WHO CONTROLS THE PAST NOW CONTROLS THE FUTURE: COUNTER-TERRORISM, DATA MINING AND PRIVACY

WAYNE N. RENKE*

Recent world events have created international security concerns and a demand for counter-terrorist measures. Information fuels counter-terrorism. "Data mining" has been touted as a means for acquiring needed information. This article describes data mining, explores its social, political and personal risks, then assesses its impact on the Charter*-protected right to privacy. The author proposes a framework for the constitutionally appropriate regulation of data mining. Data mining is portrayed as a potentially valuable counter-terrorism tool which must be governed responsibly, if its costs are not to exceed its benefits.*

Les récents événements qui se sont passés dans le monde ont soulevé des inquiétudes sur la sécurité internationale et le besoin de mesures anti-terrorisme. L'information alimente l'anti-terrorisme. « L'exploration de données » a été décrite comme étant un moyen d'obtenir l'information requise. L'article décrit l'exploration de données, en examine les risques sociaux, politiques et personnels, et ensuite l'effet sur le droit du respect de la vie privée protégé par la Charte*. L'auteur suggère un cadre de régie constitutionnellement approprié pour l'exploration de données; cet exploration est décrite comme étant un outil anti-terrorisme potentiellement utile qu'il faut gérer de manière responsable si on ne veut pas que les coûts excèdent les avantages.*

TABLE OF CONTENTS

* B.A. (Hons), M.A., LL.B., LL.M., Professor, Faculty of Law, University of Alberta.

Who controls the past now controls the future
Who controls the present now controls the past
Who controls the past now controls the future
Who controls the present now?[1]

I. INTRODUCTION

September 11, Madrid, London, Bali — twice: our world, the Western World as we live it, is no longer what it was. International terrorism has succeeded in insinuating itself into our planning, our architecture, our consciousness. If not always explicit, it is always present. It has woven itself into the fabric of risk that limits our activity. New risks impose new demands on the law enforcement, military and intelligence communities. Their counter-terrorism operations require the fuel of information. New demands require new tools. "Data mining" has emerged as one means of providing the information required for counter-terrorism operations. But despite any benefits data mining may provide, it engenders severe risks of its own. Left unchecked, its effects could be as damaging as the threats it is deployed to combat. One solution could be to put this tool aside. Another solution, which I shall pursue here, is to regulate the technology appropriately. At its root, data mining invades privacy. That privacy is constitutionally protected. Data mining should therefore be regulated by principles analogous to those constitutionally demanded for other forms of privacy-invading technologies. To elaborate my argument, I shall discuss the following issues: the information needs of counter-terrorism; data mining, as a source of information for counter-terrorism; the social, political and personal risks of data mining; whether the information that is data mining's resource supports constitutionally cognizable "reasonable expectations of privacy"; and the features of constitutionally appropriate data mining regulation.

II. COUNTER-TERRORISM'S INFORMATION HUNGER

Counter-terrorist operations have an insatiable need for information, or, more precisely, for relevant information or actionable intelligence. Any law enforcement or intelligence operation requires information — and the more the better — but counter-terrorist demands are perceived to be particularly acute. Some features of contemporary terrorism explain its professional opponents' peculiar hunger for information.

The strongest motivation for information acquisition rests on terrorism's risks. Risk must be assessed in terms of both the probability of harm and the magnitude of harm. From the perspective of probability, for a North American (as opposed to, say, a resident of Israel or Iraq) the risk of injury as a result of terrorism is low. Residents of Canada or the U.S. are more likely to be injured by ordinary crime and far more likely to be injured in automobile accidents than by terrorist attacks.[2] From the perspective of magnitude of harm, terrorism

---

1 Rage Against the Machine, "Testify," *The Battle of Los Angeles* (New York: Sony BMG Music Entertainment, 1999), lyric by Zack de la Rocha.

2 Nicholas D. Kristof, "117 Deaths Each Day" *New York Times* (13 March 2004), online: Wired New York Forum <www.wirednewyork.com/forum/showthread.php?t=4651>; Louis Hugo Francescutti, Tracey M. Bailey & Trevor L. Strome, "Injuries: Public Health's Neglected Epidemic" in Tracey M. Bailey, Timothy Caulfield & Nola M. Ries, eds., *Public Health Law and Policy in Canada* (Markham, Ont.: LexisNexis Canada Inc., 2005) 219 at 222.

poses severe risks, in multiple dimensions. Like ordinary crimes and natural disasters, terrorism produces risks of physical harm. The harm may be relatively localized, as when a suicide bomber detonates himself or herself in a crowd. The harm may be more broadly distributed, as when suicide bombers or terrorists use explosive devices to attack public structures or facilities. Oklahoma City, Madrid and London provide examples; September 11 saw the destruction of multiple structures by suicidal terrorists employing aircraft as explosive devices. Furthermore, modern terrorism poses the threat of using weapons of mass destruction such as fission, fusion, chemical or biological weapons.[3] Terrorism is distinguished from other types of violent crime by the sheer scale of its potential destructiveness.

The distribution of terrorist risk has another aspect: everyone and anyone is a target, a subject of risk, a risk-bearer. In contrast, much of the physical violence of traditional organized crime is directed at other criminals or their associates. And in contrast to some older forms of political terrorism, physical violence is not directed at government, military, police or opponent group targets. Modern terrorism of the al Qaeda sort makes targets of individuals whether they are on the street, in a subway or in an office building. It does not discriminate on the basis of social, economic, cultural, ethnic or religious status. Moreover, modern terrorism does not discriminate in terms of location. The location attacked might be New York, Washington, Madrid, London or Bali; it could as easily be Cincinnati, Hawaii, Sydney, Warsaw or Montreal.

Terrorists' suicidal attacks heighten risk by nullifying many types of general defensive tactics, including preventative environmental "target-hardening" techniques. General defensive tactics may rely on sending a message to a potential perpetrator that he or she will be caught in the act, will be tracked down quickly or will be injured if he or she penetrates a defensive perimeter. Suicidal motivation renders these messages irrelevant. In an area that permits liberty of movement, general passive defensive tactics cannot guarantee freedom from attack. All pedestrians cannot be blocked or contained. Active prevention, not static defence, is required. If pre-emptive defensive tactics cannot be relied on, information is needed to identify and stop potential bombers.[4]

Because of the risks posed by terrorism, counter-terrorist operations must emphasize prevention over reaction. Better than an after-the-fact response is preventing the terrorist event from occurring at all. Early intervention reduces the risk borne by individuals. The

---

3 Report of the Technology and Privacy Advisory Committee, *Safeguarding Privacy in the Fight Against Terrorism* (Washington, D.C., March 2004), online: Center for Democracy and Technology, <www.cdt.org/security/usapatriot/20040300tapac.pdf> at 11 [TAPAC]. Nonetheless, the probabilities of a chemical, biological or nuclear terrorist attack should not be exaggerated. For a practical assessment of the probabilities of these sorts of threats see Linda Rothstein, Catherine Auer & Jonas Siegel, "Rethinking Doomsday" (November/December 2004) *Bulletin of the Atomic Scientists* at 36-41, 44-47, 73, online: Bulletin of the Atomic Scientists <www.thebulletin.org/article.php?art_ofn=nd04rothstein>. Lower-tech attacks and attacks on soft targets such as nuclear power plants and, generally, our electrical infrastructure, remain serious concerns (Rothstein, *ibid.*); and Gregory S. McNeal, "The Terrorist and the Grid" *New York Times* (13 August 2005), online: NYtimes.com <www.nytimes.com/2005/08/13/opinion/13mcneal.html?ex=1281585600&en=da3009d44b59a224&ei=5090&partner=rssuserland&emc=rss>.

4 TAPAC, *ibid.*; Bruce Hoffman, "The Logic of Suicide Terrorism" *The Atlantic Monthly* (June 2003), online: The Atlantic Online <www.theatlantic.com/doc/200306/hoffman>.

risks are too great to allow individuals to bear the risks themselves. Prevention, of course, is not unique to counter-terrorism. The criminal law does have a preventative aspect, and policing has always had a preventative function.[5] Yet the State, in large measure, does not attempt to stop most crimes before they occur. The emphasis of ordinary criminal justice is on reaction: the crime occurs, the State investigates and arrest and prosecution follow (or not).

The severity and distribution of the terrorist risk and the emphasis on prevention contribute to the information needs of counter-terrorist operations. While reactive operations start from the facts of a particular event and radiate out from the crime scene, preventative operations lack this starting point. They begin with minimum knowledge. It is not known exactly when an attack will occur, where an attack will occur, or who will be the targets. It is not known whether there will be more-or-less simultaneous attacks. In the words of a Markle Foundation report, "the decentralized nature of the terrorist threat thus leads to exponentially more — and widely scattered — information to process and share."[6]

The information needs of counter-terrorism are aggravated by three additional features of modern terrorism. First, modern terrorist groups tend to lack formal structure.[7] In contrast, traditional organized criminal groups, such as La Cosa Nostra, had hierarchies.[8] Their organizational structure restricted the scope of informational scans. Wiretaps could be placed at the home of a crime family boss. A social club where crime family members and associates met could be placed under surveillance. Al Qaeda and sympathetic groups lack hierarchy. Bin Laden doubtless holds a position of power and influence, but even he has been described as more of a mediator or catalyst than a "boss" of lower-level units. Following Madrid, some commentators suggested that "al Qaeda" was unified by ideology, not command structure.[9] Al Qaeda — if it "is" anything at all — is a network of individuals, relatively autonomous cells and ideas, not a hierarchy.[10] The London bombings, moreover,

---

5 I explain this more fully in "Criminal Justice and Public Health" in Bailey, Caulfield & Ries, *supra* note 2, 429 at 433-34.

6 Second Report of the Markle Foundation Task Force, *Creating a Trusted Information Network for Homeland Security* (New York City, December 2003), online: Markle Foundation <www.markletaskforce.org/reports/TFNS_Report2_Master.pdf> at 14 [Second Markle Report]. Because attacks cannot be easily predicted and because terrorists give little or no advance warning, information processing must be done quickly to avert threats (*ibid.* at 14); TAPAC, *supra* note 3 at 11.

7 Lack of formal structure is not the same as lack of preparation or *ad hoc* organization; video surveillance showed the London bombers making a practice run; the September 11 terrorists were highly organized.

8 See Joseph F. O'Brien & Andris Kurins, *Boss of Bosses: The Fall of the Godfather — the FBI and Paul Castellano* (New York: Simon & Schuster, 1991); Jerry Capeci & Gene Mustain, *Gotti: Rise and Fall* (Toronto: Penguin Books Canada, 1996); Ralph Blumenthal, *The Gotti Tapes* (Toronto: Random House Canada, 1992).

9 Elaine Sciolino, "Europe Meets the New Face of Terrorism" *New York Times* (1 August 2005), online: NYTimes.com, <www.nytimes.com/2005/08/01/international/europe/01threat.html?ei=5088&en=ce7493e14d834cb3&ex=1280548800&adxnnl=1&partner=rssnyt&emc=rss&adxnnlx=1129676838-dUTqRMLrminf6FnoKnOtYg>; see Bruce Hoffman, "What Can We Learn from the Terrorists?" (Washington, D.C., 2004) *Global Agenda*, online: Rand Corporation <www.rand.org/commentary/011604GA/learn_from_al-qaeda.pdf>.

10 John Poindexter, Robert Popp & Brian Sharkey, Defense Advanced Research Projects Agency (DARPA), Hicks and Associates Inc., "Total Information Awareness (TIA)" (March 2003), online: Institute of Electrical and Electronics Engineers (IEEE XPlore) <http://ieeexplore.ieee.org.login.ezproxy.library.ualberta.ca/iel5/8735/27673/01235220.pdf?tp=&arnumber=1235220&isnumber=27673> at 1 [TIA]; Janice Gross Stein, "Network Wars" in Ronald J. Daniels, Patrick Macklem & Kent

taught us that the terrorist personnel are not other than us, not "from away." They may be people who have lived with us all their lives. Thus, we know neither the target nor the terrorist. We do not know where to look. To be safe, we must look everywhere.

Second, terrorists use modern communication tools.[11] La Cosa Nostra belonged to an earlier technological age, when meetings and telephone calls were the main means of transmitting information. Al Qaeda elements may communicate by e-mail or through the use of Internet websites. They may communicate from anywhere, to anywhere, routed through anywhere. Again, to be safe, we must look everywhere.

Third, modern terrorist groups tend not to be amenable to penetration and tend not to produce informers, reducing the take of human intelligence. Penetrating groups like al Qaeda with law enforcement or intelligence personnel is difficult, if only because of the shortage of agents with the requisite linguistic skills. In contrast, law enforcement has had some notable successes in planting undercover operatives in traditional organized crime groups.[12] Similarly, law enforcement has had some notable successes in turning members of traditional organized crime groups into informers.[13] If individuals' primary motive is salvation or matrydom (as is the case for at least front-line al Qaeda terrorists) as opposed to self-interest (as is the case for most members of organized crime) finding the leverage or angle to motivate treachery against their comrades is difficult.[14] If informers or agents are not available, information must be sought from other sources.

Do the foregoing considerations compel the conclusion that the realities of the contemporary terrorist threat create qualitatively new information demands? Some might not accept this conclusion. Instead, they might argue that the appetite for information is not a function of terrorism, but of the public-private apparatus of counter-terrorism and of surveillance society. Some persons have vested interests in selling information technology tools and in advancing their careers, and therefore in maintaining terrorism hysteria. On a level deeper than mere self-interest, our technologies, our expectations and our conceptual framework fabricate demands for ever more complete information, regardless of our actual

---

Roach, eds., *The Security of Freedom: Essays on Canada's Anti-Terrorism Bill* (Toronto: University of Toronto Press, 2001) 73; Andre DeMarce, "Qaeda Link Seen in Bali Suicide Bombings" (12 October 2005), online: Terrorism Research Center <www.terrorism.com/modules.php?op=modload&name=WarReports&file=index&view=691>; "But it was the amorphous quality of the network — fluid, rapidly evolving — that made it so difficult to combat. You couldn't infiltrate it. You couldn't listen in on it, except by accident. You couldn't locate it geographically because it wasn't in any one place. In truth, the network represented a radically new kind of opponent, and one that required radically new techniques to combat it" (Michael Crichton, *State of Fear* (New York: Avon Books, 2004) at 283).

11 TAPAC, *supra* note 3 at 12.

12 Rick Cowan & Douglas Century, *Takedown: the Fall of the Last Mafia Empire* (New York: Berkeley Books, 2002); Joseph D. Pistone (with Richard Woodley), *Donnie Brasco: My Undercover Life in the Mafia* (Markham, Ont.: Penguin Books Canada, 1987).

13 Peter Maas, *The Valachi Papers* (New York: HarperCollins, 1968); *Underboss: Sammy the Bull Gravano's Story of Life in the Mafia* (New York: HarperCollins, 1997).

14 Although not impossible. On occasion individuals connected with terrorist groups will provide information to the State about their colleagues. See for example William K. Rashbaum & Benjamin Weiser, "Scheme by 2 to Train Terrorists is Outlined in U.S. Court Papers" *New York Times* (31 May 2005), online: SITE Institute <www.siteinstitute.org/bin/articles.cgi?ID=news95305&Category=news&Subcategory=0>.

needs and whether additional information might serve those needs better.[15] I can neither explore nor evaluate these criticisms here. I would only observe that the September 11, Madrid, London and Bali attacks did occur. They were not manufactured, they were not television fiction.[16] We have no reason to think that further attacks will not occur. We have no reason to think that the terrorists will seek to minimize casualties or that they will forego weapons of mass destruction or the production of mass casualties. We have no reason to think that we are safe. We have every reason to take all personally, legally and politically appropriate steps to defend ourselves against further attacks. One might be forgiven for hypothesizing that part of our defence could entail better information management.

