Data mining and surveillance in the post-9.11 environment

Oscar H. Gandy, Jr.
Herbert I. Schiller Professor
Annenberg School for Communication
University of Pennsylvania

For presentation to the
Political Economy Section, IAMCR
Barcelona, July, 2002

Introduction

In his wildly successful book on the future of cyberspace[1], Lawrence Lessig responded to a general challenge to privacy activists: tell us what is different about surveillance in the computer age. Lessig suggests that the difference is to be seen in the ease with which the data generated from the routine monitoring of our behavior can be easily stored, and then searched at some point in the future.

That is, because more and more of our daily life involves interactions and transactions that generate electronic records, our lives become fixed in media that can be examined and reviewed at will. Lessig and others who are concerned about threats to privacy[2] have identified the countless ways in which our behavior in public places, as well as in the privacy of our homes, generates records that reside in the computers of corporations and government agencies.

While a sales clerk in the local store might take note of your interest in different pieces of jewelry, or clothing as you make your way from counter to counter, their monitoring does not generate a searchable record of each of your visits to the store. Indeed, unless they are security guards, and you are looking particularly suspicious that day, they don't follow you from floor to floor. It is only when you purchase those socks or gloves that a searchable record is made. However, the generation of transaction records in cyberspace is many times more extensive than it is in the world of bricks and mortar. Web servers generate a record each time a visitor clicks on a banner ad, or follows a link in order to learn more about some commodity or service.

In addition, because of the ways in which Web technology facilitates the linkage of records, the click streams, or mouse droppings that you leave behind as you browse around much of the web, makes it easy for marketing service providers like Double Click to develop a cumulative record.[3] Because Double Click manages the serving of ads for several thousand publishers on the web, your profile may contain information about a broad range of goods and services about which you have revealed some interest.

Because the cost of storing data in electronic form continues to drop (one theoretical estimate, based on what engineers refer to as Moore's law, suggests that the cost of storage drops by 50 % every 18 months),[4] there is less of an incentive for organizations to discard **any** transaction-generated information. The only problem that businesses and government agencies face is how to make sense of these growing mountains of data.[5]

Enter the mathematical wizards who brought us both the bell curve and the ballistic missile, and voila! we have the science of data mining, or as the specialists would prefer, the science of Knowledge Discovery in Databases or KDD.

Data mining, as a tool for the discovery of meaningful patterns in data is the product of some rapidly developing techniques in the field of applied statistical analysis. Of particular importance for those of us who are concerned about the implications of data mining for individual and collective privacy, is the fact that data mining software, products, and services are being introduced into the marketplace by a number of competing vendors. The increasing sophistication of the software packages, and the rapidly declining prices for custom, and off-the-shelf data mining products means that the techniques will soon be in widespread use.

In addition, the government's heightened concern with security following the events of 9.11 means that an infusion of tax dollars for research and development is likely to attract a swarm of competitors. We can expect this activity to support an even more rapid expansion in the capacity of these systems to produce strategic "intelligence" from what would ordinarily be meaningless bits of data stored in computers all around the globe.

What I would like to do in this paper is provide a thumbnail sketch of data mining as a technology, identify some of the leading firms and the nature of their data mining products, and then identify some of the things that trouble privacy advocates, and others who are concerned about civil liberties.

Data Mining

As I have suggested, data mining is an applied statistical technique. The goal of any data mining exercise is the **extraction** of meaningful intelligence, or knowledge from the patterns that emerge within a database after it has been cleaned, sorted and processed. The routines that are part of the data mining effort are in some ways, similar to the techniques that are used to extract precious minerals from the soil. However, whereas the extraction of precious metals is often labor intensive, and represents risks to both workers and the environment, the extraction of intelligence from computer databases is increasingly being automated, in ways that reduce the direct risks to labor, while amplifying the risks to society in general. Indeed, as I hope to demonstrate, the impact of this technology on the social environment in the long run may be as destructive as strip mining.

