, meaning a different address is assigned to a user each time she logs on to the network; or static, that is assigned to a computer by an ISP to be its permanent Internet address. The question of whether an IP address constitutes "personal data" has been much debated in the EU.[42] It is equivalent to asking whether "435 Fifth Avenue, New York, New York" or "++1(212)435-2170" constitutes personal data. The answer depends on whether the address might be linked to an "identified or identifiable natural person" through reasonable means.[43] Clearly, a static address is more "personal" than a dynamic address; and in either case, an address is more "personal" in the possession of an ISP, which has the capacity to link it to a specific user's registration information, than in

[40] Article 2(a) of Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995 O.J. (L 281) 31 [hereinafter EU Data Protection Directive].

[41] *See* IP Address, WIKIPEDIA, available at http://en.wikipedia.org/wiki/IP_address.

[42] *See* CHRISTOPHER KUNER, EUROPEAN DATA PRIVACY LAW AND ONLINE BUSINESS (2nd Ed. Oxford University Press 2007), at p. 91-95; also *see* Article 29 Working Party Working Document, Privacy on the Internet – An Integrated EU Approach to On-line Data Protection, November 2000, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2000/wp37en.pdf.

[43] *See supra* note 40 and accompanying text.

the hands of other parties.[44] The European data protection watchdog, the Article 29 Working Party,[45] has recently opined that even dynamic IP addresses constitute "personal data." It stated that "unless the ISP is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data, to be on the safe side."[46] Consequently, even if Google could not link an IP address (and therefore her search log) to a specific individual, the fact that ISPs have such capability and that the government may order them to do so renders search logs "personal data" for privacy purposes. It is the capacity to link, not the actual linking, that makes the data personal.

Second, to overcome the difficulty of profiling users who access search engines using a dynamic IP address, search engines set "cookies" which tag users' browsers with unique identifying numbers.[47] Such cookies enable search engines to recognize a user as a recurring visitor to the site and amass her search history, even if she connects to the Internet via a different IP address. As a result of pressure by EU data protection regulators, Google has recently announced it would shorten the duration of its cookie, which was initially set to expire in 2038, to a period of two years after a user's last Google search.[48] The privacy benefits of such a move are doubtful, however, since as long as Google remains the Internet's leading search engine, users are bound to renew

[44] This is typically the case, although in certain circumstances, such as a user logging on to the Internet anonymously in an Internet café, even the ISP cannot link the address to an individual user.

[45] The Article 29 Working Party is the group of national data protection commissioners created by Article 29 of the Data Protection Directive and charged with its interpretation. *See* generally Joel Reidenberg, *Resolving Conflicting International Data Privacy Rules in Cyberspace*, 52 STAN. L. REV. 1315, 1364-66 (2000).

[46] See Article 29 Working Group Opinion 4/2007 on the concept of personal data, Jun. 20, 2007, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf, at p. 17.

[47] The Google privacy policy states: "When you visit Google, we send one or more cookies - a small file containing a string of characters - to your computer that uniquely identifies your browser. We use cookies to improve the quality of our service by storing user preferences and tracking user trends, such as how people search. Most browsers are initially set up to accept cookies, but you can reset your browser to refuse all cookies or to indicate when a cookie is being sent. However, some Google features and services may not function properly if your cookies are disabled." *See* Google Privacy Policy, *supra* note 32. As a matter of fact, few users change their browser's default settings to reject cookies. *See* Jessica J. Thill, *The Cookie Monster: From Sesame Street to Your Hard Drive*, 52 S.C. L. REV. 921 (2001).

[48] *See* Cookies Expiring Sooner to Improve Privacy, Official Google Blog, Jul. 16, 2007, available at http://googleblog.blogspot.com/2007/07/cookies-expiring-sooner-to-improve.html.

the two-year period on a daily basis.[49] One of the major weaknesses of a cookie as a tracking device is the fact that it is accessibly only by the web server that placed it on a user's computer. In other words, the New York Times cookie is read by the New York Times web site, but not by Yahoo or Wikipedia. You might therefore think of a cookie as a device that helps one snoop after a guest in her own house, but not in neighboring houses or public areas. However, this weakness has been overcome by Google in its recent takeover of advertising powerhouse DoubleClick.[50] DoubleClick is the leading provider of Internet-based advertising, tracking users' behavior across cyberspace and placing advertising banners on web sites. The company is a long-time nemesis of privacy advocates. In February 2000, EPIC filed a complaint with the FTC alleging that DoubleClick was unlawfully tracking the online activities of Internet users and combining surfing records with detailed personal profiles contained in a national marketing database.[51] The case ended in a settlement, pursuant to which DoubleClick undertook a line of commitments to improve its data collection practices, increase transparency and provide users with opt out options.[52] Doubleclick continues to utilize third-party cookies[53] as well as its "DART" (Dynamic, Advertising, Reporting, and Targeting) technology to track user activity across multiple web sites. In its recent complaint to the FTC about the Google-DoubleClick merger, EPIC alleged that by purchasing Doubleclick, Google expanded its ability to pervasively monitor users not only on its web site but also on cyberspace as a whole.[54]

[49] *See*, *e.g.*, Ryan Singel, *Google Changes Cookie Policy but Privacy Effect is Small*, WIRED BLOG NETWORK, July 16, 2007, available at http://blog.wired.com/27bstroke6/2007/07/google-changes-.html.

[50] Elinor Mills, Google buys ad firm DoubleClick for $3.1 billion, CNET NEWS.COM, Apr. 13, 2007, available at http://www.news.com/Google+buys+ad+firm+DoubleClick+for+3.1+billion/2100-1024_3-6176079.html?tag=st.rn.

[51] In the Matter of DoubleClick, Complaint and Request for Injunction, Request for Investigation and for Other Relief, before the Federal Trade Commission (Feb. 10, 2000), available at http://www.epic.org/privacy/internet/ftc/DCLK_complaint.pdf.

[52] Joel Winston, Acting Associate Dir., Div. of Fin. Practices, FTC, Letter to Christine Varney, Esq., Jan. 22, 2001, available at http://www.ftc.gov/os/closings/staff/doubleclick.pdf.

[53] "While cookies are only sent to the server setting them or one in the same Internet domain, a web page may contain images or other components stored on servers in other domains. Cookies that are set during retrieval of these components are called third-party cookies. Advertising companies use third-party cookies to track a user across multiple sites." *See* HTTP cookie, WIKIPEDIA, available at http://en.wikipedia.org/wiki/HTTP_cookie.

[54] Complaint and Request for Injunction, *supra* note 11; Also *see* In the Matter of Google and DoubleClick, Supplemental Materials in Support of Pending Complaint and Request for Injunction, Request for Investigation and for Other Relief, (Jun. 6, 2007), available at http://www.epic.org/privacy/ftc/google/supp_060607.pdf; Canadian Internet Policy and Public Interest Clinic, Section 9 Application for an Inquiry into the Proposed Merger of Google, Inc. and DoubleClick Inc. (Aug. 2, 2007) (addressed to Canadian Competition Bureau), available at http://www.cippic.ca/uploads/Google-DC_s.9_CompAct_complaint_FINAL.pdf.

Third, much like IP addresses, cookies are arguably not "personal data" because they identify a specific browser (typically, a computer) as opposed to an individual person. Yet, if a cookie and related search log could be cross-referenced with an individual's name, the cookie itself would become personal data. Think of the cookie as a label on a "box of personal data" of an unnamed person, who is under investigation by a private investigator. Typically, the label says something like "740674ce2123e969," and thus does not implicate anyone's privacy. Yet, once the private investigator comes across the person's name, she immediately affixes it to the label, rendering the contents of the box "personal data." The box of personal data is of course analogous to a user's search log and Google to the private investigator. And there are plenty of instances in which Google comes across a user's real name. In addition to its search engine, Google provides users with a wide array of online services, many of which require registration using real name and e-mail address credentials. First and foremost is Gmail, the ubiquitous web based e-mail service launched in April 2004 as a private beta release by invitation only and opened to the public in February 2007. Gmail gained its prominence and notoriety by providing a simple bargain for users: *get* an unprecedented amount of online storage space; *give* Google the opportunity to scan your e-mails' contents and add to them context-sensitive advertisements.[55] The launch of Gmail turned out to be one of the most controversial product launches in the history of the Internet and placed Google at the center of a fierce privacy debate.[56] Privacy advocates criticized the precedent set by Google of eliminating a person's expectation of privacy in the contents of her communications, as well as the consequential violation of non-subscribers' privacy interests in their correspondence.[57] This article does not address the serious privacy issues raised by Gmail itself, but rather the synergetic privacy risk created by cross-referencing user search logs with

[55] When Gmail was initially launched in 2004 with 1GB of storage space, Hotmail, its leading competitor, provided users with 2MB (that is, 0.2% of what Gmail gave). *See* Kim Zetter, *Free E-Mail With a Steep Price?*, WIRED NEWS, Apr. 1, 2004, available at http://www.wired.com/news/business/0,1367,62917,00.html, stating that "[t]he size of the storage would blow away offerings from rivals like Yahoo and Microsoft's Hotmail."

[56] *See* Matthew A. Goldberg, *The Googling of Online Privacy: Gmail, Search-Engine Histories and the New Frontier of Protecting Private Information on the Web*, 9 LEWIS & CLARK L. REV. 249, 250 (2005); Jason Isaac Miller, *Note, "Don't Be Evil": Gmail's Relevant Text Advertisements Violate Google's Own Motto and Your E-Mail Privacy Rights*, 33 HOFSTRA L. REV. 1607 (2005).

[57] *See* Thirty-One Privacy and Civil Liberties Organizations Urge Google to Suspend Gmail, Privacy Rights Clearinghouse, Apr. 19, 2004, available at http://www.privacyrights.org/ar/GmailLetter.htm; also *see* Gmail Privacy Page, EPIC web site, available at http://www.epic.org/privacy/gmail/faq.html.

information collected by Gmail as part of the registration process. In other words, registration to Gmail or additional Google services such as Google Talk (instant messaging service), Google Reader (RSS feeds), Google Calendar (a user's schedule), or Google Checkout (credit card/payment information for use on other sites),[58] places the missing "name tag" on a user's search log, thereby rendering its contents highly combustive from a privacy perspective.[59] Notice that cross-referencing user search logs with registration information is distinct from Google correlating search logs with users' e-mail contents, the prospect of which is an additional cause of concern for privacy advocates.[60] It simply means Google can pick the name of a user off of her registration form and attach it to a cookie, which serves as the key to her search log. In other words, because Google uses the same cookie to maintain a particular user's search history and to identify her when she logs-on to her Gmail account, the anonymous nature of the cookie is lost and the search log becomes sensitive personal data.

Finally, as demonstrated by the New York Times in the AOL case,[61] even thoroughly anonymized search logs can be traced back to their originating user. This can be done by combing search queries for personal identifiers, such as a social security numbers or credit card details. It becomes simpler yet by the tendency of users to run "ego searches" (also known as "vanity searches" or "egosurfing"), the practice of searching for one's own name on Google (once, twice, or many times per day).[62] In fact, in its effort to quash the government subpoena issued in *Gonzales v. Google*, Google itself posited that "search query contents can disclose identities and personally identifiable

[58] There is also Google Web History, of course, which provides consenting users with a personalized search experience linked to a personal account. Hence, Google Web History explicitly de-anonymizes one's search log.

[59] While it is true that users may register for services such as Gmail with a false or pseudonymous name, I suspect few do. I use Gmail as my main e-mail account due to its geographic and chronological versatility (you do not have to change e-mail addresses each time you relocate or switch jobs) and storage space. I use my real name, since I would not want colleagues or friends to receive e-mails from "Dr. No" or "Omer1970" and have to guess that I am the sender.

[60] See, e.g., Goldberg, *supra* note 56, at p. 252. While Google has said it had no plans to correlate e-mail and searches, it maintains the ability to do so and does not rule out doing so in the future. *Ibid*, at p. 254.

[61] *See* Barbaro & Zeller, *supra* note 35.

[62] Egosurfing, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Egosurfing.

information such as user-initiated searches for their own social security or credit card numbers, or their mistakenly pasted but revealing text."[63]

To sum, the contents of user search logs are clearly personal in nature. The question is whether such contents may be traced to a specific user. Google's ability to combine IP addresses, persistent cookies and user registration information renders the data in search logs not only personal but also personally identifiable.

III. Use of data

Why do search engines maintain search logs? What is the information used for, and by whom? Who else may access the information and under what conditions? The answers to these questions will affect the privacy analysis of user search logs. This part distinguishes between use of information by the search engine itself and use by other parties.

c) Use by search engine

The recent investigation launched by the Article 29 Working Party into Google's privacy and data retention practices[64] prompted Google to publicly explain its need to maintain user search logs. In his response to the Article 29 Working Party, Peter Fleischer, Google's chief privacy officer, explains that retention of search logs is critical to Google's ability to operate and improve its services, and to provide adequate security for its users.[65] Google faces the daunting task of having to guess what users intend, essentially "read their minds," based on two or three words they enter as a search query. As Google co-founder Larry Page puts it, "[t]he perfect search engine would understand exactly what you mean and give back exactly what you want."[66] What complicates matters even more is that a single query may indicate

[63] See, Gonzales v. Google, Trial Motion, Memorandum and Affidavit, Google's Opposition to the Government's Motion to Compel, 2006 WL 728287 (Mar. 13, 2006)
[64] Article 29 Working Party Letter, *supra* note 9.
[65] Letter of Mr. Peter Fleischer, Global Privacy Counsel, Google, to Mr. Peter Schaar, Chairman, Article 29 Data Protection Working Party (Jun. 10, 2007), available at http://www.epic.org/privacy/ftc/google/gres_a29_061007.pdf.
[66] *See* Google.com, Corporate Information, Our Philosophy, Never settle for the best, available at http://www.google.com/corporate/tenthings.html. James Grimmelmann observes: "Divining user intent

different intentions in different contexts.[67] For example, the words "Paris Hilton video" might be entered by a user searching for accommodation in the French capital, or (more likely) by one eager to follow the celebrity heiress' latest antics. Similarly, a search for "king of France," which would normally call for information about French monarchy, might have a different meaning during the summer months, when users search for information about the *Tour de France* and its onetime *roi*, Lance Armstrong. Google's Fleischer states that "analyzing log data is an important tool to help our engineers refine search quality and build helpful new services."[68] He points out, for example, that Google Spell Checker, which automatically looks at a query and checks to see if the user entered the most common (and therefore, typically correct) version of a word’s spelling, is based on search log analysis. For example, if a user enters the words "Condoleza Rice," her search results would be preceded by the question: "Did you mean: Condoleezza Rice?"

Google emphasizes the use of search logs in preventing fraud and abuse. Fleischer states that "it is standard among Internet companies to retain server logs with IP addresses as one of an array of tools to protect the system from security attacks . . . Historical logs information can also be a useful tool to help us detect and prevent phishing, scripting attacks, and spam, including query click spam and ads click spam."[69] Google uses search logs in its technical arms race against web sites and their agents that employ illicit means in order to improve search results' placement[70]; practices collectively known as "black hat" search engine optimization (SEO).[71] Analysis of search logs may help detect "black hat" SEO methods, such as "link farms" (a group of web sites hyperlinking each other, also known as "spamdexing,"

from a search query is a notoriously difficult problem and the same query may indicate a different intent in different contexts." Grimmelmann, *supra* note 2, at p. 7.

[67] In a piece recently written for the Financial Times, Mr. Fleischer writes: "There was a survey conducted in America in the 1980s that asked people a deceptively simple question: 'Who was shot in Dallas?' For many who had lived through the national trauma of 1963 . . . there was only one answer: JFK. For others, who followed every twist of the Ewing family . . . there was also only one answer: JR." Peter Fleischer, *Google's search policy puts the user in charge*, FIN. TIMES, May 25, 2007, available at http://www.ft.com/cms/s/2/560c6a06-0a63-11dc-93ae-000b5df10621.html.

[68] Fleischer Letter, *supra* note 65.

[69] *Ibid.*

[70] High placement among search results is one of the main determinants of success for a business today. As Nissenbaum and Introna put it, "to exist [online] is to be indexed by a search engine." Introna & Nissenbaum, *supra* note 3, at p. 173.

[71] *See* Search engine optimization, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Search_engine_optimization. For SEO generally *see* SEMPO, Search Engine Marketing Professional Organization, available at http://www.sempo.org/home.

i.e., spamming the search engine index),[72] Google Jacking (creating a rogue copy of a popular web site which shows contents similar to the original, but redirects users to an unrelated or malicious web site),[73] or Keyword stuffing (loading a web page with keywords, for example, by coloring text to blend with the background).[74]

To be sure, few if any users would disapprove of optimizing search results and combating fraud. Yet Google also analyzes search logs for revenue generating purposes, particularly for targeting and maximizing the effectiveness of advertisements. Google, after all, is an advertising company.[75] As James Grimmelmann notes, "the overwhelmingly predominant model for web search today is contextual advertising, in which, the search engine, in addition to showing its users results, shows them advertisements, most commonly textual ones."[76] The name of the game in online advertising, which is dominated by the pay-per-click (PPC) method of billing,[77] is maximizing click-through rate (CTR), that is, the number of times users who visit a web page featuring an advertisement actually click the ad.[78] And in order to maximize CTR, Google gauges user tastes, preferences, interests and needs. Google's chief executive officer, Eric Schmidt, stated: "If we target the right ad to the right person at the right time and they click it, we win."[79] Targeting "the right ad to the right person at the right time" requires knowing the users; and knowing the users means analyzing their search history.

