

THE COMING TECHNOLOGY OF KNOWLEDGE DISCOVERY: A FINAL BLOW TO PRIVACY PROTECTION?

Charles Weiss*

Expected advances in information technology for pattern recognition (“connecting the dots”) are likely to greatly increase the scope and sophistication of already powerful techniques for zeroing in on individual citizens and generating a detailed profile of their lives. These techniques may well prove to be valuable tools for the fight against terrorism. At the same time, they are likely to sweep away what little is left of U.S. privacy law, which has already been weakened by the dramatic advances in technology for surveillance and data mining, by the elimination of barriers between data acquisition by government and the much less regulated private sector, and by the whittling away of Fourth Amendment protections by the courts.

These dramatic developments demand a thorough reexamination of existing legal precedent. Four critical steps in pattern-based searches offer possible intervention points for the protection of civil liberties in such pattern-based searches: data acquisition, data sharing, pattern identification, and deanonymization of previously anonymized data. In each of these areas, however, legal protections have withered in the face of

* Charles Weiss is Distinguished Professor and Chair of Science, Technology, and International Affairs at the Edmund A. Walsh School of Foreign Service at Georgetown University, a post he has held since 1997. A Harvard-trained biochemical physicist (B.A., chemistry and physics, *summa cum laude*, 1959; Ph.D., chemical physics and biochemistry, 1965), he was the first Science and Technology Adviser to the World Bank, and served in that capacity from 1971 to 1985. From 1985 to 1997, he helped launch and taught in the Program in Science and Technology Policy at the Woodrow Wilson School at Princeton University. He also has held positions as Visiting Professor at the University of Pennsylvania, Visiting Scholar at the University of California (Berkeley), Course Director at the Foreign Service Institute of the U.S. Department of State, and Professorial Lecturer at the Nitze School of Advanced International Studies at Johns Hopkins University. He is the author of numerous publications on international science and technology policy. This Article was originally prepared for the conference on “The Future of Privacy,” held at the Center for International and Security Studies of the School of Public Affairs at the University of Maryland on October 30, 2003. The author would like to thank Neal Pollard, Joseph Onek, and Julie Cohen for helpful advice; Jonathan Lalley for his excellent research assistance; and Bruce Canetti for going beyond the call of duty as an Articles Editor.

advances in technology. A series of U.S. Supreme Court decisions have whittled away at the doctrine of “reasonable expectation of privacy” that once placed limitations on data acquisition from the enormous variety of sensors now available for government surveillance. Recent legislation has eliminated remaining barriers on the sharing of information between government and the private sector, thereby undermining previous legal distinctions between government surveillance and private data mining for commercial purposes. There is little jurisprudence on data aggregation or on the government’s ability to identify patterns in data to which it legally has access. In the future, deanonymization of personally identifiable data is likely to be a critical means of privacy protection against possible abuse of pattern-based searches. For example, it was to have been a key element in the federal government’s defunct Total Information Awareness research program.

*If this protection is to be effective, deanonymization must be controlled by a neutral third party, such as a FISA court, and must be subject to specific standards of proof. For this purpose, a three-tiered standard of proof is suggested for deanonymization of patterns thought to be indicative of possible terrorist activity. There is a danger that the recent Supreme Court decision in *Hübel v. Sixth Judicial Court of Nevada* may create a possible precedent for “reasonable suspicion” as a general deanonymization standard. To the contrary, the most usual standard should be “reasonable indication,” the standard used by the FBI to justify initiating an investigation. In cases involving the most serious forms of terrorist activity, “reasonable suspicion”—the standard for a minimally intrusive Terry stop—should apply. Patterns that rely on databases of sensitive personal information and that do not involve the most serious dangers should meet a standard of “reasonable belief” before they can be deanonymized.*

*The internet has eliminated the “de-facto privacy” shield provided by the sheer inconvenience of assembling publicly available data from a variety of scattered sources. This makes it essential to change the long-standing doctrine of *U.S. v. Miller* that information, once revealed to anyone for any reason, is no longer private. This doctrine is inimical to American notions of privacy and is likely to give rise to a dangerous public backlash once its implications for civil liberties come to be widely understood. Until this happens, however, Congress and the courts are unlikely to take a stand on this issue. To avoid this backlash, the intelligence community, working with Congress, should urgently consider what access it really needs to material on the personal lives of U.S. citizens, and what limitations it can accept on the use of this material, independently of what is now allowed by legal precedent.*

I. INTRODUCTION

Research now underway may soon make it possible for the government to zero in on any citizen and create—easily, quickly, and cheaply—a detailed picture of his or her life by integrating information from a vast array of sensors. These sensors range from video cameras¹ to intercepted telephone or Internet conversations² to financial and medical records.³ This knowledge-discovery technology would be much more powerful and wider in scope than the data-mining technology now in common use.⁴ The full development of this technology is most likely five to ten years away; therefore it is impossible to know its exact capabilities or whether it will even work at all.⁵ Still, it is not too early to begin considering the implications of this technology for civil liberties and privacy.

The threat of terrorism is real, and this kind of intelligence-support technology may well be an essential and effective tool for detecting terrorists before they have a chance to act. If this technology can be made to work as its promoters hope, it could provide governments with capabilities of unprecedented power and scope. If implemented without scrupulous attention to civil liberties, however, it could become the technological underpinning of an efficient totalitarian state. Even if this technology were implemented without mistake and without abuse, such a capability would call not only for a reevaluation of the balance between national security and individual privacy, but also for a systematic reexamination of the legal meaning of privacy and of the legal, constitutional, and practical safeguards that protect it.

