

THE PANOPTIC GAZE OF WEB 2.0:
HOW WEB 2.0 PLATFORMS ACT AS INFRASTRUCTURES OF DATAVEILLANCE

Presented at:
Social Software and Web 2.0: Critical Perspectives and Challenges for Research and Business

Aalborg University, Aalborg, Denmark
October 6, 2007

Michael Zimmer
Department of Culture & Communication
New York University

Preface

I would like to first thank Anders Albrechtslund, Thomas Ryberg, the Doctoral School of Human Centered Informatics, the e-Learning Lab, and everyone else Aalborg University for planning this excellent and important seminar, and for inviting me attend so I can share my thoughts, as well as learn from everyone else.

Title

The title of my talk today is “The Panoptic Gaze of Web 2.0: How Web 2.0 platforms act as Infrastructures of Dataveillance”

Rhetoric of Web 2.0

The rhetoric surrounding Web 2.0 technologies presents certain claims about media, identity, and technology. It suggests that everyone can and should use new information

technology to organize and share information, to interact within communities, and to express oneself. It promises to empower creativity, to democratize media production, and to celebrate the individual while also relishing the power of collaboration and social networks.

Externalities 2.0

But Web 2.0 also embodies a set of unintended consequences – externalities – including the increased flow of personal information across networks, the rise in data mining to aggregate data across the network, the drive for intelligent agents that predict your needs, and the underlying philosophy of placing these powerful tools in hands of all users.

Externalities 2.0

I argue that these externalities of Web 2.0 systems represent a new and powerful “infrastructure of dataveillance,” which brings about a new kind of panoptic gaze of both users’ online *and even their offline* activities.

This panoptic gaze of Web 2.0 manifests itself in many ways. Today I will focus on two particular infrastructures of dataveillance: the drive towards “Search 2.0” – or the “perfect search” – and the rise of “Amateur data mining & P2P surveillance” across Web 2.0 properties.

The Panoptic Gaze of Dataveillance

But first, what do we mean by the “panoptic gaze of dataveillance”? Let’s start with the more familiar part, the “panoptic gaze.”

Panopticon

Conceived in 1791, Jeremy Bentham's Panopticon prison was designed to maintain (by allusion, if not by fact) *perpetual surveillance* of its inhabitants: by placing prison guards in central tower with a one-way observation system surrounded by rooms for those to be watched, the subjects were unable to determine when, or if, they were being watched. Through this unique architectural design, Bentham believed that the constant threat that one *could* be surveilled at any time would force the subjects to internalize the effects of surveillance:

The more constantly the persons to be inspected are under the eyes of the persons who should inspect them, the more perfectly will the purpose of the establishment have been attained. ... This being impossible, the next thing to be wished for is, that, at every instant, seeing reason to believe as much, and not being able to satisfy himself to the contrary, he [the watched] should *conceive* himself to be so.

Panopticon (2)

Through such an arrangement of the threat of perpetual surveillance, Bentham believed disciplinary power would be *automatic*, and thus exercised with minimal effort, or, as Michel Foucault later reflected, the Panopticon would "induce in the inmate a state of conscious and permanent visibility that assures the automatic functioning of power" (Foucault, 1977, p. 197).

Panoptic Gaze

This automatic functioning of disciplinary power, then, manifested itself through what Foucault describes as a *panoptic gaze*:

There is no need for arms, physical violence, material constraints. Just a gaze. An inspecting gaze, a gaze which each individual under its weight will end by internalizing to the point that he is his own overseer, each individual thus exercising this surveillance over, and against, himself. A superb formula: power exercised continuously and for what turns out to be a minimal cost. (Foucault, 1980, p. 155)

Panoptic Gaze (2)

Foucault viewed the panoptic gaze as the quintessential disciplinary apparatus of modern society, where the functioning of power extended beyond Bentham's specific architectural form, and manifested itself in various contexts of everyday life: the home, the school, the hospital, the workplace, and so on. Thus, the gaze of the Panopticon expands to become "a whole complex mechanism, embracing ... stricter methods of surveillance [and] more efficient techniques of locating and obtaining information" (Foucault, 1977, p. 77).

Power via Panoptic Gaze

Oscar Gandy identified one source of the everyday functioning of power via the panoptic gaze when he warned of the "panoptic sort" (Gandy, 1993), a means whereby individuals are continually identified, assessed and classified through surveillance of their day-to-day commercial and financial transactions, for the purpose of coordinating and controlling their access to consumer goods and services.