No company evaluates user preferences as well as Google. Research shows that users click advertisements 50 percent to 100 percent more often on Google than they do on its main competitor, Yahoo.[80] The cream of the crop in PPC advertising programs is Google AdWords,[81] the company's main source of revenue. And AdWords gives

[72] *See* Link Farm, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Link_farm.

[73] *See* Page hijacking, WIKIPEDIA, available at http://en.wikipedia.org/wiki/302_Google_Jacking.

[74] *See* Keyword stuffing, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Keyword_stuffing.

[75] Saul Hansell, *Google Wants to Dominate Madison Avenue, Too*, NY TIMES, Oct. 30, 2005, available at http://tinyurl.com/b3w6t.

[76] Grimmelmann, *supra* note 2, at p. 8.

[77] "Pay per click (PPC) is an advertising model used on websites, advertising networks, and search engines where advertisers only pay when a user actually clicks on an ad to visit the advertiser's website." *See* Pay per click, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Pay_per_click.

[78] Click-through rate, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Click-through_rate.

[79] Hansell, *supra* note 75.

[80] *Ibid.*

[81] *See* Google AdWords, available at http://adwords.google.com/select/Login.

priority to advertisements that bring in the most money, based not only on an advertiser's bid per click but also on the number of times users click the ad.

Google argues (while others dispute[82]) that it engages in less "behavioral targeting" (*i.e.*, tailoring advertisements to individual users based on their preferences gleaned from search logs) than its main competitors, Microsoft's adCenter and Yahoo's SmartAds.[83] Yet Google admits that its system "incorporates a large number of signals (such as the user's query, the user's location, type of site, contents, and the advertiser's landing page) when targeting and ranking ads."[84] And to quote Google co-founder Sergei Brin, "I don't think it's a big deal to show opera glasses to someone searching for binoculars that you somehow infer is a woman."[85] This, in other words, is what behavioral targeting is all about.

Google, Microsoft, Yahoo, and other major search engines, harbor giant – and ever growing – warehouses of user search logs. John Battelle called this great body of knowledge "The Database of Intentions,"[86] meaning "the aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result."[87] Taking a step back, one might ask: why should search engines *not* retain user search logs? Given the increasingly small costs of data warehousing,[88] relative dearth of regulation, and potentially lucrative use of the information, search engines have little incentive to delete users' search logs. This treasure trove of information is a "massive

[82] *See*, *e.g.*, Google's New Behavioral Targeting For AdWords Reviewed, Searchenginewatch.com, Aug. 2, 2007, available at http://blog.searchenginewatch.com/blog/070802-125836; Vishesh Kumar, *Google Shuns Behavioral Ad Targeting – for Now*, THESTREET.COM, Aug. 6, 2007, available at http://tinyurl.com/2gpnjh, stating "a close look at Google's carefully crafted position about behavioral targeting suggests that the company may be much more inclined to use the technique than the headlines suggest."

[83] Declan McCullagh, *In their own words: Search engines on privacy*, CNET NEWS.COM, Aug. 13, 2007, available at http://news.com.com/In+their+own+words+Search+engines+on+privacy/2100-1029_3-6202047.html; also *see* Eric Auchard, *Google wary of behavioral targeting in online ads*, REUTERS, Aug. 1, 2007, available at http://tinyurl.com/2rptw3.

[84] McCullagh*, ibid.*

[85] *Ibid.*

[86] *See*, generally, Battelle, *supra* note 5.

[87] John Battelle, *The Database of Intentions*, John Battelle's Searchblog, Nov. 13, 2003, available at http://battellemedia.com/archives/000063.php.

[88] Battelle notes that "the average cost per megabyte for storage has plummeted, and it will continue to drop until the point where it essentially reaches zero." Battelle, *supra* note 5, at p. 10. Also *see* John Markoff, *Reshaping the Architecture of Memory*, NY TIMES, Sep. 11, 2007, available at http://www.nytimes.com/2007/09/11/technology/11storage.html?ref=technology, stating "if an idea that Stuart S. P. Parkin is kicking around in an I.B.M. lab here is on the money, electronic devices could hold 10 to 100 times the data in the same amount of space."

database of desires, needs, wants, and likes that can be discovered, subpoenaed, archived, tracked, and exploited to all sorts of ends."[89] It is these additional uses I turn to now.

d) Use by third parties

The Database of Intentions is a valuable asset, a virtual honey pot for various third parties, ranging from national security and law enforcement personnel to hackers and identity thieves. At present, search engines do not sell users' personal data to third parties,[90] yet they retain the ability to do so in the future.[91] Search engines do share user data with subsidiaries, affiliated companies and other "trusted" business partners for the purpose of data processing and the provision of services.[92] In addition, they retain the right to transfer data to a third party in case a merger or consolidation.[93]

Certain third parties may – and in fact do – try to obtain user personal data from search engines through legal process. First and foremost, the government may use search logs for national security and law enforcement purposes, including the prevention, detection and prosecution of crimes.[94] Clearly, a user searching for terms such as "illegal child pornography" or "prepare pipe bomb" warrants law enforcement intervention. And indeed, governments tend to emphasize the most severe criminal activities, such as pedophilia, terrorism and organized crime, when seeking authority to access user search logs.[95] Few would dispute the imperative to provide government with all necessary tools to combat such heinous acts. Yet the picture becomes murkier when the government seeks access to search logs of individuals who search for ""murder husband" or even "how to cheat IRS." And it is certainly more complex

[89] Battelle, *ibid*.
[90] *See* Google Privacy Policy (Oct. 14, 2005), available at http://www.google.com/intl/en/privacypolicy.html#information; Yahoo! Privacy Policy (Nov. 22, 2006), available at http://info.yahoo.com/privacy/us/yahoo/details.html; Microsoft Online Privacy Statement (Jan. 2006), available at http://privacy.microsoft.com/en-us/fullnotice.aspx#use.
[91] *See* discussion *infra* notes 210-12 and accompanying text.
[92] *See* Privacy Policies, *supra* note 90.
[93] *Ibid.* The term "trusted" appears in the Google and Yahoo privacy policies.
[94] *See* generally Michael D. Birnhack & Niva Elkin-Koren, *The Invisible Handshake: The Reemergence of the State in the Digital Environment*, 8 VA. J.L. & TECH. 6 (2003).
[95] *See*, *e.g.*, Declan McCullagh, *Terrorism invoked in ISP snooping proposal*, CNET NEWS.COM, May 30, 2006, available at http://news.com.com/2100-1028_3-6078229.html; Prepared Remarks of Attorney General Alberto R. Gonzales at the National Center for Missing and Exploited Children (NCMEC) (Apr. 20, 2006), available at http://www.usdoj.gov/ag/speeches/2006/ag_speech_060420.html.

when search terms, such as "Falun Gong" or "democracy Tiananmen," are criminalized in one jurisdiction yet entirely legitimate in another.[96]

All major search engines declare in their privacy policies that they comply with legal process and government requests for information. Google, for example, affirms it will disclose user data to "satisfy any applicable law, regulation, legal process or enforceable governmental request."[97] A full search warrant, supported by an affidavit showing probable cause, would in all cases enable law enforcement officers to access search engine data.[98] The New York Times recently reported that AOL alone responds to approximately 1,000 such criminal search warrants each month.[99] In most cases, however, much less than a full search warrant would suffice.[100] Search engines have been forthcoming in complying with government requests for users' personal data even where the consequences for identified users have been dire.[101] Police is increasingly using search engine records as incriminating evidence in a variety of cases, ranging from homicide[102] to wireless hacking.[103]

This "Invisible Handshake" between search engines and law enforcement is troubling.[104] To be sure, Government surveillance is justified in limited circumstances. For example, under the Federal Wiretap Act,[105] interception of voice and electronic communications is permitted under certain conditions. Yet such

[96] *See* Clive Thompson, *Google's China Problem (and China's Google Problem)*, NY TIMES MAGAZINE, Apr. 23, 2006, available at http://tinyurl.com/re2cn.
[97] Google Privacy Policy, *ibid.*
[98] Grimmelmann, *supra* note 2, at p. 16.
[99] Saul Hansell, *Increasingly, Internet's Data Trail Leads to Court*, NY TIMES, Feb. 4, 2006, available at http://tinyurl.com/dpty3; Adam Liptak, *In Case About Google's Secrets, Yours Are Safe*, NY TIMES, Jan. 26, 2006, available at http://tinyurl.com/cmnzg.
[100] *See* discussion *infra* notes 296-301 and accompanying text; also *see* Gonzales v. Google, Amicus Brief of Center for Democracy & Technology in Support of Google's Opposition to the Motion to Compel of Attorney General Gonzales, 2006 WL 733757 (Mar. 13, 2006), stating: "That companies that receive subpoenas from the government routinely comply with even broad requests for information further demonstrates why the government must be required to show a sufficiently specific reason for demanding the information (...)"
[101] *See*, *e.g.*, Jim Kerstetter, *Group says Yahoo helped jail Chinese journalist*, CNET NEWS.COM, Sep. 6, 2005, available at http://news.com.com/Group+says+Yahoo+helped+jail+Chinese+journalist/2100-1028_3-5851705.html; but *see* recently, AP, *Brazilian prosecutors say Google has not provided Orkut user information regarding crimes*, INT'L HERALD TRIB., Aug. 22, 2007, available at http://www.iht.com/articles/ap/2007/08/22/business/LA-FIN-Brazil-Google.php.
[102] *See* Lester Haines, *Alleged techie killer Googled 'neck snap break'*, THE REGISTER, Nov. 14, 2005, available at http://www.theregister.com/2005/11/14/techie_murder_evidence.
[103] U.S. v. Schuster, *supra* note 34.
[104] Birnhack & Elkin-Koren, *supra* note 94.
[105] 18 U.S.C. §§ 2510-2522 (2000).

interventions are very limited in scope and require a strict evidentiary showing of "probable cause plus."[106] The prospect of pervasive surveillance by private sector corporations over the activities of all Internet users, globally, at all times, coupled with regular access by law enforcement, is menacing.

Government access to user search logs raises the additional risk of function creep. Data intercepted in a search for terrorists may eventually be used by the government to prosecute tax offenders or collect debt. Privacy invasions which may be deemed necessary to combat serious crime or national security risks appear disproportional when used for fiscal administration. And preventive law enforcement tests the limits of legitimate government action in a democratic society. Nabbing a terrorist before he realizes his plot to bomb a passenger jet is one thing.[107] It is quite another thing to arrest a teenager who runs Google searches for "kill guns," "prozac side effects," "brutal death metal bands," and "blood gore," and is therefore profiled by a data mining program as a potential "Columbine shooter." Indeed, you might not want him as a classmate of your daughter or son; but incarceration on the basis of Google searches, essentially thoughts as opposed to deeds, is surely problematic.[108]

In addition to criminal activity, search engine logs may be useful for litigants in civil cases, including copyright infringement, divorce, libel, employment disputes, and shareholder actions.[109] The recording industry has been particularly aggressive in its attempts to identify users who violate copyright law through service of subpoenas on

[106] Jayni Foley, *Are Google Searches Private? An Originalist Interpretation of the Fourth Amendment in Online Communication Cases*, 22 BERKELEY TECH. L.J. 447, 454 (2007).
[107] *See* MARKLE FOUNDATION, MOBILIZING INFORMATION TO PREVENT TERRORISM: THIRD REPORT OF THE MARKLE FOUNDATION TASK FORCE (July 2006), available at http://www.markle.org/downloadable_assets/2006_nstf_report3.pdf.
[108] Consider the following exchange, from the film Minority Report:
"Knock, knock.
'Who's there?'
'FBI. You're under arrest.'
'But I haven't done anything.'
'You will if we don't arrest you,' replied Agent Smith of the Precrime Squad." MINORITY REPORT (20th Century Fox 2002), cited in K.A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 1 (2003); also *see* Clive Thompson, *Open Source Spying*, NY TIMES MAGAZINE, Dec. 3, 2006, available at http://tinyurl.com/2tewer; COLLEEN MCCUE, DATA MINING AND PREDICTIVE ANALYSIS: INTELLIGENCE GATHERING AND CRIME ANALYSIS (Butterworth-Heinemann 2006).
[109] Fred von Lohmann, *Could Future Subpoenas Tie You to 'Britney Spears Nude'?*, LAW.COM, Feb. 6, 2006, available at http://www.law.com/jsp/article.jsp?id=1138961111185.

online intermediaries, mainly ISPs.[110] While such cases have not yet been extended to search engines, the mega-lawsuit recently brought by Viacom against YouTube and its corporate parent, Google, for contributory and vicarious copyright infringement may have the effect of drawing search engines into the fray.[111]

Third party subpoenas (*subpoena duces tecum*) are issued as a matter of course in civil litigation based on the relevance of evidence held by the intermediary.[112] The relevancy requirement is liberally construed to permit the discovery of information which ultimately may not be admissible at trial.[113] An employer, for example, may seek to summon an employee's search logs to prove he had used his computer for private purposes or, worse yet, to seek pornographic material on the job. A couple engaged in divorce proceedings may subpoena each other's search logs; the husband to prove his wife planned a secret vacation getaway; the wife to prove her husband sought homosexual escort services. Shareholders may subpoena corporate insiders' search queries to prove that they had engaged in insider trading.

Overbroad subpoenas seeking irrelevant information may be quashed or modified. A court must modify a subpoena if it subjects a non-litigant to an undue burden.[114] In *Gonzales v. Google*, Google argued that the government subpoena of user search logs constituted an undue burden, based on the time and resources allegedly required to gather the requested information, as well as the risk to Google trade secrets and confidential commercial information. Google further claimed that the information requested by the government was irrelevant and that it imposed on Google the risk of responding to inadequate process based on the Electronic Communications Privacy

[110] *See* Recording Industry Association of America, Inc. v. Verizon Internet Services, Inc., 351 F.3d 1229 (D.C. Cir. 2003); the question in the Verizon case was whether Section 512(h) of the Digital Millennium Copyright Act (DMCA), Pub. L. No. 105-304, 112 Stat. 2860 (Oct. 28, 1998), allows a copyright owner to subpoena the identity of an ISP subscriber where the alleged infringing material is present on the ISP's server. Also *see* Sony Music Entertainment Inc. v. Does 1-40, 326 F. Supp. 2d 556 (S.D.N.Y. 2004).

[111] *See* Viacom v. YouTube and Google, Complaint for Declaratory and Injunctive Relief and Damages (D.Ct. S.D.N.Y., Mar. 13, 2007), available at http://news.com.com/pdf/ne/2007/ViacomYouTubeComplaint3-12-07.pdf.

[112] Fed. R. Civ. Proc. 26(b), 45.

[113] *See*, *e.g.*, Shoen v. Shoen, 5 F.3d 1289 (9th Cir. 1993).

[114] Fed. R. Civ. Proc. 45(c)(3)(A)(iv). *See* Mattel Inc. v. Walking Mountain Prods., 353 F.3d 792 (9th Cir. 2003).

Act (ECPA).[115] Users' privacy rights were raised by neither the government nor Google in their arguments in the case. In fact, the court explicitly stated it "raises, *sua sponte*, its concerns about the privacy of Google's users apart from Google's business goodwill argument."[116] Google *did* argue that "the production of the requested data will result in a chilling effect on Google's business and user trust." However, the user trust argument was based on Google allegedly having to "compromise its privacy principles and produce [data] to the government on such a flimsy request." The premise of such an argument apparently is that if the subpoena had not been based on a "flimsy request," Google *would* comply. Indeed, such a conclusion may be drawn from Google's privacy policy. In any event, the user privacy argument was raised to establish the potential commercial harm to Google itself, and not as an independent basis to quash the subpoena.

Users' fundamental rights have occasional been raised by online intermediaries seeking to resist third party subpoenas. ISPs have mostly relied on users' free speech (but not privacy) interests in litigation intended to reveal users' identity.[117] Yet courts have yet to determine what speech interests, if any, users have in anonymous *search*.[118]

In addition to government and private actors serving legal process, Google's information goldmine is bound to attract hackers and data thieves. Valuable databases get infiltrated regardless of the robustness of security measures in place. Security breaches abound even in highly guarded industries such as financial services, health services and telecommunications. Rogue employees sell data to criminals; negligent employees lose laptops; computers are stolen and back up tapes lost; passwords compromised and firewalls lowered.

[115] Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended at 18 U.S.C. §§ 2510-2520, 2701-2711, 3121-3127 (2000)) [hereinafter ECPA].

[116] Gonzales v. Google, *supra* note 30, at p. 687.

[117] This line of cases is based on the Supreme Court ruling in McIntyre v. Ohio Elections Comm'n, 514 U.S. 334, 115 S.Ct. 1511, 131 L.Ed.2d 426 (1995), establishing the right to anonymous speech. *See, e.g.*, In re Subpoena Duces Tecum to America Online, Inc., 2000 WL 1210372, 52 Va. Cir. 26 (2000), rev'd on other grounds, 542 S.E.2d 377 (2001); but *see* Doe v. Cahill, 884 A.2d 451, 33 Media L. Rep. 2441 (Del. Sup. 2005). Also *see* 2TheMart.Com, 140 F.Supp.2d at 1090 (granting motion to quash subpoena seeking identities of anonymous ISP subscribers in shareholder derivative suit).