Before the information revolution, the government could assemble a comprehensive picture of the life of an individual only by investing major amounts of time, money, and skilled personnel in acquiring and aggregating data from widely scattered sources—even if these sources

1. See *United States v. Biasucci*, 786 F.2d 504, 510–12 (2d Cir. 1986) (finding video surveillance constitutional when the affidavit supporting the warrant showed its necessity for a difficult criminal investigation of organized crime). *But see United States v. Nerber*, 222 F.3d 597, 603–06 (4th Cir. 2000) (finding video surveillance unconstitutional when defendants were left alone in an informant's hotel room because they had a legitimate expectation of privacy).

2. See, e.g., *Griggs-Ryan v. Smith*, 904 F.2d 112, 118–19 (1st Cir. 1990).

3. *Cal. Bankers Ass'n v. Schultz*, 416 U.S. 21, 75–78 (1974). For a general discussion of surveillance and the Department of Homeland Security, see MATTHEW BRZEZINSKI, *FORTRESS AMERICA: ON THE FRONT LINES OF HOMELAND SECURITY—AN INSIDE LOOK AT THE COMING SURVEILLANCE STATE* (2004). See also ROBERT O'HARROW, *NO PLACE TO HIDE: BEHIND THE SCENES OF OUR EMERGING SURVEILLANCE SOCIETY* (2005) (discussing the interaction between private information, anti-terror governmental practices, and knowledge technology companies).

4. See generally JEFFREY ROSEN, *THE NAKED CROWD: RECLAIMING SECURITY AND FREEDOM IN AN ANXIOUS AGE* (2004). See also K.A. Taipale, *Data Mining and Domestic Security: Connecting the Dots To Make Sense of Data*, 5 *COLUM. SCI. & TECH. L. REV.* 2, 53 n.223 (2003) [hereinafter Taipale, *Data Mining and Domestic Security*], at <http://www.stlr.org/cite.cgi?volume=5&article=2> (examining advanced technologies for aggregating and analyzing data in the context of domestic security).

5. For an example of a company using and developing this technology, see Convera, *About Convera's RetrievalWare*, at <http://www.convera.com/products/> (last visited Dec. 28, 2004).

were open to the public.⁶ The government's authority to do so was further limited by a piecemeal and disjointed, yet modestly effective patchwork of constitutional, statutory, and institutional safeguards that each focused on a particular step in the acquisition and integration of knowledge.⁷

The situation is now radically different. Advances in information technology, bolstered by legislation and court decisions, have eroded the effectiveness of existing constitutional and statutory protections.⁸ The technologies in development could soon overwhelm the effectiveness of what is left of those protections. There has already been an enormous increase in the power of the private sector to use data-mining and knowledge-discovery technology to gather information about individual citizens and to analyze the data so obtained for patterns of behavior that will make marketing more efficient.⁹ Furthermore, the institutional limitations on sharing data among government agencies—which in many cases amount to inefficiencies in the conduct of government—are being revised to make the fight against terrorism more effective.¹⁰ At the same time, this new efficiency puts additional pressure on the privacy protections that remain.¹¹

6. Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137, 1140–55 (2002) [hereinafter Solove, *Access and Aggregation*].

7. See The Privacy Act of 1974, 5 U.S.C. § 552a(b) (2000); *Kyllo v. United States*, 533 U.S. 27, 35–40 (2001); *In re Sealed Case*, 310 F.3d 717, 727 (Foreign Intel. Surv. Ct. Rev. 2002).

8. See generally SAMUEL DASH, *THE INTRUDERS: UNREASONABLE SEARCHES AND SEIZURES FROM KING JOHN TO JOHN ASHCROFT* (2004); Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083 (2002) [hereinafter Solove, *Digital Dossiers*]. For a critical analysis of present electronic surveillance law, see Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1264, 1278–98 (2004).

9. This Article distinguishes among three different concepts: data mining, search technology, and knowledge discovery. (1) Data mining is a well-established technology that applies one or more algorithms to well-known, structured data—for example, commercial records of a customer's buying or traveling habits. It is “analyzing large and complex databases to discover useful, previously unknown knowledge.” David Jensen, *Data Mining in Networks* (Dec. 11, 2002), at <http://kdl.cs.umass.edu/people/jensen/papers/nrcdbsse02/slide10.html> (containing an excellent overview of data-mining technology in an annotated slide presentation). It should be noted that Jensen uses the term “knowledge discovery” as synonymous with data mining, contrary to the usage in this Article. (2) Search technology applies a single well-defined algorithm to very large quantities of unstructured data, typically the content of documents and Web sites, to seek a well-established pattern. The Google algorithm, for example, uses a “relevance ranking” that includes a measure of how many links connect to a given Web site. Google, *Our Search: Google Technology*, at <http://www.google.com/technology/> (last visited Nov. 3, 2004). (3) Knowledge discovery is an evolving technology that uses multiple algorithms to integrate information presented in a variety of representations, including both structured and unstructured databases, geo-spatial data, images, and “stored meta-knowledge” (data about the nature or content of a data file that facilitates search and filtering). Knowledge-discovery technology seeks to recognize concepts even when they are expressed in different ways. See E-mail from Alianna J. Maren, Chief Scientist, Eagle Force University, to Charles Weiss, Professor of Science, Technology, and International Affairs, Georgetown University (May 14, 2004, 11:47:51 EST) (on file with author).

10. See *In re Sealed Case*, 310 F.3d at 722–27.

11. The age and quality of hardware and software available to government officials, even in the intelligence community, constitutes yet another such inefficiency.