Imagine if you can, the mountains of transactional data that are generated each time each of us purchases commodities that are marked with universal product codes. When we use credit or check verification cards, or any of a number of retail vendors' discount cards, individually identifiable information is captured and linked with those purchases. There is little wonder that global retail chains, like Wal-Mart have invested substantial resources in the development of data warehouses to manage the details and extract the value in the terabits of data being generated each day throughout their networks.[6]

Our interactions with government agencies, as well as with the component parts of the massive health care system, also generate detailed records. However, because these data are not gathered in standard forms with classification schemes akin to the UPC code, there are tremendous pressures within these industries to move toward greater standardization and comparability across transactions.[7]

Although progress is being made somewhat more slowly in translating voice messages into text for automated processing, no such barriers exist for classifying e-mail text, or the posts that are made to newsgroups. The textual components of web pages are also relatively easy to classify and describe, although the graphics on those pages still represent something of a problem for developers.

Even more problematic in terms of the need to develop common codes and classification standards, is the digitized output of surveillance cameras. However, it seems likely that the rate of success in developing classification techniques will increase substantially in response to research and development initiatives rushed through the legislature in response to the events of 9.11.

The goals of data mining

In general, data mining efforts are directed toward the **generation of rules** for the classification of objects. These objects might be people who are assigned to particular classes or categories, such as “that group of folks who tend to make impulse buys from those displays near the check out counters at the supermarket.”

The generation of rules may also be focused on discriminating, or distinguishing between two related, but meaningfully distinct classes, such as “those folks who nearly always use coupons,” and “those who tend to pay full price.”

Among the most common forms of analysis are those that seek to discover the **associative rules** that further differentiate between clients and customers. For example, video rental stores seem to be interested in discovering what sorts of movies tend to be rented together, and what sorts of movies tend to be associated with the sale of microwave popcorn or candy. Designers are especially interested in being able to predict the response of individuals to different offers or appeals. In an attempt to develop reliable sorting tools, data miners seek to discover patterns of association between demographic characteristics and commercial behaviors**. Discriminant analyses** are especially concerned with differentiating between high value and low value customers. In the case of urban retailers considering the placement of radio advertisements, these analyses help them to determine what sorts of ads are more likely to generate prospects, rather than suspects.[8]

Businesses seek to maximize profits by minimizing risk. They do this by identifying individuals, who, by virtue of their profiles, ratings, or comparative scores, should probably be ignored, avoided, or treated with the utmost deference and respect. Some business service providers may also rely upon the pattern recognition features of data mining programs to determine whether a credit card is likely to have been stolen, even though a theft has not been reported to the bank. For example, the search for "outliers" within a data set is often useful for predicting when a previously good customer is at risk of default, or about to make a fraudulent claim.

Similarly, insurers may utilize risk avoidance programs in an attempt to identify patterns of behavior that suggest elevated risk.[9] Starbucks reportedly makes use of a data-mining package called RiskIntelligence to help it reduce its exposure to lawsuits based on in-store accidents. In order for their risk minimization efforts to pay off for an organization like Starbucks with thousands of outlets, these sophisticated models have to incorporate representations of dozens of different floor plans, in addition to assessing the distribution of accidents by time of day, and season of the year. Of course, these data have different implications depending upon part of the country in which a store is located.[10]

Data mining specialists at the MITRE Corporation recently described one project that involved developing a strategy for identifying, or targeting vehicles for inspection by law enforcement officers.[11] The goals of this project were similar to another project that was supposed to mine aircraft accident and incident reports in order to identify what they referred to as "precursors" of dangerous situations in the air.