[118] Grimmelmann, *supra* note 2, at p. 16.

California's Security Breach Information Act ("SB 1386") of 2003,[119] which was followed by a spate of state legislation across the U.S.,[120] has led to the disclosure of security breaches in companies such as Citigroup, Bank of America, CardSystems, Merrill Lynch, T-Mobile, LexisNexis, Choicepoint, and Time Warner, as well as in dozens of colleges and universities, hospitals and federal, state and municipal government departments.[121] The number of people whose persona data have been affected by security breaches through September 2007 is estimated at more than 150 million, including, for example, 40 million Visa and MasterCard accounts compromised by a hacking incident at data processor CardSystems Solutions; 28 million veterans whose names, Social Security numbers, dates of birth, phone numbers and addresses were stored on a laptop computer stolen from a government employee's home; 3.9 million accountholders whose data have been compromised by Citigroup when it lost a shipment of computer backup tapes sent via UPS; and approximately 145,000 individuals whose personal data were sold by data aggregator Choicepoint to criminals posing as legitimate businesses. The point is that no matter what security measures are in place, data stored will eventually be data breached. The best method of dealing with data security, and consequently data subject privacy, is not storing personal data in the first place. The larger the storage base and more valuable the data therein, the more attractive and lucrative it becomes for hackers, thieves and cash-strapped employees.[122]

IV. Privacy problems

Any discussion of the right to privacy ultimately reaches the most basic of questions, namely "what does privacy *mean*?" What does it mean when I say that the collection and use of search logs may be privacy invasive? Numerous attempts have been made

[119] Cal. Civ. Code § § 1798.29, .82, .84 (West Supp. 2006).

[120] *See* generally John Kennedy, *Slouching Towards Security Standards: The Legacy Of California's SB 1386*, 865 PLI/PAT 91 (2006) (reviewing legislation); also *see* Gramm-Leach-Bliley Act § 5, 15 U.S.C. § 6801, 6805 (2000), and the Interagency Guidance issued pursuant thereto: INTERAGENCY GUIDANCE ON RESPONSE PROGRAMS FOR UNAUTHORIZED ACCESS TO CUSTOMER INFORMATION AND CUSTOMER NOTICE, 70 Fed. Reg. 15,736, 15,743 (Mar. 29, 2005); INTERAGENCY GUIDANCE ESTABLISHING INFORMATION SECURITY STANDARDS, 69 Fed. Reg. 77,610 (Dec. 28, 2004).

[121] *See* Privacy Rights Clearinghouse, A Chronology of Data Breaches, available at http://www.privacyrights.org/ar/ChronDataBreaches.htm.

[122] *See* generally, Lynn M. LoPucki, *Human Identification Theory and the Identity Theft Problem*, 80 TEX. L. REV. 89 (2001).

to define privacy and many are no doubt forthcoming.[123] For the purposes of this article, it is sufficient to build on one of the existing frameworks for analyzing privacy. I chose Daniel Solove's *Taxonomy of Privacy*, which is comprehensive, topical and robust.[124] Solove bases his taxonomy on *activities that invade privacy*. As I show below, collection, aggregation, storage, use and transfer of search logs by search engines raise many of the privacy problems surveyed by Solove and thus fit neatly into his framework.

g) Aggregation

Solove defines aggregation as the "gathering together of information about a person."[125] He explains that "combining information creates synergies. When analyzed, aggregated information can reveal new facts about a person that she did not expect would be known about her when the original, isolated data was collected."[126] Part I above discussed the privacy invasive practice of aggregation in the context of search targets' privacy.[127] Yet user search logs too raise a significant data aggregation problem. User search logs aggregate vast amounts of data from tiny bits of information revealed by users gradually over time. Entering a search query for "French mountains," may not give much away; "French mountains" and "ski vacation" is more telling; add to that "Christmas deals," "gift to grandchild," "NY Paris flights," "category D car rentals," "five star hotels," and "disabled access" – and a lucid picture begins to emerge. Search by search, click by click, the profile and identity of a user becomes discernable.[128] And if this is evident after half a dozen searches, consider the wealth and depth of information collected in a search log containing thousands and thousands of searches over a period of months or years. Even the few users who are aware of search engines' data compilation practices

[123] Some of the notable works are Ruth Gavison, *Privacy*, 89 YALE L.J. 421 (1980); ALAN F. WESTIN, PRIVACY AND FREEDOM (1967); Charles Fried, *Privacy*, 77 YALE L.J. 475 1968); Prosser, *supra* note 13; Solove, Taxonomy, *supra* note 13;

[124] Solove, Taxonomy, *supra* note 13.

[125] Solove, *ibid*, at p. 507.

[126] *Ibid*, id.

[127] *See supra* note 17 and accompanying text.

[128] Barbaro & Zeller, *supra* note 35.

probably underestimate the impact of search logs on their privacy, effectively making them "transparent" over time.[129]

What complicates matters even more is the highly concentrated nature of the search engine industry.[130] Data aggregation by dispersed actors is less troubling from a privacy perspective. Although certain professionals, such as your doctor or banker, maintain a large aggregation of your personal data, doctors and bankers abound, so an investigator (or government agency) looking for your personal data would have to incur significant search costs surveying many different data collectors. Furthermore, you may "diversify" your personal data portfolio by spreading information among different healthcare providers and bank accounts. With search, you not only know that voluminous data are being compiled, but also who is compiling them. Government, private litigants, and hackers alike know that Google and, to a lesser extent, Yahoo and MSN are where the information is.[131]

If privacy invasive prospects of search logs had already been serious prior to Google's recent DoubleClick merger, the mega-transaction has raised the stakes even more. In its statement to the FTC supporting EPIC's complaint, the New York State Consumer Protection Board states that "[t]he combination of DoubleClick's Internet surfing history generated through consumers' pattern of clicking on specific advertisements, coupled with Google's database of consumers' past searches, will result in the creation of 'super-profiles,' which will make up the world's single largest repository of both personally and non-personally identifiable information."[132] This, no doubt, is data aggregation at its best (or worst).

[129] DAVID BRIN, THE TRANSPARENT SOCIETY – WILL TECHNOLOGY FORCE US TO CHOOSE BETWEEN PRIVACY AND FREEDOM? (1998).

[130] *See* NIELSEN//NETRATINGS, Nielsen//NetRatings Announces July U.S. Search Share Rankings (Aug. 20, 2007), available at http://www.nielsen-netratings.com/pr/pr_070820.pdf, listing Google with 53.3% of all searches, Yahoo with 20.1%, MSN/Windows Live Search with 13.6%; the next seven leading search engines combined have less than 10% of the market; an estimated 4.1 billion search queries conducted at Google during the month of July 2007. Also *see* Tair-Rong Sheu & Kathleen Carley, *Monopoly Power on the Web: A Preliminary Investigation of Search Engines*, available at http://arxiv.org/ftp/cs/papers/0109/0109054.pdf.

[131] *See*, *e.g.*, Battelle, *supra* note 5, at p. 6, noting that "the Database of Intentions … lives in many places, but three or four places in particular – AOL, Google, MSN, Yahoo (…)"

[132] Letter to FTC Chair Deborah Platt Majoras from Mindy Bockstein, Chairperson and Executive Director, State of New York, State Consumer Protection Board regarding "DoubleClick Inc. and Google. Inc. Merger" (May 1, 2007), available at http://www.epic.org/privacy/ftc/google/CPB.pdf.

h) Distortion

Information in search logs may be highly misleading, with potentially harsh results for users. Search queries such as "assassinate US president" do not necessarily imply criminal intent, but may rather point to a student writing a history seminar. Similarly, if you search for "growing marijuana," you are not necessarily a teenager considering a career modeled after Mary-Louise Parker's television character, but may rather be a parent concerned with growing drug use in schools.

A real-life example of the elusive distinction between fact and fiction in user search logs was presented by the New York Times reporters who exposed Thelma Arnold as the face behind the randomly assigned "AOL Searcher No. 4417749."[133] Although the reporters were able to glean Ms. Arnold's identity from her search log, they were also led astray by many of her search queries, such as "hand tremors," "nicotine effects on the body," "dry mouth," and even "bipolar," which appear to imply a wide range of ailments (or fear thereof). Ms. Arnold set the record straight by explaining that "she routinely researched medical conditions for her friends to assuage their anxieties. Explaining her queries about nicotine, for example, she said: 'I have a friend who needs to quit smoking and I want to help her do it.'"[134] Ms. Arnold, who is a 62-year-old widow, also searched for the terms "dances by laura," "dances by lori," "single dances" and "single dances in Atlanta." She explained these entries as follows: "A woman was in a [public] bathroom crying. She was going through a divorce. I thought there was a place called 'Dances by Lori' for singles."[135] In user search logs, therefore, what you see is not always what you get.

Solove defines distortion as "the manipulation of the way a person is perceived and judged by others, and involves the victim being inaccurately exposed to the public."[136] Recognizing the potentially harmful effects of inaccurate information, the EU Data Protection Directive provides that personal data must be "accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they

[133] Barbaro & Zeller, *supra* note 35.
[134] *Ibid*.
[135] *Ibid*.
[136] Solove, Taxonomy, *supra* note 13, at p. 550.

were collected or for which they are further processed, are erased or rectified."[137] In addition, European data subjects must be provided with the right to access their personal data without delay and at reasonable cost, as well as the right to rectify, erase or block data that are inaccurate or incomplete.[138] The combination of inaccurate and misleading data, ease of government access, and lack of transparency and accountability to users, makes user search logs highly problematic from a privacy perspective.

i) Exclusion

The prohibition against secret databases is a basic feature of EU data protection law, the learning of decades of totalitarian regimes that used information in secret databases to police and terrorize citizens into conformity and submission.[139] A corollary of the basic prohibition is the right of a European data subject to obtain information about data collected about her, the identity of the entity collecting the data, and the purposes for which they will be used.[140] Data subjects are entitled to access their personal data and, if necessary, correct or amend them.[141] Solove refers to "the failure to provide individuals with notice and input about their records as exclusion."[142] He explains that "exclusion creates a sense of vulnerability and uncertainty in individuals . . . [I]n a world where personal information is increasingly used to make important decisions about our lives, powerlessness in this arena can be significantly troublesome."[143]

Public awareness to the extent of data retention by search engines is minimal. A survey held pursuant to the government's request for Google search records reveals that "89% of respondents believe that their web searches are kept private, and 77%

[137] Article 6(1)(d) of the Data Protection Directive.

[138] Article 12 of the Data Protection Directive.

[139] For a good recent exposé *see* the German film *Das Leben der Anderen* (The Lives of Others) (Bayerischer Rundfunk, Germany 2006) (documenting the activities of the omniscient East German Stasi). MICHEL FOUCAULT, DISCIPLINE AND PUNISH (Vintage Books, 2d ed. 1995); Spiros Simitis, *Reviewing Privacy in an Information Society*, 135 U. PA. L. REV. 707 (1987).

[140] Articles 10-11 of the EU Data Protection Directive.

[141] Article 12 of the EU Data Protection Directive.

[142] Solove, Taxonomy, *supra* note 13, at p. 523.

[143] *Ibid*, at p. 523-24.

believe that Google web searches do not reveal their personal identities."[144] To a great extent, Google's collection of search queries is a secret database.

In its complaint to the FTC concerning the Google-DoubleClick merger, EPIC points out that a user must click on four links from the Google homepage in order to obtain information concerning the company's data collection practices.[145] First, on the Google homepage, a user must click on "About Google." Second, the user must click on "Privacy Policy," which displays the "Google Privacy Policy Highlights" page. Third, the user has to click on the link to Google's full Privacy Policy, which outlines the information Google collects and how it uses it. Included in this list is the term "log information," which is described in text that contains the hyperlinked term "server logs." A fourth click on the term "server logs" leads the user to a FAQ entry for "What are server logs?" It is only there that the user can learn that Google retains her IP address in connection with her search queries.[146]

Google's privacy policy is thus difficult to decipher.[147] And even the full privacy policy fails to explain clearly what Google *does* with the information in search logs. In addition, it is not clear whether and to what extent users have access to their search logs.[148] Such access is now provided to users who subscribe to Google's recently

[144] Linda Rosencrance, *Survey finds solid opposition to release of Google data to feds*, COMPUTERWORLD, Jan. 24, 2006, available at http://www.computerworld.com/securitytopics/security/privacy/story/0,10801,107993,00.html. This article will (hopefully) be read by students and lawyers interested in privacy or cyberlaw; yet I urge you to consider when *you* first became aware of search engines' data aggregation practices. Given that I assume the answer will be "not too long ago" (if that), consider the lack of knowledge by the general public.

[145] EPIC complaint, *supra* note 11, at p. 7.

[146] *Ibid*, at p. 7-8. *See* Google Privacy Policy Highlights, available at http://www.google.com/intl/en/privacy.html; Google's full Privacy Policy, available at http://www.google.com/intl/en/privacypolicy.html; Google's explanation for the term "search logs," Google Privacy FAQ, available at http://www.google.com/intl/en/privacy_faq.html#serverlogs.

[147] Recently asked why the Google homepage does not contain a link to the company's privacy policy, Mr. Fleischer explained: "Google has a very sparse homepage. It's one of the things that we're very proud about. It's kind of clean and Zen-like. Last I counted I think we had something like 35 words on our homepage. On ours with only 35 words, we had to keep it very sparse. Now of course we're a search engine, so anybody who wants to *see* our privacy policy can type Google privacy policy and, trust me, it will come up as result number one. It's not hard to find. We're a search company. We don't believe in pushing things into people's face." Matthew Magee, *Google privacy chief talks*, OUT-LAW RADIO, Jul. 5, 2007, available at http://www.out-law.com/page-8285.

[148] Google's privacy policy states: "When you use Google services, we make good faith efforts to provide you with access to your personal information and either to correct this data if it is inaccurate or to delete such data at your request if it is not otherwise required to be retained by law or for legitimate business purposes. We ask individual users to identify themselves and the information requested to be accessed, corrected or removed before processing such requests, and we may decline to process

launched service, Google Web History.[149] According to Google, the service permits users to view and search across web pages they have visited in the past, including their Google searches; provides trends on their web activity; and helps deliver more personalized search results based on what users searched for and which sites they visited.[150] Users of Google Web History may access their web history and edit or delete items therein. Yet such access comes at a significant privacy cost, because Google stores not only the search queries of Web History users, but also the web pages they have visited. Mr. Fleischer himself admits that "personalized search does raise privacy issues. In order for it to work, search engines must have access to your web search history. And there are some people who may not want to share that information because they believe it is too personal. For them, the improved results that personalized search brings are not matched by the 'cost' of revealing their web history."[151] The question is whether Google users who do not subscribe to Google Web History, ostensibly due to that very "cost," are not already paying a similar price given Google's retention of their search logs and access to DoubleClick's web use profiles. And counter to Web History users, Google search users are not provided with the opportunity to edit or delete items from their search logs (at least not by simple means).

j) Secondary use

One of the fundamental principles of data protection law in OECD and EU instruments[152] is the principle of purpose specification. Under the purpose

requests that are unreasonably repetitive or systematic, require disproportionate technical effort, jeopardize the privacy of others, or would be extremely impractical (for instance, requests concerning information residing on backup tapes), or for which access is not otherwise required. In any case where we provide information access and correction, we perform this service free of charge, except if doing so would require a disproportionate effort." Google's full Privacy Policy, *supra* note 32.

[149] Google Web History, available at www.google.com/psearch; *see* Margaret Kane, *Your Web history, courtesy of Google*, CNET NEWS.COM, Apr. 20, 2007, available at http://news.com.com/8301-10784_3-9710855-7.html; Tom Espiner, *Google launches Web History tool in U.K.*, CNET NEWS.COM, Aug. 3, 2007, available at http://news.com.com/2100-1030_3-6200619.html.

[150] What is Web History?, Google Accounts Help, available at http://www.google.com/support/accounts/bin/answer.py?answer=54068&topic=10472.

[151] Fleischer Letter, *supra* note 65.

[152] Section 9 of the OECD Guidelines; Article 5(b) of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data, Council of Europe Treaties No. 108 (Jan 28, 1981), available at http://conventions.coe.int/Treaty/en/Treaties/html/108.htm; Article 6(1)(B) of the EU Data Protection Directive, which provides that personal data must be "collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes."

specification principle, personal data obtained for one purpose must not be used or made available for another purpose without the data subject's consent. In the EU, the purpose specification principle is based on the underlying belief that personal data "belongs" to the data subject and may be collected, used and transferred (collectively, "processed") by the user of the data (in the EU, "data controller"), strictly for those purposes consented to by the data subject or prescribed by law.

Solove explains that secondary use of personal data "creates a dignitary harm . . . emerging from denying people control over the future use of their data, which can be used in ways that have significant effects on their lives."[153] He points out that "secondary use resembles breach of confidentiality, in that there is a betrayal of the person's expectations when giving out information."[154]

The case of user search logs is instructive. When you enter a search term in Google, you consent to that information being used to respond to your query, and no more. You do not (knowingly, necessarily) agree that Google will aggregate your current query with all of your past searches and mine the data in order to improve its service. Nor do you probably expect Google to make use of this information to target you with effective advertisements or analyze your ad viewing behavior.[155] You most certainly do not expect Google to disburse this information to the government or private parties engaged in litigation against you.