In Section II, the Article reviews the history of privacy law and the diminishing limits the law places on intrusions by the government and the private sector. Section III summarizes recent and likely future developments in the four aspects of knowledge discovery: data acquisition, data sharing, pattern identification, and deanonymization. Section IV explores the application of existing privacy law to each of these four aspects of knowledge discovery. Section V suggests a new set of legal standards affording a more appropriate balance between privacy and security, including standards of proof applicable to deanonymization in different situations. Section VI discusses the implications of the interaction between knowledge-discovery technology and privacy law for the intelligence and law enforcement communities. Finally, Section VII concludes that the legislative and judicial branches, working together with the intelligence and law enforcement communities, need to establish new law to protect civil liberties from this technology even as the nation uses it to fight the terrorist threat.

II. PRIVACY LAW PROTECTION AND ITS LIMITS

American jurisprudence regarding privacy historically has distinguished between (1) privacy as it may be invaded by the efforts of the private sector to gather information for commercial gain, and (2) privacy as it may be invaded by the government to obtain information for purposes of surveillance of possible criminal activity,¹² or of possible actions that could endanger national security or public order, such as terrorist activities. While the concepts underlying these two kinds of protections have begun to influence each other, they rest on quite different philosophical and constitutional foundations.

A. *Privacy Vis-à-Vis the Private Sector*

Until a few decades ago, privacy in U.S. jurisprudence was treated as a branch of tort law.¹³ Four torts were defined involving intrusion into privacy: intrusion upon seclusion, appropriation of name or likeness, publicity given to private life, and publicity placing a person in false light.¹⁴ There has been a recurring exploration of the tension between public and private rights to privacy.¹⁵ The Supreme Court also has varied

12. See *Katz v. United States*, 389 U.S. 347, 350–51 n.5 (1967).

13. RESTATEMENT (SECOND) OF TORTS § 652 (1977).

14. William Prosser, *Privacy*, 48 CAL. L. REV. 383, 389, 407 (1960). This classification forms the basis for the treatment of the subject by the American Law Institute. RESTATEMENT (SECOND) OF TORTS §§ 652(A)–652(I) (1977).

15. See, e.g., Daniel Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393 (2001) [hereinafter Solove, *Privacy and Power*] (discussing the database privacy problem as it applies to public and private databases); Rachel K. Zimmerman, *The Way the “Cookies” Crumble: Internet Privacy and Data Protection in the Twenty-First Century*, 4 N.Y.U. J. LEGIS. & PUB. POL’Y 439 (2000–2001) (discussing the difficulty traditional privacy law has had in keeping pace with technological advances).

its definition of privacy.¹⁶ Privacy has been conceived as a right of control, or the “claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.”¹⁷ This theory logically leads to the idea of a market for personal information in which individuals may disclose information to a company in exchange for a consideration they find satisfactory—such as a discount, the chance to win a prize, a subscription to a monthly newsletter, notification of future products, or even nothing at all. In this model, individuals exercise their rights over their personal information by doing business with companies of whose information policies they approve. Once information has been communicated for whatever reason, it becomes a business asset belonging to the company. It can be shared, sold, or given away in accordance with the business needs of the company, subject to statutes restricting the sharing of specific kinds of information of special sensitivity.¹⁸

This concept of a market for personal information has proven to be unworkable because of irreparable market imperfections.¹⁹ For the most part, statutes purporting to assure an informed consumer of his privacy offer only very limited protection.²⁰ Individuals have neither the time to absorb the mountains of privacy brochures they receive (e.g., from corporations covered by the Fair Credit Reporting Act), nor the technical expertise or the market power to force changes in policies of which they disapprove.

The fundamental difficulty here lies in the long-standing legal doctrine that business records are the property of the business, not of the person that they concern.²¹ Personal checks, for example, are the property of the issuing bank and are not subject to a reasonable expectation of privacy on the part of the check-writer.²² Similarly, the Supreme Court has held that a pen register can be installed in a telephone company exchange to record the numbers called from a particular home telephone without a search warrant, on the grounds that the telephone company could be recording these numbers for business reasons (this in the days before caller ID became common).²³

16. TECHNOLOGY AND PRIVACY ADVISORY COMMITTEE, SAFEGUARDING PRIVACY IN THE FIGHT AGAINST TERRORISM 21–32 (Mar. 2004), available at <http://www.cdt.org/security/usapatriot/20040300tapac.pdf> [hereinafter TAPAC REPORT].

17. ALAN F. WESTIN, *PRIVACY AND FREEDOM* 7 (1967).

18. See, e.g., Right to Financial Privacy Act, 12 U.S.C. §§ 3401–3422 (2000); Fair Credit Reporting Act, 15 U.S.C. § 1681 (2000).

19. Julie E. Cohen, *DRM and Privacy*, 18 BERKELEY TECH. L. J. 575, 605–09 (2003), available at <http://www.law.berkeley.edu/journals/btlj/articles/vol18/Cohen.stripped.pdf> (last visited Nov. 10, 2004).

20. See generally Solove, *Privacy and Power*, *supra* note 15; Zimmerman, *supra* note 15; TAPAC REPORT, *supra* note 16.

21. See *United States v. Miller*, 425 U.S. 435, 440–43 (1976).

22. See *Cal. Bankers Ass’n v. Schultz*, 416 U.S. 21, 75–78 (1974). In response to this decision, Congress enacted the Right to Privacy Act of 1978, setting forth procedures for government access to the customer records of financial institutions. 12 U.S.C. §§ 3401–3403 (2000).

23. *Smith v. Maryland*, 442 U.S. 735, 737–46 (1979).