Even if one does accept that counter-terrorism does have legitimate information needs, one might argue that no further types of information need to be gathered. The problem is not to obtain more information, but to share it better.[17] In the pithy words of the American Civil Liberties Union (ACLU), "[y]ou don't find a needle in a haystack by bringing in more hay."[18] The need for better information sharing has been emphasized by the 9/11 Commission[19] and in the Second Markle Report.[20] Yet while I believe that we must accept the conclusion that the counter-terrorism communities must share information better, this conclusion does not entail (as the ACLU suggests) that the response to terrorism would not be enhanced through the acquisition of further information:

> The decentralized nature of the terrorist threat ... leads to exponentially more — and widely scattered — information to process and share. The reality is that every hour of every day, our intelligence and law enforcement agencies, health care providers, private companies, and numerous other players receive information that might be relevant to uncovering a terrorist plot and preventing an attack.[21]

At this point data mining makes its entrance, as a means to provide that "further information."

---

15 Richard V. Ericson & Kevin D. Haggerty, *Policing the Risk Society* (Toronto: University of Toronto Press, 1997) at 8; Kevin D. Haggerty & Amber Gazso, "Seeing beyond the ruins: Surveillance as a Response to Terrorist Threats" (2005) 30 Canadian Journal of Sociology 169 at 182-83; Priscilla M. Regan, "Privacy as a Common Good in the Digital World" (Paper presented at the Annual Meeting of the American Political Science Association, August 1999), online: William Ball, Department of Political Science, The College of New Jersey <http://ball.tcnj.edu/pols291/readings/001004 reganprisc.pdf>.

16 See Christopher Norris' response to Jean Baudrillard's perspective on the first Gulf War: "Baudrillard and the War that Never Happened" in Christopher Norris, *Uncritical Theory: Postmodernism, Intellectuals and the Gulf War* (Amherst: University of Massachusetts Press, 1992) 11, "Postscript," *ibid.* at 192.

17 Ann Cavoukian, *National Security in a Post-9/11 World: The Rise of Surveillance ... the Demise of Privacy?* (May 2003), online: Information and Privacy Commissioner of Ontario <www.ipc.on.ca/userfiles/page_attachments/nat-sec.pdf> at 26 [Cavoukian, "National Security"].

18 American Civil Liberties Union, *Q & A on the Pentagon's "Total Information Awareness" Program* (20 April 2003), online: American Civil Liberties Union <www.aclu.org/Privacy/Privacy.cfm?ID=13652&c=130> at para. 12 [Q & A on TIA].

19 U.S., National Commission on Terrorist Attacks, *The 9/11 Commission Report: Final Report of The National Commission on Terrorist Attacks upon the United States*, (New York: W.W. Norton, 2004) at 416-19.

20 *Supra* note 6 at 14.

21 *Ibid.*

III. DATA MINING AS AN INTELLIGENCE SOURCE

The view that data mining is a useful source of information for counter-terrorist operations is based on a series of facts and a hypothesis. The facts are as follows:

The late twentieth and early twenty-first centuries have been marked by significant advances in computer-related technology. There has been massive growth in communication, in networking or connectedness and in data management tools; in caching or data storage (its cost is decreasing); and in computing power, which is doubling every 18 to 24 months (and its cost is decreasing, too).[22]

As a result of these advances, vast electronic holdings of information about individuals and their transactions have been established by the private sector and government.[23] Business and government have always maintained records — neither is conceivable without record-keeping techniques and record making. Our lifetimes, however, have witnessed the growth of electronic records sometimes supplementing, sometimes supplanting paper records.

Private sector records include information publicly available on the Internet, whether uploaded by subject individuals (on personal websites), or uploaded by third parties (for example, in accounts of public events) and commercial records, such as retailer records, which could include what was purchased, how much was purchased or how many items were purchased, when items were purchased, price information, the method of payment (cash/debit/charge), where the items were purchased (store location) or where the product (electricity or power) was used; or financial records, which could include aggregated information concerning credit or debit card use. Private sector records also include video rental, library, car rental and flight-booking records.

Some public organizations that are not part of the apparatus of government, such as universities,[24] hold information such as individuals' personal contact information, emergency contact information, educational records, library borrowing records, swipe card use records (respecting access to buildings or parts of buildings), health information or criminal records (for students or staff doing work with protected persons).

The State, in its federal, provincial and municipal manifestations, maintains a tremendous number of records about individuals, such as records respecting driver's licence information, including name, phone number, address and physical details; vehicle registration records, including vehicle identification numbers, vehicle make, model and year information;

---

22 Markle Foundation Task Force, *Protecting America's Freedom in the Information Age* (October 2002), online: Markle Foundation <www.markletaskforce.org/documents/Markle_Full_Report.pdf> at 12 [First Markle Report]; Jeffrey W. Seifert, "Data Mining: An Overview," Congressional Research Service Report for Congress (updated 7 June 2005), online: Federation of American Scientists <www.fas.org/sgp/crs/intel/RL31798.pdf> at 2.

23 The growth in electronic record keeping has led to the claim that "a defining characteristic of the information age is 'the disappearance of disappearance,'" the elimination of practical obscurity (Michael Levi & David S. Wall, "Technologies, Security, and Privacy in the Post-9/11 European Information Society" (2004) 31 J.L. & Soc'y 194 at 206); Kevin D. Haggerty & Richard V. Ericson, "The surveillant assemblage" (2000) 51 British Journal of Sociology 605 at 619.

24 *McKinney v. University of Guelph*, [1990] 3 S.C.R. 229 at 268ff, La Forest J. [*McKinney*].

municipal tax records, including street address and home value; income tax and goods and services tax records; court records, including information about charges, judicial interim release dispositions, trial results and sentences; immigration records, including information respecting entry into and exit from country; visa status; and forensic and other records gathered through law enforcement processes, including fingerprint records and DNA records.

Information has not merely been stored in these records. It has been organized and analyzed. Because of the "volume challenge," the high "data ingestion rates," of modern data storage, forms of automated analysis are required.[25] "Data mining" appears here.[26]

The term "data mining" has a broad vernacular use, meaning "searches of one or more electronic databases of information."[27] In this broad sense, data mining includes two types of procedures — query-based information retrieval and automated pattern discovery. Query-based information retrieval reveals information that is already expressly or explicitly in a database or set of databases.[28] The queries or hypotheses on which analysis is based are developed by users.[29] This type of analysis, K.A. Taipale has commented, may be "slow, expensive and highly subjective."[30] Automated pattern discovery, or "data mining," in its more specialized sense, is the "non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data."[31] It differs from query-based information retrieval in that the information revealed is not expressly in the database or databases analyzed. The terms of analysis are not initially dictated by users. The search patterns are initially detected not by humans, but by algorithms applied to training data. The detected patterns were previously unknown to human users, and are, in that sense, "new."[32] The detected patterns are then tested on fresh data. The resulting patterns are applied to other data sets to draw inferences from that data or to make predictions based on that data.[33] Human input is not absent from the data mining process. While automated discovery may yield patterns, the value or significance of patterns must be evaluated by skilled technical and

---

25 David B. Cousins, Doyle J. Weishar & J. Brian Sharkey, "Intelligence Collection for Counter Terrorism in Massive Information Content" (23 October 2003), 2004 IEEE Aerospace Conference Proceedings, vol. 5 at 3273; Colleen McCue, Emily S. Stone & Teresa P. Gooch, "Data Mining and Value-Added Analysis" (November 2003) FBI Law Enforcement Bulletin, online: <www.fbi.gov/publications/leb/2003/nov03leb.pdf> at 1. Data volume includes not only size or number of records, but dimensionality or the number of fields of data recorded: K.A. Taipale, "Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data" (2003) 5:2 Colum. Sci. Tech. L. Rev. 1 at 14.

26 What follows is not a technical description — which, I confess, would be beyond my powers — but a non-technical description that should suffice for the purposes of later legal analysis.

27 See TAPAC, *supra* note 3 at viii (asterisked footnote); Taipale, *supra* note 25 at 6, n. 11.

28 Lee Tien, "Privacy, Technology and Data Mining" (2004) 30 Ohio N. U. L. Rev. 389 at 393-94; Seifert, *supra* note 22 at 1.

29 Seifert, *ibid.*

30 *Supra* note 25 at 14 [footnote omitted].

31 Usama M. Fayyad, Gregory Piatetsky-Shapiro & Padhraic Smyth, "From Data Mining to Knowledge Discovery: An Overview" in Usama M. Fayyad *et al.*, eds., *Advances in Knowledge Discovery and Data Mining* (Menlo Park, Cal.: AAAI Press/MIT press, 1996) 1 at 6; Taipale, *ibid.* at 22, 23, 28.

32 Taipale, *ibid.* at 22; Ann Cavoukian, "Data Mining: Staking a Claim on Your Privacy" (Toronto: Information and Privacy Commissioner of Ontario, 1998), online: <www.ipc.on.ca/docs/datamine.pdf> at 5 [Cavoukian, "Data Mining"]; Jay Stanley, "ACLU Statement on Terrorist Information Awareness Before the Department of Defense Technology and Privacy Advisory Committee" (19 June 2003), online: American Civil Liberties Union <www.aclu.org/safefree/general/16854leg20030619.html> [Stanley, "ACLU Statement"].

33 Tien, *supra* note 28 at 394; Cavoukian, "Data Mining," *ibid.* at 4; Taipale, *ibid.* at 21.

analytical specialists.[34] Humans must decide whether the application of the pattern meets an acceptable "confidence interval" or has an acceptable error rate.[35]

Data mining is one step in a "knowledge discovery" process, which involves pre-processing, data mining and post-processing. Pre-processing includes data collection, selection and warehousing. Warehousing may be actual, as when a party makes its own copy of data; or it may be virtual, as when a party has network access to "legacy" databases held by other custodians.[36] Pre-processing also includes data cleansing and transformation, to eliminate "noise" in data, to deal with missing data and generally to ensure that data is in a form that will permit analysis.[37] Post-processing includes the interpretation and evaluations of patterns, and decision making and action.[38]

The private sector has used data mining for a variety of purposes. Credit-granting firms use data mining to assess credit risk and to detect fraud.[39] Retailers use it to aid in product selection and placement and for coupon offers.[40] Medical and pharmaceutical firms use it to improve the effectiveness of products and treatments.[41] Private sector data mining has, in fact, been identified as a significant privacy concern for Canadians.[42]

The public sector has used data mining to detect fraud and waste and to improve unit performance.[43] The public sector has also used data mining (in the broad sense) to aid law enforcement. These initiatives generally involve traditional query-based searches, run through extensive and previously separate data sets. For example, in the Multi-State Anti-Terrorism Information Exchange (MATRIX) project, the databases searched included State-owned databases and open public sources.[44] The COPLINK project, in its "Connect" aspect,

---

34 Seifert, *supra* note 22 at 3; Taipale, *ibid.* at 24.

35 Taipale, *ibid.* at 31.

36 Taipale, *ibid.* at 25, 26, 42, 43.

37 Cavoukian, "Data Mining," *supra* note 32 at 4-5; Judith A. Miller, "Intelligence Collection and Civil Liberties: Technology and Privacy in Intelligence Collection" (30 October 2003), online: Federation of American Scientists <www.fas.org/irp/congress/2003_hr/103003miller.pdf> at 2.

38 Taipale, *supra* note 25 at 25, 30; Seifert, *supra* note 22 at 2.

39 McCue, Stone & Gooch, *supra* note 25 at 2; Seifert, *ibid.* at 3; Jay Stanley, "Is the Threat from 'Total Awareness' Overblown?" (18 December 2002), online: American Civil Liberties Union <www.aclu.org/Privacy/Privacy.cfm?ID=11501&c=130> [Stanley, "Threat"]; Cousins, Weishar & Sharkey, *supra* note 25 at 5; Cavoukian, "Data Mining," *supra* note 32 at 4.

40 McCue, Stone & Gooch, *ibid.* at 2; Seifert, *ibid.* at 3-4.

41 Seifert, *ibid.* at 3.

42 Cavoukian, "Data Mining," *supra* note 32 at 2; Renee M. Pomerance, "Redefining Privacy in the Face of New Technologies: Data Mining and the Threat to 'Inviolate Personality'" (2005) 9 Can. Crim. L. Rev. 273 at 284 [Pomerance, "Redefining"]; Arthur J. Cockfield, "Who Watches the Watchers? A Law and Technology Perspective on Government and Private Sector Surveillance" (2003) 29 Queen's L.J. 364 at 375 [Cockfield, "Watchers"].

43 Seifert, *supra* note 22 at 4; Cavoukian, "Data Mining," *ibid.* at 4.

44 William J. Krouse, "The Multi-State Anti-Terrorism Information Exchange (MATRIX) Pilot Project" Congressional Research Service Report for Congress (18 August 2004), online: Federation of American Scientists <www.fas.org/irp/crs/RL32536.pdf>. MATRIX-accessed databases included criminal history, Department of Corrections, sexual offender registry, driver's licence and motor vehicle registry databases, as well as bankruptcy, federal aviation, domain names and professional licence registries ("Seisint FACTS™ For The MATRIX Project" (29 September 2003), online: American Civil Liberties Union <www.aclu.org/FilesPDFs/seisint_facts_83.pdf> at 12-14; "Frequently Asked Questions [about MATRIX]," online: American Civil Liberties Union <www.aclu.org/FilesPDFs/official%20 matrix%20faq.pdf>). The MATRIX program was terminated on 15 April 2005 (American Civil Liberties

uses information and knowledge management system technologies to capture, access, analyze, visualize and share "law enforcement-related information" — information available to the police, but which had been scattered across different information sources.[45]

These are the facts. The hypothesis is this: if terrorists intend to attack the U.S. or Canada, their operatives will engage in transactions. Those transactions will produce electronic records and those transactions will leave a "signature in the information space."[46] The terrorist signature can be discerned by data mining. Identification of the terrorist signature will permit the early detection of terrorist activity and support preventative counter-terrorism measures.[47] Furthermore, because the patterns relied on — particularly those that are computer-detected — will be non-obvious, terrorists will find it difficult to engage in counter-surveillance tactics. They will not know the patterns to avoid. Unless cash-only transactions were engaged in, a transactional signature would be unavoidable, even if that signature were distributed amongst proxies.

The hypothesis assumes that the "information space" shall be very large. Precursor acts are likely to appear entirely legitimate, viewed in isolation. Those precursor acts could take place in virtually any area of electronically recorded transactional activity. It may be necessary to process significant quantities of the information space to find signatures of interest.[48]

The identification of terrorist signatures could be accomplished through traditional query-based searches. Analysts would develop patterns or models, test them against historical data and use them to make predictions in relation to new data. The patterns or models could be based on the study of past attacks, or on wargaming or "red teaming," in which analysts think through terrorist attack possibilities.[49] Alternatively, patterns could be developed through automated data mining processes.

The hypothesis was pursued, most famously or infamously, by the 2002 Total Information Awareness (later Terrorist Information Awareness) project (TIA) of the Information Awareness Office of the Defense Advanced Research Projects Agency (popularly, DARPA). Following negative publicity, public outcry and bungled public relations efforts, all related

---

Union, "Second Major Snoop Program Shut Down by Privacy Opposition" (15 April 2005), online: American Civil Liberties Union <www.aclu.org/privacy/spying/15324prs20050415.html>).

45 Hsinchun Chen *et al.* "COPLINK: Managing Law Enforcement Data and Knowledge" (January 2003) 46:1 Communications of the ACM 28, online: <http://forum1.knowledgeboard.com/download/1798/CACMpdf.pdf>.

46 TIA, *supra* note 10 at 2, 3. Note that the data mining hypothesis does not require that an individual have a law enforcement "record" to be identifiable as a terrorist suspect — identification is based on transactional pattern recognition.