A possible retort is that you do indeed consent, implicitly at least, to all of these uses, which are specified in Google's privacy policy. However, the implicit consent argument is tenuous at best. First, consent is based in this case on a browse-wrap agreement,[156] which is hard to assemble[157] and harder to comprehend. Second,

[153] *Ibid*, at p. 521-22.
[154] *Ibid*, at p. 522.
[155] *See* Grimmelmann, *supra* note 2, at p. 15, noting that "even the display of advertising keyed to a user's query is arguably a purpose other than that intended by the user."
[156] A browse-wrap agreement is one that is typically presented at the bottom of a web site and where acceptance is based on "use" of the site. Hence, there is no affirmative signal of the user's assent to the contract's terms. Browse-wrap agreements are distinguished from, and obviously more problematic than "click-through agreements," which require an offeree to click on an acceptance icon, manifesting assent to be bound. *See* Specht v. Netscape Communications Corp., 306 F.3d 17 (2d Cir. 2002). Also *see* Ian Rambarran & Robert Hunt, *Are Browse-Wrap Agreements All They Are Wrapped Up To Be?*, 9 TUL. J. TECH. & INTELL. PROP. 173 (2007); Terry Ilardi, *Mass Licensing – Part 1: Shrinkwraps,*

Google's privacy policy remains constructively opaque concerning the *primary* use of search logs, rendering secondary use all the more difficult to accept.

Google's use of search data for secondary purposes and the privacy issues it raises expose a broad rift between U.S. and EU data protection. The purpose specification principle, so deeply ingrained in EU law, is not at all evident in the U.S. In the U.S., the assumption traditionally underlying relationships between data subject and controller is that the controller owns the data and may use, reuse or sell it to third parties at will.[158]

k) Breach of confidentiality

Ever since Warren and Brandeis "reinvented" the right of privacy in their seminal article in 1890, privacy has been closely intertwined with the law of confidentiality.[159] English courts to this day hesitate to declare an independent right of privacy, preferring to seek the comfort of traditional breach of confidence law.[160] They do so even at a price of "stretching" the confidentiality doctrine to account for practically nonexistent relations between the parties.[161]

Solove distinguishes breach of confidentiality from the tort of public disclosure of private facts, which has been classified as a privacy cause of action under William Prosser's classic taxonomy.[162] He explains that "[b]oth involve the revelation of secrets about a person, but breaches of confidentiality also violate the trust in a specific relationship. In this way, the tort emerges from the concept of a fiduciary

Clickwraps & Browsewraps, 831 PLI/PAT 251 (2005); Christina Kunz et al., *Browse-Wrap Agreements: Validity of Implied Assent in Electronic Form Agreements*, 59 BUS. LAW. 279 (2003).

[157] *See supra* notes 145-46 and accompanying text.

[158] *See* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373 (2000).

[159] Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193 (1890). Warren and Brandeis cite a line of 19th Century English cases, the most well known of which is Prince Albert v. Strange [1849] 2 De G & Sm 293, 1 Mac & G 25, which imposed liability on disclosures of information protected under an implied contract or a trust or confidence.

[160] *See* Wainwright v. Home Office [2004] 2 A.C. 406.

[161] *See* Campbell v. MGN Ltd [2004] 2 A.C. 457; Douglas v. Hello! Ltd [2001] QB 967.

[162] William L. Prosser, *Privacy*, 48 CAL. L. REV. 383 (1960); also *see* Restatement (Second) of Torts § 652D (1976); WILLIAM L. PROSSER & W. PAGE KEETON, LAW OF TORTS 856-63 (1984 & Supp.1988).

relationship."[163] Hence, "[t]he harm from a breach of confidence . . . is not simply that information has been disclosed, but that the victim has been betrayed."[164] In other words, the fundamental rationale of confidentiality law is not the protection of privacy but, rather, the protection of a relationship of confidence.[165]

Are Google users "betrayed" by the company when it makes secondary use of their personal data or divulges information to third parties? Is Google in a fiduciary relationship with its users? Courts have traditionally applied the confidentiality paradigm to professionals in fiduciary roles,[166] such as lawyers,[167] doctors,[168] therapists[169] and banks.[170] Yet English law has gradually expanded the confidentiality doctrine to protect data subjects against disclosure of personal data by non-fiduciaries, including the press.[171]

As Lord Nicholls observes in the Naomi Campbell case, "[t]his cause of action has now firmly shaken off the limiting constraint of the need for an initial confidential relationship . . . Now the law imposes a 'duty of confidence' whenever a person receives information he knows or ought to know is fairly and reasonably to be regarded as confidential."[172] Ironically, this paradigm shift, influenced by European

[163] Solove, Taxonomy, *supra* note 13, at p. 526-27.
[164] *Ibid*, at p. 527.
[165] *See* John D. McCamus, *Celebrity Newsgathering and Privacy: The Transformation of Breach of Confidence in English Law*, 39 AKRON L. REV. 1191, 1209 (2006). As one court explains: "To promote full disclosure, the medical profession extends the promise of secrecy. The candor which this promise elicits is necessary to the effective pursuit of health; there can be no reticence, no reservations, no reluctance." Hammonds v. Aetna Cas. & Sur. Co., 243 F. Supp. 793, 801 (N.D. Ohio 1965).
[166] *See*, generally, Susan M. Gilles, *Promises Betrayed: Breach of Confidence as a Remedy for Invasions of Privacy*, 43 BUFF. L. REV. 1 (1995); G. Michael Harvey, *Comment, Confidentiality: A Measured Response To the Failure of Privacy*, 140 U. PA. L. REV. 2385 (1992); Alan B. Vickery, *Note, Breach of Confidence: An Emerging Tort*, 82 COLUM. L. REV. 1426 (1982).
[167] *See* Lee A. Pizzimenti, *The Lawyer's Duty to Warn Clients About Limits on Confidentiality*, 39 CATH. U. L. REV. 441, 463-71 (1990).
[168] See, *e.g.*, South Carolina State Board of Medical Examiners v. Hedgepath, 325 S.C. 166, 480 S.E.2d 724 (1997); McCormick v. England, 494 S.E.2d 431, 432 (S.C. Ct. App. 1997); Hammonds v. Aetna Casualty & Sur. Co., 243 F. Supp. 793 (N.D. Ohio 1965); Mull v. String, 448 So. 2d 952 (Ala. 1984); also *see* Joseph White, *Physicians' Liability for Breach of Confidentiality: Beyond the Limitations of the Privacy Tort*, 49 S.C. L. REV. 1271 (1998).
[169] *See*, *e.g.*, Doe v. Roe, 93 Misc.2d 201, 400 N.Y.S.2d 668 (NY Sup. 1977); MacDonald v. Clinger, 446 N.Y.S.2d 801, 805 (App. Div. 1982).
[170] *See*, *e.g.*, Young v. U.S. Dept. of Justice, 882 F.2d 633 (2nd Cir. 1989); Rush v. Maine, 387 A.2d 1127 (Me. 1978); Peterson v. Idaho First Nat'l Bank, 367 P.2d 284, 290 (Idaho 1961). *See* generally Edward L. Raymond, *Annotation, Bank's Liability Under State Law For Disclosing Financial Information Concerning Depositor or Customer*, 81 A.L.R. 4th 377 (1992).
[171] *See* Attorney General v. Guardian Newspapers Ltd (No 2) [1990] 1 AC 109.
[172] Campbell v. MGN, *supra* note 161, at p. 464-65.

legal instruments, has had the effect of bringing the English concept of "confidentiality" closer to the U.S. notion of privacy, captured in Justice Harlan's celebrated "reasonable expectation of privacy" test.[173] As Lord Nicholls holds, "[e]ssentially the touchstone of private life is whether in respect of the disclosed facts the person in question had a reasonable expectation of privacy."[174] Incrementally, the basis for protection of confidential information in the UK is becoming the confidential nature of information itself, rather than a fiduciary relationship between the parties.

Whether based on an implied confidentiality term of contract between Google and its users or on the private nature of the information itself, Google should account to users in case of disclosure of information to third parties. Users who confide in Google their fears and wishes, needs and interests, may, arguably, expect their search logs to be used by the company for improvement of its services or prevention of fraud. But they do not expect their personal data to be transferred to third parties and must be compensated if Google breaches their trust. Below, I further develop the case for use of confidentiality to protect the privacy of search engine users.[175]

l) Additional problems and chilling effect

Aggregation, exclusion, distortion, secondary use and breach of confidentiality are the main privacy problems raised by Google's retention and use of search logs. Additional privacy problems classified by Solove's taxonomy are implicated, yet to a lesser degree. Consider *disclosure*, which occurs when certain true (but embarrassing) information about a person is revealed to the public.[176] Disclosure would take place if Google disclosed potentially embarrassing user search logs to third parties. For example, the disclosure of a husband's search logs in divorce proceedings to prove he had sought adult gay entertainment. Such disclosure may also be classified as a breach of confidentiality.[177] Google's retention and use of search logs raises the problem of *surveillance*. It is analogous to a constant,

[173] Katz v. United States, 389 U.S. 347, 360-61; 88 S. Ct. 507; 19 L. Ed. 2d 576 (1967) (Harlan, J., concurring) [hereinafter Katz].
[174] Campbell v. MGN, *supra* note 161, at p. 466. Baroness Hale too refers to "the 'reasonable expectation of privacy' [as] a threshold test." *Ibid*, at p. 496.
[175] *See* discussion *infra* notes 328-38 and accompanying text.
[176]*See* Restatement (Second) of Torts § 652D (1976).
[177] Solove, Taxonomy, *supra* note 13, at p. 531.

indiscriminate, omnipresent search exposing not only unlawful action but also lawful activity. Clearly, the compilation by the government of such a "Database of Intentions" would ring alarm bells. There is no inherent reason why such surveillance should be treated differently when undertaken by a private company. As Battelle put it, "we need not live in fear of an all-knowing Big Brother. Instead, we should live in fear of any entity that possesses the ability to know whatever it wishes to know, should the need ever arise."[178] A nearly priceless asset with unparalleled scope and depth, Google's database attracts rogue actors seeking to obtain useful information concerning users' activities, interests and commercial needs. This of course raises the problem of "glitches, security lapses, abuses, and illicit uses of personal information," which Solove calls "*insecurity*."[179]

Finally, while not a privacy problem under Solove's taxonomy, Google's data retention and use practices my have a *chilling effect* on online search. I have argued that most users are not aware of Google's privacy practices. Increased public awareness will mean decreased use of search engines, or, at least, self-censored search. Google itself made this point in its response to the government's subpoena of search queries. Google argued that "the production of the requested data will result in a chilling effect on Google's business and user trust."[180] According to Google, "[i]f users believe that the text of their search queries into Google's search engine may become public knowledge, it only logically follows that they will be less likely to use the service . . . this chilling effect on Google's business is potentially severe."[181] Needless to say, search engine users in China and other totalitarian regimes must think hard before looking for information about unpopular opposition groups or historic events.[182] A user entering a search query such as "free Taiwan" in China or "Islamic Jihad" in Egypt may pay a dear price for her curiosity. Yet self censorship will afflict not only societies in which democracy and freedom of speech are scarce,

[178] Battelle, *supra* note 5, at p. 203. Also *see* Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815, 852 (2000), stating "the private sector is as much the potential enemy of privacy as the State and, as a result, we must fear the government's inaction as much as its action."

[179] Solove, Taxonomy, *supra* note 13, at p. 517.

[180] Google's Opposition, *supra* note 63.

[181] *Ibid.*

[182] *See*, *e.g.*, Lester Haines, *Egyptian blogger jailed for four years*, THE REGISTER, Feb. 22, 2007, available at http://www.theregister.co.uk/2007/02/22/egyptian_blogger_jailed. Also *see* Thompson, *supra* note 96.

but Western democracies as well.[183] In order to avoid potential embarrassment and remain above suspicion, users will refrain from intimate or potentially unpopular search queries such as "impotence drugs," "S&M whore," or, for fear of government surveillance, "Bin Laden 9/11." As Julie Cohen thoughtfully observes, "[p]ervasive monitoring of every first move or false start will, at the margin, incline choices toward the bland and the mainstream . . . The condition of no-privacy threatens not only to chill the expression of eccentric individuality, but also, gradually, to dampen the force of our aspirations to it."[184]

V. Privacy solutions

This part outlines the main solutions available to the search logs privacy problem. Technological solutions permit users to mask their identity and browse anonymously, yet are complicated to implement and not entirely foolproof. Privacy policies are drafted by lawyers to protect search engines from liability, not users' privacy, and are based on user consent that is neither informed nor freely given. Constitutional doctrine in the U.S. is flawed insofar as it affords no protection for personal data held by third parties. Statutory provisions are difficult to decipher and provide a surprisingly low level of protection for the contents of communications, as long as such communications are not intercepted in transit. Emerging data retention requirements advanced by national security and law enforcement agencies further restrict user privacy by compelling service providers to maintain traffic data for extended periods of time. A return to the law of confidentiality and evidentiary privileges may reinforce user privacy without eliminating the ability of search engines themselves to make use of the data they collect.

g) Technological solutions

Technological problems often have technological solutions and search privacy is no exception. Privacy invasive technologies are met by an array of privacy enhancing technologies (PETs) that enable users achieve a degree of (though rarely complete)

[183] *See* Susan W. Brenner & Leo L. Clarke, *Fourth Amendment Protection for Shared Privacy Rights in Stored Transactional Data*, 14 J.L. & POL'Y 211, 265 (2006).
[184] Cohen, *supra* note 158, at p. 1426.

online anonymity.[185] The European Commission has recently published a communication to promote the use of PETs to counter online and offline privacy threats.[186] PETs cover a range of different technologies, including encryption tools, cookie management, Internet browser settings, and anonymization schemes.[187] Unfortunately, the vast majority of search users remain oblivious to PETs. In the context of search engines, users may implement various technological measures, ranging from simple steps providing partial protection to more complicated procedures providing greater relief.[188]

To begin with, search users may avoid logging in to their search engine or any related services, or using their ISP's search tool. As long as users manage to separate registration information from search logs, it is difficult to link their identity to their search history. Hence, the label on the "box of personal data,"[189] which constitutes a user's search log, would specify a cookie number or IP address but not an actual name. Once a user registers to services such as Gmail or Google Web History (through the same browser that she uses to conduct search), she compromises her privacy as registration information may be cross-referenced with the apparently anonymous cookie or IP address. And since a user's ISP knows who she is, it will be able to link her identity to any searches conducted on the ISP search facility. Separating registration information from search, however, will not suffice to protect users from retention of search logs based on persistent cookies,[190] which may, at a later point in time, be de-anonymized.

[185]Joel Reidenberg has argued that digital environments can be regulated by their technological capabilities and the design choices made by computer engineers. *See* Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553 (1998).

[186] Press Release, Promoting Data Protection by Privacy Enhancing Technologies (PETs) (May 2, 2007), available at http://tinyurl.com/2jogvz.

[187] The government, in turn, is diligent in devising "responsive" surveillance technologies to counter PETs, such as encryption and anonymization tools, which might be put to use by organized crime or terrorists. *See*, *e.g*, Ric Simmons, *Why 2007 is Not Like 1984: A Broader Perspective on Technology's Effect on Privacy and Fourth Amendment Jurisprudence*, 97 J. CRIM. L. & CRIMINOLOGY 531 (2007); Orin S. Kerr, *The Fourth Amendment in Cyberspace: Can Encryption Create a "Reasonable Expectation of Privacy?"*, 33 CONN. L. REV. 503 (2001).

[188] For a review of options and practical tips, *see* EFF, Six Tips to Protect Your Online Search Privacy, available at http://www.eff.org/Privacy/search/searchtips.php; Ethan Zuckerman, *A Technical Guide to Anonymous Blogging: Security measures for hiding your identity online*, TECHSOUP, Dec. 15, 2006, available at http://www.techsoup.org/learningcenter/internet/page6042.cfm.

[189] *See* discussion *supra* notes 55-60 and accompanying text.

[190] *See* HTTP cookie, *supra* note 53.

To combat this problem, users may set their browsers to block cookies from search engines or allow only session cookies, *i.e.*, cookies erased each time the browser shuts down. More sophisticated users will use anonymous proxy servers and anonymizing software. A proxy server is a buffer between a user's computer and the web.[191] It may be used to accumulate and save files that are requested by a large number of users (caching server),[192] or help in cases where a web site imposes restrictions on users from certain countries or geographical regions (circumventing server). A proxy server that removes identifying information from user requests for the purpose of anonymity is called an anonymizing proxy server or anonymizer. Anonymizers do not provide web sites with the user's IP address and effectively hide from third parties any information about the user and her search and browsing habits.[193] However, the anonymizer itself may collect information concerning the user, and there have been instances of malicious proxy servers recording sensitive personal data, including users' unencrypted logins and passwords.