In the aviation example, an analysis of accident descriptions from on-board data recorders, or in the transcripts of witnesses, might reveal a critical pattern. For example, an analysis of accident reports might produce an indication that accidents often occur following an indication in air traffic controller tapes that the plane “veered to the left following takeoff.” Further analyses of accident records might indicate that such references were also more likely when particular malfunctions emerged during winter weather conditions. Each accident that is accompanied with a voice record that can be transformed into text can be searched for keywords, or terms that become increasingly meaningful with each accident. Over time, these terms become associated with some technical problem, at the same time they become more distinct from other terms or phrases that indicate a mistake, or a pilot error. The analysis of each additional accident adds to the power of the explanatory models that data mining helps to produce.

In the case of motor vehicle targeting, the data mining effort is directed toward identifying the attributes of the driver as well the vehicle that increase the ability of the resultant profile to identify targets. Each time a driver is subject to a more detailed inspection, more information is added to the pools of data about drivers and vehicles. One problem in developing these profiles is that the sorts of data that can be gathered during a full inspection, to say nothing of the data that might be gathered after an arrest, are substantially different from data that can be gathered on the basis of remote observation, or a cursory inspection.

We have no doubt heard quite a bit about the use of race as an element in the profiles used by State Police to identify vehicles they believe are likely to be transporting drugs.[12] Because of the adverse public response to the use of race as a predictor, data mining efforts are being directed toward finding other cues that the police believe to be equally useful, but less politically sensitive indicators of sufficient probable cause to justify a search.

Data mining technology

The technology of data mining becomes more sophisticated with each passing day. Neural networks are just one of the more sophisticated analytical resources being used more frequently in data mining applications. Neural nets are said to mimic the ways in which the brain processes information. These systems learn, or become increasingly accurate over time. The statistical learning model attaches and adjusts the weights that are applied to different attributes, or variables, and these weights are adjusted in response to each correct, and incorrect prediction or determination.

One application for neural networks is in support of fraud detection by insurers. In deciding whether a reported accident is likely to have been staged, an analytical model will have been developed based on the assessment of the detailed records of thousands of reported accidents, some of which have been determined to be fraudulent. Varying weights will be assigned to such things as the age and gender of the driver. As the model is developed other, perhaps more important indicators, such as whether the person called an attorney, and whether the claimant's **physician** was called before, or after the **attorney** was called, will be added to the model. The goal is to be able to make a prediction, and provide an estimate of its accuracy. It asks "how **likely** is this case to be a fraudulent claim?" the client or user of the application needs to decide whether they should just pay the claim, or whether they should risk angering the client by requesting additional information?

Some applications systems allow its users to view, navigate through, and generate a variety of recommendations derived from an analysis of multidimensional databases stored in remote locations.

We can imagine a sophisticated user in the not too distant future pulling out her combination PDA/telephone, in order to choose between some competing strategies that she had been considering using with the client she is about to meet for a working lunch.

While that scenario might be a bit far fetched in today's environment, according to one industry analyst, in five years time, e-business firms will have moved "business intelligence" of this sort to the average desktop in the same way that Microsoft Word and Excel have become commonplace.[13]

Commercially available data mining software

A number of firms have begun to offer data mining services and software products that make it easier for web-based marketers to transform transaction generated data into intelligence that they can use for customer segmentation. Well-defined segments often become the primary resource of a marketing campaign. Among the leaders of this emerging market are firms with names like digiMine, Accrue, NetGenesis and Personify. These firms provide analytical services to web-based companies. The emerging market also include familiar providers of statistical software like SPSS, that includes neural networks and rule induction features in its Clementine Service resource.

Some comprehensive software packages, or client service products are designed to facilitate customer relationship management (CRM). As I understand it, the philosophy of CRM is no longer one of capturing the largest share of the **market**; rather it is capturing the largest share of the most valuable customer's business.[14] This is an orientation to the market that reflects a belief that 20 percent of a firm's customers will provide 80 percent of their revenue. Corporate strategists believe that capturing that revenue can be assured only by "growing the customer" through "cross -selling." It also requires identifying those customers that are unlikely to grow, and then finding subtle ways to suggest that "it's time for them to go" because "they are the weakest link" in the relationship between you and your customers.