Gandy's concern with panoptic sorting has been expanded beyond the consumer realm into a broader social milieu, where the notion of "social sorting" highlights the growing drive in our modern surveillance society for identification and classification of citizens, often as a powerful means of creating and reinforcing long-term social differences.

Since sorting and classification has been shown to be closely entwined with the exercise of power (Bowker & Star, 1999; Foucault, 1971; Suchman, 1997), the consequences of panoptic and social sorting – and the panoptic gaze which forms their foundation – creates effects that, as sociologist David Lyon describes, concern "deep discrimination...and social justice" (Lyon, 2003b, p. 1).

Panoptic Gaze of Dataveillance

Given this, we can now understand the panoptic gaze, the automatic functioning of power, the consequences of panoptic and social sorting the emerge from its discriminatory gaze.

Panoptic Gaze of Dataveillance

...and we can not turn our attention to the role of dataveillance in our thesis.

Dataveillance

The catalyst triggering both Gandy and Lyon's anxiety with the panoptic gaze was the rapid emergence of a complex set of technologies and practices that involve "the collection, processing, and sharing of information about individuals and groups that is generated through their daily lives as citizens, employees, and consumers. This panoptic apparatus resembles what Roger Clarke has refers to as *dataveillance*, defined as both "the massive collection and storage of vast quantities of personal data" and "the systemic use of [such] data...in the investigation or monitoring of one or more persons".

For example, everyday interactions with health care providers, online retailers, highway tollbooths, local grocery stores and libraries result in the collection, analysis, storage and sharing of information about one's address, purchasing habits, age, education, health status, travel activity, employment history, phone numbers and much more, forming detailed "digital dossiers" of nearly every individual within its gaze.

Digital Dataveillance

Since Clarke's first conceptualization of dataveillance almost twenty years ago, advances in digital networking, data storage capacity and processing power have enabled previously unimaginable levels of interconnectivity, aggregation, and real-time analysis of a wide array of personal information. A new age of digital dataveillance has seemingly emerged.

The role of modern information and communication technologies within infrastructures of dataveillance cannot be understated: frequent shopping cards connect purchasing patterns to customer databases, intelligent transportation systems enable the tracking and recording of vehicles as they travel the highways, electronic key cards manage access to locations while creating a record of one's movements, RFID tags track inventories and purchases, and biometric technologies digitize one's intrinsic physical or behavioral traits for automated identification and authentication.

Internet Dataveillance

Recently, the Internet has emerged as not only a revolutionary technology for communication, commerce and the distribution of information, but also as an ideal infrastructure of dataveillance, enabling the widespread monitoring and collection of personal and identifiable information about its millions of users. The privacy and surveillance concerns with various Internet technologies have been well documented and debated, ranging from the use of web cookies and tracking bugs, the emergence of spyware and digital rights management systems, workplace monitoring of electronic communications, the aggregation and data-mining of users clickstream data, and the widespread monitoring of Internet traffic by law enforcement agencies.

Panoptic Gaze of Dataveillance

The emergence of digital & Internet dataveillance technologies reveals how the gaze of the panopticon has extended from Bentham's physical prison tower into the realm of advanced information technologies and computer databases that facilitate the collection and exchange of information about individuals.

Panoptic Gaze of Dataveillance

Yet, the rising ubiquity of digital dataveillance in everyday life has concrete material effects on par with the most pervasive panoptic gaze envisioned by Bentham or Foucault. The resultant effects relate not just to personal privacy, but also issues of discrimination, social justice, and personal freedom. Law professor Michael Fromkin (2000) summarizes these effects best:

Reams of data organized into either centralized or distributed databases can have substantial consequences beyond the simple loss of privacy caused by the initial data collection, especially when subject to advanced correlative techniques such as data mining. Among the possible harmful effects are various forms of discrimination, ranging from price discrimination to more invidious sorts of discrimination. Data accumulation enables the construction of personal data profiles. When the data are available to others, they can construct personal profiles for targeted marketing, and even, in rare vases, blackmail. For some, just knowing that their activities are being recorded may have a chilling effect on conduct, speech, and reading.