From these precedents, there emerged the doctrine that once a piece of personal information is revealed to another person or organization for whatever purpose, it is no longer private or subject to the control of the person concerned.²⁴ In Europe, by contrast, an individual retains some control over his or her personal information even after it has been disclosed.²⁵ In the United States, with specific exceptions provided by law, it is not possible to share information with a specific institution for a specific purpose and be sure that the information will not be further disclosed and will not be made available to the government without a warrant. On the contrary, voluminous amounts of personal information can be made available upon seemingly innocent disclosure—financial information disclosed to a bank or mortgage company; medical information disclosed to a doctor, laboratory, or insurance company; information on books or videos purchased or borrowed from a bookstore, library, or rental agency; records of communications or television viewing in the hands of a telephone or cable company—subject to each company’s information policies and a variety of statutory restrictions that are specific to each sector, and that, in the aggregate, fall considerably short of full protection.²⁶ Whereas statutes at least provide a stable legal regime,²⁷ company policies are subject to change—especially if the company finds itself in financial difficulties or is sold, in which case its store of accumulated information can become an important source of cash.

24. *Miller*, 425 U.S. at 443 (“[T]he Fourth Amendment does not prohibit the obtaining of information revealed to a third party and conveyed by him to Government authorities, even if the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in a third party will not be betrayed.”); see also Taipale, *Data Mining and Domestic Security*, *supra* note 4, at 53 n.223; TAPAC REPORT, *supra* note 16, at 31–32. *But cf.* Patricia L. Bellia, *Surveillance Law Through Cyberlaw’s Lens*, 72 GEO. WASH. L. REV. 1375, 1403–09 (2004) (arguing that *Miller* and similar cases do not hold that all communications stored by third parties lack an expectation of privacy, but rather that these cases only offer certain factors (for example, the user’s purpose in conveying the information) that determine when the expectation is no longer valid).

25. In brief, the Organization for Economic Co-operation and Development (“OECD”) rules, which have been enacted with variations by the various European countries, require a well-defined purpose for the collection of information by private firms, a limit on the amount of information to be gathered, and external oversight of the process, with exclusions or exceptions for national security or intelligence. ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT, OECD GUIDELINES ON THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA 13–18 (2002) [hereinafter OECD GUIDELINES].

26. Peter P. Swire, *The System of Foreign Surveillance Law*, 72 GEO. WASH. L. REV. 1306, 1323 n.112, 1332–33 nn.177–80, 1356–59 [hereinafter Swire, *Foreign Surveillance*], available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=586616 (last visited Nov. 3, 2004). Moreover, the FBI can obtain access to many of these records by means of a “national security letter,” without a warrant and without informing the subject of the investigation. *Id.* It is actually illegal to inform the subject. *Id.* This led to the “revolt of the librarians,” who in many parts of the country refused to make their borrowers’ records available to law enforcement officials. *Id.* See generally GINA MARIE STEVENS, CONG. RES. SERVICE, PRIVACY: TOTAL INFORMATION AWARENESS AND RELATED INFORMATION ACCESS, COLLECTION AND PROTECTION LAWS (Mar. 21, 2003), available at <http://www.fas.org/irp/crs/RL31730.pdf>.

27. For a summary of the statutory limitations on data sharing, see TAPAC REPORT, *supra* note 16, at 111 n.177.

B. Privacy Vis-à-Vis Government Surveillance

The limits on the power of the government to intrude on privacy arise from the First, Fourth, and Fifth Amendments to the U.S. Constitution. One may distinguish among numerous aspects of privacy. First, secrecy lies in the privacy of place—a person's home and its extension to places or data repositories where he or she has a "reasonable expectation of privacy," such as a closed telephone booth,²⁸ a public rest room,²⁹ and (perhaps) a computer or a pager.³⁰ Autonomy, on the other hand, concerns the privacy of personal behavior—the "right to be left alone" in one's "private space"—as defined by Supreme Court decisions establishing the rights to contraception,³¹ abortion,³² and consensual sex between adults.³³ Finally, anonymity is the privacy of anonymous communication, especially in political speech, and of anonymous membership in associations.³⁴

Most of the legal literature concerning privacy has focused on private databases and has sought to achieve a balance between the commercial advantages of data-mining techniques for both business and customers on the one hand, and the individual's right to privacy on the other.³⁵ Implicit in these limitations is the assumption that government surveillance is separate and distinct from commercial intelligence. This assumption, however, has been made nearly obsolete by the Patriot Act.³⁶ The Patriot Act extends previous legislation and allows the government to purchase or lease access to private databases in order to search them for patterns of illegal behavior, as long as "a significant purpose" of the search is to collect information related to foreign intelligence—including signs of possible terrorist activity.³⁷

28. *Katz v. United States*, 389 U.S. 347, 352–53 (1967).

29. *United States v. Delaney*, 52 F.3d 182, 188 (8th Cir. 1995).

30. *Trulock v. Freeh*, 275 F.3d 391, 403 (4th Cir. 2001) (finding a reasonable expectation of privacy in password-protected computer files); *United States v. Chan*, 830 F. Supp. 531, 534–36 (N.D. Cal. 1993) (finding a reasonable expectation of privacy in numbers stored in pager). *But see* *Muick v. Glenayre Elecs.*, 280 F.3d 741, 743 (7th Cir. 2002) (finding no reasonable expectation of privacy by employee in laptop computer lent by employer when employer announced that laptops could be inspected).

31. *Griswold v. Connecticut*, 381 U.S. 479, 485–86 (1965).

32. *Roe v. Wade*, 410 U.S. 113, 153 (1973).