47 TIA, *ibid.* at 2.

48 TIA, *ibid.* at 3; Stanley, "Threat," *supra* note 39; R. Popp *et al.*, "Countering Terrorism Through Information Technology" (March 2004) 47:3 Communications of the ACM 36, online: Association for Computing Machinery <http://delivery.acm.org/10.1145/980000/971642/p36-popp.html?key1=971642&key2=8174789311&coll=GUIDE&dl=ACM&CFID=68419960&CFTOKEN=12760131>.

49 Stanley, "ACLU Statement," *supra* note 32.

to the adverse privacy impacts of TIA, Congress prohibited funding for the program.[50] The Computer Assisted Passenger Prescreening System, which went through two iterations and now carries the name Secure Flight, also used data mining elements, and also has had its funding blocked because of privacy concerns.[51] While the tales of these projects are fascinating, they lie outside the scope of this article. Of greater moment is that these are not the only data mining projects — others are ongoing. Although TIA's funding was terminated, DARPA was permitted to pursue related classified projects.[52] The U.S. Department of Defense is pursuing several data mining projects, as is the Advanced Research and Development Activity Center operated out of the National Security Agency.[53] The Central Intelligence Agency (CIA) is reported to be pursuing some data mining projects.[54] The U.S. General Accounting Office has reported that 52 U.S. federal agencies "are using or planning to use data mining, 'factual data analysis,' or 'predictive analysis,' in some 199 different efforts," of which at least 29 relate to detecting terrorist or criminal activities.[55] Less spectacularly, but perhaps more effectively, the U.S. Department of Justice and the Federal Bureau of Investigation (FBI) have purchased and used information from ChoicePoint, a U.S. data mining firm.[56] According to the Technology and Privacy Advisory Committee (TAPAC),[57] "TIA was not the tip of the iceberg, but rather one small specimen in a sea of icebergs."[58]

One might observe that the terrorism-data mining hypothesis appears to have been of interest largely in the U.S., and therefore suggest that whatever might be the benefits or risks of counter-terrorist data mining projects, none of this concerns Canadians. One might add the observation that the Canadian federal government has not emphasized data mining initiatives in its counter-terrorism program. Data mining does not figure, for example, in the *Lawful Access* consultation paper.[59] Data mining, however, is coming to a database near you. Canada does share information (such as flight information) with other governments, including the U.S. government. Canadian information, then, could be mined along with other

---

50 Seifert, *supra* note 22 at 5-7; TAPAC, *supra* note 3 at viii; Cockfield, "Watchers," *supra* note 42 at 389; Gina Marie Stevens, "Privacy: Total Information Awareness Programs and Related Information Access, Collection, and Protection Laws" (21 March 2003), Congressional Research Service Report for Congress, online: Federation of American Scientists <www.fas.org/irp/crs/RL31730.pdf> at 10.

51 Jay Stanley & Barry Steinhardt, "Bigger Monster, Weaker Chains: The Growth of American Surveillance Society" (New York: ACLU, 2003), online: American Civil Liberties Union <www.aclu.org/privacy/gen/15162pub20030115.html>.

52 TAPAC, *supra* note 3 at vii; Cavoukian, "National Security," *supra* note 17 at 6; First Markle Report, *supra* note 22.

53 TAPAC, *ibid.* at viii, 2-3; First Markle Report, *ibid.* at 10.

54 First Markle Report, *ibid.*

55 Stephen E. Fienberg, "Homeland Insecurity: Datamining, Terrorism Detection, and Confidentiality," online: National Institute of Statistical Sciences <www.niss.org/dgii/TR/Fienberg-ISI-Confidentiality-Terrorism-12-3-04.pdf> at 1.

56 Stanley & Steinhart, *supra* note 51 at 8; Stevens, *supra* note 50 at 4.

57 Appointed by Secretary of Defence Donald Rumsfed in February 2003 "to examine the use of 'advanced information technologies to identify terrorists before they act'" (TAPAC, *supra* note 3 at vii).

58 *Ibid.* at 5.

59 Department of Justice, Industry Canada & Solicitor General Canada, *Lawful Access — Consultation Document* (25 August 2002), online: Justice Canada <http://canada.justice.gc.ca/en/cons/la_al/consultation_index.html>; Nevis Consulting Group Inc., General Editor, *Summary of Submissions to the Lawful Access Consultation* (28 April 2003), online: Justice Canada <http://canada.justice.gc.ca/en/cons/la_al/summary/index.html>; Pomerance, "Redefining," *supra* note 42 at 286.

information.[60] Data mining capabilities are being built into computer, database and networking products.[61] David Loukidelis, the Information and Privacy Commissioner for British Columbia, made the following comments:

> [I]t would be naïve to think that Canadian governments will ignore for long the ever-richer trove of digital personal information that exists in the private sector. I would not be surprised, for example, if the recently announced Canadian no-fly list is populated in part using commercially-acquired data. Overall, as digital databases proliferate, become more comprehensive and become life-long, it will be very difficult to resist using this information.[62]

Resisting the impulse to do what can be done will indeed be difficult, particularly since the technological imperative appears to be that if it can be done it should be done.[63] Put more positively, "[i]nnovation in technology is an important part of our nation's competitive edge against terrorist organizations."[64] If information technology can help to prevent a terrorist attack in Toronto, Montreal or Vancouver, it would be irresponsible for Canada not to even consider its use.

If we do consider the use of data mining and the counter-terrorism benefits that may flow from it, we must also consider data mining's potential costs. I turn to that inquiry next.

IV. THE RISKS OF DATA MINING

Data mining, at root, involves invasions of privacy. By "privacy" I mean privacy in relation to personal information:

> As the Task Force put it...: "This notion of privacy derives from the assumption that all information about a person is in a fundamental way his own, for him to communicate or retain for himself as he sees fit."[65]

Data mining seeks out "personal information" or information about identifiable individuals.[66] If information could not be linked to identifiable individuals, it would not assist in identifying terrorists and preventing terrorist attacks. Individuals provided the information to private,

---

60 Pomerance, "Redefining," *ibid.* at 275-76. This issue has been a concern of the B.C. Information and Privacy Commissioner. See the documents at <www.oipc.bc.ca/sector_public/usa_patriot_act/patriot_act_resources.htm>.

61 Taipale, *supra* note 25 at 15.

62 David Loukidelis, "The National Security Imperative — Is the Private Sector Becoming An Arm of the State?" (Speech delivered to the Canadian Bar Association Annual Meeting, Vancouver, B.C., 16 August 2005), online: Office of the Information and Privacy Commissioner of British Columbia <www.oipcbc.org/pdfs/Speeches/CBA-AGMSpeech.pdf> at 5. See also David Loukidelis, "Information Technology, National Security & Privacy Protection" (Paper presented to Annual Conference, Canadian Institute for the Administration of Justice, 29-30 September 2005), online: Office of the Information and Privacy Commissioner of British Columbia <www.oipcbc.org/publications/speeches_presentations/CIAJSpeech(RevisedFinal)(Oct3-2005).pdf> at 5 [Loukidelis, "Information Technology"]; Cockfield, "Watchers," *supra* note 42 at 368, 385, 391.

63 "Technology drives uses. Where there is a way there is a will"(Ericson & Haggerty, *supra* note 15 at 34).

64 Second Markle Report, *supra* note 6 at 10; TAPAC, *supra* note 3 at viii.

65 *R. v. Dyment*, [1988] 2 S.C.R. 417, La Forest J. at 429 [*Dyment*].

66 *Personal Information Protection and Electronic Documents Act*, S.C. 2000, c. 5, s. 2(1) [*PIPEDA*]; *Personal Information Protection Act*, S.A. 2003, c. P-6.5, s. 1(j) [*PIPA*].

public non-governmental, or governmental custodians for, usually, non-counter-terrorism or law enforcement purposes. Generally, individuals are entitled to assume that the information will be used and disclosed only for the purposes for which it was collected.[67] Without consent and without notification to the individuals, data mining acquires the information, analyzes it and produces information that may be relied upon by the State for counter-terrorism purposes.

The justification of these privacy limitations by data mining is utilitarian or consequentialist, and is based on the potential for data mining to deliver high-grade intelligence that can be used to stop terrorists before they can strike and to bring terrorists to justice. A consequentialist justification of data mining faces three main types of challenge. First, data mining must be founded on a reasonable judgement that it will actually deliver the benefits promised. Second, data mining must not itself produce risks that outweigh any benefits it may deliver. Third, data mining must limit other valuable interests (such as privacy) only to the degree necessary to achieve its objectives. Any additional limitation would be, from the standpoint of achieving the objectives, superfluous, excessive or unnecessary — and, to that extent, unjustified. This part shall consider the first two types of challenges — (A) the risks that data mining cannot deliver on its promises, and (B) the social, political and personal risks created by data mining. The third type of challenge shall be considered in the next part.

A. THE RISKS OF NON-DELIVERY

Data mining is beset by some inherent weaknesses that undermine its ability to provide accurate and useful intelligence or, alternatively, that dispose it to provide erroneous information. These weaknesses stem from data mining's reliance on recorded information and from its use of profiling techniques to extract information from records.

1. PROBLEMS WITH DATA

Data mining faces the perennial database problem: garbage in, garbage out. The data accessed for mining could suffer from one or more of four weaknesses.[68] The data may be incomplete, missing fields or records. It may be incorrect, involving non-standard codes, incorrect calculations, duplication, linkage to the wrong individual or other mistaken inputting; the initial information provided may have been incorrect. It may be incomprehensible, involving (for example) bad formatting or the inclusion of multiple fields in one field. It may be inconsistent, involving overlapping codes or code meanings that change over time.[69] Furthermore, even if data is recorded accurately and properly, different databases may use different formatting standards, making data sharing or the

---

67 Substantiating this claim is the burden of Part V below.

68 Terrence A. Maxwell, "Information Policy, Data Mining, and National Security: False Positives and Unidentified Negatives" (Paper presented to the 38th Hawaii International Conference on System Sciences, 2005), online: IEEE Computer Society <http://csdl2.computer.org/comp/proceedings/hicss/2005/2268/05/22680134c.pdf> at 4; Seifert, *supra* note 22 at 11, 12; TAPAC, *supra* note 3 at 37-38; Loukidelis, "Information Technology," *supra* note 62 at 7.

69 Taipale does caution against exaggerating the problem of "dirty data." Statistical methods are being developed to deal with "dirty data" issues (*supra* note 25 at 68).

"interoperability" of different databases difficult.[70] Data from different sources must be "scrubbed" or "cleaned" to permit integration and comparison.[71]

2. PROBLEMS WITH PROFILING

Pattern-based data mining could rely on profiles produced by individuals or on computer-generated profiles. Profiling generally, human-generated profiles, and computer-generated profiles are each problematic.

a. Profiling Generally

Profiling is a normal human technique for drawing conclusions about individuals, things, or events. I will consider the profiling of individuals only. Profiling has the following logic: we begin with a proposition along the lines of "If an individual X satisfies criteria A, B, C ... (or some number of these criteria), then X has the characteristic Y." We then consider a particular individual and determine whether he or she matches the requisite criteria. If so, we conclude that the individual has the character in question. The characteristic could be, for example, dangerousness, trustworthiness or competence. The linkage between the criteria and the characteristic may be based on personal experience or shared experience. Essentially, the linkage relies on observed correlations between the criteria and individuals who actually possess the characteristic. The satisfaction of the criteria provides a basis for prediction that an individual possesses the characteristic. Of course, the satisfaction of the criteria is no guarantee or is not absolute proof that the individual has the characteristic. The satisfaction of criteria is merely some more or less convincing evidence that the individual indeed has the characteristic. There may be false positives — individuals who exhibit the criteria, but do not have the characteristic (for example, a rogue may feign trustworthiness to gain and exploit trust). There may be false negatives — individuals who do not exhibit the criteria but do have the characteristic (for example, an individual who had given no indication of bravery, but was "stand-up" in the face of hostility). Furthermore, the criteria we believe are correlated with the characteristic in fact may not be, or may not be the only set of criteria that could be used to predict the characteristic. Our criteria may be accurate enough, but we may misapply them — for example, we may rely on insufficient evidence (had we known more about the individual, we would have come to a different conclusion). Hence, profiling has error rates (false positives and false negatives), depends on the reliability of our criteria, and depends on the proper application of those criteria.

The respectability of profiling is confirmed by its use in criminal litigation. For example, an accused may seek to tender good character evidence to raise a doubt about whether he or she committed a crime. In his or her testimony, the accused might relate particular good works he or she has accomplished. In effect, the accused is providing evidence to show that he or she meets the criteria of the good character trait (for instance, honesty or non-violence); from the inference that the accused has that character, the inference may be drawn that the accused is not the sort of individual who would have committed the offence, and so the

---

[70] Seifert, *supra* note 22 at 11.

[71] Maxwell, *supra* note 68 at 3.

inference may be drawn that the accused did not commit the offence.[72] An accused may tender profile evidence respecting potential perpetrators to raise a doubt about whether he or she committed a crime. The accused might call an expert to testify that the crime in question could only be committed by an individual with a particular psychiatric condition; that condition is correlated with a set of criteria; and the accused does not satisfy those criteria. Hence, either some third party committed the alleged act, or it did not occur at all.[73] "Similar fact" cases relating to the identity of perpetrators also involve the use of profile evidence tendered by the Crown. The overall Crown strategy is to show that the accused committed the offence charged — and that the offence was not committed by a third party or the allegations were not just made up — because the accused has been linked to very similar offence circumstances: it would defy common sense, it would be beyond mere coincidence, for the accused to be "correlated" with the same sorts of criminal conduct on multiple occasions and not be the perpetrator. The Supreme Court has made profiling an explicit step in the evidential process. To rely on the evidence of other criminal acts, the Crown must establish that the other acts and the acts relevant to the offence charged were likely committed by the same perpetrator (whoever that might be).[74] That is, the profile of both sets of acts is such that a reasonable person would conclude that the acts were the products of a single individual. In all three types of criminal cases, the profile evidence is "circumstantial" at best and not determinative. Its probative value depends on the reliability of the profile employed[75] and on the appropriateness of the application of the profile criteria.

The error rates associated with profiling create difficulties for data mining. Data mining deals with large data sets and large populations. Even a small percentage of false positives would produce a dramatically large number of misidentified but innocent individual targets. For a U.S.-sized population of 250 million, a false positive error rate of 0.01 percent would still lead to the misidentification of 25,000 individuals.[76] Error rates will be a product not only of the profile deployed but of the underlying data. The "dirtier" the data mined, the higher the error rates.

Developing reliable profiles for terrorists may be extremely difficult. We can validate our profiles concerning good character through personal experience. Profiles of offences used in similar fact cases are based on highly restricted and manifest fact sets. Credit card companies can develop profiles to identify fraudulent uses through the collection of vast sets of transactional information involving legitimate transactions and the identification of anomalous outlier activities. The companies are also aware of actual cases of fraud and can ensure that hypothesized suspicious patterns correlate with actual fraudulent conduct.[77]

---

72 *R. v. McNamara (No. 1)* (1981), 56 C.C.C. (2d) 193 (Ont. C.A.).

73 *R. v. Mohan*, [1994] 2 S.C.R. 9 [*Mohan*]; *R. v. J.-L. J.*, [2000] 2 S.C.R. 600, 2000 SCC 51 [*J.-L.J.*].

74 *R. v. Arp*, [1998] 3 S.C.R. 339.

75 The common sense profile of the upstanding citizen, for example, has been found to be unreliable (evidence of good character has diminished weight) in the case of pedophiles, who may masquerade as good and decent individuals to prey on children (*R. v. Profit*, [1993] 3 S.C.R. 637). A common difficulty in the *Mohan* and *J.-L. J.* cases (*supra* note 73) was that the profiles asserted were not reliable, in the sense of having an appropriate scientific/empirically-tested basis. For an amusing *reductio ad absurdum* of the profiling of drug couriers, see *United States v. Hooper*, 935 F.2d 484 at 499-500 (2d Cir. 1991), Pratt J. dissenting.