...A further danger is that the government or others will attempt to use the ability to construct persona profiles in order to predict dangerous or antisocial activities before they happen. People whose profiles meet the criteria will be flagged as dangerous and perhaps subjected to increased surveillance, searches, or discrimination. (pp. 1469-1471)

Or, as Clarke notes,

[The] impact of dataveillance is a reduction in the meaningfulness of individual actions, and hence in self-reliance and self-responsibility. ...In general, mass dataveillance tends to subvert individualism and the meaningfulness of human decisions and actions.

Panoptic Gaze of Web 2.0

These effects of the panoptic gaze of dataveillance are intensified as the infrastructures of dataveillance become digital and web-based. They are amplified even further, I argue, with the introduction of Web 2.0 infrastructures, such as “the perfect search” engine and platforms which facilitate amateur data mining and P2P surveillance.

Panoptic Gaze of Web 2.0

Let's first look at Search 2.0

Web Search as the Center of Gravity

As the Internet has become increasingly important to modern citizens in their everyday lives, *web search engines* have emerged as today's prevailing information interface for accessing the vast amount of information available on this global network. 84% of American adult Internet users have used a search engine to seek information online. On any given day, more than 60 million American adults send over 200 million information requests to web search engines, making searching the web second the second most popular online activity (behind using e-mail). Originally designed to provide easy access to Internet websites, search engines now provide gateways to online images, news reports, Usenet archives, financial information, video files, e-mail and even one's desktop files. Recently, search engine providers, such as Google, have started to digitize items in the “material” world, adding the contents of popular books, university libraries, maps, and satellite images to their growing, searchable indices. Reflecting on the rapid emergence of search-related applications, Silicon Valley venture capitalist Roger McNamee

noted that “search is the new center of gravity for the computer industry” (McNamee, 2005). The same can be said more generally for the role of search engines as today’s dominant information interface: *Search engines have become the center of gravity for people’s everyday information-seeking activities.*

Search 2.0 – the Perfect Search

Since the first search engines started to provide a way of interfacing with the content on the Web, there has been a drive for the “perfect search engine,” one that has indexed all available information and provides fast and relevant results. A perfect search engine would deliver intuitive results based on users’ past searches and general browsing history, knowing, for example, whether a search for the keywords “Washington” and “apple” is meant to help a user locate Apple Computer stores in Washington, D.C. or nutritional information about the Washington variety of the fruit. A perfect search engine must also be able to provide results that suit the “context and intent” of the search query, and must have “perfect recall” in order to deliver personalized and relevant results that are informed by who the searcher is.

When asked what a perfect search engine would be like, Google co-founder Sergey Brin replied quite simply, “like the mind of God.”

Dataveillance in Search 2.0

Attaining such an omnipotent and omniscient ideal, requires search engine providers to collect as much information about their users as possible – they must engage in dataveillance of their users’ online activities. To accomplish this, most web search engines maintain detailed server logs recording each web search request processed through their search engine and the

results clicked. In addition to including the IP address of individual queries in their server logs, most search engines also rely heavily on web cookies to help differentiate users and track activity from session to session.

However, as users increasingly take advantage of software and browser features that make it easier to view, delete and block web cookies received from the sites they visit, web providers increasingly urge users to create an account with the website and login when using the services.

Google, for example, started experimenting with products and services that required users to register and login in early 2004, including personalized search results, e-mail alerts when sites about a particular topic of interest are added to Google's index. Many other services requiring a Google Account soon followed, even if some cookies are blocked, any activities performed while logged into your Google account can be associated across their various services. Google's encouragement of the creation of Google Accounts, combined with its use of persistent web cookies, provides an infrastructure of dataveillance for the creation of detailed server logs of users' search activities.

Dataveillance in Search 2.0 (2)

The majority of web searchers are not aware that search engines have the ability to actively track users' search behavior. For many, this ability first became apparent when news broke of the Department of Justice's legal battle with Google in early 2006, followed by the release of thousands of search records by AOL later that year. In response to the growing anxiety over the surveillance of search engine use, various tools and strategies emerged to help hide or obfuscate one's web search activity from being easily tracked and monitored. Such efforts, while

important, tended to either require a certain level of technological sophistication (i.e., the use of anonymous routing services such as Tor), suggest the blocking of search engine cookies which could prevent other desirable services from working properly, or focus only on keeping one's *web searches* from being tracked, which ignores the myriad of other opportunities available for web search engine providers to collect detailed information about its users.