Increased demand for data mining tools

While the firms that are providing these products and services will continue to try and stimulate demand through aggressive marketing, the American government has provided a substantial increase in the incentives available for expanding the research and development of data mining applications.

Prior to September 11th, the most pressing concern was the threat of a deepening recession. The U.S. Department of Labor and a web-based employment service, Monster.com, entered into an agreement to merge public and private databases. These partners assumed that data mining might prove to be a valuable resource in the analysis and description of workforce trends.[15] The assumption was that variations in the demand for **temporary** workers might serve as a useful barometer of trends in the labor force more generally.

Obviously, more pressing concerns emerged after 9.11. Less than a week after the assault, an article in business section of USA TODAY asked reflexively, "What can tech companies do?" Expanded use of data mining was high on the list of possibilities. One CEO from one of the few surviving Internet communications firms suggested that we "are experts at data mining and we have vast resources of data to mine. We have used it to target advertising. We can probably use it to identify suspicious activity or potential terrorists."[16] There is little doubt that "volunteers" like him provided the information that was used to identify several of the so-called "material witnesses" who are still the guests of the FBI. Much more cautious responses were offered by technology developers who suggested that we were probably still years away from the kinds of data mining technology that might have allowed us to predict and interrupt the plans of the hijackers.[17]

Nevertheless, in response to what they perceives to be a continuing threat of terrorism, the Pentagon announced a major initiative to speed the development of new technologies that could be actually be deployed within 12-18 months.[18]

At the top of the wish list was an appeal for "ideas to identify and track down suspected terrorists, and to predict their future behavior." This goal was linked with a desire to "develop an integrated information base and a family of data mining tools and analysis aids." What the Pentagon was looking for was an analytical resource that would assist in the "identification of patterns, trends, and models of behavior of terrorist groups and individuals…. The system would allow 'what if' type modeling of events and behavioral patterns and result in predictive analysis products." What they are hoping for is a system that can scan data in the nation's computer networks and if they "discover that a member of an extremist group also brought explosives and visited a Web site about building demolition, they might be able to halt a potential attack."[19]

And finally, commercial providers were invited to "develop a deception detection device for use with counter-terrorism-based structured interviews for passengers of the various modes of transportation." In this particular case, the Pentagon was apparently seeking a reliable, portable, and efficient polygraph device.[20]

I am concerned that the development of these systems for use by defense and security agencies increases the likelihood that once these advanced systems come to be used routinely by a host of government agencies, they will soon become available as off-the shelf commodities for use in the commercial sector.[21]

The spread of the technology to the commercial sector is also likely to be accelerated because of increased pressure and latitude for engaging in privacy invasive activities in defense of this thing called "homeland security," under the expanded scope provided by the USA PATRIOT Act.[22] There is particular concern about the availability of details about individual's searching of the web, in that the capture of URLs from public terminals and private computers provides easy access to the content of files accessed by individual users.[23]

Social implications of data mining

So, why should this concern us?

As we have noted, data mining systems are designed to facilitate the identification and classification of individuals into distinct groups or segments. From the perspective of the commercial firm, and perhaps for the industry as a whole, we can understand the use of data mining as a discriminatory technology in the rational pursuit of profits. However, as a society organized under different principles, we have come to the conclusion that even relatively efficient techniques should be banned or limited because of what we have identified as unacceptable social consequences, or **externalities**.

For example. Marsha Stepanek of Business Week referred to the application of data mining techniques in electronic commerce as "Weblining."[24] I suspect she chose this term precisely because she hoped it would activate the collective distaste that Americans have expressed toward spatial, or geo-demographic discrimination against neighborhoods and communities defined by race. Indeed, these are techniques that the courts and legislatures have banned as "redlining" when used by banks and mortgage firms.