76 Levi & Wall, *supra* note 23 at 207; Miller, *supra* note 37; Stanley, "ACLU Statement," *supra* note 32.

77 Cousins & Weishar, *supra* note 25 at 5; Stanley, "Threat," *supra* note 48.

Terrorist conduct is unlike credit card use. Terrorist conduct preceding an attack is likely to be designed to appear legitimate. Unlike fraud, it will not be outlier conduct in the midst of legitimate conduct, but apparently legitimate conduct in the midst of legitimate conduct. Furthermore, terrorist conduct is rare. The evidential basis for constructing models of terrorist behaviour is perilously small.[78] Finally, terrorism is fluid in its methods. While past types of attacks may be repeated (nothing succeeds like success), terrorists have large scope for innovation. Even if profiles are derived from the evidence, they may more measure the past than predict the future.[79] If we are prepared for the attack that was, we may miss the attack to come.

The argument, then, might be that in addition to developing profiles based on known terrorist conduct, profiles should be developed respecting "lower level," "frequently repeated events" correlated with terrorism, such as illegal immigration, large funds transfers, the use of front businesses and recruitment activities.[80] These sorts of lower level activities, however, may be correlated with offences other than terrorism or may be pursued for their own sake; indeed, the activities (for example, large funds transfers) may be completely legal, or even constitutionally protected (for example, activities that might be caught under the "recruitment" rubric). Because lower level activities have linkages to wide varieties of conduct, the false positive rate of profiles based on these activities must necessarily be high.

b. Human-Generated Profiles

Human-generated profiles should be the product of objective, dispassionate analysis of evidence. They may be, however, mere hunches or speculation, propped up by bias or prejudice.[81] We have already seen examples of racial profiling in the war against terror.[82] Running improperly derived profiles through digitized information does not cleanse them of error.

c. Computer-Generated Profiles

Profiles generated through computer operations should avoid some of the error, bias or prejudice of human-generated profiles.[83] A difficulty is that prejudice and bias may simply be buried deeply in code — which must issue from humans.[84] The embedding of prejudice in code may resist easy exposure and it will be shielded from scrutiny by the "mystique of science."[85] In any event, while patterns produced automatically may be valid, they may not be useful. The determination of usefulness requires human assessment.

---

78 Taipale, *supra* note 25 at 35; Q&A on TIA, *supra* note 18; Maxwell, *supra* note 68 at 7.

79 Maxwell, *ibid.* at 5.

80 Taipale, *supra* note 25 at 35.

81 Stanley, "Threat," *supra* note 48.

82 See Kent Roach, "Making Progress on Understanding and Remedying Racial Profiling" (2004) 41 Alta. L. Rev. 895; David M. Tanovich, "E-Racing Racial Profiling" (2004) 41 Alta. L. Rev. 905.

83 Taipale, *supra* note 25 at 33, n. 118.

84 Maxwell, *supra* note 68 at 7.

85 *R. v. Béland*, [1987] 2 S.C.R. 398 at 434, La Forest J. [*Béland*]; *Mohan*, *supra* note 73 at 21.

3. CONSEQUENCES OF DATA MINING INACCURACIES

Data mining's risks of inaccuracy are not merely risks that the technique will fail to provide reliable information. If data mining produces unreliable information, there will be individual costs and national security itself will be threatened. If data mining produces unreliable information, individuals will be targeted for State intervention. The intervention might range from an interview at an airport, to arrest and the laying of criminal charges, to a full-scale Special Weapons and Tactics team-supported raid on a residence. The innocent will suffer the imposition of severe burdens. Moreover, burdens are not wholly borne alone: individuals connected with a wrongly targeted individual will also suffer to varying degrees — whether through the inconvenience of a delayed flight or a flight forced to return to the ground, or through the loss on arrest of a parent or spouse. Data mining inaccuracy threatens national security by focusing attention on the wrong individuals and by diverting human and technological resources from other inquiries and investigations. While the State is distracted and the innocent are being tracked down and processed, terrorists will work unmolested.

B. RISKS DELIVERED BY DATA MINING

Data mining generates some risks of its own. Some of these risks relate to the potential misuse of the technology. Other risks are inherent in even the best-intentioned uses of data mining.

1. RISKS OF MISUSE

Data mining may be abused through its use outside of the contexts in which it might be justified and through unauthorized uses.

A frequently voiced concern is that data mining will be beset by "mission creep."[86] State (as opposed to private) data mining has been advanced to address the special needs of counter-terrorism. An extraordinary threat calls for an extraordinary tool. But if data mining is deployed in the counter-terrorism field, it is argued that State officials will inevitably use data mining for other offences, which do not involve the risks or the informational demands of counter-terrorism.

Information collected through data mining could be improperly disclosed and the data mining system could be abused if authorized users fail to protect confidentiality or if third parties obtain unauthorized access to the system or system records.[87]

Data mining capabilities could be abused by authorized users, who might perform searches and disclose results for illegitimate purposes. Regrettably, this is not a merely theoretical worry. In Michigan, police officers with access to a database used it to help friends, stalk women, threaten motorists and track estranged spouses.[88] Some members of the Edmonton

---

86 TAPAC, *supra* note 3 at 39; Miller, *supra* note 37 at 4; Loukidelis, "Information Technology," *supra* note 62 at 9.

87 TAPAC, *ibid.* at 40.

88 Q&A on TIA, *supra* note 18.

Police Service have abused the Canadian Police Information Centre (CPIC) system, running searches on individuals for illegitimate purposes.[89] The FBI was infamously involved in wide-scale snooping using the wiretap and hidden microphones, the surveillance technology of the day.[90]

2. INHERENT RISKS

Even if not misused, the methods of data mining generate social, political and personal risks. Pattern-based data mining threatens the relationship between individuals and the State. As a general rule, the State is permitted to obtain access to personal information for law enforcement purposes only on the basis of "individualized suspicion." Evidence is sought respecting a particular individual in connection with a particular offence; or, at least, information about individuals is sought in connection with a particular offence. Pattern-based data mining contemplates the State having broad access to many individuals' personal information, when there is no basis for even a suspicion of wrong-doing (since individual records may be of legal activities), to develop profiles and to run profiles against that information. Instead of the State having to justify access to personal information, it begins with irresistible access. The State intrudes in our private lives to an unprecedented degree. It becomes our silent, observant shadow. This level of intimacy with the State is contrary to our traditions of individual liberty and limitations of State power.[91]

A critical risk posed by data mining is the loss of what has been called "practical obscurity"[92] or what might be called "privacy through inefficiency." As we go through our daily lives, we know — although we may seldom think of this — that we are creating records held by various private and public bodies. We know, as well, that if we were to become criminal suspects, the authorities could and would gain access to some of these records. The records, however, are held by separate parties. They are not assembled into informational mosaics of our transactional lives. In practice, we are obscure, since no record custodian has more than a context-specific glimpse of us. We have privacy as against the State, since it does not have custody of all of our transactional information, and it must make particular inquiries with custodians to obtain information, following the applicable due process rules. Our privacy has been protected by systemic inefficiency.[93] The networked assemblage of records presupposed by data mining negates practical obscurity by itself. Our transactional records are all available for viewing. It is as if the State has an actual or virtual dossier

---

89 See *e.g.* Frank Borsato, "Report on Investigation into the Use of Personal Information," Investigation Report F2005-IR-001 (27 April 2005), online: Office of the Information and Privacy Commissioner of Alberta <www.oipc.ab.ca/ims/client/upload/F2005_IR_001.pdf>; and Florence Loyie, "Police won't reveal who initiated queries" *Edmonton Journal* (10 December 2005) B3.

90 Timothy Lynch, "Breaking the Vicious Cycle: Preserving our Liberties while Fighting Terrorism," *Policy Analysis* No. 443 (26 June 2002), online: Cato Institute <www.cato.org/pubs/pas/pa-443es.html>; Electronic Privacy Information Center, "The Attorney General's Guidelines," online: <www.epic.org/privacy/fbi/>; Richard Hack, *Puppetmaster: The Secret Life of J. Edgar Hoover* (Beverly Hills: New Millennium Press, 2004).

91 Q&A on TIA, *supra* note 18; Stanley, "ACLU Statement," *supra* note 32; Tien, *supra* note 28 at 401, 402; Stanley & Steinhardt, *supra* note 51 at 12; Jean-François Blanchette & Deborah G. Johnson, "Data retention and the panoptic society: The social benefits of forgetfulness," online: Graduate School of Education & Information Studies (UCLA) <http://polaris.gseis.ucla.edu/blanchette/papers/is.pdf> at 3.

92 Levi & Wall, *supra* note 23 at 206.

93 Stanley, "Threat," *supra* note 48; Taipale, *supra* note 25 at 58-59.

assembled on us all.[94] In Kevin D. Haggerty and Richard V. Ericson's memorable phrase, a "surveillant assemblage" is constituted that operates as a "functional entity."[95] It is as if the State is constantly viewing us. This is the Panopticon imposed through information technology, what some commentators call "dataveillance."[96]

The consequences of this loss of privacy cannot be properly predicted now. If individuals understand that they are constantly under surveillance, a "chilling" effect may occur — particularly if individuals perceive data mining to be just one of multiple State surveillance techniques. Individuals may constrain their freedoms of belief, expression or association, for fear of generating suspicious patterns.[97] On the level of commercial transactions alone, this chilling effect could be registered in relation to books purchased, flights to particular locations for particular conferences, purchases of items over the Internet, or even attendances at particular locations in a city (that might be evidenced through meal-purchase credit card records).

Individuals' perceptions of constant surveillance may have an even more profoundly corrosive effect. Priscilla M. Regan, for example, has argued that privacy is a social value, in the sense that our social relations depend on a measure of privacy to which corresponds a measure of trust.[98] I would develop this thought as follows: if we had perfect knowledge of others, we would not need to "trust" them because we could predict what they would do and we would take appropriate steps. Because others are private to us, and we are private to them, we must trust. Furthermore, because we have our privacy, because we know that others lack perfect knowledge about us, most of us are "trust-worthy," in the sense that we take responsibility for doing the right thing towards our neighbours. And because we trust and we are trust-worthy, we find ourselves able to cooperate — whether on our highways, on our sidewalks, or in our gatherings and organizations. State surveillance does not give ordinary individuals perfect knowledge, but it does expand the knowledge of the State. To that extent, the State need not "trust" us. To that extent as well, we do not know how much of our lives has been communicated — we do not know who knows what about us; and we do not know who knows what about others. While we may not have the State's surveillance-enhanced information, the fact that it exists tends to undermine the need for trust. If, because our privacy is lessened, our trust in others becomes lessened too, the nature of our lives together could become very different.

V. THE CONSTITUTIONAL STATUS OF THE DATA MINED

Given this grim catalogue of risks, one might wonder whether data mining is a technique that should be employed at all. Perhaps the Congressional decision to block funding for TIA points in the right direction. I suggest, however, that prohibition is likely to prove more dangerous in the long run than regulation. Data mining as a counter-terrorism tactic is not

---

94 Q&A on TIA, *supra* note 18.

95 *Supra* note 23 at 608.

96 Levi & Wall, *supra* note 23 at 200.

97 Tien, *supra* note 28 at 399; Stanley & Steinhardt, *supra* note 51 at 14; TAPAC, *supra* note 3 at 35.

98 Cavoukian, "National Security," *supra* note 17 at 47, citing Priscilla M. Regan, *Legislating Privacy: Technology, Social Values, and Public Policy* (Chapel Hill: University of North Carolina Press, 1995); see also *R. v. Mills*, [1999] 3 S.C.R. 668 at paras. 81-82, McLachlin & Iacobucci JJ. [*Mills*].

likely to disappear, just because overtly dedicated government funding is not available. It is likely to be pursued by the private sector. Governmentally sponsored research and development could continue under opaque budget lines. Even Congress permitted data mining research and development to continue under classified budgets. The counter-terrorism community is not likely simply to ignore a potentially powerful tool because of apocalyptic speculation. If the research and development does continue *sub rosa*, technical concerns are likely to dominate. There would be no guarantee that data mining would be established in a form that maximizes privacy protection. We could wake up one fine day to find ourselves well mined, and not have a legislative regime in place to deal with what is all around us. It is better to anticipate the legislative scheme for a technology in its infancy than to try to catch up to and constrain a mature and powerful technology.

If it were accepted that data mining should be regulated, an important question concerns the foundation for the regulation: would the regulation be political or would it rest on constitutional rights? The risks of data mining may be social, political and personal — but are they constitutional risks as well? If data mining would not limit any constitutional rights, the use and form of data mining would be political issues only. This might incline some politicians to take the view that the ends justify the means.[99] Characterizing data mining as a political issue only would not necessarily entail that data mining would be left unregulated, or that any regulation would not seek to limit the ill effects of data mining. The forum for addressing data mining, though, would only be Parliament. If data mining would limit constitutional rights, the use and form of data mining would have to conform to constitutional standards, arguably some version of the rigorous *Hunter v. Southam*[100] standards. Citizen control and oversight of data mining could be effected not merely at the ballot box, but in court. Parliament would be required to craft legislation to conform to constitutional standards.

The critical question, then, is whether data mining limits constitutionally protected privacy rights. Do individuals have "reasonable expectations of privacy" in the information to be mined? If not, the State has a free hand. If so, even if the reasonable expectation of privacy is "diminished," the State must follow constitutional standards in data mining.

I will approach this problem by considering some aspects of the constitutional analysis that I believe are non-controversial; and whether the type of information that would be data mined supports reasonable expectations of privacy.

---

[99] Cavoukian, "National Security," *ibid.* at 46.

[100] *Hunter v. Southam Inc.*, [1984] 2 S.C.R. 145 [*Hunter v. Southam*].

A. NON-CONTROVERSIAL CONSTITUTIONAL ASPECTS OF DATA MINING

A few constitutional points are tolerably clear:

(a) Because the active party in data mining is the State, which seeks information for counter-terrorism or law enforcement purposes, the *Charter*[101] applies to data mining, through s. 32.

(b) Generally, the *Charter* does protect informational privacy,[102] and may, in the right circumstances, extend that protection even to commercial information.[103]

(c) If individuals do have constitutionally protected interests in the mined information, and on the assumption that data mining would be done without individuals' consent, the State's data mining activities for counter-terrorism purposes would amount to searches of databases for information and seizures of information: "[T]he essence of a seizure under s. 8 is the taking of a thing from a person by a public authority without that person's consent"[104] for a purpose contrary to the interests of the person.[105] Queries are searches for information. As a result of a query, the information sought is revealed to the person making the query. The information retrieval is a seizure of information. The person making the inquiry would (at least if a search were successful) make a copy or create a record of the information retrieved. Again, making copies would amount to seizures of information.[106]

(d) If data mining would amount to search or seizure, it should be expressly authorized in legislation.[107] In the absence of statutory authorization to search and seize, as a general rule, we maintain the right to be left alone by State agents.[108]

(e) If the State conduct would amount to a search or seizure, the appropriate *Charter* provision on which to base the analysis is s. 8, under which "[e]veryone has the right to be secure against unreasonable search or seizure."[109] The present analysis

---

101 *Canadian Charter of Rights and Freedoms*, Part I of the *Constitution Act, 1982*, being Schedule B to the *Canada Act 1982* (U.K.), 1982, c. 11 [*Charter*].

102 *Dyment, supra* note 65 at para. 22; *R. v. Tessling*, [2004] 3 S.C.R. 432, 2004 SCC 67 at para. 23, Binnie J. [*Tessling*]; *R. v. Law*, [2002] 1 S.C.R. 227, 2002 SCC 10 at para. 16, Bastarache J. [*Law*].