Google's Broader Dataveillance Infrastructure

So while public attention has recently focused on the ability of search engine companies such as Google or AOL to track user's search history in detailed server logs, and some limited solutions to avoid tracking of one's web searches have emerged, less attention has been paid to how Google can also monitor and aggregate activity *across the myriad products and services the make up their larger web search information infrastructures*.

In all, Google has amassed an extensive web search information infrastructure comprising nine distinct information-seeking contexts. Inherent in this infrastructure is the ability to collect and aggregate a wide array of personal and intellectual information about its users: to engage in widespread digital dataveillance.

Logo slides...

These nine contexts include: general information inquiries, academic research, news and political information, communication and social networking, personal data management, financial data management, shopping and product research, computer file management, and Internet browsing.

Google's Broader Dataveillance Infrastructure (summary)

So, by striving to create the *perfect search engine*, Google has assembled over two dozen interconnected products and services within their growing information infrastructure. Google's encouragement of the creation of Google Accounts, combined with its use of persistent web cookies, provides the necessary architecture for the creation of detailed server logs of user's browsing history both on Google web properties and beyond. The result is an infrastructure of dataveillance, arming Google with the ability to collect and aggregate a wide array of personal and intellectual information about its users, extending beyond just what website they search for, but also including what news they read, what interests they have, the blogs they follow, the books they enjoy, the stocks in their portfolio, and perhaps even every website they visit.

Search 2.0's Gravitational Pull

So, while it is easy to think of search engines like Google as one-way information interfaces, where you enter a search term, and you get results, there is an important *feedback loop*; the interface is two-way. More than just the center of gravity of information seeking online, Google's information infrastructure also acts as a black hole, if we extend the metaphor, using its gravitational forces to pull as much information about its users *into* its domain as possible.

By monitoring and aggregating the results of every web search performed, every image result clicked, every website bookmarked, or every page visited with the Toolbar, Google has created sophisticated infrastructure of dataveillance. The result is what John Battelle calls a "database of intentions":

This information represents, in aggregate form, a place holder for the intentions of humankind - a massive database of desires, needs, wants, and likes that can be

discovered, subpoenaed, archived, tracked, and exploited to all sorts of ends. Such a beast has never before existed in the history of culture, but is almost guaranteed to grow exponentially from this day forward. This artifact can tell us extraordinary things about who we are and what we want as a culture. (Battelle, 2003)

While many of our day-to-day habits – such as using credit cards, ATMs, cell phones, or automated toll collection systems – leave countless “virtual footprints” of our activities, the panoptic gaze of Google’s infrastructure of dataveillance tracks our search histories, e-mails, blog posts or general browsing habits, providing “an excellent source of insight into what someone is *thinking*, not just what that person is doing.”

Panoptic Gaze of Search 2.0

The full effects of the panoptic gaze of Google’s infrastructure of dataveillance have yet to fully materialize and are difficult to predict. But there are some examples to consider: Instances of how users of Google’s infrastructure were made visible for the exercise of disciplinary power include a court ordering Google to provide the complete contents of a user’s Gmail account, including e-mail messages he thought were deleted; or the introduction of evidence that a suspected murderer performed a Google search for the words “neck snap break”.

Google appears to recognize, at least partially, the disciplinary threats of storing such robust records of its users activities. The company announced it would move user data collected from its Chinese site outside of that country in order to prevent China’s government from being able to access the data without Google’s consent. But while Google resisted the US Government’s request for random search records, the company recently agreed to comply with a Brazilian court order to release data on users of its Orkut social networking site to help Brazilian authorities investigate use of the site related to illegal activities.

Panoptic Gaze of Search 2.0

Google's infrastructure of dataveillance also spawns instances of "panoptic sorting" where users of Google are identified, assessed and classified for economic purposes. Google, like most for-profit search engine providers, is financially motivated collect as much information as possible about each user: they can charge higher advertising rates when ads are accurately placed before the eyes of users with relevant needs and interests. And if they know you like Porsche sports cars, they won't waste time showing you ads of a station wagon. Through the panoptic gaze of its diverse suite of products, Google collects as much information as possible about an individual's behavior, and considers it to be potentially useful in the profiling and categorization of a user's potential economic value. Recognizing that targeted advertising will be the "growth engine of Google for a very long time", Google CEO Eric Schmidt stressed the importance of collecting user information, acknowledging that "Google knows a lot about the person surfing".