However, because the Internet is only marginally defined by geography[25], the "neighborhoods" that will be excluded from access to goods and services are primarily conceptual, rather than spatial. Because of this fact, the victims of weblining are less likely to be aware of their status as victims of discrimination. As a result, they will be even less likely to organize into an aggrieved group in order to challenge their exclusion from opportunities in the marketplace.

Let me be clear. There are some people who argue that even the use of race, gender, and age as elements within evaluative models should be allowed because they are economically efficient.[26] That is, because of the relatively high correlation between personal attributes like race, class and gender, and a host of other status and behavioral indicators, these data facilitate rational discrimination. In addition, because these indicators are relatively inexpensive to capture and interpret it would be inefficient, and irrational to bar their use in decision-making.

On the other hand, those of us who argue against this sort of discrimination are concerned that if we allow decision makers to use race, and gender, and other markers of group identity as the basis for exclusion from opportunity, then we will only strengthen the correlation between group membership and social status.

African American males provide the most obvious examples. The use of race as a bar to employment leads quite reasonably to a belief among African Americans that investment in education is makes little sense. Indeed, examples abound which suggest that more promising alternatives can be found in the underground economy.

Of course, it is not only discrimination on the basis of race, gender, or age that should concern us. Our concerns should be based more generally on what we understand to be the social consequences that flow from using a decision system that systematically bars members of groups or segments of the population from acquiring the informational resources that are essential to their individual development and their collective participation in the economy and in the public sphere.

As communications researchers we should be especially sensitive to the use of discriminatory technologies like data mining to determine which people will have access to the information that they need to make sense of the world. When data mining systems are used by companies in the communication or information fields to segment their audiences in the service of the bottom line, we should assume that disparities in access to information would emerge.[27]

While segmentation and targeting may be efficient, and it may serve the competitive and strategic interests of media organizations and their clients, I believe that it is destructive of the social fabric.

Segmentation reinforces difference, while it obscures those things we share in common. This is a point that has been made recently by Cass Sunstein, in his book, Republic.com.[28] Although Sunstein suggests that the increasing polarization we observe is the product of consumer choice, I believe he underestimates the influence of strategic marketing.

It is important to remember that access to information is often determined the kinds of **subsidies** that advertisers are willing to provide in order to gain access to the consumers (and I might add, the voters) whom they value the most. Financial support from advertisers not only provides subsidies to the people who need it least, but by withholding support from the less desirable audiences, the media which continue serve these disfavored groups must either deliver a lower quality product, or seek advertisers with less wholesome commodities for sale. Once again, the differences between us are drawn more sharply, and ironically, they seem to make even more sense.[29]

So, what are we to do?

The standard responses of governmental agencies like the Federal Trade Commission (FTC) are, in my view, unlikely to provide much protection from the dangers that will accompany the widespread use of data mining.

Most recently, the orientation of policy makers in the United States has been toward corporate self-governance, and away from regulation. The FTC has emphasized the value of a much-modified standard of “fair information practices” that are supposed to ensure that the public enjoys “notice and choice” regarding the collection and use of personal information.

While consumer oriented legislation may provide some increased security for individuals in their dealings with the health care establishment, and with regard to their children’s exploration of the Web, these regulations are for the most part meaningless as a defense against the social harms that data mining represents.

First of all, the dominant privacy framework is one that emphasizes “individually identified information.” Although much of the talk in policy circles is about the development and use of consumer profiles, the power of data mining lies not in its ability to target specific individuals, but in its ability to increase the benefits to be derived from controlling the behavior, on the average, of members of well-defined groups. Individuals readily provide the sorts of indexical details that allow decision makers to assign them to the right groups at the right time, in order for a critical decision to be made.

Second, citizens and consumers cannot expect to be meaningfully informed about the uses to which their transaction-generated information will be applied. This is the case, in large part, because those who manage these data warehouses have only the most general awareness of those future uses themselves. Individuals are generally provided with a meaningless choice between doing without, and providing a blanket license for whatever uses of information that a data manager decides is appropriate.