Another anonymizing option is Tor, also known as the "Onion Router," an Electronic Frontier Foundation (EFF) project, originally sponsored by the US Naval Research Laboratory.[194] Tor is a software product that encrypts users' Internet traffic and then sends it through a series of randomly selected computers, thus obscuring the source and route of the data request. It allows a user to communicate with computers on the Internet without those computers or computers on route knowing where or who the user is. Yet Tor, too, is not foolproof.[195] Researchers have presented techniques allowing analysts with only a partial view of the network to infer which nodes are being used to relay the anonymous streams and therefore greatly reduce the

[191] *See* Proxy server, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Proxy_server.
[192] *See*, *e.g.,* Squid, available at www.squid-cache.org/.
[193] For a variety of anonymous browsing options, *see* Anonymizer web site, available at http://www.anonymizer.com/. Another popular option is Privoxy, which strips out hidden identifying information from Internet traffic, blocks advertisements and can be configured to manage cookies. Privoxy is available at http://www.privoxy.org/.
[194] TOR, available at http://tor.eff.org/. *See* Tor (anonymity network), WIKIPEDIA, available at http://en.wikipedia.org/wiki/Tor_(anonymity_network).
[195] See, e.g., Ryan Naraine, *Hacker builds tracking system to nab Tor pedophiles*, ZDNET, Mar. 6, 2007, available at http://blogs.zdnet.com/security/?p=114.

anonymity provided by Tor.[196] In addition, Tor slows down the browsing experience rendering it far less attractive for users.

While onion routers, anonymizers and cookie management are used to anonymize traffic across a range of Internet activities, TrackMeNot, a lightweight (41K) browser extension invented by NYU law professor Helen Nissenbaum and researcher Daniel C. Howe, addresses search engines specifically.[197] TrackMeNot periodically issues randomized search queries to leading search engines, thereby hiding users' actual search trails in a cloud of "ghost" queries. Hence, if one searches for "herpes treatment" (as in one's troubles), "restaurants 10012" (as in one's zip code), and "John Doe" (as in one's vanity search), TrackMeNot will drown such queries in randomized queries such as "back pain" (as in someone else's troubles), "hospitals 78521" (as in another area's zip code), and "Britney Spears" (as in another name). Nissenbaum and Howe remark that TrackMeNot works "not by means of concealment or encryption (*i.e.*, covering one's tracks), but instead, paradoxically, by the opposite strategy: noise and obfuscation."[198]

h) Privacy policies and the limits of consent

In the absence of federal law governing the collection, retention and use of search logs, it has fallen to search engines to craft their own privacy policies.[199] Google's privacy policy declares that "privacy is important" and promises to protect users' personal data.[200] Privacy policies are incorporated by reference into search engines' terms of use, which are service agreements "agreed" to by users by mere use of the companies' services (*i.e.*, browse-wrap agreements).[201] To see how this is done, consider Google's Terms of Service, which state:

[196] *See* Steven J. Murdoch & George Danezis, *Low-Cost Traffic Analysis of Tor*, paper presented at 2005 IEEE Symposium on Security and Privacy, Oakland CA, May 2005, available at http://www.cl.cam.ac.uk/~sjm217/papers/oakland05torta.pdf.

[197] TrackMeNot, available at http://mrl.nyu.edu/~dhowe/trackmenot/.

[198] *Ibid*.

[199] See Privacy Policies, *supra* note 90.

[200] Google Privacy Policy, *supra* note 32.

[201] *See*, *e.g.*, Google Terms of Service (Apr. 16, 2007), available at http://www.google.com/accounts/TOS; Yahoo! Terms of Service, available at http://info.yahoo.com/legal/us/yahoo/utos/utos-173.html; Microsoft Service Agreement (May 2007), available at http://tou.live.com/en-us/default.aspx.

> "*2.1 In order to use the Services, you must firstly agree to the Terms. You may not use the Services if you do not accept the Terms.*
>
> *2.2 You can accept the Terms (. . .) (B) by actually using the Services. In this case, you understand and agree that Google will treat your use of the Services as acceptance of the Terms from that point onwards."*
>
> *(…)*
>
> *7.2 You agree to the use of your data in accordance with Google's privacy policies."*[202]

Hence, users are held to have read and consented to Google's privacy policy.

Reliance on industry self regulation and user consent is ill advised in the search engine context. EPIC, for example, opposes the architecture of Google's privacy policy, which places information concerning user search logs at a distance of four links from the company's homepage.[203] In addition, certain terms in Google's privacy policy may be interpreted in more than one way. For example, Google's Privacy Policy Highlights state: "We may also share information with third parties in limited circumstances, including when complying with legal process, preventing fraud or imminent harm, and ensuring the security of our network and services."[204] "Limited circumstances" is certainly a broad enough term to encompass a host of data transfers that are detrimental to user privacy. And what does "legal process" mean in this context? In *Gonzales v. Google*, Yahoo, Microsoft and AOL complied with the government's request for URLs and search queries without requiring a search warrant. As discussed below, the standard for government requests for information is not entirely clear. The term "legal process" has vastly different privacy implications depending on whether the standard is "probable cause" (Fourth Amendment standard for search warrants),[205] "specific and articulate facts giving reason to believe" (Stored Communications Act standard for access to certain stored records),[206] or simply

[202] Google Terms of Service, *ibid.*

[203] *See* discussion *supra* notes 145-46 and accompanying text.

[204] *See* Google Privacy Policy Highlights, *supra* note 146.

[205] Wiretaps require a higher standard sometimes referred to as "probable cause plus." *See* Foley, *supra* note 106, at p. 454; James X. Dempsey, *Digital Search & Seizure: Updating Privacy Protections to Keep Pace with Technology*, 865 PLI/PAT 505 (2006).

[206] 18 U.S.C. § 2703(d). Title II of ECPA, the Stored Communications Act (SCA), is codified at 18 U.S.C. §§ 2701-2711.

"relevance" to an investigation (ECPA standard for pen registers).[207] A recent Report on Search Privacy Practices by the Center for Democracy & Technology concludes that "industry self-regulation by itself will never provide strong enough privacy safeguards . . . In particular, whatever information is retained is available to the government under a mere subpoena, issued without a judge's approval. Companies will continue to face the intricacies and loopholes of our nation's patchwork of privacy laws so long as no federal standard exists."[208]

Even if a user finds a satisfactory privacy policy, she should be wary of relying on the company's promise to protect her rights. Search engines typically reserve the right to modify and amend their browse-wrap agreements unilaterally, at any time and without notice. Although Google warrants that it "will not reduce your rights under this [Privacy] Policy without your explicit consent," deciding whether a given policy modification "reduces" a user's right may be controversial. And in any case, Google stands out in this respect among other leading search engines, which do not restrict their right to modify privacy policies.[209] In addition, the ability to modify privacy practices to reduce user rights may be concealed in apparently innocuous language in Google's privacy policy. For example, Google claims it does not correlate users' e-mail and search records.[210] Yet Google's Privacy Policy Highlights provide that "Google collects personal information when you register for a Google service or otherwise voluntarily provide such information. We may combine personal information collected from you with information from other Google services or third parties to provide a better user experience, including customizing contents for you."[211] Thus, Google reserves the right to correlate users' e-mail and search data and it may do so under the current privacy policy without changing its terms to "reduce" users' rights.

[207] *See* federal wiretap law, codified as 18 U.S.C. § § 2510-2522. The federal wiretap act is often referred to as "Title III", since it was initially enacted as Title III of the Omnibus Crime Control & Safe Streets Act of 1968, Pub. L. 90-351, 82 Stat. 212.

[208] Center for Democracy & Technology, Search Privacy Practices: A Work In Progress, CDT Report (August 2007), available at http://www.cdt.org/privacy/20070808searchprivacy.pdf.

[209] Consider the Microsoft Live Search privacy statement, which provides: "We will occasionally update this privacy statement to reflect changes in our services and customer feedback. When we post changes to this Statement, we will revise the 'last updated' date at the top of this statement . . . We encourage you to periodically review this statement to be informed of how Microsoft is protecting your information." Microsoft Online Privacy Statement (Jan. 2006), available at http://privacy.microsoft.com/en-us/fullnotice.aspx.

[210] Goldberg, *supra* note 56, at p. 254.

[211] Google Privacy Policy Highlights, *supra* note 146.

The fleeting nature of privacy protections under self imposed (and generally self serving) privacy policies, as well as companies' retention of the right to unilaterally modify their agreements, raise broader contractual issues related to browse-wrap agreements. As discussed above,[212] a browse-wrap agreement is typically presented at the bottom of the web site and user acceptance is inferred from use of the site.[213] In a growing number of cases, customers have challenged the enforceability of browse-wrap agreements, based on insufficient notice, lack of consent, or unconscionable terms. In *Specht v. Netscape*,[214] the Second Circuit held that "[r]easonably conspicuous notice of the existence of contract terms and unambiguous manifestation of assent to those terms by consumers are essential if electronic bargaining is to have integrity and credibility."[215] The *Specht* court was unwilling to enforce the terms of an agreement which has not been seen by one of the parties. Consequently, the less conspicuous the notice of the existence of the contract, the harder it is to imply user consent. The fact that Google's privacy policy and terms of use do not appear on the search engine's homepage arguably casts a shadow over their enforceability. Search engine privacy policies are also problematic according to *Specht's* additional rationale, the doctrine of unconscionability. Courts usually inquire into the manner in which the parties entered the contract and the quality of consent (procedural unconscionability), as well as into the fairness of the resulting terms (substantive unconscionability). Substantive unconscionability is present where there are manifestly unjust contractual terms, such as terms that are immoral, conflict with public policy, or "bizarre or oppressive."[216] The more substantively oppressive a contract term is the less evidence of procedural unconscionability will be required to

[212] *Supra* note 156 and accompanying text.

[213] See, generally, Rambarran & Hunt, *supra* note 156; Kunz et al., *supra* note 156; Sharon K. Sandeen, *The Sense and Nonsense of Web Site Terms of Use Agreements*, 26 HAMLINE L. REV. 499 (2003); Robert A. Hillman & Jeffery J. Rachlinski, *Standard-Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429 (2002); Comment, *Into Contract's Undiscovered Country: A Defense of Browse-Wrap Licenses*, 39 SAN DIEGO L. REV. 1363 (2002).

[214] Specht v. Netscape Communications, *supra* note 156; also *see* Brazil v. Dell Inc., 2007 WL 2255296 (N.D.Cal. 2007); Ticketmaster Corp. v. Tickets.com, Inc., 2003 WL 21406289 (C.D. Cal., 2003); Mary E. DeFontes, et al. v. Dell Computers Corporation, et al., 2004 R.I. Super, Lexis 32 (Sup. Ct. R.I. 2004).

[215] Specht v. Netscape Communications*, ibid*, at p. 35. *Cf.* Register.com, Inc. v. Verio, Inc., 356 F.3d 393 (2d Cir. 2004).

[216] *See* Restatement (Second) Contracts § 211(3) (1981); Hillman & Rachlinski, *supra* note 213, at p. 457.

conclude that the term is unenforceable, and *vice versa*.[217] A unilateral modification clause, which appears in a browse-wrap agreement that is not prominently posted on a company's homepage, is consequently subject to attack.[218]

Perhaps the greatest shortcoming of privacy policies is their grounding on user consent. After all, if users agree to their search queries being logged, retained, analyzed and possibly disclosed, who is to complain? Yet too much is made of consent in this context. To be meaningful, consent must be informed and freely given. However, most users probably are not aware that their transactions with Google leave a personally identifiable, permanent track record, much less agree to such a result. To the contrary, users operate under a false sense of anonymity and security, neither knowing nor intending their queries to assemble, bit by bit, into a rich personal profile. Thus, user consent is not well informed and nor is it freely given. Freely given consent assumes voluntary choice. However, given that Google and its main competitors implement similar privacy practices,[219] search engine users do not have any real choice. The choice between using search engines under current policies and forgoing use altogether is no choice at all. Not using Google means not participating in today's information society. It is tantamount to never using a telephone, not riding a car, or residing in a secluded cabin in the woods. Google has become ubiquitous, practically a public utility. "Consent" is illusory where it is given (implicitly) by a captive audience to an agreement few if any users have ever read, which includes provisions that the vast majority of users are not aware of, that were unilaterally drafted to serve corporate interests. A privacy protection regime based on such consent provides no privacy protection at all.

i) Constitutional protection – and the lack thereof

The Fourth Amendment provides that "[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall

[217] Brazil v. Dell Inc., *supra* note 214, at *7.

[218] *See* Goldberg, *supra* note 56, at n. 33.

[219] See Declan McCullagh & Elinor Mills, *How search engines rate on privacy*, CNET NEWS.COM, Aug. 13, 2007, available at http://www.news.com/How+search+engines+rate+on+privacy/2100-1029_3-6202068.html.

not be violated, and no warrants shall issue, but upon probable cause . . ."[220] In its foundational 1967 decision in *Katz*, the Supreme Court established a two part test measuring whether a person has a "reasonable expectation of privacy" that is entitled to constitutional protection. In his famous concurrence, Justice Harlan held that the appropriate inquiry is composed of a subjective prong, checking whether "a person [has] exhibited an actual (subjective) expectation of privacy", and an objective prong, verifying whether "the expectation [is] one that society is prepared to recognize as 'reasonable'."[221] The Supreme Court concluded in *Katz*, that the government's act of wiretapping a public telephone booth to listen to *Katz's* conversations violated *Katz's* "reasonable expectation of privacy," and, where performed without an adequately issued search warrant, violated the Fourth Amendment.

The Supreme Court's decision in *Katz* became a fortress for privacy protection over the past four decades. However, two Supreme Court decisions dating from the late 70's destabilized the one of the fortress' foundations, eroding privacy protection where personal data is held by third parties, such as Google.[222] In the first case, *United States v. Miller*,[223] the Supreme Court held in 1976 that bank customers had no "reasonable expectation of privacy" in financial records held by their bank. The Court reasoned that a customer who voluntarily reveals her financial data to a third party (the bank) "assumes the risk" that that third party would pass the information on to the government.[224] The Court reached this conclusion notwithstanding the fact that "the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in the third party will not be betrayed."[225] The Court's rationale follows the proverb attributed to Benjamin Franklin, "three may keep a secret, if two of them are dead."[226] Once the "secret" is out, even if revealed in confidence as part of a banker-customer relationship, the customer can expect no privacy and should not be surprised if the data are passed on to third parties.[227]

[220] U.S. Const. amend. IV.
[221] Katz v. United States, *supra* note 173, at p. 361 (Harlan, J., concurring).
[222] *See* generally Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102 MICH. L. REV. 801 (2004).
[223] United States v. Miller, 425 U.S. 435 (1976) [hereinafter Miller].
[224] Miller, *ibid*, at p. 443.
[225] *Ibid*, id.
[226] Benjamin Franklin, WIKIQUOTE, available at http://en.wikiquote.org/wiki/Benjamin_Franklin.
[227] *See* Patricia L. Bellia, *Surveillance Law Through Cyberlaw's Lens*, 72 GEO. WASH. L. REV. 1375, 1402 (2004).

Miller's assumption of risk analysis was extended in 1979 in *Smith v. Maryland*, which held that telephone users lack a reasonable expectation of privacy in the telephone numbers they dial.[228] Once again, the Court reasoned that users cannot maintain a reasonable expectation of privacy in information conveyed to a third party, the phone company, because the company may use the data for a variety of purposes. The Court held that "a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties."[229] Distinguishing *Katz*, the Court held that the pen registers at issue in *Smith*, which capture numbers dialed, "do not acquire the 'contents' of communications."[230] Hence, Fourth Amendment protection continues to apply insofar as personal data held by a third party include the "contents" of a communication. Constitutional protection is extinguished where no contents are involved.[231]

Courts have extended the *Miller* and *Smith* "assumption of risk" paradigm to a wide variety of circumstances involving the disclosure of personal data to trusted third parties, who then proceed to transfer the data to the government.[232] In a line of cases, courts authorized warrantless government access to ISP's customer records, including names, screen names, addresses, birthdates, and passwords.[233] The Ninth Circuit Court of Appeals recently applied the *Miller* doctrine to a government request for additional ISP subscriber information, including to/from addresses of e-mail messages, IP addresses of websites visited and total amount of data transmitted to or

[228] Smith v. Maryland, 442 U.S. 735, 743-44 (1979) [hereinafter Smith].

[229] *Ibid*, id.

[230] *Ibid*, at p. 747-48.

[231] Justice Stewart, dissenting in Smith, questions the sharp contents-non contents distinction, observing that "[th]e numbers dialed from a private telephone—although certainly more prosaic than the conversation itself—are not without 'contents.' . . . [They] easily could reveal the identities of the persons and the places called, and thus reveal the most intimate details of a person's life." *Ibid*, at p. 748. Analogizing electronic communications to postal mail, Orin Kerr refers to the distinction as one between "contents" (Constitutionally protected) and "envelope" (not Constitutionally protected). Orin S. Kerr, *Internet Surveillance Law After the USA Patriot Act: The Big Brother That Isn't*, 97 NW. U. L. REV. 607, 611-16 (2003) [hereinafter Kerr, Patriot Act]; *cf.* Daniel Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1264, 1288 (2004).

[232] See, e.g., United States v. Jacobsen, 466 U.S. 109 (1984) (a package of drugs sent via Federal Express); SEC v. Jerry T. O'Brien, Inc., 467 U.S. 735 (1984) (financial records held by broker-dealer); California v. Greenwood, 486 U.S. 35 (1988) (garbage bags left at the curb); United States v. Phibbs, 999 F.2d 1053 (6th Cir. 1993) (credit card statements and phone records).