Panoptic Gaze of Search 2.0

Perhaps the most potent consequence of the panoptic gaze of Google's infrastructure of dataveillance is how, like Bentham's original Panopticon, it might induce "a state of conscious and permanent visibility". There is rising anxiety among searchers about the presence of such systematic monitoring of their online information-seeking activities, and how they are made "visible" within this seemingly anonymous activity. When the New York Times tracked down a web searcher from the recent release of AOL search data, her reaction was to exclaim "My goodness, it's my whole personal life...I had no idea somebody was looking over my shoulder".

This rising consciousness and visibility has lead to a rapid emergence of various tools and strategies to help hide or obfuscate one's web search activity from being easily tracked and

monitored. Examples include FoxyProxy used to disguise your IP address, and TrackMeNot, an extension that send ghost search queries to search engines in order to obfuscate your search history.

Resisting Search 2.0

Yet, the adoption rates of such tools remains unknown, and as Google continues to expand its information infrastructure (Picasa photo sharing, Spreadsheets, Writely, Earth, etc) it becomes increasingly difficult for everyday users to recognize the data collection threats of these services, and easier to take the design Google's infrastructure of dataveillance merely "at interface value," and more difficult to resist the expanding infrastructure of dataveillance within Search 2.0 platform.

Resisting Search 2.0

Greg Elmer, in his book *Profiling Machines*, warns of the dangers of such an environment where the collection of personal information becomes a prerequisite of participation which inevitably entrenches power in the hands of the technology designers:

Ultimately, what both requesting and requiring personal information highlight is the centrality of producing, updating, and deploying consumer *profiles*.... And although Foucault warns of the self-disciplinary model of punishment in panoptic surveillance, computer profiling, conversely, oscillates between seemingly rewarding participation and punishing attempts to elect not to divulge personal information. (Elmer, 2004, pp. 5-6)

This blurring of punishments and rewards – subtle *requests* and not so subtle *commands* for personal information – reoccurs in Google's information interface where the default settings and arrangement of services make the collection of personal information automatic and difficult to resist.

Panoptic Gaze of Search 2.0

To summarize, the emergence of Search 2.0 platforms represent complex information infrastructures that connect various products and services through the use of persistent web cookies or universal logins. As search engines continue become the center of gravity of user's information-seeking needs, their gravitational pull captures an increasing amount of personal information about those users, resulting in a robust *infrastructure of dataveillance* focusing a panoptic gaze on users as they engage in their everyday information-seeking activities online.

This continued sophistication and expansion of Search 2.0 – epitomized by Google's suite of information-seeking products – represent the emergence of a new “social, political, and technical infrastructure that renders visualization meaningful for the basis of disciplinary social control”, whose “methodical, technology-driven, [and] impersonal” panoptic gaze is quickly becoming “a primary mechanism of surveillance and, by extension, social control in our society” (Staples, 2000, p. 5). These search-based infrastructures of dataveillance increasingly contribute to a rapidly emerging “soft cage” (Parenti, 2003) of everyday surveillance, where they, like other dataveillance technologies before them, contribute to the curtailing of individual freedom, affect users' sense of personal identity and concept of self, and present issues of “deep discrimination...and social justice” (Lyon, 2003b, p. 1).

Panoptic Gaze of Web 2.0 (transition)

So here we begin to understand the threats of the panoptic gaze of web 2.0 as it relates to Search 2.0 – the drive for the perfect search engine.

Panoptic Gaze of Web 2.0 (transition)

My time is starting to run out, but I would like to quickly discuss a 2nd important externality of web 2.0: the rise of amateur data mining and what I call “peer to peer surveillance.” Let me preface these remarks by noting that this is a new area of exploration for me, and that I haven’t yet spent enough time theorizing about their importance; I have no grand theories to espouse. But I would like to just point out some trends I see in web 2.0 and social software that have implications in the ability of everyday users to surveil one another.

Don, the Camera Thief

Let us begin with an interesting case of amateur data mining made possible through Web 2.0. This is the case of the stolen camera:

The blog BoingBoing blogged about this story of a woman lost her camera while on vacation, but the family who found it refused return it because their child liked it so much. A few days later, BoingBoing received an e-mail from someone who claimed his name was “Don Deveny,” purportedly a Canadian lawyer, who implied that the post was illegal and that BoingBoing was liable for making it. The folks at BoingBoing doubted the legitimacy of the email (the word “lawyer” was misspelled, for example), and decided to see what he could find out about “Don.”