33. *Lawrence v. Texas*, 539 U.S. 558, 570–73 (2003).

34. *Whalen v. Roe*, 429 U.S. 589, 599 (1977) (quoting Philip B. Kurland, *The Private I*, U. CHI. MAG., Autumn 1976, at 7–8); Cohen, *supra* note 19, at 598; Taipale, *Data Mining and Domestic Security*, *supra* note 4, at 56; cf. Helen Nissenbaum, *Privacy As Contextual Integrity*, 79 WASH. L. REV. 119, 128 (2004) (making a somewhat similar distinction among privacy against excessive intrusion, privacy of sensitive information, and privacy of spaces or spheres deemed to be confidential).

35. *See, e.g.*, Solove, *Privacy and Power*, *supra* note 15, at 1398.

36. *Uniting and Strengthening America By Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001* ("the Patriot Act"), Pub. L. 107-56, 115 Stat. 272.

37. *Id.* at § 218 (codified as amended at 50 U.S.C. § 1804(a)(7)(B) (Supp. 2001)). *See generally* TAPAC REPORT, *supra* note 16, at 26 (summarizing the Patriot Act); Swire, *Foreign Surveillance*, *supra* note 26, at 1330–34 (criticizing the Patriot Act). For a review of statutes that concern private information in commercial databases and government access to that information, see James X.

Most Americans instinctively feel that the government should not know more about them than is absolutely necessary. However, there is, in fact, no overall limit on the quantity of information that the government may gather on any individual from public sources, even in the absence of an articulated need for this information, let alone a warrant from a neutral third party.³⁸

From a technical point of view, information searches may be divided into two basic types: subject-based searches, which seek information on a particular individual or organization; and pattern-based searches, which are “fishing expeditions” to seek patterns associated with terrorism or other criminal behavior, but not based on a previous “particularized suspicion.”³⁹

The limitations on subject-based searches of databases by the government historically have been practical rather than legal or institutional.⁴⁰ It has long been possible to assemble a detailed picture of a person’s life and activities by examining a variety of scattered public records: real estate, court, vital statistics, etc. Until recently, this was expensive in terms of personnel, money, and time, resulting in a sort of “de facto privacy.”⁴¹ The Internet has enabled the government to overcome these limitations by substantially reducing search costs. Meanwhile, the growth of commercial databases has greatly expanded the quantity of available information.⁴²

The limitations on pattern-based searches are both institutional and practical. The major users of such searches are intelligence and law enforcement agencies. Intelligence agencies seek to provide advance warning of events that might endanger national security.⁴³ Their searches are subject to a variety of constraints, collectively known as “minimization,” which require, for example, that the names of U.S. citizens not suspected of activities detrimental to the United States be deleted from intelligence agency records.⁴⁴ Law enforcement agencies use pattern-based searches to identify patterns that may be indicative of criminal activity, drawing attention to situations requiring further

Dempsey & Lara M. Flint, *The Future of Internet Surveillance Law, Commercial Data, and National Security*, 72 GEO. WASH. L. REV. 1459, 1476–83 (2004).

38. See Taylor v. Miller, 853 F. Supp. 305, 306–08 (W.D. Wis. 1994).

39. Taipale, *Data Mining and Domestic Security*, *supra* note 4, at 57–58.

40. See *id.* at 58–59.

41. Jensen, *supra* note 9, at commentary accompanying slide 13.

42. See, e.g., Solove, *Access and Aggregation*, *supra* note 6, at 1150–55.

43. See generally JEFFREY T. RICHELSON, *THE U.S. INTELLIGENCE COMMUNITY* (4th ed. 1999) (examining in detail all aspects of intelligence work by the U.S. government).

44. See, e.g., Exec. Order No. 12,333, 46 Fed. Reg. 59,941 (Dec. 4, 1981), available at <http://www.cia.gov/cia/information/eo12333.html>; Memorandum from John Ashcroft, U.S. Attorney General, to Heads of Department Programs, Guidelines for Disclosure of Grand Jury and Electronic and Oral Interception Information Identifying United States Persons (Sept. 23, 2002), at <http://www.usdoj.gov/olp/section203.pdf>.

scrutiny.⁴⁵ There are no legal limits to such searches as long as they are based on data to which police have legal access.

III. THE TECHNOLOGY OF KNOWLEDGE DISCOVERY

In order to analyze the implications for privacy, this Article divides the acquisition and integration of personal information into four steps. Each step represents both a critical step in knowledge discovery and a possible intervention point for the protection of civil liberties. The steps are: (1) data acquisition; (2) data sharing; (3) pattern identification, which involves querying available databases and identifying patterns; and, if the resulting patterns are anonymous, (4) deanonymization, or the “selective revelation” of the identity of individuals connected to these patterns.

In each of these steps, technology is advancing rapidly. To further accelerate this advance, the federal government has placed a major priority on funding the development of this technology.⁴⁶ The technical and legal developments relating to each of these steps have been thoroughly reviewed in articles by Taipale⁴⁷ and by Rosenzweig.⁴⁸ These developments are summarized briefly in the next four sections of this Article.

A. Data Acquisition

The technical means of data acquisition have vastly expanded in recent decades. These new methods include: high resolution photography taken by satellites or unmanned aerial drones in visible and infrared wavelengths; centrally monitored video surveillance in public places; face and gait recognition technology; technology for the monitoring of telecommunications and e-mail traffic; fingerprint scanners

45. For a discussion of recent efforts by law enforcement agencies to use pattern-based searches, see Dempsey & Flint, *supra* note 37, at 1468–69. See also William Welsh, *States, Feds Working to Fix MATRIX (Data Sharing)*, GOV'T COMPUTER NEWS, June 21, 2004, at 19 (discussing the success and failure of the Multistate Antiterrorism Regional Information Exchange System, a federal pilot program to share information among state law enforcement agencies).