103 *Thomson Newspapers Ltd. v. Canada (Director of Investigation and Research, Restrictive Trade Practices Commission)*, [1990] 1 S.C.R. 425 at 506-507, 517-18, La Forest J. [*Thomson Newspapers*]; *R. v. Jarvis*, [2002] 3 S.C.R. 757, 2002 SCC 73 at paras. 72, 95, Iacobucci & Major JJ. [*Jarvis*].

104 *Dyment, supra* note 65 at 431; *R. v. Colarusso*, [1994] 1 S.C.R. 20 at 41, 58, La Forest J. [*Colarusso*]; *R. v. Dersch*, [1993] 3 S.C.R. 768 at 778, Major J. [*Dersch*]; *Law, supra* note 102 at para. 15.

105 *Colarusso, ibid.* at 56-57; *R. v. Evans*, [1996] 1 S.C.R. 8 at para 11, Sopinka J.

106 *Mills, supra* note 98 at para. 77. Electronic surveillance constitutes "search or seizure"(*R. v. Thompson*, [1990] 2 S.C.R. 1111 at 1137-38, Sopinka J. [*Thompson*]).

107 "Powers to search the person and premises of a suspect, save with respect to the right of search incident to arrest, are entirely statutory in Commonwealth countries" (*R. v. Rao* (1984), 46 O.R. (2d) 80 (C.A.) at 94, Martin J.A., leave to appeal to S.C.C. refused, [1984] 2 S.C.R. ix [*Rao*]; see also *R. v. Turcotte* (2005), 200 C.C.C. (3d) 289, 2005 SCC 50 at paras. 51, 41, Abella J. [*Turcotte*]; *R. v. Mann*, [2004] 3 S.C.R. 59, 2004 SCC 52 at paras. 37and 45, Iacobucci J.).

108 *Hunter v. Southam, supra* note 100 at 159, Dickson J., as he then was.

109 Section 8, and not s. 7, is the appropriate section in these circumstances (*R. v. S.A.B.*, [2003] 2 S.C.R. 678, 2003 SCC 60 at para. 36, Arbour J. [*S.A.B.*]).

need not stray into s. 1 of the *Charter*, since what is sought is the proper constitutional regulation of data mining. If that proper regulation is determined, it will comply with s. 8 and resort to s. 1 would be unnecessary.

(f) The s. 8 analysis has two main components. First, it must be determined whether data mining engages any constitutionally protected interests — in other words, whether any reasonable expectations of privacy are at stake. Second, if so, the constitutional constraints on data mining — the conditions that must be met for it to be reasonable — must be determined.

I shall turn to the first component of the s. 8 analysis.

B. REASONABLE EXPECTATIONS OF PRIVACY IN TARGET INFORMATION

One might argue that the "reasonable expectation of privacy" doctrine that has grown in the Supreme Court jurisprudence does not provide satisfactory criteria for assessing when checks may be constitutionally imposed on State conduct. But some test is necessary to distinguish cases in which constitutional regulation is necessary from cases in which it is not. The reasonable expectation of privacy jurisprudence does provide a threshold test and does provide guidelines for legal argument. Given the current state of our doctrine, it is the only test we have.

Whether or not reasonable expectations of privacy will be recognized depends on consideration of the "totality of the circumstances."[110] The Supreme Court has recognized that purported expectations of privacy must meet a normative rather than a descriptive standard.[111] The question, reflected through the lens of the judiciary, is whether, "in a society such as ours," individuals who have allowed information to be collected in electronic form by record custodians *should* be judged to have reasonable expectations of privacy in the information.[112] The question is an "objective" one, rather than one dictated by the subjective expectations of individuals.[113]

The following factors are relevant to the assessment of whether an individual has a "reasonable expectation of privacy" in relation to personal information:

(a) the conduct of the individual;[114]

---

110 *Tessling*, *supra* note 102 at para. 31; *R. v. Buhay*, [2003] 1 S.C.R. 631, 2003 SCC 30, at para. 18, Arbour J. [*Buhay*]; *R. v. Edwards*, [1996] 1 S.C.R. 128 at para. 45, Cory J. [*Edwards*].

111 *Tessling*, *ibid.* at para. 42.

112 *Buhay*, *supra* note 110 at para. 19. The Supreme Court jurisprudence is in tension on the issue of whether the "social standard" should be interpreted empirically, as relating to the actual state of public opinion, or normatively, as relating to how standards *should* be set in a fundamentally just environment. Although the Supreme Court may from time to time slip into a quasi-empirical mode of analysis, the proper approach is the normative approach — especially in the absence of public opinion evidence, and given that the Court's focus should be on justice, which is sometimes in conflict with public opinion. For a discussion of the approaches in tension, see James A.Q. Stringham, "Reasonable Expectations Reconsidered: A Return to the Search for a Normative Core for Section 8?" (2005) 23 C.R. (6th) 245.

113 *Tessling*, *supra* note 102 at para. 42.

114 *Ibid.* at paras. 46, 48.

(b) disclosure of the information to third-party custodians;[115]

(c) the nature of the information;[116]

(d) the nature of the custodian's relationship with the individual;[117]

(e) the nature of the custodian's relationship with the State;[118]

(f) the "place" where the search occurred;[119] and

(g) the nature of the search technology and the manner in which the information was obtained.[120]

Data mining involves a series of steps. The privacy analysis must therefore be layered. Different considerations are engaged at different steps in the process. Privacy should be considered from the standpoint of access to (1) a single database and (2) aggregated databases. Privacy should also be considered from the standpoint of (3) the totality of interests affected by the search and seizure.

1. MINING OF RECORDS HELD BY A SINGLE CUSTODIAN: DISCUSSION OF FACTORS

For data mining to occur, an individual must have disclosed personal information to a custodian. Assume, for the moment, that the manner of search of a single database is not constitutionally troublesome, so factor (g) need not be considered.

a. Conduct

The conduct of an individual is relevant to the assessment of the expectation of privacy.[121] If an individual abandoned a personal object or even a bit of bodily matter, or if an individual exposed an object to public view, then the individual gave up any reasonable expectation of privacy in relation to that thing.[122] What should be the result if an individual voluntarily provided personal information to a custodian with knowledge that the information would be disclosed to the public? Helen Nissenbaum has correctly argued that we must be sensitive to context. We must pay attention to "contextual integrity."[123]

In one sort of context, an individual may provide information to a public or private entity for entry on a public registry. The point of having the registry is to make the information

---

115 *Ibid.* at para. 49.
116 *Ibid.* at para. 59; *R. v. Plant*, [1993] 3 S.C.R. 281 at 293, Sopinka J. [*Plant*].
117 *Plant*, *ibid.*
118 *Ibid.* at 294-95.
119 *Tessling*, *supra* note 102 at para. 44; *Plant*, *ibid.* at 293.
120 *Tessling*, *ibid.* at paras. 50, 56; *Plant*, *ibid.*
121 *Thomson Newspapers*, *supra* note 103 at 506-07.
122 *Thompson*, *supra* note 106 at 1142-44.
123 Helen Nissenbaum, "Privacy as Contextual Integrity" (2004) 79 Wash. L. Rev. 119 at 125, 139.

available to members of the public or to State officials for various State purposes. Such registries would include Alberta's Personal Property Registry, the office of the Registrar of Motor Vehicles Services, the office of the Registrar of Corporations and the Land Titles Office.[124] By disclosing information to these registries, the individual would arguably have given up any reasonable expectation of privacy in the information. The *Plant* case supports this approach. If information in a database is accessible to the public generally, State agents should have the same access to the information as other members of the public.[125]

In another sort of context, the individual may provide the information for use on a website for an organization or on a personal website (supported by a commercial service provider). One might be inclined to rely on *Plant*, and argue that just as members of the public can read the website, State agents should have the same unfettered access.[126]

A counter-argument is available. The mere fact that information is publicly accessible does not entail that State agents should have automatic access to the information for State purposes. The *Plant* approach assumes that two options are available: information is private and not available to the public; and information is public, and so available to anyone, including State agents. One might argue that this assumption ignores other possibilities. There is the possibility, for instance, that information is made available to the public for use as members of the public. This use does not include use for the purposes of law enforcement or counter-terrorism. It is true that if a person who is a police officer accesses publicly available information, the physical act is the same, whether he or she is acting only as a private citizen or as a State agent. Acting as a State agent, though, makes a great deal of difference to the legal significance of the access to the information. The State agent is the medium through which the information is moved into the context of State investigation, prosecution and coercive consequences. Moreover, the State is intervening in relation to the information with an electronic search and retrieval of the information. The State's purposes for access differ from individuals' purposes, and the State's use of information differs from individuals' uses.

The *Duarte*[127] and the *Hebert*[128] cases recognize the critical differences between an act performed by an ordinary citizen (having a conversation) and an act performed by a State agent (respectively, recording the conversation and serving as a State informer). We may reasonably be understood to bear the risk of a reader who alerts the authorities about our message, but not the risk that State authorities will — unannounced and warrant-free — review our work for investigatory purposes. The issue at stake is the same as that concerning

---

124 See the *Freedom of Information and Protection of Privacy Act*, R.S.A. 2000, c. F-25, s. 4(1)(l) [*FOIPP*]. On the issue, for example, of the disclosure of motor vehicle information, see *Access to Motor Vehicle Information Regulation*, Alta. Reg. 140/2003.

125 *Plant*, *supra* note 116 at 294-95.

126 Cockfield, "Watchers," *supra* note 42 at 374.

127 "[T]he law recognizes that we inherently have to bear the risk of the 'tattletale' but draws the line at concluding that we must also bear, as the price of choosing to speak to another human being, the risk of having a permanent electronic recording made of our words" (*R. Duarte*, [1990] 1 S.C.R. 30 at 48, La Forest J. [*Duarte*]).

128 *R. v. Hebert*, [1990] 2 S.C.R. 151 at 184, McLachlin J., as she then was [*Hebert*]. The *Hebert* protections of the "right to silence" apply only if an individual has been arrested or detained.

warrantless attendance of State agents at public lectures or symposia or in our lecture theatres.[129]

If the counter-argument is cogent, contrary to the view of Sopinka J. in *Plant*, the State should not necessarily have unfettered access to information, just because the general public has such access; and the mere fact that an individual has posted information on a public website does not entail that the information is free to be mined.

b. Disclosure to Third-Party Custodians

The mere fact of providing the information to a third-party custodian does not entail that the expectation of privacy is lost. The individual need not even maintain a property or a possessory interest in the information or the records that set out the information.[130] This point might seem fairly evident, but it marks a point of divergence between Canadian and U.S. jurisprudence. In *United States v. Miller*, the U.S. Supreme Court held that once information has been disclosed to a third party, the subject ceases to have a constitutionally protected privacy interest in the information.[131] In contrast, the *Charter* jurisprudence acknowledges the persistence of constitutionally protected interests in information disclosed to third parties in a variety of contexts, including health information disclosed to a physician;[132] "therapeutic" information disclosed by a sexual assault complainant or witness to a third-party services provider;[133] and information provided by a sexual assault complainant to the Crown.[134]

c. Nature of the Information

The nature of the information disclosed is a critical factor in the expectation of privacy analysis. Some information, such as electricity consumption information or external patterns of heat distribution from a house, by itself supports little or no reasonable expectation of privacy.[135] The greater the relevance of the information to the "biographical core" of the individual, to the "intimate details" of the individual's life or his or her "personal lifestyle or private decisions," the stronger the expectation of privacy.[136] This does not mean that business records, even those kept in accordance with statutory requirements, bear no expectation of privacy; the level of privacy protection for this sort of information, however, is lower than that for more intimate information.[137] The mere fact that information has a financial aspect does not entail a low expectation of privacy. Information collected from credit card or debit card information custodians, it should be noted, could contain biographical core information. A credit card, for example, could be used to purchase

---

129 The U.S. "Attorney General's Guidelines" impose some policy-based constraints on FBI conduct in this area (see *supra* note 90).

130 *Edwards*, *supra* note 110 at para. 29; *McInerney v. MacDonald*, [1992] 2 S.C.R. 138.

131 425 U.S. 435 (1976); TAPAC, *supra* note 3 at 22-23.

132 *Dersch*, *supra* note 104 at 778.

133 *Mills*, *supra* note 98 at para. 77.

134 *Ibid.* at para. 108.

135 *Plant*, *supra* note 116 at 293 and *Tessling*, *supra* note 102 at para. 63, respectively.

136 *Plant*, *ibid.*

137 *Thomson Newspapers*, *supra* note 103 at 506-507, 517-18; *Jarvis*, *supra* note 103 at paras. 72, 95.

prescriptions and prescriptions could disclose an individual's medical condition. The fact of purchasing from certain establishments or of purchasing certain goods could also disclose significant information about an individual's "biographical core."

The application of this factor bears out Dickson C.J.C.'s insight that s. 8 might protect "interests beyond the right of privacy."[138] The types of "choices" or "decisions" that are "personal" or form part of an individual's biographical core may be choices or decisions that are constitutionally protected freedoms under s. 2 of the *Charter*: freedom of conscience and religion; freedom of thought, belief, opinion and expression; freedom of peaceful assembly; and freedom of association. To the extent that a search or seizure of information exposes the non-public exercises of these freedoms, or, put another way, to the extent that a search or seizure tends to "chill" the non-public exercises of these freedoms, to that extent, the search or seizure limits a reasonable expectation of privacy.[139] On this approach, video rental, library borrowing, or book or other information-media purchasing records should be granted a high measure of privacy protection.

Telephone number information — the telephone number or location from which a telephone call originates or at which it is received or intended to be received — might not seem like "biographical core" information. This information is in the custody of third-party telephone service providers. Nonetheless, the *Criminal Code* establishes a procedure for the issuance of a warrant to install a device and record this information.[140] Similarly, an individual's movements in public, whether in a motor vehicle, in some other mode of transport or on foot, are seen by many others. Nonetheless, the *Criminal Code* establishes a procedure for the issuance of a warrant to install a device to identify the location of an individual.[141] The existence of these procedures is some evidence supporting the proposition that even information that does not manifestly implicate the biographical core should receive some constitutional protection.

d. Relationship between the Custodian and the Individual

If information has been legitimately collected by a law enforcement, military or intelligence organization for law enforcement or counter-terrorism purposes, it would be expected that these organizations would and should be free to use and disclose this information for these purposes. Yet even these organizations, as well as most private and public organizations, are statutorily required to protect privacy.[142] Generally, private organizations in Alberta are bound by *Personal Information Protection Act* (*PIPA*)[143] or the *Health Information Act* (*HIA*);[144] and Canadian international, inter-provincial and federal works, undertakings and businesses are bound by the *Personal Information Protection and*

---

138 *Hunter v. Southam*, *supra* note 100 at 159.

139 See *Thomson Newspapers*, *supra* note 103 at 517-18.

140 R.S.C. 1985, c. C-46, s. 492.2.

141 *Ibid.*, s. 492.1; see *R. v. Wise*, [1992] 1 S.C.R. 527, Cory J. [*Wise*].

142 Some registry information is not caught by, *e.g.*, *FOIPP*, which allows the information to be disclosed to the investigatory State without statutory hindrance (see *FOIPP*, *supra* note 124, s. 4(1)(l)).

143 *Supra* note 66.

144 R.S.A. 2000, c. H-5 [*HIA*]. I shall assume that the *HIA* is effective as providing "substantially similar" protection to that provided under *PIPEDA*, *supra* note 66, s. 26.