They first contacted many of the law societies in Canada, none of whom had any record of a “Don Deveny” licensed to practice law in Canada. (BTW, it is illegal to pretend to be a lawyer). From their e-mail exchange, they were able to isolate the writer’s real e-mail address from the message headers, and through a Google search, find other pages that contain that address. That led them to a profile page for a user of the website called “Canada Kick A**” who

shared the very same e-mail address. That profile page had a different person's name (perhaps "Don's" real name?), and also listed a location and profession for the user (he's not a lawyer). It didn't take much to figure out (or at least get a better clue) as to who this e-mailer was, and his profile page on a Web 2.0-inspired discussion board made it much easier.

Readers of BoingBoing did some amateur data mining of their own: a commenter at the original camera owner's blog who shared many of the same sentiments of "Don," along with many of the same spelling errors. This commentor used a different screen name, but when asked to identify himself, also said he was a lawyer. Another reader then discovered that a user with that same screen name recently bid on memory cards at eBay that would have been used in the stolen camera. More amateur data mining ensued, and discovered another user profile at a different discussion forum with the same user name and same "favorite sites" listed in the signature file. And this page included a photo of the user: Is this "Don" our camera thief?

Lonelygirl15

And then there's the story of Lonelygirl15, the mysterious girl leaving video confessions on YouTube with a huge following of devoted fans, yet know one knew who she was or if they were really just a kid's video diary or perhaps a large hoax or advertising campaign. After some amateur data mining, the truth came out:

A reader was surfing an article on Lonelygirl15 at a random website when he came across a comment that linked to a private MySpace page that was allegedly that of the actress who plays Lonelygirl15. As the profile was set to "private," there was no real info one could glean from the page. However, when he queried Google for that particular MySpace user name, "jeesss426," I found a Google cache from the page a few months ago when it was still public.

A lot of the details of the girl's background quickly emerged: She was an actress from a small city in New Zealand who had moved to Burbank recently to act. The name on the profile was "Jessica Rose." When he happened to query Google image search for "Jessica Rose New Zealand" he was instantly rewarded with two cached thumbnail photos of Lonelygirl15, a.k.a. Jessica Rose, from a New Zealand talent agency that had since removed the full size versions. A later search on Yahoo on "jeessss426" also turned up a whole load of pictures from her probably forgotten ImageShack photo sharing account. Lonelygirl15 was revealed.

Little effort was needed to link up the various e-mails, user names, personal data and photos shared across blogs, discussion forums and other Web 2.0-style sites to track down "Don the camera thief" or "LoneyGirl15". Moving more and more of our activities to Web 2.0 makes it harder to remain anonymous, and the myth of "security through obscurity" seems to be disappearing as various crumbs of our true identity are being scattered across the Web 2.0 landscape.

Farrand Field

Another example is Farrand Field. Every year, thousands of students at the University of Colorado celebrate April 20th with a large marijuana smoking event on Farrand Field in the center of campus. University authorities, obviously, don't condone the illegal activity, but have had a hard time tracking down the violators. Identification has been made easier since the police can now look at user's Flickr streams to find photographs that have been "tagged" or labeled with student's real names, or YouTube video streams where, along with identifying labels, students might also call out each others' names for easy identification. Campus police can also

just search through user's myspace, Live Journal or Facebook accounts to find mention of participation in the event (which the police did do). By sharing this personal information on Web 2.0, students are making law enforcement's job much easier.

Farrand Field

This year, the university also installed numerous video surveillance cameras in order to capture the faces of students, and posted them to a website offering rewards to anyone who can help identify them. Hundres were identified through this peer-surveillance technique.

Amateur Facial Recognition

But rather than relying on other students to snitch, this would be easier if the University could just process the photos through a facial recognition system to identify each smoker. Or what if we could take the photo of Don the Camera thief or Lonelygirl, and just using their images, figure out who they were? We would need a pretty powerful facial recognition system, and we'd need to be sure they were in the database. This is a difficult task, as most facial recognition and surveillance systems are quite expensive, owned and operated by law enforcement agencies, and only include the names & faces of criminals or terrorists. But thanks to Web 2.0, these data-mining and surveillance tools are increasingly being put in the hands of people like you and me – we can now engage in our own “peer to peer surveillance” and amateur facial regonition.