46. See, e.g., Stephen M. Davis & Robert G. Hanski, *Governments Fund Technology Ventures in Homeland Security*, N.Y. L.J., June 16, 2003, at 9, available at <http://www.hewm.com/use/articles/technologyhomeland.pdf>; see also O'HARROW, *supra* note 3, at 247–80.

47. See Taipale, *Data Mining and Domestic Security*, *supra* note 4, at 2; K.A. Taipale, *Technology, Security and Privacy: The Fear of Frankenstein, the Myth of Privacy and the Lessons of King Ludd*, 7 YALE J. L. & TECH. 123 (2004–2005) [hereinafter Taipale, *Technology, Security, and Privacy*], available at <http://www.yjolt.org/>.

48. Paul Rosenzweig & Michael Scardaville, *The Need to Protect Civil Liberties While Combating Terrorism: Legal Principles and the Total Information Awareness Program*, The Heritage Foundation (Feb. 5, 2003) [hereinafter Rosenzweig & Scardaville, *The Need to Protect Civil Liberties*], available at <http://www.heritage.org/Research/HomelandDefense/lm6.cfm>; Paul Rosenzweig, *Proposals for Implementing the Terrorism Information Awareness System*, The Heritage Foundation (Aug. 7, 2003) [hereinafter Rosenzweig, *Terrorism Information Awareness*], available at <http://www.heritage.org/Research/HomelandDefense/lm8.cfm>.

and other biometrics; cell phone and automobile locators; implantable radio frequency chips; thermal imaging technology; and technology that analyzes DNA samples from bits of shed skin.⁴⁹ All of these technologies are under active development, but some work better than others.⁵⁰

B. Data Sharing

One may project the likely growth of data-sharing technology by examining current plans for two rather separate technologies. The first is technology to obtain diverse information stored in different formats by querying (or “mediating” among) various sources. The second is an architecture for knowledge discovery that retrieves data from both structured and unstructured sources in a form suitable for pattern detection. To be sure, alternative technical approaches are currently in development, but there is no way of knowing in advance which will succeed. Even so, these two plans illustrate the kind of policy problems that are likely to be raised by improved information technologies now on the horizon.

The first of these technologies was planned for development under the government’s recently canceled Total Information Awareness (“TIA”) Program of the Defense Advanced Research Projects Agency (“DARPA”).⁵¹ TIA was to have financed the development of technology to mediate among diverse sources of information, i.e., to query information collected by video surveillance, aerial photography, telecommunications or e-mail intercepts, internal intelligence reports, open print sources, and many other kinds of sources—each in a different electronic format—so that patterns could be extracted in support of intelligence analysis with minimum human intervention.⁵² The resulting technology would also have had wide application in fields other than

49. JAY STANLEY & BARRY STEINHARDT, BIGGER MONSTER, WEAKER CHAINS: THE GROWTH OF AN AMERICAN SURVEILLANCE SOCIETY 2, 3, 6, 7 (Jan. 15, 2003), at <http://www.aclu.org/Privacy/Privacy.cfm?ID=11573&c=39>.

50. *Id.*; ROSEN, *supra* note 4, at 98.

51. For a summary of the various research programs being financed by DARPA under the TIA umbrella, see TAPAC REPORT, *supra* note 16, at 106–08 nn.78–79. See also ROSEN, *supra* note 4, at 99–100. According to its former director, TIA research was intended to: 1) lead to technologies to help the intelligence analyst to assess and process data; 2) define what data to seek; 3) find relevant information; 4) understand what this information means; 5) define options regarding what actions to take; and 6) present the resulting analysis to the decision maker. See John M. Poindexter, *Security with Privacy*, available at <http://www.maxwell.syr.edu/campbell/Library%20Papers/Event%20papers/ISHS/Poindexter.pdf> (adapted from John M. Poindexter, *Finding the Face of Terror in Data*, N.Y. TIMES, Sept. 10, 2003, at A25) (last visited June 11, 2005). The TAPAC Report further notes that the classified addendum to the law abolishing the TIA program “makes it clear that TIA-like activities could continue to be pursued outside of the public’s view.” TAPAC REPORT, *supra* note 16, at 3. The committee was “informed of a variety of other activities in DoD [Department of Defense] and elsewhere that raise similar privacy issues [to those raised by TIA].” *Id.* at 3. For a non technical description of the TIA program, see O’HARROW, *supra* note 3, at 190–213.

52. Rosenzweig, *Terrorism Information Awareness*, *supra* note 48, at 14.

intelligence, such as medical research, in which data appear in similarly diverse formats.⁵³

The second plan, still pending with the Department of Defense (“the DoD”), calls for the development and implementation of an architecture for knowledge discovery as part of a much larger plan for so-called “net-centric enterprise services.”⁵⁴ Such architecture would make it possible to retrieve data from both structured and unstructured databases, as well as from nonverbal sources such as images and geospatial data, in a form suitable for pattern detection.⁵⁵ The DoD’s immediate objective is to make possible quick retrieval of data from a variety of internal sources in support of the DoD’s operational and logistical functions.⁵⁶ The added scope of this architecture, should it be successfully developed, would add considerably to the range of data available to an intelligence support system.