Consumers think they are protected in some way by regulations that guarantee them access to the information about them. The idea is that this access will allow them to challenge the accuracy of the data that have been recorded in these files. I doubt that there is any meaningful way for an individual consumer to challenge the cumulative score they have been assigned by some data mining operation based on neural net technology. Indeed, I doubt that few, if any us would understand the complex algorithm that produced it.

I recall the classic case of one Claire Cherry, a White woman in Georgia who claimed that she had been a victim of discrimination because Amoco denied her a gasoline credit card. It seems that her application had been denied in part because she lived within a zip code that included a high proportion of African Americans.

The problem that Ms. Cherry faced was that the scoring system used by Amoco made use of a multivariate model that included 38 variables. Understandably, she was unable to specify the impact that her zip code, and its underlying racial component actually had on the determination of her credit status.[30] Contemporary scoring models use hundreds of variables, and even more problematic from the perspective of today's consumers, many of these analytical models are adjusted continuously in order to incorporate the latest information that recent transactions provide.

It may be possible for privacy advocates to limit the storage of transaction data for longer than is absolutely necessary. They may also attempt to limit the use of this information for purposes unrelated the initial transaction. It seems unlikely, however, that one sector of government would seek the elimination of data in its files at the same time that other sectors are trying to **require** their secure storage, and increased sharing with any who can claim a legitimate interest. Indeed, following the bad press that the security agencies and president Bush has received of late, the Attorney General has been pushing for greater sharing between the U.S. and foreign governments. Recent developments within the European Community suggest that there is greater willingness on the part of governments to require more extensive data storage and sharing by public and private organizations.[31]

In the final analysis, the best strategies available to those of us who are concerned about the social costs of discrimination, may involve the mobilization of public opinion. People tend to be outraged when they discover, or are informed that they have been discriminated against. There is some value, therefore, in supplying the press with egregious examples of individuals, or communities, or classes of people, who have been victimized by data mining, and by the use of profiles based on irrelevant attributes like race or ethnicity. Of course, it is hard to predict how the public will respond to a growing awareness that discrimination is widespread, and that they are clearly at risk. In part it is a question about what people believe to be fair, and what they believe the circumstances demand.

Unfortunately, I remained concerned that the use of data mining in the so-called "war against terrorists" will soften us up for its use in the war against global competitors, or against the threat to shrinking profits, and a few "horror stories" about some "so-called victims" of discrimination will do little to shift the tide.

[1] Lawrence Lessig. Code and other laws of cyberspace. New York: Basic Books, 1999.

[2] David Lyon. Surveillance society. Monitoring everyday life. Philadelphia: Open University Press, 2001.

[3] Jeff Sovern. Opting in, opting out, or no options at all: The fight for control of personal information. 74 Washington Law Review, (1999),1033-1118.

[4] William Gates. Business at the speed of thought. New York: Warner books, 1999, p. 143.

[5] Heather Green, Linda Himelstein, Robert Hof & Irene Kunii. The information gold mine. Business Week (July 26, 1999), pp.EB16-23.

[6] Ibid., p. 232. Gates notes that Wal-Mart spent millions of dollars on its data mining system as a "high-end" customer.

[7] Geoffrey Bowker & Susan Star. Sorting things out: Classification and its consequences. Cambridge: MIT Press, 1999.

[8] Advertisers have been charged with avoiding the placement of advertisements for retail with Black-format radio stations because their clients believe that African Americans are likely to steal. See Kofi Ofori, When being No. 1 is not enough. The impact of advertising practices on minority owned & minority formatted broadcast stations. Report to the Federal Communications Commission. Washington, DC: Civil Rights Forum on Communications Policy, 1999.

[9] Ann Cavoukian. Data mining: Staking a claim on your privacy. Information and Privacy Commissioner, Ontario, Canada. January, 1998.

[10] Sara Roberts-Witt. Gold diggers. PC Magazine (on-line)February 20, 2001.