Riya

Enter “Riya”. Riya hopes to take the image tagging of Flickr one step further by compiling a facial recognition profile of every user-submitted photo on the web.

Riya is a photo sharing and search site that lets you tag and search images based on facial recognition technology. Here’s how it works: you upload your photo library to Riya and “tag” the faces in your photos by putting a box around them and labeling it with the person’s name. After you have named a few faces, Riya’s facial recognition technology will take over and attempt to automatically tag different faces it recognizes so that you don’t have to.

The plan for Riya is to take that facial recognition profile you’ve created for the people in your photos, and allow you to even search the web for other photos with those same faces.

If you used Google Image Search to find photos, you would only get results of photos where the person who put the image on the website actually labeled it with that person’s name. But with Riya, you can easily find *any* photo on the web that *looks* like that facial profile, whether or not the person who posted actually labeled the image or not.

Riya

For example, let’s say your girlfriend told you she was going away to visit her grandmother. But instead, she went to a party with an ex-boyfriend. You’d have no way of knowing she did that. Now lets say someone took a group photo and posted it on their Flickr page. They labeled the photo with some people’s names, but not your girlfriend’s, because she didn’t even know who that was. Without a text label of your girlfriend’s name, you wouldn’t be able to find the photo if you searched on Flickr or through Google Image Search. *But Riya can*

find her, and when you search for your girlfriend using their system, that photo appears, with her face highlighted.

Riya

So, while users of Flickr upload and tag photos of friends (and strangers), all of which are searchable, Riya is applying the facial recognition profiles it learns from your photos and actively scans the entire web, finding faces of people in all kinds of photos. And once you find one photo of a person, you just click their face, and you'll find a whole bunch more, whether they wanted them to be identifiable or not.

Conclusion

So, these quick examples reveal how the growing power and ubiquity of Web 2.0 applications might contribute to the ease of amateur data mining and facilitate peer-to-peer surveillance. Certainly, these are not the intended uses of such technologies, but an externality *that must be confronted*.

Similarly, identifying the panoptic gaze of Search 2.0 applications does not mean that Google is necessarily "being evil" in construction such tools. Most are incredibly innovative and quite useful. But we must address the consequences of this growing infrastructure of dataveillance.

My point here today is to stand before you as a scholar committed to the ideal that technologies can be designed to further not only instrumental values such as communication, efficiency or even entertainment, but also ethical and human values, such as privacy, autonomy and liberty. An important first step is to understand how the design and use of technologies

might impact values of moral and ethical import. I hope my presentation has helped to start the process of gaining a critical understanding of the value implications of Web 2.0 technologies.

The next step is more difficult: what do we do about it? Can we resist these tools? Appropriate them to protect values? Design them in value-sensitive ways? That is the challenge before us, and one I hope we can discuss it more over the rest of the day.

Thank you.

Table 1: Google Suite of Products and Services

Product	Description	Information Collected	Notes
<i>1. General Information Inquiries</i>			
Web search	Query-based website searches	<ul style="list-style-type: none"> • Web search queries • Results clicked 	
Personalized Homepage	Customized Google start page with content-specific modules	<ul style="list-style-type: none"> • News preferences • Special interests • Zip code 	<ul style="list-style-type: none"> • Use in conjunction with Google Account is encouraged
Alerts	E-mail alerts of new Google results for specific search terms	<ul style="list-style-type: none"> • News preferences • Special interests • E-mail address 	<ul style="list-style-type: none"> • Alerts for a user's own name (vanity search) are common
Image Search	Query based search for website images	<ul style="list-style-type: none"> • Search queries • Results clicked 	
Video	Query based search for videos hosted by Google	<ul style="list-style-type: none"> • Search queries • Videos watched/downloaded • Credit card information for purchased videos • E-mail details for shared videos 	<ul style="list-style-type: none"> • Google Video Player available for download with additional DRM technology
Book Search	Full text searches of books scanned into Google's servers	<ul style="list-style-type: none"> • Search queries • Results clicked • Pages read • Bookseller pages viewed 	<ul style="list-style-type: none"> • Google Account required in order to limit the number of pages a particular user can view
<i>2. Academic Research</i>			
Scholar	Full text searches of scholarly books and journals	<ul style="list-style-type: none"> • Search queries • Results clicked • Home library (Optional) 	
<i>3. News and Political Information</i>			
News	Full text search of recent news articles	<ul style="list-style-type: none"> • News search queries • Results clicked 	<ul style="list-style-type: none"> • With a Google Account, users can create customized keyword-based news sections
Reader	Web-based news feed reader	<ul style="list-style-type: none"> • Feed subscriptions • Usage statistics 	
Blog Search	Full text search of blog content	<ul style="list-style-type: none"> • Search queries • Results clicked 	
<i>4. Communication and Social Networking</i>			
Gmail	Free web based e-mail service with contextual advertising	<ul style="list-style-type: none"> • Text of email messages • E-mail searches performed • Email address or cellphone number (used for account creation) 	<ul style="list-style-type: none"> • Creation of GMail account automatically results in activation of Google Account • Logging into Gmail also logs user into their Google Account