C. Pattern Identification

The TIA and knowledge-discovery technologies mentioned in the previous subsection also illustrate the likely path of technologies for identification of patterns indicative of terrorist activity.⁵⁷ The TIA program envisioned a powerful search engine capable of carrying out pattern-based searches by simultaneously querying a large number of databases, ranging from clandestine communication intercepts to commercial records, looking for patterns typical of terrorist activity.⁵⁸ These patterns were to be derived from the transactions (such as air ticket purchases, hotel and car rentals, increased communications traffic, etc.) required for hypothetical attacks (“templates”), simulated by so-called “red teams.”⁵⁹ From the point of view of an intelligence agency, the purpose of such a system would be to reduce the flood of incoming data to proportions manageable by a human analyst and to allow a more rational allocation of intelligence resources.

53. Taipale, *Technology, Security, and Privacy*, *supra* note 47, at 217–20.

54. Defense Information Systems Agency, U.S. Dep’t of Def., Core Services: Net-Centric Enterprise Services, at http://www.disa.mil/main/prodsol/cs_nces.html (last updated Aug. 19, 2004).

55. Maren, *supra* note 9.

56. *See id.*

57. Poindexter, *supra* note 51.

58. *See id.* at 2–3. Since the bandwidth and processing capacity of the overall TIA system would inevitably be limited, the search would probably begin by querying databases considered to be of highest priority. Databases may be labeled as high priority either because they contain information regarding transactions common to a large number of possible terrorist activities, or because they constitute “flashing beacons” pointing to a high likelihood of a specific kind of terrorist act (e.g., purchase of explosives or fissile material).

59. Pattern-based searches for advance indications of terrorist activity present particular difficulty because there are, fortunately, few terrorist incidents on which to model the template for such searches. *See* Poindexter, *supra* note 51. TIA researchers therefore employed “red teams” to simulate terrorist action and search for common features. *Id.* These red team exercises were to lead to the identification of those databases that are especially critical to the anti-terrorism effort. *See id.*

Knowledge-discovery technology would amplify these capabilities in that it would recognize concepts in many different verbal and nonverbal formulations. Knowledge-discovery technology is to operate at five steps or levels: (1) concept extraction, which involves identifying the object being observed; (2) concept correlation, or determining which concepts are typically found to occur together; (3) syntactic discovery, which concerns determining the word that best describes the connection between the concepts found to be correlated in the second step; (4) context-based discovery, meaning the identification of the broader context of the connection identified in step three; and (5) semantic discovery, or providing a coherent prose statement of the meaning of the pattern indicated by steps one through four.⁶⁰ Remarkably, technologies already exist that are capable of carrying out each of these five steps, albeit in limited contexts.⁶¹ This technology is likely to supplant the simpler database search methods now in use for civilian purposes, such as Google,⁶² within a few years.⁶³ Whereas search technologies can only identify situations in which certain key words occur together, successful knowledge-discovery technology would be able to identify both the reasons for the correlation and the context in which it arises. This would greatly enhance the ability of an intelligence support system to identify patterns.⁶⁴

TIA also included research on technology that could be used to protect civil liberties.⁶⁵ One possibility was a so-called “privacy appliance” that would intervene between the TIA engine and each database to be queried, and into which a variety of safeguards to individual privacy would be programmed.⁶⁶ This appliance was also supposed to authenticate the identity and verify the authorization of the person or organization making the query, as well as to encrypt data in such a way as to preserve relationships without revealing the identity of the individuals concerned.⁶⁷ The appliance was also to provide an

60. These steps are part of sensitive and protected, but not classified, information in several intelligence agencies. Respected professionals in the field acknowledge these steps, but official documents discussing them are unavailable at this time. For a company currently using and developing these tools, see generally EagleForce Associates, EagleForce Current News, at <http://www.theeagleforce.com/News.shtml> (last visited Dec. 28, 2004). The chief scientist behind these products at EagleForce is Dr. A.J. Maren, who has discussed these services and the steps of knowledge discovery with the author. Dr. Maren’s biography is available at <http://www.theeagleforce.com/Leadership.shtml#04> (last visited Mar. 6, 2005).

61. See, e.g., Doug Beizer, *Intelligence Community Looks at Convera Software*, WASH. POST, Nov. 29, 2004, at E4; see also Convera, *supra* note 5.

62. Google, *supra* note 9.

63. Maren, *supra* note 9; see also MARY DEROSA, CENTER FOR STRATEGIC AND INT’L STUDIES INSIGHTS, DATA MINING AND DATA ANALYSIS FOR COUNTER-TERRORISM 3 (Mar. 2004).

64. DEROSA, *supra* note 63, at 3–5; Maren, *supra* note 9.

65. DEF. ADVANCED RESEARCH PROJECTS AGENCY, FACT FILE: A COMPENDIUM OF DARPA PROGRAMS AUGUST 2003, at 7 (2003), at <http://www.darpa.mil/body/pdf/final2003factfilerev1.pdf>; see DEROSA, *supra* note 63, at 16–20.