[233] *See* Guest v. Leis, 255 F.3d 325 (6th Cir. 2001); United States v. Kennedy, 81 F. Supp. 2d 1103 (D. Kan. 2000); United States v. Hambrick, 55 F.Supp.2d 504 (4th Cir. 2000); United States v. Cox, 190 F. Supp.2d 330, 332 (N.D.N.Y. 2002); United States v. Bach, 310 F.3d 1063 (8th Cir. 2002).

from an account.[234] The Court concluded that "these surveillance techniques are constitutionally indistinguishable from the use of a pen register that the Court approved in *Smith*."[235] It reasoned that "e-mail and Internet users, like the telephone users in *Smith*, rely on third-party equipment in order to engage in communication" and that "e-mail to/from addresses and IP addresses constitute addressing information and reveal no more about the underlying contents of communication than do phone numbers."[236] The Court did set aside discussion of government access to a list of URLs visited by ISP subscribers, noting that "[s]urveillance techniques that enable the government to determine not only the IP addresses that a person accesses but also the uniform resource locators ('URL') of the pages visited might be more constitutionally problematic. A URL, unlike an IP address, identifies the particular document within a web site that a person views and thus reveals much more information about the person's Internet activity."[237] Hence, *Smith*, with its "assumption of risk" analysis, applies to government access to non-contents information, whereas *Katz* continues to hold for communication contents.

Are user search logs entitled to Fourth Amendment protection?[238] Under the "assumption of risk" doctrine, users may be held to have relinquished any reasonable expectation of privacy in search queries once they are typed into Google. Such users have "voluntarily turned over information to a third party" and are therefore not entitled to Fourth Amendment protection. Alternatively, search queries may be characterized as the contents of a communications, reasserting Fourth Amendment protection under the *Smith* exception. The question of search queries as contents of communications is addressed below.[239] I concentrate here on the shortcomings of the constitutional doctrine.

[234] United States v. Forrester, ___ F.3d ___, 2007 WL 2120271 (9th Cir. 2007). But *see* Steven Warshak v. United States, ___ F. 3d. ___, 2007 WL 1730094 (6th Cir. 2007), holding Fourth Amendment protection does apply to the contents of e-mail stored on an ISP's server. Here too, the critical distinction is between contents and non-contents information, an elusive concept in the context of Internet communications. For prior decisions discussing Fourth Amendment protection for e-mails that have reached their destination, *see* United States v. Charbonneau, 979 F. Supp. 1177, 1184 (S.D. Ohio 1997); Smyth v. Pillsbury Co., 914 F. Supp. 97, 101 (E.D. Pa. 1996).
[235] United States v. Forrester, *ibid*, at *6.
[236] *Ibid*, id.
[237] *Ibid*, id, at n. 6.
[238] *See* Matthew D. Lawless, *The Third Party Doctrine Redux: Internet Search Records and the Case for a "Crazy Quilt" of Fourth Amendment Protection*, 2007 UCLA J. L. & TECH. 1.
[239] *See* discussion *infra* notes 283-87 and accompanying text.

Commentators have often criticized the *Miller* and *Smith* "assumption of risk" doctrine.[240] One basic problem emanates from the *Katz* two-pronged test itself. The *Katz* test is cyclical, because the greater the expectation one has of being subject to surveillance, the less constitutional protection one has. The Court in *Smith* was well aware of this shortcoming, stating that "if the Government were suddenly to announce on nationwide television that all homes henceforth would be subject to warrantless entry, individuals thereafter might not in fact entertain any actual expectation of privacy regarding their homes, papers, and effects. Similarly, if a refugee from a totalitarian country, unaware of this Nation's traditions, erroneously assumed that police were continuously monitoring his telephone conversations, a subjective expectation of privacy regarding the contents of his calls might be lacking as well."[241] Hence, the *Katz* test as applied in *Miller* and *Smith*, becomes a self-fulfilling paranoid prophecy, where one's suspicion of government surveillance strips one of constitutional protection. In other words, what you expect is what you get (and you are probably right to expect the worst).

Commentators note that the *Miller* and *Smith* application of the *Katz* formula is fatally flawed, because it treats the objective prong of the *Katz* test as a positive rather than normative question.[242] Furthermore, in his dissent in *Smith*, Justice Marshall states that it is idle to speak of voluntary "assumption of risk" where, as a practical mater, individuals have no realistic choice. Justice Marshall observes that "[i]mplicit in the concept of assumption of risk is some notion of choice . . . By contrast here, unless a person is prepared to forgo use of what for many has become a personal or professional necessity, he cannot help but accept the risk of surveillance."[243] This observation reverberates in the search engine context. As discussed above, Google users have no plausible alternative to using the leading Internet search engine or one

[240] See, e.g., Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1135 (2002); Bellia, *supra* note 227, at p. 1397-1412; Gavin Skok, *Establishing a Legitimate Expectation of Privacy in Clickstream Data*, 6 MICH. TELECOMM. TECH. L. REV. 61, 78 (2000); Ric Simmons, *From Katz to Kyllo: A Blueprint for Adapting the Fourth Amendment to Twenty-First Century Technologies*, 53 HASTINGS L.J. 1303 (2002); Raymond Shih Ray Ku, *The Founders' Privacy: The Fourth Amendment and the Power of Technological Surveillance*, 86 MINN. L. REV. 1325 (2002).

[241] Smith, *supra* note 228, at p. 740 n. 5.

[242] Brenner & Clarke, *supra* note 183, at p. 247-50; also see Lawless, *supra* note 238 (advocating an "operational realities test").

[243] Smith, *supra* note 228, at p. 749-50.

of its competitors, which apply similar privacy policies. "Assumption of risk" analysis is misleading in this context. Users do not convey personal data to Google because they have chosen to do so after careful deliberation and cost-benefit analysis. They do so because they have to.

An additional problem concerns the scope of constitutional protection. The Fourth Amendment protects individuals from *government* search and seizure. It curtails the investigatory power of government officials. It does not apply to the private sector at all and therefore does not limit Google from collecting, using, retaining or transferring data to third parties.[244] The private sector, so the theory goes, will self regulate to reach an efficient equilibrium based on consumers' privacy preferences and companies' information needs.[245] Yet commentators question both the fairness and efficiency of a market based solution.[246] They point out that privacy invasions typically cause many small, individualized injuries that might be difficult to vindicate through litigation.[247] They argue that consumers in information transactions are hampered by cognitive limitations, which Michael Froomkin dubbed "privacy myopia," causing them to "sell their privacy bit by bit for frequent flyer miles."[248] In addition, even assuming perfect information, customer choice is restricted, given that it is Google that decides which terms to offer in the first place.[249] Thus, it is a take-it-or-leave-it proposition for users, which in the context of search engines, they either "take" or revert to another era.

[244] *See* United States v. Jacobsen, *supra* note 232, at p. 113, holding that "[the Fourth Amendment] is wholly inapplicable to a search or seizure, even an unreasonable one, effected by a private individual (…)"

[245] See, e.g., Peter P. Swire, *Markets, Self-Regulation and Government Enforcement in the Protection of Personal Information*, in U.S. DEP'T OF COMMERCE, PRIVACY AND SELF-REGULATION IN THE INFORMATION AGE (1997), available at http://www.ntia.doc.gov/reports/privacy/selfreg1.htm. The classic law and economics analyses of privacy are Richard A. Posner, *The Right of Privacy*, 12 GA. L. REV. 393 (1978); and George J. Stigler, *An Introduction to Privacy in Economics and Politics*, 9 J. LEGAL STUD. 623 (1980). But *see* Richard S. Murphy, *Property Rights in Personal Information: An Economic Defense of Privacy*, 84 GEO. L.J. 2381 (1996); and Cohen, *supra* note 158.

[246] *See* Paul M. Schwartz, *Property, Privacy, and Personal Data*, 117 HARV. L. REV. 2055, 2076-94 (2004).

[247] Neil M. Richards, *The Information Privacy Law Project*, 94 GEO. L.J. 1087, 1099 (2006).

[248] A. Michael Froomkin, *The Death of Privacy*, 52 STAN. L. REV. 1461, 1502 (2000).

[249] Cohen, *supra* note 158, at p. 1397.

The U.S. constitutional approach to privacy is based on the longstanding American ethos of hostility to the state and big government.[250] Privacy (particularly in one's home)[251] is based on the concept of liberty and individual freedom from government intervention. Nowhere is this more evident than in the uniquely American strand of "decisional privacy" cases,[252] beginning with Justice Douglas' landmark decision in *Griswold v. Connecticut*,[253] continued in the controversial *Roe v. Wade*[254], and recently synthesized in *Lawrence v. Texas*.[255] Writing for the Court in *Lawrence*, Justice Kennedy begins his decision by writing that "*[l]iberty* protects the person from unwarranted government intrusions into a dwelling or other private places. In our tradition the State is not omnipresent in the home. And there are other spheres of our lives and existence, outside the home, where the State should not be a dominant presence."[256]

Contrast the narrow scope of constitutional privacy protection in the U.S. to the situation in Europe, where privacy has been recognized as a fundamental right in constitutional instruments ranging from the 1950 ECHR[257] to the 2004 Treaty Establishing a Constitution for Europe.[258] In Europe, not only privacy but also data protection is a constitutional right;[259] and both rights apply to the private as well as the public sector. Counter to American constitutional privacy, which is based on *liberty*, the European constitutional approach is grounded on the underlying value of

[250] *See* JEFFREY ROSEN, THE UNWANTED GAZE: THE DESTRUCTION OF PRIVACY IN AMERICA 5 (2000); THOMAS PAINE, COMMON SENSE 65 (Penguin Classics, Penguin Books 1986) (1776).

[251] *See* Boyd v. United States, 116 U.S. 616 (1886). Also *see* Note, *The Right to Privacy in Nineteenth Century America*, 94 HARV. L. REV. 1892 (1981).

[252] Gormley, *supra* note 13, at p. 1391-1406.

[253] 381 U.S. 479 (1965).

[254] 410 U.S. 113 (1973).

[255] 539 U.S. 558 (2003).

[256] *Ibid*, at p. 562 (emphasis added). *See* Jacob Strahilevitz, *Consent, Aesthetics, and the Boundaries of Sexual Privacy after Lawrence v. Texas*, 54 DEPAUL L. REV. 671 (2005); Laurence H. Tribe, *Lawrence v. Texas: The "Fundamental Right" that Dare Not Speak Its Name*, 117 HARV. L. REV. 1893 (2004).

[257] Article 8 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (1950), 213 UN Treaty Ser. 221 (1955) ("ECHR") provides: "Everyone has the right to respect for his private and family life, his home and his correspondence."

[258] Article II-67 of the Treaty establishing a Constitution for Europe, Official Journal 2004/C 310/01 (Dec. 16, 2004) (the "Constitutional Treaty"), provides: "Everyone has the right to respect for his or her private and family life, home and communications."

[259] Article II-68(1) of the Constitutional Treaty provides: "Everyone has the right to the protection of personal data concerning him or her."

human dignity.[260] Article 1 paragraph 1 of the German Constitution, for example, declares that "[h]uman dignity shall be inviolable." And Article 2 of the German Constitution establishes the right of "[e]very person . . . to free development of his personality insofar as he does not violate the rights of others." This right of personal autonomy, or "right of personality" (*Persönlichkeitsrecht*), is the basis for the fundamental right of "informational self determination," *i.e.*, one's right to control one's personal data.[261] The fundamental value of human dignity is clearly not restricted to interaction with the government; rather it applies in equal force to the private sector. Dignitary harms, such as unlawful discrimination or invasion of privacy, may be inflicted by individuals and businesses as well as the government.[262]

Nowhere is the difference between the U.S. and European constitutional frameworks more striking than in the context of the *Miller* and *Smith* doctrine. Under *Miller* and *Smith*, where personal information is voluntarily turned over to a third party, constitutional analysis ends. Conversely, in Europe, where personal information is turned over to a third party constitutional analysis just begins. Indeed, the whole thrust of European data protection law, which affords data subjects control over their personal data, pertains to the fair and lawful use of information *by third parties*. Data protection is mandated in the EU by the EU Data Protection Directive, which requires Member States to establish an intricate statutory framework governing all aspects of collection, use and transfer of personal data, and to set up independent regulatory authorities to enforce the law. And the EU Data Protection Directive applies not only to government data controllers but also to private sector entities.

Consequently, while search engines' collection, use and retention of search logs does not raise a constitutional issue in the U.S., at least insofar as the government is not

[260] For a fascinating account *see* James Q. Whitman, *The Two Western Cultures of Privacy: Dignity Versus Liberty*, 113 YALE L.J. 1151 (2004); also *see* Robert C. Post, *Three Concepts of Privacy*, 89 GEO. L.J. 2087 (2001); Matthew W. Finkin, *Menschenbild: The Conception of the Employee as a Person in Western Law*, 23 COMP. LAB. L. & POL'Y REV. 577 (2002); EDWARD J. EBERLE, DIGNITY AND LIBERTY: CONSTITUTIONAL VISIONS IN GERMANY AND THE UNITED STATES (2002).

[261] The landmark decision is the German Constitutional Court's 1983 holding in the famous census case, BVerfGE 65, 1, available at http://www.datenschutz-berlin.de/gesetze/sonstige/volksz.htm (in German).

[262] See, e.g., Paul M. Schwartz, *German and U.S. Telecommunications Privacy Law: Legal Regulation of Domestic Law Enforcement Surveillance*, 54 HASTINGS L.J. 751 (2003).

involved, it falls squarely within the ambit of European constitutional law. In this context too, search engine users' privacy is inadequately protected in the U.S.[263]

j) Statutory protection – a cobweb full of holes

Enacted in 1986 as an amendment to the federal wiretap statute,[264] the ECPA[265] is a highly complex, dense piece of legislation.[266] Originally intended to adapt federal privacy protections to new and emerging technologies, ECPA has become technologically outdated itself, setting legal categories based on technological distinctions that are no longer relevant.[267] ECPA consists of three statutes, the Wiretap Act,[268] the Pen Register Act,[269] and the Stored Communications Act (SCA).[270] The SCA, which applies to communications stored by third parties, is most relevant to search engine users' privacy.[271]

The level of privacy protection set forth by the SCA depends on whether we deal with (a) voluntary or compelled disclosure of information; (b) by an "electronic communication service" or a "remote computing service;" (c) that offers services "to

[263] *See* Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531 (2006), stating "[c]ommon wisdom teaches that the Fourth Amendment exists to protect privacy – and that it does a miserable job of it."

[264] Title III of the Omnibus Crime Control and Safe Streets Act of 1968.

[265] *Supra* note 115.

[266] Orin Kerr promises that "[a]lthough the rules found in § 2702 and § 2703 can seem maddeningly complicated at first, they prove surprisingly straightforward in practice." Orin S. Kerr, *A User's Guide to the Stored Communications Act, and a Legislator's Guide to Amending it*, 72 GEO. WASH. L. REV. 1208, 1222 (2004) [hereinafter, Kerr, SCA]. Surveillance powers under the ECPA were expanded by the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act of 2001, Pub. L. No. 107-56, 115 Stat. 272. *See* generally Marc Rotenberg, *Foreword: Privacy and Secrecy after September 11*, 86 MINN. L. REV. 1115 (2002); Sharon H. Rackow, *Comment, How the USA PATRIOT Act Will Permit Governmental Infringement upon the Privacy of Americans in the Name of "Intelligence" Investigations*, 150 U. PA. L. REV. 1651 (2002) (criticizing the impact of the Patriot Act on civil rights); *cf.* Note, *The Patriot Act's Impact on the Government's Ability to Conduct Electronic Surveillance of Ongoing Domestic Communications,* 52 DUKE L.J. 179 (2002) (arguing that the Patriot Act will have less impact on privacy than feared). For a discussion of state surveillance statutes, *see* Charles H. Kennedy & Peter P. Swire, *State Wiretaps and Electronic Surveillance After September 11*, 54 HASTINGS L.J. 971 (2003).

[267] *See* Dempsey, *supra* note 205, at p. 521; Bellia, *supra* note 227, at p. 1423-24, stating that "under the government's approach, seemingly trivial choices by a subscriber among different technical options a service provider offers have tremendous legal consequences." Kerr, SCA, *supra* note 266, at p. 1216-17.

[268] 18 U.S.C. §§ 2510-2522 (2000 & Supp. II 2002).

[269] 18 U.S.C. §§ 3121-3127 (2000 & Supp. II 2002).

[270] 18 U.S.C. §§ 2701-2711 (2000 & Supp. II 2002).

[271] For a good exposition *see* Kerr, SCA, *supra* note 266; Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1701 (2004).

the public" or not; (d) of the "contents of a communication" or of non-contents; (e) of communications that are in "electronic storage" or in transit.

The SCA applies to two types of communications service providers, providers of "electronic communication service" (ECS) and providers of "remote computing services" (RCS). An ECS means "any service which provides to users thereof the ability to send or receive wire or electronic communications."[272] An RCS means "the provision to the public of computer storage or processing services by means of an electronic communications system."[273] The RCS provisions were originally conceived to cover data processing outsourcing services,[274] yet are currently applicable to the activities of search engines.[275] While Google does not provide "computer storage" services,[276] it does offer "processing services." Much like in traditional data processing, a user transmits data (a search query) to Google via an electronic communication; Google processes the data according to its proprietary algorithm and sends the result (a list of hyperlinks) back to the user. Substantially, Google maintains a log of its communications with users, which is the precisely the aspect of RCS that raised privacy concerns for drafters of the SCA.