Electronic Documents Act (*PIPEDA*) in relation to information respecting commercial activities.[145] In Alberta, public bodies are bound by the *Freedom of Information and Protection of Privacy Act* (*FOIPP*)[146] and the *HIA*; and federal public bodies are bound by the *Privacy Act* (*PA*).[147] On a very general level, this privacy legislation shares commitments to a common set of privacy principles, which include the following:[148]

(i) "identifying purposes" — "[t]he purposes for which personal information is collected shall be identified by the organization at or before the time the information is collected";[149]

(ii) "consent" — "[t]he knowledge and consent of the individual are required for the collection, use, or disclosure of personal information, except where inappropriate";[150]

(iii) "limiting collection" — "[t]he collection of personal information shall be limited to that which is necessary for the purposes identified by the organization";[151] and

(iv) "limiting use, disclosure, and retention" — "[p]ersonal information shall not be used or disclosed for purposes other than those for which it was collected, except with the consent of the individual or as required by law. Personal information shall be retained only as long as necessary for the fulfillment of those purposes."[152]

Privacy legislation does not displace legal authority to acquire information through non-consensual or coercive processes.[153] Outside of these legally coercive contexts, however, the fundamental orientation of privacy legislation is to require organizations to collect information only with the informed consent of individuals, and to use and disclose that information only in accordance with the express purposes for which the information was collected. Statutory privacy regimes, then, provide a measure of support for expectations of privacy concerning information held by custodians subject to those regimes.

Furthermore, aside from statutory protections, the Supreme Court has recognized that the information collected by a third party should only be used or disclosed for the purposes for which the information was obtained: "[T]he limited purpose for which [the information] was obtained cannot be ignored."[154] According to McLachlin J. (as she then was), "[p]rivacy is not an all or nothing right.... Privacy interests in modern society include the reasonable

---

145 *PIPEDA*, *ibid.*, s. 4.

146 *Supra* note 124.

147 R.S.C. 1985, c. P-21 [*PA*].

148 For ease of exposition, reference is made only to the *PIPEDA* schedule. These principles may be traced in the other legislation. For a similar list, see Loukidelis, "Information Technology," *supra* note 62 at 11-12.

149 *PIPEDA*, Sch. 1, 4.2 Principle 2.

150 *Ibid.*, 4.3 Principle 3.

151 *Ibid.*, 4.4 Principle 4.

152 *Ibid.*, 4.5 Principle 5.

153 See, *e.g. FOIPP*, *supra* note 124, ss. 3(c), (d), 40(1)(g).

154 See *Colarusso*, *supra* note 104 at 55; *Dyment*, *supra* note 65 at 432; *Law*, *supra* note 102 at para. 22.

expectation that private information will remain confidential to the persons to whom and restricted to the purposes for which it was divulged."[155]

e. Relationship of the Custodian and the State

The statutory privacy regimes uniformly allow for the disclosure of information for a variety of purposes. A body must disclose information if so required by court process.[156] Under s. 32 of *FOIPP*, the head of a public body has an obligation to disclose "to the public, to an affected group of people, [or] to any person" information that concerns "a risk of significant harm to the environment or to the health or safety of the public, [or] of the affected group of people, [or] of the person." This provision has analogues in other privacy statutes.[157] Moreover, under s. 40(1)(q) of *FOIPP*, a public body may disclose personal information "to a public body or a law enforcement agency in Canada to assist in an investigation (i) undertaken with a view to a law enforcement proceeding, or (ii) from which a law enforcement proceeding is likely to result." This provision has analogues in other privacy statutes.[158] No constitutional principle of use immunity or derivative use immunity prevents information collected without a dominant purpose of investigation from being transmitted to the State for investigatory or law enforcement purposes.[159] The presence of these latter "cooperation clauses," however, does not negate expectations of privacy.

Even confidential relationships (including solicitor-client privilege) are subject to exceptions.[160] Exceptions do not entail that individuals should not be entitled to expect that their information will be kept, generally, confidential. The fact of possible disclosure does not altogether negate the expectation of privacy.[161]

Furthermore, the "cooperation clauses" create discretions to provide information, not obligations to provide information. One would expect custodians to exercise this discretion with a measure of caution and with due regard for privacy interests — otherwise information in the hands of public bodies would be virtually in the hands of law enforcement. There would only be law enforcement databases, one big government collection of information.

f. Place

In *Plant*, Sopinka J. suggested that State access to third-party computer records is not as significant an interference with privacy as would be an interference with the individual's home or office computer.[162] One might agree that we have obvious reasonable expectations

---

155 *Mills*, *supra* note 98 at para. 108.

156 *FOIPP*, *supra* note 124, ss. 40(1)(g), 3(d).

157 *PA*, *supra* note 147, s. 8(2)(m); *PIPEDA*, *supra* note 66, ss. 7(2)(b), 7(3)(e); *PIPA*, *supra* note 66, s. 20(f); *HIA*, *supra* note 144, ss. 35(k), (l), (m).

158 *PA*, *ibid.*, s. 8(2)(e); *PIPEDA*, *ibid.*, ss. 7(2)(a), (3)(c.1), (c.2), (d); *PIPA*, *ibid.*, s. 20(e); *HIA*, *ibid.*, s. 35(1)(j).

159 *Jarvis*, *supra* note 103 at para. 95.

160 *Smith v. Jones*, [1999] 1 S.C.R. 455.

161 *A.M. v. Ryan*, [1997] 1 S.C.R. 157 at para. 24, McLachlin J.; *Tessling*, *supra* note 102 at para. 42; *Buhay*, *supra* note 110 at paras. 22, 23.

162 *Plant*, *supra* note 116 at 295.

of privacy respecting our dwellings, offices and home computers, but it does not follow that because a search and seizure does not involve one of these obvious places, we have no reasonable expectation of privacy or only a diminished expectation of privacy. Justice McLachlin commented in her minority concurring decision in *Plant* that Sopinka J. begged the question.[163] Justice Sopinka's view might be considered muddled, since s. 8 — as is well known — protects people, not places.[164]

2. MINING OF RECORDS HELD BY A SINGLE CUSTODIAN: CONCLUSIONS

The application of these factors to personal information held by a single database custodian does not permit many generalizations:

(i) If information has been legitimately acquired by law enforcement, the military or intelligence authorities for law enforcement or counter-terrorism purposes, it may be used and disclosed for those purposes, and an individual has no reasonable expectation of privacy that might block this use or disclosure. This information could be mined.

(ii) If information held by a custodian is not subject to a statutory privacy regime and if that information is available to the public and the State for a variety of public purposes, then an individual has no reasonable expectation of privacy that might block disclosure to the State for law enforcement or counter-terrorism purposes.

(iii) If an individual voluntarily places records in a publicly accessible area, an argument, supported by authority, is that the individual has no reasonable expectation of privacy in the information, which would leave it free to be mined. A counter-argument, based on "contextual integrity," is available, though; and if that argument prevails, the State should not have automatic access even to this information.

(iv) If information has little to do with an individual's biographical core (such as electricity consumption information), then again, based on authority, the individual has no reasonable expectation of privacy in the information. However, it cannot be assumed that just because information is in a commercial database the information cannot support a reasonable expectation of privacy.

(v) If none of the exceptional situations described in (i) - (iv) applies, an individual should be recognized as having a reasonable expectation of privacy in the personal information.

The situations described in (iii) and (iv) raise an important practical point: how should it be determined whether or not a reasonable expectation of privacy exists respecting the information? In typical criminal search and seizure circumstances, the evidence would be introduced at trial, and the propriety of the State conduct could be challenged. Data mining,

---

163 *Ibid.* at 303-304.

164 *Edwards*, *supra* note 110 at para. 45.

though, may grind on in silence, without information ever being tendered in any public forum. The results of data mining may be used to take action against an individual, without notice to the individual that data mining was a source of information. The important decision about whether information supports a reasonable expectation of privacy should not be left to State agents alone. Some neutral intervention is required.

I suggest that the legislation governing data mining should provide that if data mining of any information other than that described in (i) or (ii) is contemplated, an application should be made to a judge for a determination of whether reasonable expectations of privacy exist in relation to the information. If no reasonable expectations of privacy should be recognized, data mining may proceed. If reasonable expectations of privacy should be recognized, then an application procedure such as that described in the next section is engaged. Alternatively, Parliament may attempt to provide a legislated definition of records which support reasonable expectations of privacy, along the lines of the definition of "record" in the third-party production application provisions in the *Criminal Code*.[165] Because of the very wide scope of interests that might arise in relation to information that might be data mined, and because of the complications introduced by data mining itself (addressed in the next two sections), a legislated definition is likely not feasible.

3. AGGREGATED RECORDS

The issue at this stage is the constitutional effect of aggregration, the actual or virtual assemblage of databases to permit a subject-based query; that is, whether aggregation itself entails qualitatively significant, constitutionally cognizable effects. The inquiry concerns the Canadian constitutionality of projects like COPLINK and MATRIX. In terms of the factors referred to above, this inquiry involves factor (g), respecting the nature of the search technology and the manner in which the information was obtained.

One hypothesis is that aggregation is not constitutionally significant. Aggregation only involves a series of searches of particular databases. If there is no reasonable expectation of privacy in relation to a single database, the repetition of single searches would not generate any new facts that would alter the constitutional assessment. Aggregation itself poses no new privacy risks.[166]

Another hypothesis is that aggregation is constitutionally significant and does, by itself, pose new privacy risks. As indicated in the discussion of the demise of "practical obscurity" in Part IV.B.2 above, if public and private databases are linked, the result is the generation of a new form of surveillance — "dataveillance." Although individual records revealed by searches might not relate to an individual's "biographical core," an assemblage of records in response to a single search may disclose much of an individual's ordinary life. A detailed account of our daily lives — a transactional narrative — can be assembled from the electronic records we leave in our wake. Information technology supplies the transactional

---

[165] *Supra* note 140, s. 278.1.

[166] Shane Ham & Robert D. Atkinson, "Using Technology to Detect and Prevent Terrorism" *Policy Brief* (January 2002), online: Progressive Policy Institute <www.ppionline.org/documents/IT_terrorism.pdf> at 4.

equivalent of a video record. Dataveillance re-constitutes the biographical core out of transactions. In theory, the record could be in "real time," or as close to real time as data entry and recovery will permit. Dataveillance is therefore also at least the transactional equivalent of a tracking device. The use of tracking devices requires a warrant, issued by a justice.[167] Dataveillance should demand no less. Furthermore, data mining does not merely track movements; it does not only produce a sort of physical record, instead, the information it produces has a semantic quality. Transactional records support inferences respecting the meaning of our actions. Data mining produces new personal information for State. The State could not obtain this information about us directly, without a warrant.[168]

4. THE PERSPECTIVE OF NON-TARGETS

Factor (g), respecting the nature of the search technology and the manner in which the information was obtained, is also engaged by the scope of data mining. Data mining does not involve searches and seizures of only the information relating to suspects or "persons of interest." In the course of data mining, the information of many individuals is assessed. This is particularly true if profile-based searches are employed. The profile-based search would involve running all individuals' information in the searched databases against the profile. Data mining engages the expectations of privacy of not merely the few terrorists among us, but of all of us, or at least of all of us who have disclosed information that may be mined. Numbers matter. If not merely one individual's but thousands of individuals' information is being searched, this is surely a matter of constitutional concern. Chief Justice Dickson referred in *Hunter v. Southam* to the *public's* interest in being left alone.[169] In *Thompson*, Sopinka J. considered the invasion of privacy of third parties to be constitutionally relevant to the issue of whether there has been an "unreasonable" search or seizure.[170] With the qualification that the circumstances would be "somewhat rare," the *Edwards* majority agreed with the *Thompson* finding.[171] The quantity of intrusions should be relevant, however, not merely to the reasonableness of the search, but to the issue of whether reasonable expectations of privacy were violated. While we might be prepared to accept that certain single records should not have constitutional protection, that should not commit us to the view that everyone's similar records should be laid open all at once or in series. This view would transform the records into State investigatory records and would erase meaningful distinctions between the custodians and the coercive arm of the State. It is in the public interest to keep warrantless seizures of records to a minimum.

VI. THE CONSTITUTIONAL REGULATION OF DATA MINING

If I am wrong, and the personal information held by third-party custodians supports no reasonable expectations of privacy, then data mining may proceed without constitutional impediment. Even so, to forestall the risks identified in Part IV above, a regime like the one

---

167 *Criminal Code*, *supra* note 140, s. 492.1.

168 See Renee M. Pomerance, "Shedding Light on the Nature of Heat: Defining Privacy in the wake of *R. v. Tessling*" (2005) 23 C.R. (6th) 229; Pomerance, "Redefining," *supra* note 42 at 284-85, 289; Loukidelis, "Information Technology," *supra* note 62 at 6.

169 *Supra* note 100 at 159.

170 *Supra* note 106 at 1143-44.

171 *Supra* note 110 at para. 38.

I shall describe is still appropriate from a policy perspective. If I am right that some personal information held by third-party custodians does support reasonable expectations of privacy, the search or seizure of the information must be authorized by law, the law must be reasonable, and the manner of the search or seizure must be reasonable.[172] Since my concern is with the design of legal regulation of data mining, and not with a particular search conducted under the purported authority of data mining legislation, only the second element of this test is relevant. I shall assume that some significant quantity of information to be data mined does support reasonable expectations of privacy and pursue the issue of how data mining might be rendered reasonable through regulation.

A. BACKGROUND

The minimum constitutional standards for the reasonableness of criminal search and seizure legislation were established in *Hunter v. Southam*.[173] The standards have procedural and substantive elements. The procedural elements are as follows: authorization for the search and seizure must be obtained before execution, and must be granted by an independent and impartial judicial officer with the discretion to authorize or not and to authorize on conditions. The authorization must be applied for on the basis of sworn evidence.[174] The substantive elements are that the judicial officer must be satisfied that it is likely that an offence has been committed and there is evidence to be found at the place of the search.[175] As Dickson C.J.C. indicated, "[t]he state's interest in detecting and preventing crime begins to prevail over the individual's interest in being left alone at the point where credibly-based probability replaces suspicion."[176]

The *Hunter v. Southam* standards should be applied to data mining, unless there are reasons for elevating or diminishing the constitutional protection. These standards have been adapted to different types of search and seizure. The following considerations are relevant to adapting the standards to data mining:

(1) the purpose served by data mining legislation;

(2) the "politico-epistemology" of data mining;

(3) the intrusiveness of data mining;

(4) the effectiveness of data mining;

(5) the legitimate uses of information derived from data mining;

(6) the potential misuses of data mining; and

---

172 *R. v. Collins*, [1987] 1 S.C.R. 265 at 278, Lamer J.
173 *Supra* note 100.
174 *Ibid.* at 160, 162.
175 *Ibid.* at 167, 168.
176 *Ibid.* at 167.

(7) the data mining oversight processes.

For data mining to be reasonable, its legislative framework must acceptably minimize the risks identified in Part IV above. If these risks were not minimized, the likely risks of social, political and personal injury that would be caused by data mining would outweigh its anticipated benefits. The risks of data mining are addressed through the seven considerations below.

B. CONSIDERATIONS BEARING ON ELEVATING OR REDUCING THE *HUNTER* V. *SOUTHAM* STANDARDS

1. PURPOSE

If, for example, search and seizure provisions are embedded in a commercial regulatory context as opposed to a law enforcement context, the level of procedural and substantive protection for expectations of privacy may legitimately be reduced. Less stringent standards reflect the diminished expectations of privacy attaching to many forms of business records in regulatory contexts.[177] Data mining, however, does not occur in a "regulatory" context.

The purposes served by data mining are, generally, the collection of information relevant to preserving national security and in particular, the collection of information to identify terrorists, aid in their prosecution and, most importantly, identify terrorist threats and prevent terrorist acts from occurring. Some might argue that the achievement of these purposes requires a relaxation or lowering of constitutional standards. Counter-terrorist operations must move quickly to prevent attacks from occurring. The threats posed by terrorists are grave. Counter-terrorist measures should therefore not be burdened with the constitutional procedures applying to ordinary law enforcement. The nation and the host of potential victims do not have the luxury of time to deal with standard warrant procedures.