Groups	Free web based discussion forums	<ul style="list-style-type: none"> • Search queries • User interests • Usage statistics • Profile information 	<ul style="list-style-type: none"> • Includes complete Usenet archives dating back to 1981 • Google Account required for creation of new Group; User encouraged to create detailed profiles, including name, location, industry, homepage, etc
Talk	Web-based instant messaging and voice calling service	<ul style="list-style-type: none"> • Contact list • Chat messages • Usage statistics 	<ul style="list-style-type: none"> • Google Account and Gmail e-mail address required
Blogger	Web-based blog publishing platform	<ul style="list-style-type: none"> • Weblog posts and comments • Profile information • Usage statistics 	<ul style="list-style-type: none"> • Google Account required • Users encouraged to create detailed profiles, including name, location, gender, birthday, etc
Orkut	Web-based social networking service	<ul style="list-style-type: none"> • Profile information • Usage statistics • E-mail address and content of invitations 	<ul style="list-style-type: none"> • Invitation-only • Google Account required • Users encouraged to create detailed profiles, including name, location, gender, birthday, etc
Dodgeball	Location-based social networking service for cellphones	<ul style="list-style-type: none"> • Profile information • E-mail address • Location • Mobile phone information • Text messages sent 	<ul style="list-style-type: none"> • User location when messages sent are tracked by Google
<i>5. Personal Data Management</i>			
Calendar	Web-based time-management tool	<ul style="list-style-type: none"> • Profile information • Events • Usage statistics 	
<i>6. Financial Data Management</i>			
Finance	Portal providing news and financial information about stocks, mutual funds; Ability to track one's financial portfolio	<ul style="list-style-type: none"> • Financial quotes • Discussion group posts • Discussion group views • Portfolio (optional) • Profile information 	<ul style="list-style-type: none"> • Google Account required for posting to discussion board • Names and e-mails are displayed with discussion posts
<i>7. Shopping and Product Research</i>			
Catalog Search	Full text search of scanned product catalogs	<ul style="list-style-type: none"> • Product search queries • Results clicked 	
Froogle	Full text search of online retailers	<ul style="list-style-type: none"> • Product search queries • Results clicked • Sites visited • Shopping list 	<ul style="list-style-type: none"> • Google Account required for shipping lists

Local / Maps	Location specific web searching; digital mapping	<ul style="list-style-type: none"> • Search queries • Results clicked • Home location 	<ul style="list-style-type: none"> • Search queries might include geographic-specific information • Default location stored via web cookie
<i>8. Computer File Management</i>			
Desktop Search	Keyword based searching of computer files	<ul style="list-style-type: none"> • Search queries • Computer file index (Optional) 	<ul style="list-style-type: none"> • Search queries visible to Google under certain circumstances • Desktop file index is stored on Google's services if using Search Across Computers
<i>9. Internet Browsing</i>			
Bookmarks	Online storage of website bookmarks	<ul style="list-style-type: none"> • Favorite websites • When visited 	<ul style="list-style-type: none"> • Google Account required
Notebook	Browser tool for saving notes while visiting websites	<ul style="list-style-type: none"> • Notes and clippings • Sites annotated 	<ul style="list-style-type: none"> • Google Account required
Toolbar	Browser tool providing access to various Google products without visiting Google websites	<ul style="list-style-type: none"> • Search queries • Websites visited 	<ul style="list-style-type: none"> • Use of some advanced features routes all browsing traffic through Google servers • Some features require Google Account
Web Accelerator	Software to speed up page load times for faster web browsing	<ul style="list-style-type: none"> • Websites visited 	<ul style="list-style-type: none"> • All browsing traffic is routed through Google servers