A fundamental distinction in the SCA is that between voluntary disclosure, where a service provider chooses to disclose information to the government or a third party,[277] and compelled disclosure, where the government seeks information from a service provider and uses the law to force disclosure.[278] The rules concerning voluntary disclosure turn on the distinction between contents and non-contents information and between government and non-government transferees.[279] Voluntary disclosure of

[272] 18 U.S.C. § 2510(15).

[273] 18 U.S.C. § 2711(2).

[274] The statute's legislative history explains that RCS exist to provide sophisticated and convenient data processing services to subscribers and customers, such as hospitals and banks, from remote facilities. *See* S.Rep. No. 99-541 (1986), reprinted in 1986 U.S.C.C.A.N. 3555, 3564.

[275] As a provider of webmail services, Google seems to fit the definition of ECS. *See* Goldberg, *supra* note 56, at p. 267-69. Orin Kerr observes that "[t]he distinction between providers of ECS and RCS is made somewhat confusing by the fact that most network service providers are multifunctional." Kerr, SCA, *supra* note 266, at p. 1215.

[276] Indeed, one of the privacy *problems* pointed out above is that Google users typically have no access to their search logs. *See* discussion *supra* notes 139-51 and accompanying text.

[277] 18 U.S.C. § 2702.

[278] 18 U.S.C. § 2703.

[279] Another critical distinction in this context is between providers that that offer services to the public and those that do not. The SCA's voluntary disclosure limitations apply strictly to providers that offer services to the public. 18 U.S.C. § 2702. Google clearly belongs to this category.

communication contents is prohibited, whether the information is disclosed to a government or non-government entity, subject to a list of exceptions specified in Section 2702(b).[280] Service providers are free to disclose non-contents information to non-government entities,[281] whereas disclosure to a government entity, even of non-contents, is banned.[282] Determining whether a transferee is a government or non-government entity is straightforward. I therefore turn to the question of whether the data disclosed, in our case consisting of user search queries, constitute contents or non-contents information.

The definition of "contents" applicable throughout the SCA appears in the Wiretap Act.[283] Section 2510(8) provides: "'contents', when used with respect to any wire, oral, or electronic communication, includes any information concerning the substance, purport, or meaning of that communication."[284] Kerr simplifies this rather cryptic definition, explaining that the contents of a communication consist of information that a person wishes to share with or communicate to another person, whereas non-contents (sometimes referred to as "envelope" information) is information about the communication that the network uses to deliver and process the contents information.[285] In other words, contents are what you write in a letter and non-contents are what you write on the envelope. Unfortunately, in the online context, the distinction becomes blurred.[286]

Does a search query constitute "the contents of an electronic communication"? As discussed above, the question is relevant not only for SCA analysis but also for Fourth Amendment purposes. No court has yet addressed the question squarely.[287] On the

[280] 18 U.S.C. § 2702(a).

[281] 18 U.S.C. § 2702(c)(6).

[282] 18 U.S.C. § 2702(a).

[283] 18 U.S.C. § 2711(1), providing that “the terms defined in section 2510 of this title have, respectively, the definitions given such terms in that section."

[284] 18 U.S.C. § 2510(8). Non-contents information is labeled by the SCA "a record or other information pertaining to a subscriber to or customer of such service (not including the contents of communications). 18 U.S.C. § 2703(c)(1).

[285] *See* Kerr, Patriot Act, *supra* note 231, at p. 611-16.

[286] Kerr argues that the conceptual difficulty in distinguishing between contents and non-contents information online "is that the legal categories of 'contents' and 'addressing information' are straightforward in the case of human-to-human communications, but can be quite murky when considering human-to-computer communications." Kerr, Patriot Act, *supra* note 231, at p. 645-46; also *see* Forrester case, *supra* note 234.

[287] But *see* Forrester case, *supra* note 234, at *6 n. 6 (discussing the proper classification of a list of URLs). Also *see* In re United States for an Order Authorizing the Use of a Pen Register & Trap, 396

one hand, search queries do appear to consist of "substance, purport, or meaning," because they convey a person's interests, passions, needs or fears, information that goes well beyond routing or addressing data. On the other hand, a search query may be regarded as a signpost pointing the search engine at the required contents. The contents, the argument would go, are the pages referred to in Google's search result; the search query is merely a tool to get to the contents.

In my opinion, search queries constitute "contents of communications." The information conveyed in search logs is far more revealing than typical "envelope" addressing data, such as telephone numbers or to/from fields of e-mail correspondence. It cuts to the very core of a person's thoughts and feelings, telling much about what she wants to buy or sell; where she plans to go on vacation; what kind of job, husband, music, or shoes she might be interested in; whom she adores and which diseases she abhors; what her political opinions are and which religious faith she subscribes to. Such information, while not the contents of a communication between a user and another person, is most certainly the contents of a communication between a user and the Google server. And if the 1980's featured extension of federal wiretapping laws to electronic communication networks, the natural progression for the new millennium is to extend protection of communication contents to the contents of communications between man and machine.

Assuming that search queries constitute contents of a communication and that Google is an RCS provider, voluntary disclosure by Google of user search queries is prohibited, regardless of whether such disclosure is to a government or non-government entity. Section 2702(b) sets forth seven exceptions to this rule.[288] Most pertinent to the disclosure of user search logs by Google are the exceptions in Sections 2702(b)(2) and 2702(b)(3).[289] Under Section 2702(b)(3), a provider may divulge the contents of a communication to a government or non-government entity

F.Supp.2d 45, 49 (D. Mass. 2005), holding that "there is the issue of search terms. A user may visit the Google site. Presumably the pen register would capture the IP address for that site. However, if the user then enters a search phrase, that search phrase would appear in the URL after the first forward slash. This would reveal contents . . . The 'substance' and 'meaning' of the communication is that the user is conducting a search for information on a particular topic."

[288] Section 2702(c) sets similar exceptions for disclosure of non-contents information, the major difference being that non-contents can be disclosed to non-government entities without restriction. 18 U.S.C. § 2702(c)(6).

[289] 18 U.S.C. §§ 2702(b)(2) –(b)(3).

"with the lawful consent of the . . . subscriber in the case of remote computing service."[290] Google may rely on user consent to its privacy policy to justify voluntary disclosure under Section 2702(b)(3). I argued above that user consent is neither informed nor freely given in this context, and is at best tenuously inferred from use of the Google site.[291] It is therefore an unacceptable basis for disclosure of contents data under the SCA.

Section 2702(b)(2) sanctions disclosure of information to the government "as otherwise authorized in Section 2517, 2511(2)(a), or 2703 of this title."[292] The relevant provision in our case is Section 2703, which governs compelled disclosure of communications data to the government.[293] Thus, the standards for government compelled disclosures become intertwined with those applicable to voluntary disclosure of contents of communications to a government entity. The main distinctions drawn by Section 2703 are between disclosure of contents (Section 2703(a) and (b)) and non-contents (Section 2703(c)); by a provider of ECS (Section 2703(a)) and RCS (Section 2703 (b)). For disclosure of contents, Section 2703 further distinguishes between contents in "electronic storage" for 180 days or less (Section 2703(a)); in "electronic storage" for more than 180 days (Section 2703(b)); or permanently held by an RCS provider (Section 2703(b)). Under Section 2703, a full search warrant is required only to access un-retrieved and unopened e-mail messages and other temporarily stored files held *pending transmission* for 180 days or less.[294]

Section 2703(b) establishes the requirements that the government must meet to compel disclosure of the contents of communications (such as user search logs) held

[290] 18 U.S.C. § 2702(b)(3).

[291] *See* discussion *supra* note 219 and accompanying text.

[292] 18 U.S.C. § 2702(b)(3).

[293] 18 U.S.C. § 2703.

[294] The statute defines "electronic storage" as "any temporary, intermediate storage of a wire or electronic communication incidental to the electronic transmission thereof," and "any storage of such communication by an electronic communication service for purposes of backup protection of such communication." 18 U.S.C. § 2510. The prevailing government approach is that only messages that have not yet been opened or retrieved by a customer are in "electronic storage." Once a message is opened, its storage is no longer "incidental to the electronic transmission thereof." Such a message is therefore "exiled" from the rather strict privacy protections of Section 2703(a) to the lax standards of Section 2703(b). *See* COMPUTER CRIME AND INTELLECTUAL PROPERTY SECTION, U.S. DEP'T OF JUSTICE, MANUAL ON SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS (2001). *Cf.* Theofel v. Farey-Jones, 359 F.3d 1066 (9th Cir. 2004).

by an RCS provider (such as Google). Under Section 2703(b), the government may compel an RCS provider to disclose the contents of a communication using one of five tools: (a) a criminal search warrant; (b) an administrative subpoena; (c) a grand jury subpoena; (d) a trial subpoena; or (e) a court order issued under Section 2703(d). A court order issued under Section 2703(d) is not equivalent to a search warrant, which requires a showing of "probable cause." Instead, a court may issue a Section 2703(d) order if the government offers "specific and articulable facts showing reasonable grounds to believe" that the communications sought are "relevant and material" to an ongoing criminal investigation.[295]

The SCA's authorization of a subpoena or Section 2703(d) order rather than a full search warrant reflects the premise that a user retains no "reasonable expectation of privacy" in the contents of communications stored by an RCS provider.[296] Consequently, Kerr notes that "[t]he most obvious problem with the current version of the SCA is the surprisingly weak protection the statute affords to compelled contents of communications under the traditional understanding of ECS and RCS."[297] This is evident particularly in the case of subpoenas, which are issued with no prior judicial approval and are enforced on a mere showing of relevance. Worse yet, when a subpoena is served on the data subject herself, she at least has notice and an opportunity to file a motion to quash or modify.[298] But where a subpoena is served on a third party, such as Google, that third party typically has little or no reason to object, and notice to the person whose privacy is being compromised may be deferred for long periods of time under Section 2705.[299] Consequently, statutory protection under ECPA follows the weak constitutional doctrine and perpetuates the vulnerability of search engine users' privacy rights.[300]

[295] 18 U.S.C. § 2703(d).

[296] Bellia, *supra* note 227, at p. 1422.

[297] Kerr, SCA, *supra* note 266, at p. 1233; also *see* Note, *Email Privacy after United States v. Councilman: Legislative Options for Amending ECPA*, 21 BERKELEY TECH. L.J. 499 (2006).

[298] *See* Christopher Slobogin, *Subpoenas and Privacy*, 54 DEPAUL L. REV. 805 (2005).

[299] 18 U.S.C. § 2705. Section 2705 states that a "supervisory official" (defined in Section 2705(a)(6)) can order notice to be delayed by up to 90 days if there is "reason to believe that notification of the existence of the subpoena may have an adverse result," (18 U.S.C. § 2705(a)(1)(B)) such as anything that "seriously jeopardizing an investigation or unduly delaying a trial." (18 U.S.C. § 2705(a)(2)(E)).

[300] For the inadequacy of statutory protection, also *see* Peter Swire, *Katz Is Dead. Long Live Katz*, 102 MICH. L. REV. 904 (2004); Deirdre K. Mulligan, *Reasonable Expectations in Electronic Communications: A Critical Perspective on the Electronic Communications Privacy Act*, 72 GEO. WASH. L. REV. 1557 (2004).

k) Data retention v. data protection

As if constitutional and statutory impediments are not enough, online privacy is increasingly meeting a formidable enemy in the shape of data retention requirements advocated by national security and law enforcement agencies worldwide. Ostensibly, the best privacy solution for user search logs would be their immediate deletion. The mere existence of the so called Database of Intentions constitutes a magnet for government investigators, private litigants, data thieves and commercial interests. In its November 2006 London Resolution, the Article 29 Working Party required that "[a]fter the end of a search session, no data that can be linked to an individual user should be kept stored unless the user has given his explicit, informed consent to have data necessary to provide a service stored (*e.g.* for use in future searches)."[301] But are search engines even *allowed* to delete users' search logs?

Governments increasingly impose data retention requirements to make online and telecom activity traceable by law enforcement agencies.[302] Data retention laws compel telecommunications companies and ISPs to collect and store customers' data. Typically, retention is restricted to non-contents data, such as subscriber information and traffic and location data. The legal and technical differences between data retention standards across EU Member States posed difficult dilemmas for service providers with European-wide operations. In addition, data retention legislation created conflicts with data protection laws, since one of the core principles of the EU Data Protection Directive is that data must be "kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed."[303] Similarly,

[301] 28th International Data Protection and Privacy Commissioners' Conference, London, UK, Resolution on Privacy Protection and Search Engines (Nov. 2-3, 2006).

[302] See, e.g., Regulation of Investigatory Powers Act 2000, 2000 Chap. 23, Part I, Chapter II; Anti-terrorism, Crime and Security Act 2001, 2001 Chap. 24, Part 11 (UK); also *see* Home Office, Retention of Communications Data Under Part 11: Anti-Terrorism, Crime & Security Act 2001, Voluntary Code of Practice, available at http://www.opsi.gov.uk/si/si2003/draft/5b.pdf. *Cf.* EDRI-gram, Italy decrees data retention until 31 December 2007 (Aug. 10, 2005), available at http://www.edri.org/edrigram/number3.16/Italy; EDRI-gram, Telecom data to be retained for one year in France, available at http://www.edri.org/edrigram/number4.6/franceretantion.

[303] Article 6(e) of the EU Data Protection Directive.

Article 6 of the Communications Privacy Directive[304] prohibits the storage of traffic data without user consent once the data are no longer required for the actual transmission of a communication or for billing purposes.[305]

The need for European-wide harmonization and clarification of the interplay between privacy and data retention legislation has led the EU to adopt a Data Retention Directive in March 2006.[306] Under the Data Retention Directive, providers of "electronic communications services" are required to store traffic data related to telephone calls, e-mails and online activity for a period of six months to two years, depending on the law in each Member State.[307] Traffic data include the identities of a customer's correspondents; the date, time, and duration of phone calls, VoIP calls,[308] or e-mail messages; and the location of the device used for a communication; but *not the contents* of a communication.[309] EU Member States have until September 2007 to implement the Data Retention Directive through national legislation to fixed line and mobile telephone providers, and until March 2009 to implement its application to e-mail, Internet telephony and Internet access.[310]

[304] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector, Official Journal L 201 (Jul. 31, 2002) [hereinafter Communications Privacy Directive].

[305] The prohibition is qualified by Article 15 of the Communications Privacy Directive, which permits the adoption of "legislative measures providing for the retention of data for a limited period justified (…) to safeguard national security (*i.e.* State security), defense, public security, and the prevention, investigation, detection and prosecution of criminal offences or of unauthorized use of the electronic communication system."

[306] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, L 105/54, Official Journal (April 13, 2006) [hereinafter Data Retention Directive]. *See* generally Francesca Bignami, *Privacy and Law Enforcement in the European Union: The Data Retention Directive*, 8 CHI. J. INT'L L. 233 (2007).

[307] Articles 3, 6 of the Data Retention Directive. The 6 to 24 month range is problematic, given that the Data Retention Directive is a harmonization device. As Google recently noted in its response to the Article 29 Working Party inquiry, even if only one EU jurisdiction implemented a measure based on the top of the range, the company would be forced to implement that measure across the board. Hence, the 6 to 24 month range sets off a "race to the bottom" (from a privacy perspective). Fleischer Letter, *supra* note 65, at p. 3.

[308] Voice over IP, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Voip.

[309] Articles 2, 5 of the Data Retention Directive.

[310] Article 15 of the Data Retention Directive. In March 2007, the UK issued draft regulations, The Data Retention (EC Directive) Regulations 2007, to become effective Oct.1, 2007, available at http://www.opsi.gov.uk/SI/si2007/draft/20077449.htm. Together with a number of other EU Member States, the UK will be delaying application of the Regulations to the Internet until March 2009.

Although the U.S. has not yet followed the European lead in data retention, adoption of such legislation has been advocated by politicians, including former Attorney General Alberto Gonzales.[311] Data retention legislation is arguably not as essential in the U.S. as it is in the EU, since unlike European companies that are hemmed-in by data protection laws, American services providers usually retain users' traffic data for commercial reasons even without being required to do so. Moreover, the Sarbanes-Oxley Act,[312] tax laws and accounting regulations reflect mounting data retention requirements applicable to U.S. companies even without a dedicated data retention statute. Finally, U.S. authorities benefit from a related, if less sweeping law enforcement tool, known as "data preservation."[313] Data preservation is set forth in the Electronic Communication Transactional Records Act of 1996,[314] which requires ISPs to retain any "record" in their possession for 90 days "upon the request of a governmental entity." Counter to European data retention, which applies across the board, American data preservation is targeted at the traffic data of a specific individual already under investigation.[315]

A U.S. District Court in California recently implemented an expansive approach to data preservation in a case involving TorrentSpy,[316] a BitTorrent[317] indexing web site. The Motion Picture Association of America (MPAA) sued TorrentSpy in February 2006, accusing the web site of facilitating illegal downloads of copyrighted

[311] *See* Declan McCullagh, *GOP revives ISP-tracking legislation*, CNET NEWS.COM, May 30, 2006, available at http://news.com.com/GOP+revives+ISP-tracking+legislation/2100-1028_3-6156948.html; Declan McCullagh, *Terrorism invoked in ISP snooping proposal*, CNET NEWS.COM, May 30, 2006, available at http://news.com.com/2100-1028_3-6078229.html.