On the other hand, precisely the stakes involved demand extreme care. If someone is identified as being a terrorist or as being involved in a terrorist plot, the consequences to that individual would likely be swift and devastating. If that individual has been targeted in error, great injury shall have been inflicted by the State itself. Furthermore, while focused on the wrong individual, State attention has been diverted from actual terrorists, leaving the nation vulnerable. The purposes served by data mining should attract scrupulous procedures, not sloppy procedures.

A set of assumptions about constitutionally mandated procedures underlies the view that the purpose served by data mining (or, indeed, by any counter-terrorism methods) should diminish procedural rigour. Constitutional procedures, such as the search warrant procedures, are understood as obstacles put into the path of investigation and enforcement by individuals with an unnatural and unfounded fear of the State and who harbour an excessive concern with the interests of accuseds and suspects. In contrast to this view, I suggest that constitutional procedures are not obstacles, but, in their way, confirm and empower good

---

[177] *Thomson Newspapers*, *supra* note 103 at 506-507.

investigative and enforcement work. Constitutional procedures require the State to have articulable reasons for using coercive and often violent measures. At bottom, constitutional procedures dictate only that the State adhere to the simple requirement that it use its powers rationally, according to just reasons, for good cause. The procedures — for example, swearing an information and convincing a neutral judicial officer that evidence relevant to an offence may be obtained — confirm that proposed State action has a rational basis and is not arbitrary or irrational. If the State is behaving rationally, it is more likely to achieve its objectives than if it simply behaves as it wishes. Furthermore, at least over the mid- to long-term, constitutional procedures are not significant practical hurdles. The police obtain plenty of warrants for plenty of different types of searches and seizures.

Circumstances may occur in which standard procedures cannot be followed. The law recognizes and accommodates exigent circumstances. Chief Justice Dickson recognized that the *Hunter v. Southam* procedures are to be followed only when feasible.[178] Particular steps in counter-terrorism operations may require rapid action that cannot accommodate usual procedures. Of course, there should be evidence that exigent circumstances exist, and a person seeking to rely on the excuse of exigent circumstances should be required to provide an account once the emergency abates. The existence of terrorist risks does not, by itself, constitute an emergency, any more than the risk of flood, earthquakes or epidemics constitutes an emergency. Emergencies occur when serious risks reach the point of imminence or when risks are in the process of being actualized. Prior to the onset of emergencies, we should engage in deliberate, constitutional risk management: "[C]ourts must not fall prey to the rhetorical urgency of a perceived emergency or an altered security paradigm ... we must not lose sight of the particular aims of the legislation."[179]

The implication is that counter-terrorist methods should not fear constitutional standards. Good counter-terrorism will be constitutional counter-terrorism. The purpose served by data mining does not entail any relaxation of the *Hunter v. Southam* standards, outside of actual emergencies.

2. THE POLITICO-EPISTEMOLOGICAL STATUS OF DATA MINING

The use in data mining of pattern-based searches, whether those are generated automatically (as in "true" data mining) or by human analysts, does not in any way lower the standards that ought to apply to data mining. The use of pattern-based searches should elevate the standards applicable to data mining.

The substantive elements of *Hunter v. Southam* establish a standard of proof and a description of what must be established. The granting of a warrant requires evidence giving rise to more than mere suspicion. A warrant may be granted only on a showing of credibly based probability or likelihood.[180] That which is to be established as likely is that an offence has occurred and that evidence relating to that offence may be gathered through the search

---

178 *Hunter v. Southam*, *supra* note 100 at 161; *Colarusso*, *supra* note 104 at 40.

179 *Re Application under s. 83.28 of the Criminal Code*, [2004] 2 S.C.R. 248, 2004 SCC 42, at para. 39, Iacobucci & Arbour JJ. [*Application under s. 83.28*].

180 See *Baron v. Canada*, [1993] 1 S.C.R. 416 at 446-47, Sopinka J.

and seizure for which authorization is sought.[181] Furthermore, the procedural elements of *Hunter v. Southam* require that searches and seizures be justified before they are carried out; they are not justified by the evidence they unearth. Profile-based data mining runs afoul of these prescriptions.

Data mining is to run even though no offence has happened yet. It may gain no evidence respecting a crime that has occurred, but may only provide information relating to a crime that may happen. If data mining does provide evidence, because of the weaknesses of profiling, it could give rise to suspicions of potential wrongdoing only, not probabilities. In any event, the justification for data mining lies in the information it produces about individuals after data is mined. Before the mining occurs, it cannot be known whether any information of interest will be produced. Our law has set its face against this sort of trolling for suspects, or "random virtue checking."[182]

Data mining, then, is highly irregular, highly exceptional. It should not only meet *Hunter v. Southam* standards, but additional standards designed to mitigate its peculiar risks.

3. INTRUSIVENESS OF DATA MINING

The degree of intrusiveness of search and seizure technology is relevant to the assessment of its reasonableness.[183]

Data mining is highly intrusive. In its full-blown, TIA form, it involves searches across the "information space," tracking through the personal information of many individuals. Data mining entails a qualitatively and quantitatively immense interference with privacy interests. Interference with others' privacy rights may be considered respecting whether or not a search was conducted in a reasonable manner.[184] Justice Sopinka indicated in *Thompson* that if the numbers of individuals whose privacy is implicated is large enough, a search may be found to be unreasonable because its manner was unreasonable:

> In my view, the extent of invasion into the privacy of these third parties is constitutionally relevant to the issue of whether there has been an 'unreasonable' search or seizure. To hold otherwise would be to ignore the purpose of s. 8 of the *Charter* which is to restrain invasion of privacy within reasonable limits. A potentially massive invasion of the privacy of persons not involved in the activity being investigated cannot be ignored.[185]

Moreover — and this is connected with the effectiveness issue — unlike DNA testing, data mining threatens to inculpate individuals falsely early in the investigative process. Because of its intrusiveness, "there is no reason to consider applying lesser minimum requirements

---

181 *Colarusso*, *supra* note 104 at 39-40.

182 *R. v. Mack*, [1988] 2 S.C.R. 903 at 941, 956, Lamer J., as he then was.

183 *Tessling*, *supra* note 102 at para. 56; *S.A.B.*, *supra* note 109 at paras. 44, 45.

184 *Edwards*, *supra* note 110 at para. 36. In the context of a criminal prosecution, this sort of argument encounters a "standing" problem: an accused can raise issues relating to violations of his or her own rights, but generally cannot raise issues relating to violations of others' rights (at para. 34).

185 *Thompson*, *supra* note 106 at 1143-44.

to [it]" than the *Hunter v. Southam* standards.[186] The intrusiveness of data mining suggests that two additional conditions be satisfied before it is permitted.

First, as usually applies to applications for authorizations for electronic interceptions of private communications, the applicant should establish

> that other investigative procedures have been tried and have failed, other investigative procedures are unlikely to succeed or the urgency of the matter is such that it would be impractical to carry out the investigation ... using only other investigative procedures.[187]

If other modes of investigation could develop the information sought, the costs of data mining would be unnecessary and unjustified. Hence, an application for an authorization for data mining should require evidence that it is an "investigative necessity."[188] Interestingly, *Criminal Code* s. 186(1.1)(c) provides that the quoted paragraph respecting investigative necessity does not apply if an application for an authorization is in relation to (*inter alia*) a terrorism offence. One might argue, then, that since wiretap authorization applications in relation to terrorism offences need not satisfy an investigative necessity criterion, neither should data mining authorization applications. In response, I note that s. 186(1.1)(c) has been the subject of some critical comment, and it cannot be said with certainty that it will survive *Charter* scrutiny.[189] It could be that Parliament was misled by its appreciation of the purpose served by this legislative provision — Parliament took a position of the sort I argued against in Part VI.B.1 above. Regardless, intrusive as wiretaps may be, data mining is far more intrusive, involving far more personal information relating to far more individuals populating the "information space." What is appropriate for counter-terrorist wiretaps is not necessarily appropriate for counter-terrorist data mining.

Second, data mining should meet a dual "minimization" test — minimization in terms of disclosure of information linked to identifiable individuals and minimization in terms of the scope of searches. "Minimization" is not a general requirement for the constitutionality of the electronic interception scheme.[190] The Canadian position on this issue appears to be founded on practical considerations:

> Because the minimization requirement precludes open-ended, "indiscriminate" interception of private communications, resort to the use of automatic, voice-activated taping systems unattended by on-site monitoring personnel is effectively foreclosed. This naturally has the effect of substantially increasing the

---

186 *R. v. Garofoli*, [1990] 2 S.C.R. 1421 at 1444, Sopinka J. [*Garofoli*].

187 *Criminal Code*, *supra* note 140, s. 186(1)(b).

188 *S.A.B.*, *supra* note 109 at para. 54; *Garofoli*, *supra* note 186 at 1443-45; *R. v. Araujo*, [2000] 2 S.C.R. 992, 2000 SCC 65, LeBel J.; Cavoukian, "National Security," *supra* note 17 at 50; TAPAC, *supra* note 3 at 50.

189 Cockfield, "Watchers," *supra* note 42 at 382; Letter from Ann Cavoukian to the Honourable Paul Zed, M.P., Chair, House of Commons Subcommittee on Public Safety & National Security (28 February 2005) (attached to correspondence of D. Loukidelis to Paul Zed (8 April 2005), online: Office of the Information and Privacy Commissioner for British Columbia <www.oipc.bc.ca/pdfs/public/24915ATA reviewltr(April20-2005).pdf>). Martin L. Friedland, however, has opined that this relaxation of the usual standard is reasonable ("Police Powers in Bill C-36" in Daniels, Macklem & Roach, *supra* note 10, 269 at 274).

190 *Thompson*, *supra* note 106 at 1138.

> expense of electronic surveillance investigations. It is therefore submitted that a general requirement of minimization for all such investigations is too onerous a burden and is one that should not be imposed upon Canadian law enforcement officials. Judges should not, however, be precluded from inserting minimization clauses into authorization orders, where, in their discretion, circumstances so warrant.[191]

Minimization in terms of linking information to identifiable individuals would not, however, be an unduly onerous requirement to impose on data mining. Data mining uses technology that could limit the disclosure of private information without involving continuous monitoring by humans. Searches may be conducted that do not reveal the nature of queries or results to eavesdroppers or records custodians.[192] Technology permits "selective revelation." Information matching a profile can be initially revealed to analysts in a "sanitized form," which does not reveal the identity of the subject; information is initially presented anonymously.[193] Technology thus permits a "security barrier" between the private data and the analyst.[194] Before the identity is linked to the information, a further judicial authorization could be required.[195] Minimization is a practical possibility.[196] It should be a constitutional requirement, because of the vast quantity of private information that will be searched through data mining, and because of the risks of wrongful identification of individuals as being involved with terrorism.

To the extent that selective revelation is employed, the intrusiveness of data mining is significantly reduced. The image behind some of the more sinister and terrifying accounts of data mining is the all-seeing eye of the State, watching our every transactional move.[197] If, though, data is analyzed not by a human but by a machine only (and not by any science fiction sentient machine), data mining should cause less uneasiness. If only information that is deemed significant is brought to the attention of humans, we have the assurance that most information will stay out of government officials' hands. And if information can only be linked to identifiable individuals following an application and with a court order, we have a further assurance that only personal information that should be given to the State will be given to the State.[198]

---

191 *Ibid.*, quoting Stanley A. Cohen, *Invasion of Privacy: Police and Electronic Surveillance in Canada* (Toronto: Carswell, 1983) at 174 [footnotes omitted].

192 Information Sciences and Technologies Study Group, *Security with Privacy* (13 December 2002), online: Electronic Privacy Information Center <www.epic.org/privacy/profiling/tia/isat_study.pdf> at 7 [ISAT]. Custodians may also be prohibited from disclosing State queries to subjects or other third parties (Philip B. Heymann & Juliette M. Kayyem, *Long-Term Legal Strategy Project for Preserving Security and Democratic Freedoms in the War on Terrorism*, online: National Memorial Institute for the Prevention of Terrorism (MIPT): Terrorism Information Center <www.mipt.org/pdf/Long-Term-Legal-Strategy.pdf> at 89).

193 TAPAC, *supra* note 3 at 49.

194 ISAT, *supra* note 192 at 10; Second Markle Report, *supra* note 6 at 34.

195 Heymann & Kayyem, *supra* note 192 at 79; TAPAC, *supra* note 3 at 51; Taipale, *supra* note 25 at 66, 79.

196 It has been argued that the minimization technologies are not yet practical possibilities; see Fienberg, *supra* note 55. If this is so, it means that data mining cannot be justified until the requisite technology is in place.

197 Believe it or not, the all-seeing eye was the original TIA logo. DARPA should have won a Darwin award for advertising.

198 Ham & Atkinson, *supra* note 166 at 11.

Minimization in terms of scope should also be required. Applicants should demonstrate that it is necessary to search specified databases. The set of databases, when combined, should be no larger than necessary.[199] This is no more than an adaptation of the typical "place" specification that is found in warrant applications.[200] The limitation of the scope of searches addresses some of the "practical obscurity" concerns raised in Part IV.B.2 above. Data mining should not necessarily draw on our full transactional history, but only on that which is established to be relevant to potential terrorist activity.

The intrusiveness of data mining entails that no lowering of the substantive standard of credibly based probability should be allowed for data mining. In *Wise*, the majority of the Supreme Court found that tracking devices limited only a reduced expectation of privacy, so a "lower standard such as a 'solid ground' for suspicion would be a basis for obtaining an authorization from an independent authority, such as a justice of the peace, to install a device and monitor the movements of a vehicle."[201] Hence the reference in ss. 492.1 and 492.2 of the *Criminal Code* to "reasonable grounds to suspect." Regardless of whether a reduced substantive standard matches the reduced expectation of privacy at play in *Wise*, data mining engages significant expectations of privacy and multiple risks. Lowering the substantive standard would aggravate risk by enhancing the availability of the process.

4. EFFECTIVENESS

An application for a data mining authorization will require the applicant to explain why the data mining is needed — in other words, what the purpose of the search is and how data mining will provide the desired information. As seen above, data mining is a highly suspect operation. To justify the privacy limitations caused by data mining, applicants should be required to demonstrate that the data mining is likely to be effective at generating reliable results.

A requirement to establish effectiveness has four aspects. Each of these addresses a risk identified in Part IV.A.1 and 2 above. First, the applicant must establish that the data being searched supports accurate analysis. This will involve evidence relating to the purpose for which the data was collected, its age and the conditions in which it was stored.[202] Second, the applicant must establish that the data collected from various sources can be rendered into a form that permits reliable searching. Evidence must be provided respecting the reliability of the "data cleaning" process. Third, the applicant must provide evidence that the minimization technology — respecting the protection of search information, anonymization and selective revelation — is likely to work.[203] Finally, and perhaps most importantly, the applicant must

---

199 Heymann & Kayyem, *supra* note 192 at 79.

200 *Criminal Code*, *supra* note 140, s. 186(4)(c).

201 *Wise*, *supra* note 141 at 549. This lower-than-credibly-based-probability standard, which is also employed in s. 492.1, was held to be unconstitutional in *R. v. Nguyen* (2004) 20 C.R. (6th) 151, 2004 BCSC 76, Halfyard J. Superior Court Justice Cohen declined to follow Halfyard J.'s reasoning in *R. c. Whitman-Langille*, [2004] Q.J. No. 14164 (S.C.) (QL). On this lower standard, see Steve Coughlan, "Nothing Plus Nothing Equals ... Something? A Proposal for FLIR Warrants on Reasonable Suspicion" (2005) 23 C.R. (6th) 239.