[11] Neil Rothleder, Earl Harris & Eric Bloedorn. Focusing on the data in data mining: Lessons from recent experience.

[12] Angela Allen-Bell. The birth of the crime. Driving while black (DWB), 25 Southern University Law Review, (Fall, 1997):195-225.

[13] Bernard Liautaud. Q&A. PC Magazine (online) February 20, 2001.

[14] D. Peppers and M. Rogers. The 1:1 Future: Building relationships one customer at a time. New York: Currency, 1997.

[15] Chris Mortonson. Beyond Monster.com: Benefits of DOL data mining with private marketplace. HR.Com. Visited 9/14/01

[16] Tom Evslin of ITXC, quoted by Kevin Maney in, What can tech companies do? USA TODAY, September 19, 2001. 6.B.

[17] Ibid.

[18] David Streitfeld & Charles Piller. Big Brother finds ally in once-wary high tech. LA Times.com (January 19, 2002). Available on-line: http://www.latimes.com/news/nationworld/natioin/la-011902techshift.story.

[19] Mike France, Jim Kerstetter, Jane Black, Alex Salkever & Dan Carney. Privacy in an age of terror. Business Week (November 5, 2001), p82-88.

[20] Scott Veale. Word for word/Pentagon science fair; Wanted: Tools to fight terrorism. All suggestions welcome. New York Times (October 28, 2001), Sect. 4 Pg 7.

[21] Bien Perez. State agencies turn to analytics to fight terrorists. South China Morning Post, October 3, 2001, Pg. 10, citing Shaun Doyle, vice president for intelligent marketing solutions, SAS Institute.

[22] Scott Carlson & Andrea Foster. Colleges fear anti-terrorism law could turn them into Big Brother. The Chronicle of Higher Education (March 1, 2002). Available on-line: http://chronicle.com/free/v48/i25/25a03101.htm.

[23] Wayne Madsen. Homeland security, homeland profits. Corpwatch. (December 21, 2001). Available on-line: http://www.corpwatch.org/issues/PID.jsp?articleid=1108.

[24] Marsha Stepanek. Weblining. Business Week, April 3, 2000.

[25] Stephen Graham & Simon Marvin. Splintering urbanizm. Networked infrastructure, technological mobilities and the urban condition. New York: Routledge (2001).

[26] Richard Epstein. Forbidden grounds: The case against employment discrimination laws. 1992

[27] Digital Silhouettes is a product offered by Predictive Networks that uses demographics, including race, to characterize Internet consumers' orientations toward 90 content subcategories on the basis of their analysis of click stream data. Their promotional materials claim that their "artificial intelligence engine derives individual preferences from web-surfing behavior." Curiously, their promotion claims that "over time, as more and more sites are visited, the appropriate level of confidence in the accuracy of the Digital Silhouette is established allowing Predictive Networks to accurately model user preferences, affinities and demographics while protecting their privacy," p. 4. Available online: http://www.predictivenetworks.com

[28] Cass Sunstein. Republic.com. Princeton, NJ: Princeton University Press, 2001.

[29] Oscar H. Gandy, Jr. "Dividing practices: Segmentation and targeting in the emerging public sphere," in W. Lance Bennett & Robert Entman (Eds.), Mediated Politics. Communication in the Future of Democracy New York: Cambridge University Press, pp. 141-159.

[30] *Claire Cherry v. Amoco Oil Co*., 490 F.Supp. 1026 (N.D. Ga. 1980).

[31] Julia Scheeres. EU law turns ISPs into spies? Wired News (May 29, 2002). Available online: http://www.wired.com/news/print/0,1294,5289,00.htm. This article describes an upcoming vote in the European Parliament, the Communications Data Protection Directive, that would allow member countries to require their telecommunications providers to retain information on customers' use of network services. This represents a reversal of general policies which required the elimination of data after the billing cycle was completed.