[312] Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (codified in scattered sections of 11, 15, 18, 28, and 29 U.S.C.).

[313] The concept of data preservation exists in Europe. *See* Articles 16-17 of Council of Europe Convention on Cybercrime, Europ. T.S. No. 185 (Nov. 23, 2001), available at http://conventions.coe.int/Treaty/EN/Treaties/Html/185.htm.

[314] 18 U.S.C. §§ 2701-2712.

[315] Catherine Crump notes that "[t]he purpose of data retention is much broader than that of data preservation. Data retention aims to change the context of Internet activity . . . Data retention 'rearchitects' the Internet from a context of relative obscurity to one of greater transparency." Catherine Crump, *Data Retention: Privacy, Anonymity, and Accountability Online*, 56 STAN. L. REV. 191, 194 (2003).

[316] *See* TorrentSpy, available at http://www.torrentspy.com/.

[317] BitTorrent is a peer-to-peer (P2P) file sharing communications protocol, allowing for broad distribution of large amounts of data without the distributor having to incur the costs of hardware, hosting and bandwidth. *See* BitTorrent, WIKIPEDIA, available at http://en.wikipedia.org/wiki/BitTorrent.

materials.[318] As part of the discovery process, the MPAA filed a motion to compel TorrentSpy to preserve and produce server log data, including IP addresses of users seeking "dot-torrent" files. TorrentSpy, which operated its web site from servers based in the Netherlands, pointed out that it had never retained server logs because the information was not necessary for its business, and since data retention was restricted by Dutch privacy law. TorrentSpy claimed that requiring it to log user data would force it to act in a manner contrary to its privacy policy, which states that the site does not collect any personal information about its users. The court granted the MPAA's motion, holding that since the data sought by the MPAA were at least temporarily available in RAM,[319] they were covered by the rules of evidence and must therefore be logged and turned over to the plaintiff.[320] The Court's ruling is much broader than a preservation order, because it is not targeted at a specific suspect and goes so far as to require the web site to store data not ordinarily kept on its servers.[321]

The tension between data protection and data retention requirements is manifest in Google's latest privacy investigation.[322] The Article 29 Working Party claimed that Google's storage period of 18 to 24 months is excessive.[323] Google responded by shortening its retention policy to a period of 18 months, but pointed out that if certain EU Member States implement the Data Retention Directive by mandating a 24 month retention period, it would have to once again adjust its policy to comply. Thus, on the one hand, Google is slapped on the wrist by data protection regulators for its retention policies, pressured to delete search logs and anonymize retained data. On the other hand, Google is mandated by data retention requirements to store the data for lengthy periods. In his response to the Article 29 Working Party, Google's Fleischer suggests

[318] *See* John Borland, *MPAA sues newsgroup, P2P search sites*, CNET NEWS.COM, Feb. 23, 2006, available at http://tinyurl.com/36zj9n.

[319] *See* Random access memory, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Random_access_memory.

[320] Columbia Pictures Inds. V. Bunneli, ___ F.Supp.3d ___ No. CV 06-1093 FMC(JCx) (C.D. Cal. May 29, 2007).

[321] The decision has been stayed pending appeal. The Electronic Frontier Foundation and the Center for Democracy and Technology filed brief of amici curiae in support of TorrentSpy's position, available at http://www.eff.org/legal/cases/torrentspy/EFF_CDT_amicus.pdf. TorrentSpy announced it would block all search queries from U.S. users rather than logging user queries in contravention of its privacy policy. *See* Jacqui Cheng, *TorrentSpy to MPAA: Log this! Site blocks US searches*, ARS TECHNICA, Aug 27, 2007, available at http://arstechnica.com/news.ars/post/20070827-torrentspy-to-mpaa-log-this-site-blocks-us-searches.html.

[322] Article 29 Working Party Letter, *supra* note 9.

[323] *Ibid*. Google has consequently shortened its data retention period to 18 months and reset its cookie to expire two years after a user's last search query. *See* Fleischer Letter, *supra* note 65.

"[a] public discussion is needed between officials working in data protection and law enforcement."[324]

The solution to Google's quandary requires finding the golden path between data protection and data retention requirements. To be sure, massive data retention is privacy intrusive and risks turning communications service providers into data warehouses for government investigations.[325] It sets the stage for pervasive surveillance of ordinary citizens whose personal data will be mined and analyzed in huge "fishing expeditions" by security and law enforcement agencies. Nevertheless, lack of any data retention constitutes a boon for terrorists, pedophiles, crime lords and hackers, and puts law enforcement agencies at a disadvantage against an increasingly sophisticated opponent.

The Article 29 Working Party suggested criteria for making data retention requirements more amenable to privacy rights. It advised that data retention be limited to narrowly tailored purposes, such as fighting terrorism and organized crime. It suggested that there must be no further processing of retained data by law enforcement authorities for related proceedings, and no access to the data by additional government or non-government entities. It requested that prevention of terrorism not include large-scale data mining schemes; that access to data be duly authorized on a case by case basis by a judicial authority; and that systems for storage of data for law enforcement purposes be separated from systems used for business purposes.[326]

Finally, it is important to note that the Data Retention Directive might not apply to search engines and user search logs. First, it is not clear whether Google is a provider of "electronic communications services," which is required to maintain traffic data

[324] Fleischer Letter, *ibid.*

[325] See, e.g., EUROISPA and US ISPA Position on the Impact of Data Retention Laws on the Fight Against Cybercrime (Sep. 30, 2002), available at http://www.euroispa.org/docs/020930eurousispa_dretent.pdf; Home Office Voluntary Code of Practice, *supra* note 302, at s. 23-24.

[326] See Article 29 Working Party Opinion 4/2005, 21 Oct., 2005, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2005/wp113_en.pdf, at p. 8-10; also *see* Opinion 3/2006, Mar. 25, 2006, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2006/wp119_en.pdf; Opinion 9/2004, Nov. 9, 2004, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2004/wp99_en.pdf.

under Article 3 of the Data Retention Directive.[327] Even if services such as Gmail or Google Talk fall within the ambit of the new legislation, search might not. Moreover, a search query may, and in fact should be regarded as contents of a communication, which remain outside the scope of the Data Retention Directive.

l) The law of confidentiality and evidentiary privileges

Disclosure of user search logs by Google could be curtailed on the basis of the law of confidentiality and evidentiary privileges.[328] One argument not raised by Google in its tussle with the U.S. government is that the subpoena requesting user search logs would compel it to disclose privileged information. Under the Federal Rules of Civil Procedure, "the court by which a subpoena was issued shall quash or modify the subpoena if it (. . .) requires disclosure of privileged or other protected matter and no exception or waiver applies."[329] Could Google have relied on the privilege exception as a basis to quash the government's subpoena? I argue that it could and indeed should have.[330]

Viewing the disclosure of information by users to search engines as a disclosure to the public, trumping such users' reasonable expectation of privacy, may fit Fourth Amendment doctrine but is otherwise out of sync with users' beliefs and expectations. Information disclosures of this type are more akin to communications protected by evidentiary privileges than they are to a broadcast to the public.[331] As Solove puts it, "[w]hen people establish a relationship with banks, Internet service providers, phone companies, and other businesses, they are not disclosing their information to the world. They are giving it to a party with implicit (and often explicit) promises that the information will not be disseminated."[332] When you enter a search query in Google you simply do not expect this information to haunt you in criminal or civil proceedings; nor do you expect it to be transferred to third party businesses for

[327] The term "electronic communications services" is derived from the Communications Privacy Directive. *See* Fleischer Letter, *supra* note 65, at p. 3.

[328] *See* generally Vickery, *supra* note 166.

[329] Fed. R. Civ. Proc. 45(c)(3)(A)(iii).

[330] For prior arguments in this guise *see* Jessica Litman, *Information Privacy/Information Property*, 52 STAN. L. REV. 1283, 1304-13 (2000); Brenner & Clarke, *supra* note 183, at p. 266-79; Gilles, *supra* note 166.

[331] Brenner & Clarke, *supra* note 183, at p. 259.

[332] Solove, Taxonomy, *supra* note 13, at p. 529.

marketing or data mining purposes. Information revealed to search engines may be highly sensitive and similar to that covered by already existing evidentiary privileges, such as physicians' (consider the query "hypertension impotence treatment"), psychotherapists' ("zyprexa side effects"), lawyers' ("income tax misrepresentation"), or priests' ("jesus savior baptizing"). Indeed, users refer questions to search engines that they would hesitate to address to any of the traditional professionals.

.

Picture Google as a human figure, say Mr. Anish Singh, who works at the search engine's data processing center and is the employee assigned to your search. You approach Mr. Singh with search queries and expect him to reply efficiently and accurately. To achieve this, you would not want Google to reassign a new employee to you each time you open a search session. To the contrary, you enjoy working with Mr. Singh, who learns to anticipate your needs and understand the idiosyncrasies of your requests. Ideally, you would want Mr. Singh to memorize all of your prior interactions, so when you search for "apple" he does not refer you to the tech-giant's web site but rather to tips about growing your favorite fruit. All this seems to imply that retention of search logs is a good thing. However, you would be appalled to learn that Mr. Singh shares your private communications with his friends or business partners, not to mention the police or your spouse. If that were the case, and if you were aware of the unbearable ease of access by third parties to Mr. Singh's work product, I seriously doubt you would share any information with Mr. Singh, regardless of how good a professional he was.[333] To say that you do not have a reasonable expectation if privacy in information shared with Mr. Singh, since you cannot control his actions is a *non sequitur*. You cannot control the actions of your doctor but still confide in her, because you know she will not compromise your privacy.

The law of confidentiality and evidentiary privileges solves the trust problem between patient and physician, customer and banker, and additional fiduciary relationships. It can also be applied to the interaction between search engines and users. A physician divulging personal information to a third party breaches her fiduciary duty of

[333] Empirical studies confirm that people have a dramatically lower likelihood to communicate if they are informed that a conversation is not privileged. *See* Daniel W. Shuman & Myrion S. Weiner, The Privilege Study: An Empirical Examination of the Psychotherapist-Patient Privilege, 60 N.C. L. Rev. 893 (1982).

confidentiality, commits a tort, and violates professional codes of ethics. The same should apply when a search engine discloses user queries to a government or non-government entity.

The fact that a communication is made in confidence does not automatically entitle it to an evidentiary privilege, unless the parties bear some relation to each other that the law seeks to protect. Rule 501 of the Federal Rules of Evidence authorizes federal courts to define new privileges by interpreting "common law principles ... in the light of reason and experience."[334] The Senate Report accompanying the adoption of the Rule 501 indicates that the Rule "should be understood as reflecting the view that the recognition of a privilege based on a confidential relationship . . . should be determined on a case-by-case basis."[335] Dean Wigmore favored creating a privilege where the benefit gained from furthering a special relationship exceeded the harm to the judicial truth-seeking process.[336] The mere possibility of disclosure of user queries could have a chilling effect on the development of the search economy specifically and online use generally.[337] Conversely, as the Supreme Court notes in *Jaffee v. Redmond*, in the context of the psychotherapist's privilege, "the likely evidentiary benefit that would result from the denial of the privilege is modest. If the privilege were rejected, confidential [communications] would surely be chilled . . ."[338]

An evidentiary privilege for search engines need not be absolute. Search engine users' privacy rights should be balanced against the need of law enforcement agencies to apprehend criminals. So, for example, a search query that constitutes – in and of itself – a criminal offense, such as "child porn," may not be protected. Determining the optimal scope of the search engine privilege would require careful balancing between users' privacy interests and the needs of the judicial process. However, this

[334] Fed. R. Evid. 501. *See* Raymond F. Miller, *Creating Evidentiary Privileges: An Argument for the Judicial Approach*, 31 CONN. L. REV. 771 (1999); Introduction, *The Development of Evidentiary Privileges in American Law*, 98 HARV. L. REV. 1454 (1985).

[335] S.Rep. No. 93-1277, p. 13 (1974) U.S.Code Cong. & Admin. News 1974, pp. 7051, 7059.

[336] *See* Miller, *supra* note 334, at p. 783, citing 8 JOHN HENRY WIGMORE, EVIDENCE § 2285 (John T. McNaughton ed., rev. ed. 1961). The test is whether a new privilege "promotes sufficiently important interests to outweigh the need for probative evidence." Trammel v. United States, 445 U.S. 40, 51 (1980).

[337] *See Who's Afraid*, *supra* note 12.

[338] Jaffee v. Redmond, 518 U.S. 1, 11-12 (1996).

legal tool, which has not yet been utilized in the context of search engines, may yield better results than some of the currently existing measures.

VI. Conclusion

The *Gonzales v. Google* case and the AOL privacy debacle were not isolated or exceptional occurrences. They are but the tip of the iceberg of an emerging privacy problem on a grand scale, featuring Internet search engines as informational gatekeepers harboring previously unimaginable riches of personal data. Billions of search queries stream across Google's servers each month, "the aggregate thoughtstream of humankind, online."[339] Google compiles individual search logs, containing users' fears and expectations, interests and passions, and ripe with information that is financial, medical, sexual, political, in short – personal in nature. Google puts these data to secondary uses, such as improving its search service, ensuring network security and targeting ads. Users may stomach such use of their personal data as part of their transaction with a company that offers an amazing service for free. Yet they are less inclined to appreciate the sharing of their data with third parties, be they commercial, government or, of course, criminal in nature.

The collection, retention and use of personal data by search engines raise a host of privacy problems, including *aggregation*, *distortion*, *exclusion*, *secondary use*, *breach of confidentiality*, *disclosure*, *surveillance*, and *insecurity*. These problems and the public realization that they exist may have a *chilling effect* on search and online activity. Search engine users who become aware that the government may be privy to their communications – or more accurately in the context of search, to their thought process – may be cowed into submission to mainstream views and opinions.

Users may counter privacy invasive technologies with PETs in a perpetual game of "hide and seek." Yet users are often unwilling to expend the time and effort, or simply not technology-savvy enough, to fight for what many believe is a lost cause. Privacy policies, one-sided browse-wrap agreements typically not read by anyone save the lawyers who draft them, cannot be relied upon to protect users' privacy. To

[339] Battelle, *supra* note 5, at p. 6.

the contrary, privacy policies are packed with broad exemptions intended to protect the interests of search engines and shield them from potential liability. As contractual documents, they are based on user consent, which is inferred from mere use of a web site, uninformed, and not truly voluntary. Having exhausted technological and contractual privacy protections, users' fall back is the constitutional and statutory scheme provided by the state. Users are bound to be disappointed, as current doctrine is ill-suited to protect their interests.

In a world of pervasive surveillance and rapidly evolving data collection, retention and processing technologies, the American doctrine granting individuals control over their personal data only insofar as the information has not been revealed to third parties is obsolete. In this day and age, third parties, such as financial institutions, insurance companies, online service providers and government agencies, maintain databases with massive amounts of personal data, including in certain cases information not known to the data subjects themselves. The line dividing protected and unprotected personal data must be drawn elsewhere, since under current doctrine individuals have no rights whatsoever in these vast data pools. The EU data protection framework, with its set of fair informational practices and regulatory data protection authorities, provides protection for personal data held by third parties. Restricting the scope of legitimate activities by such "data controllers," the EU Data Protection Directive protects individuals even after they have parted with their personal data, obtaining a sounder balance between government and business interests and individuals' fundamental rights.

Statutory protection for search engine histories is also fundamentally flawed. Privacy in electronic communications is protected by a Byzantine statutory framework dating from the 1980's, when the Internet was in its infancy and search but a distant dream.[340] It is not clear whether search queries constitute contents of communications entitled to protection under the statutory scheme. Even if they do, protection under the SCA chapter of the ECPA is surprisingly weak, permitting access to the contents of communications with a mere administrative subpoena. In updating the ECPA for

[340] Battelle deems Archie, created in 1990 by a group of McGill University students, the first Internet search engine. *See* Battelle, *supra* note 5, at p. 39-40; also *see* Archie search engine, WIKIPEDIA, available at http://en.wikipedia.org/wiki/Archie_search_engine.

the new millennium, lawmakers should clarify the classification of search queries as contents and fortify the legal protection afforded to them by requiring a full search warrant for access thereto.

As if the current dearth of protection is not enough, information privacy is about to receive a severe blow with the advent of data retention legislation. Such laws, not only permitting service providers to retain personal data but actually compelling them to do so, are advanced by national security and law enforcement agencies with far greater political clout than privacy advocates. In the public debate about combating terrorism and serious crime, the voice of privacy advocates is often muted, their quest vilified, as if they are accomplices to the commission of a crime. A reasonable balance must be struck between the needs of law enforcement and the democratic imperative of not casting a light of suspicion on all law abiding citizens.

Search engines owe a duty of confidentiality to users, whether by contract or due to the inherently private nature of search data. Customers reveal sensitive personal data to professionals such as physicians, psychotherapists, lawyers and bankers, based on trust that confidential information will not be disclosed. Evidentiary privileges protect such professionals from having to betray customer trust when summoned to testify in court. Search engines too should benefit from an evidentiary privilege, albeit qualified, based on the same rationales underlying exiting privileges. Such a privilege would correctly balance the benefit gained from furthering the special relationship between search engines and users with the harm to the judicial process, while at the same time protecting users' privacy.