202 TAPAC, *supra* note 3 at 51.

203 *Ibid.* at 49.

establish that any profiles used to generate information are reliable predictors of terrorist activity or involvement in terrorist activities. Evidence must be provided respecting the error rates associated with the profile — the rates of false positives and false negatives.[204] An authorization should be granted only if the error rates fall within an acceptable range. To establish the "acceptable range," evidence should be adduced of the post-processing measures designed to mitigate the effects of erroneous identifications.

The requirement to establish effectiveness has practical, pre-litigation implications. Practically, proponents of data mining should be very concerned to develop data mining technology that is "privacy-friendly," that can minimize privacy intrusions while still being effective. It has therefore been recommended that a citizen advisory board be established respecting data mining research, to ensure that technology develops with sensitivity to privacy concerns.[205] Some of the research relating to data mining, particularly concerning the development of profiles, should be classified. Other elements, though, such as those concerning minimization technology, could and should be the subjects of public (and publicly funded) research and discussion.

A related point is that the regulation of data mining cannot only occur through law. Some have suggested that data mining technology is neutral and that what is important is the set of legal rules that govern its use and protect privacy.[206] The assumption of technological neutrality is misplaced. According to Lawrence Lessig, code is law.[207] If external legal regulation were relied on as the only means of regulating data mining, it would likely be ineffective, much as the law of copyright has been largely ineffective to regulate copyright violations accomplished through Internet technology. Technology, to a great extent, determines how people will actually behave.[208] Technological constraints bind users. For privacy protections to be effective, external regulation must be reflected in technologically internal regulation.[209] The point is not to rely on technology or law alone. The two should be mutually reinforcing.

The requirement to establish effectiveness also has implications for the hearing of the application. Evidence of effectiveness relating to technology and profiles would be put before a judge. Judges are qualified to assess the reliability of expert evidence and profile evidence. This sort of evidence (as indicated above respecting profiles) plays a role in a variety of cases. The difficulty in this context is that the authorization application would be made *ex parte*, with no other party in a position to challenge the reliability of the technology or the profile. Furthermore, it is likely that there will be no evidence respecting profile reliability from members of the scientific community other than the proponents of the profile, since the profile research will be classified. Hence, the judge should be assisted by a court-

---

204 *Ibid.* at 50; Heymann & Kayyem, *supra* note 192 at 80.

205 ISAT, *supra* note 192 at 9.

206 Ham & Atkinson, *supra* note 166 at 11; TAPAC, *supra* note 3 at viii; Nissenbaum, *supra* note 123 at 155.

207 Taipale, *supra* note 25 at 12, n. 30, citing Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 1999) at 83-99; Cockfield, "Watchers," *supra* note 42 at 400.

208 Moreover, technology not only resides in a social context; it is itself a social construction (Taipale, *supra* note 25 at 13).

209 *Ibid.* at 12-13.

appointed expert or a team of experts who can assess the usefulness of the profile. The expert or experts would be bound not to disclose any classified information disclosed in the hearing. A trial judge does have the inherent jurisdiction to call his or her own witnesses.[210] Presumably, the calling of witnesses to assist the court falls within the inherent jurisdiction of a superior court judge in a non-trial setting. The issue of whether or not to call for expert assistance could be left to the discretion of the judge hearing the application; and to ensure inherent jurisdiction, these hearings could be restricted to superior court judges. Alternatively, legislation could specify that data mining authorization hearings be before both a judge and a panel of experts, constituted in some prescribed manner.[211]

5. USES

Because of the inherent error rates attendant on any likely profiles, no prejudicial State action against individuals should be taken solely on the basis of data mining information. It can raise only bare suspicions. Data mining information, in this regard, is analogous to "heat imaging" information about houses: "[A]t present no warrant could ever properly be granted solely on the basis of a FLIR image."[212] The information may support or corroborate other information, forming part of a valid basis for State action.

Automatically generated profiles have a further weakness. While profiles may be valid and may concern actually existing sets of correlations, the profiles may not be significant or useful. Automatically generated profiles must be reviewed and assessed by human analysts, to confirm their significance.[213] Human analyst evidence of significance should be required before anonymized profile evidence is linked (on the authority of a court order) to an identifiable individual.[214]

6. MISUSES

Data mining may be misused. The species of misuse were canvassed in Part IV.B.1 above.

One type of misuse is "mission creep," whereby data mining, which could have a proper use in counter-terrorism operations, is extended into investigations of any type of offence.[215] Mission creep could be controlled through the following means:

(a) The procedure may be made available only for listed offences, as is the case, for example, for authorizations for electronic interceptions of private

---

210 *R. v. Finta*, [1994] 1 S.C.R. 701 at 857.

211 For an interesting and useful discussion of innovations regarding expert evidence in civil contexts, see Alberta Rules of Court Project, Discovery and Evidence Committee, *Consultation Memorandum 12.3: Expert Evidence and "Independent" Medical Examinations* (February 2003), online: Alberta Law Reform Institute <www.law.ualberta.ca/alri/pdfs/cnslt_memo/cm12-3.pdf>.

212 *Tessling*, *supra* note 102 at para. 55.

213 Taipale, *supra* note 25 at 72, 19, 32.

214 Miller, *supra* note 37 at 5.

215 Seifert, *supra* note 22 at 12.

communications,[216] forensic DNA analysis warrants,[217] or investigative detentions.[218] The last example, which restricts an extraordinary procedure to "terrorism offences," might be a good model — setting aside the multitude of concerns about the breadth of the foundational notion of "terrorist activity" and the scope of conduct characterizable as terrorism offences.[219]

(b) The procedure may be initiated only on the "chop" of a senior government official or delegate, as is the case, for example, for authorizations for electronic interceptions of private communications,[220] dangerous offender applications,[221] or investigative hearings.[222]

(c) The procedure may be executed only by a designated agency and should not be available to peace officers generally. If the tool is not available to law enforcement personnel generally, then the temptation to use the tool too broadly is somewhat abated. Currently, the *Security Offences Act* charges the RCMP with responsibility for investigating national security offences.[223] It might be argued that the responsibility for data mining, which produces more information than evidence, should be given to the Canadian Security Intelligence Service (CSIS). This institutional dedication of the procedure would assist in keeping the tool from being expanded into other law enforcement matters.

(d) The use of the information could be statutorily restricted to use in the investigation of the designated offences. This tactic is employed by the forensic DNA warrant provisions.[224] Contravention of the use restrictions is an offence.[225]

(e) A retention schedule could be established for information derived from data mining. Again, this tactic is employed by the forensic DNA warrant provisions.[226] If the records gained from data mining are destroyed after a specified period, unless the records are being used for a particular investigation, the opportunity to access the records later for other purposes is reduced.

Data mining may be misused by individuals, whether they are authorized users or unauthorized users, such as hackers. It has been argued that individual misuses may be

---

216 *Criminal Code*, *supra* note 140, s. 183, definition of "offence."

217 *Ibid.*, s. 487.04.

218 *Ibid.*, s. 2, definition of "terrorism offence." The investigative detention procedure was, with the addition of use and derivative use immunities relating to extradition and deportation hearings, held to be constitutional by the Supreme Court in *Application under s. 83.28*, *supra* note 179.

219 See *e.g.* Kent Roach, "The New Terrorism Offences and the Criminal Law" in Daniels, Macklem & Roach, *supra* note 10 at 151; Don Stuart, "The Dangers of Quick Fix Legislation in the Criminal Law: The Anti-Terrorism Bill C-36 should be Withdrawn" in Daniels, Macklem & Roach, *ibid.* at 205.

220 *Criminal Code*, *supra* note 140, s. 185(1).

221 *Ibid.*, s. 752.1(1).

222 *Ibid.*, s. 83.28(3).

223 R.S.C. 1985, c. S-7, s. 6.

224 *Criminal Code*, *supra* note 140, ss. 487.08(1) - (2.1).

225 *Ibid.*, ss. 487.08(3), (4).

226 *Ibid.*, s. 487.09.

controlled in three main ways. First, data mining systems should have "robust permissioning structures," making unauthorized access difficult.[227] Credentials could be established by password or biometrics. Systems should have different access levels, so that very few individuals would have access to entire systems. Second, systems should track usage and usage logs should be both tamper-resistent and tamper-evident.[228] This creates an "electronic paper trail."[229] Finally, systems should be periodically audited.[230] Specific technological means to prevent misuse by individuals should not be legislated, since technology changes rapidly and the legislation could be rapidly outdated. Instead, legislation might provide that data mining systems shall be subject to security measures prescribed in regulations.

7. OVERSIGHT

Even if data mining will be useful to counter-terrorism operations and even if it can be regulated in a way that mitigates the risks it poses, its very existence might still produce the chilling effect or the impairment of trust described in Part IV.B.2 above. These risks may be addressed by making the use of the process as open and accountable as possible. A Markle Foundation report accurately identified the need to engender public trust in the use of data mining. The public must understand why government needs the information, know what government will do with the information, and be confident that individuals' rights are not being abused.[231]

One technique used for electronic interceptions of private communications and for the extraordinary anti-terrorism tools is the requirement to file annual reports with Parliament or the legislatures.[232] This at least provides a foundation for political accountability. Alternatively, and perhaps more effectively, a new Parliamentary body could be constituted to monitor the use of data mining and other designated high-technology searches and seizures.[233]

Oversight by individuals would be enhanced if the legislation establishing data mining created a civil action for misuse of the process.[234] To aid plaintiffs, statutory damages for violations of privacy rights could be established.[235] Such damages would be supplemental to any recovery for false imprisonment or other injuries consequent on erroneous profiling. The legislation could establish concurrent jurisdiction for hearing the statutory cause of action

---

227 First Markle Report, *supra* note 22 at 17; Ilam & Atkinson, *supra* note 166 at 4; Second Markle Report, *supra* note 6 at 35; Taipale, *supra* note 25 at 73.

228 ISAT, *supra* note 192 at 7, 13.

229 Heymann & Kayyem, *supra* note 192 at 80; TAPAC, *supra* note 3 at 49; Taipale, *supra* note 25 at 20.

230 Loukidelis, "Information Technology," *supra* note 62 at 14.

231 Second Markle Report, *supra* note 6 at 15; Loukidelis, "Information Technology," *ibid.* at 15.

232 See *Criminal Code*, *supra* note 140, ss. 195 and 83.31, respectively; and see Cavoukian, "National Security," *supra* note 17 at 54; Heymann & Kayyem, *supra* note 192 at 86; Miller, *supra* note 37 at 5; Letter from David Loukidelis to the Honourable Anne McLellan, the Honourable Irwin Cotler & the Honourable David Emerson (8 April 2005), online: Office of the Information and Privacy Commissioner of British Columbia <www.oipcbc.org/pdfs/public/16763lawfulaccessltr(April8 -2005).pdf> at 2 [Loukidelis, letter].

233 Loukidelis, letter, *ibid.* at 3; Cockfield, "Watchers," *supra* note 42 at 402-403.

234 Taipale, *supra* note 25 at 73, 20, n. 60.

235 The precedent of statutory damages in copyright matters might be emulated (see *Copyright Act*, R.S.C. 1985, c. C-42, ss. 38.1, 38.2).

in the provincial courts and the trial division of the Federal Court of Canada.[236] In addition, the public should be clearly informed that misuse of the data mining process supports a complaint to the relevant oversight body — whether Security Intelligence Review Committee in the case of CSIS,[237] or the Public Complaints Commission in the case of the RCMP.[238]

A practical means of constraining data mining is to require that it disrupt custodians as little as is reasonably possible, that the State compensate custodians for any costs incurred in making data available for mining, and that the State indemnify custodians for any damage awards or fines incurred as a result of disclosing information in accordance with authorized data mining.[239]

C. SUMMARY OF THE APPLICATION OF THE *HUNTER* V. *SOUTHAM* STANDARDS

The preceding discussion may be summarized as follows:

1. If individuals have no reasonable expectations of privacy in information in the hands of custodians, the State may data mine the information.

2. If data mining is contemplated for any information which may or may not support reasonable expectations of privacy, an application should be made to a judge for a determination of whether the information supports reasonable expectations of privacy.

3. If individuals do have reasonable expectations of privacy in information in the hands of custodians, the information may be data mined only if the data mining is judicially authorized before the data mining takes place. The authorization procedure has two stages:

   (a) An application must be made, on the basis of sworn or affirmed evidence, setting out

      (i) the particular purpose to be served by the data mining,

      (ii) how the data to be mined will contribute to that purpose, with regard to

         (A) the state of the data to be mined,

         (B) the capacity of the data to be reliably aggregated for mining,

         (C) the contributions of the particular databases sought to be mined, and

         (D) the reliability of the profiles to be employed;

---

236 *Ibid.*, s. 37.
237 *Canadian Security Intelligence Service Act*, R.S.C. 1985, c. C-23, s. 34.
238 *Royal Canadian Mounted Police Act*, R.S.C. 1985, c. R-10, Part VI.
239 Heymann & Kayyem, *supra* note 192 at 80; TAPAC, *supra* note 3 at 51.

(iii) the necessity of data mining for the investigation and why other procedures are not reasonably available or whether other procedures are already being pursued; and

(iv) the means by which privacy will be protected in the course of the data mining, with particular regard to the anonymization of results or the de-linking of information from identifiable individuals.

The judge (with the input of any court-appointed expert or experts) must be satisfied on a balance of probabilities that the data mining will achieve the intended results while protecting privacy to the requisite degree.

(b) The results of the data mining must be reviewed by analysts. If the results are deemed significant, an application may be made, on the basis of sworn or affirmed evidence, detailing that significance and the error rates. If a judge finds on a balance of probabilities that the results are significant, the risks of error are not excessive and measures are in place to deal with errors, the judge may permit the results of the data mining to be linked to individually identifying information.

4. The legislation establishing data mining should contain additional features minimizing the risks of data mining:

(a) State action against individuals on the basis of data mining results alone should be forbidden;

(b) data mining should be available only in relation to designated offences;

(c) the data mining process should be initiated only with the consent or approval of a senior government official;

(d) data mining should be run only out of the offices of a designated agency, such as CSIS;

(e) the use of information gained through data mining should be restricted to designated offences;

(f) information gained through data mining that is not actively in use should be destroyed;

(g) data mining systems should have appropriate security features, as established by regulation;

(h) data mining activities should be subject to an annual reporting obligation, whether to Parliament (or the provincial legislatures or both) or to a new political oversight body;

(i) a civil action for misuses of data mining should be established; and

(j) data mining should interfere with custodians as little as reasonably possible, and the State should be obligated to compensate or indemnify custodians for compliance expenses.

VII. CONCLUSION

As one might gather from the preceding summary, the regulation of data mining is a formidable task, although not dissimilar to the tasks of regulating other new-technology-based searches or seizures. It may be that the technology of data mining is not now and in the foreseeable future will not be at the levels required to satisfy the applicable constitutional standards. The technology, then, should not be used. It may be that the burdens of complying with a regulatory regime of the sort I have outlined will outweigh the benefits foreseeable from data mining. The technology, then, should not be used. Data mining may hold investigatory promise; it certainly promises social, political and personal risks. It is a technology that should not be deployed unless its use is very carefully managed. The Markle Foundation has provided a good concluding description of the issues:

> Data mining can be a useful tool. But it is also a tool that invites concern about invasion of privacy. Extravagant claims have been made about the potential uses of data mining, matched by similarly extravagant notions of the vast private or public databases that should be opened to such journeys of exploration. Neither the real needs nor the real capabilities are so exotic. Though there are areas where more data may need to be collected, the immediate challenge is to make more effective use of the mountains of data that are already in government hands or publicly available. Data mining, like any other government data analysis, should occur where there is a focused and demonstrable need to know, balanced against the dangers to civil liberties. It should be purposeful and responsible.[240]

[240] First Markle Report, *supra* note 22 at 27.