66. DEROSA, *supra* note 63, at 18.

67. See *id.*

“immutable” audit trail so that the search could later be traced if abuse came to be suspected.⁶⁸

D. Deanonymization

The TIA searches were supposed to result in patterns possibly associated with terrorist activity. Each pattern would be “anonymized” (i.e., stripped of the names of individuals associated with the information in the pattern) and provided to the human analyst, perhaps accompanied by an indication of the level of uncertainty of the association between the available data and the possible terrorist attack in order to facilitate the allocation of enforcement resources.⁶⁹ The search would then be refined by consulting additional databases until the pattern matched the previously defined terrorist template with sufficient certainty to justify that the individual associated with the pattern be identified.⁷⁰ To support this anonymization process, TIA was to include 1) research on automated technology for stripping names from documents in such a way that identities could not be inferred from the data delivered to the source of the inquiry; and 2) automated technology for selective revelation (sharing the information in anonymized or deanonymized form, as appropriate to the situation).⁷¹

The names of the individuals associated with each of the patterns identified by the TIA search engine were to be kept in an escrow maintained by a trusted third party.⁷² The names were to be disclosed (or deanonymized) only upon specific authorization, presumably by an administrative judge of the Foreign Intelligence Surveillance Act (“FISA”) court⁷³ or perhaps (less desirably) by the Department of Justice. The balance between individual privacy and national security presumably was to be embodied in the standard of proof to be applied to the identification of these individuals. It would also be technically possible, and maybe desirable on policy grounds, to label certain kinds of data or databases as particularly sensitive, and to require a heightened standard of proof before the TIA engine was permitted to consult them.

68. *Id.* Despite the TIA support for this research, the TAPAC Report criticized “DARPA’s failure to build protections for privacy into TIA technologies as they were being developed. . . . [The] separate initiatives to develop privacy protecting tools . . . were not part of the TIA tools that DARPA was demonstrating in 2002 and 2003.” TAPAC REPORT, *supra* note 16, at 20. The report continues, “Ironically, [DARPA’s failure to ensure that privacy is built into the developmental process] also contributed to Congress withdrawing funding for the privacy enhancing technologies that DARPA was developing.” *Id.*

69. Rosenzweig & Scardaville, *The Need to Protect Civil Liberties*, *supra* note 48, at 8.

70. *Id.*; Taipale, *Technology, Security, and Privacy*, *supra* note 47, at 194–97. Many security needs for data analysis, including watch-list matching and pattern matching, may be accomplished within an anonymized data framework. Data analysis, including some forms of data mining, may also be possible.

71. Rosenzweig & Scardaville, *The Need to Protect Civil Liberties*, *supra* note 48, at 8.

72. Rosenzweig, *Terrorism Information Awareness*, *supra* note 48, at 13–16.

73. *Id.* at 15.

IV. PRIVACY LAW AND THE TECHNOLOGY OF KNOWLEDGE DISCOVERY

A. Data Acquisition

The limitations on data acquisition are based on the provisions of the Fourth Amendment prohibiting unreasonable search and seizure.⁷⁴ The Supreme Court has interpreted that these provisions provide protection for people if the area to be searched is one where a person would have an “actual (subjective) expectation of privacy,” and if that expectation would be accepted as “reasonable” by society.⁷⁵

For the purposes of this Article, the major limitation of this doctrine is that, on its face, it would allow a citizen’s reasonable expectation of privacy to be steadily reduced by improvements in surveillance technology.⁷⁶ The Supreme Court recognized this problem in *Kyllo*, ruling that police are not authorized to use technology “not in public use” (at issue there, infrared detectors used from a public street to detect high-intensity lamps being used to grow marijuana in a private home) in order to gather information remotely that would otherwise have required a search warrant.⁷⁷ The Court did not speak to the issue of what rule would govern situations once the technology in question had come into common use, as has already happened with caller ID and is likely soon to happen with wireless digital photography (e.g., camera phones). Eventually, the march of technology probably will force the Court to adopt a new standard.⁷⁸

B. Data Sharing

The statutory and institutional limitations on the ability of government to share data among its own agencies have largely disappeared. The U.S. Privacy Act of 1974 prohibited one government agency from disclosing personally identifiable information to another without the consent of the data subject, with the exception of law enforcement and other broadly defined “routine” purposes.⁷⁹ This

74. U.S. CONST. amend. IV.

75. *Katz v. United States*, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

76. A second major limitation, unrelated to technology, is that the government could in principle vitiate its protection by simply announcing that it could carry out a search whenever it pleased, so that no one had any reasonable expectation of privacy anywhere. Under an expansive interpretation of this rule, for example, the government might vitiate any “reasonable expectation of privacy” on the Internet by requiring every Internet provider to post a notice that the government may intercept and read all e-mail or chat room messages, or even by requiring all citizens to carry shoulder-mounted video cameras to allow the government to track their movements. ROSEN, *supra* note 4, at 150.

77. *Kyllo v. United States*, 533 U.S. 27, 40 (2001).

78. See *infra* Section V.

79. The Privacy Act of 1974, 5 U.S.C. § 552a(b) (2000); see TAPAC REPORT, *supra* note 16, at 26; see also Swire, *Foreign Surveillance*, *supra* note 26, at 1315–20 (providing historical background of national security, surveillance, and privacy); STEVENS, *supra* note 26, at 6.

modest limitation was substantially vitiated by the ability of government agencies to purchase and gain access to private databases.⁸⁰

A second major obstacle to sharing information within the government was the rule that barred the sharing of information between criminal investigators and intelligence collectors—an institutional rather than a statutory barrier.⁸¹ In particular, the barrier to sharing information between intelligence agencies and agencies investigating criminals is thought to have been a significant factor in the failure to prevent the events of 9/11.⁸² Ironically, in the declassified decision by the United States Foreign Intelligence Surveillance Court of Review (the FISA appellate court), the court ruled that this barrier was not required by statute and, in retrospect, need never have existed.⁸³ In practice, this barrier seems to have had as much to do with personalities, “turf” protection, and bureaucracy as with any legal restrictions.

C. Pattern Identification

There is little jurisprudence relating to the right of government to “connect the dots”—to identify patterns by assembling and integrating data from a variety of sources, and thereby form a detailed picture of an individual, as long as it has the right of access to the individual pieces of data. Two cases separated by almost two decades, one involving a public company and the other a private citizen, point in different directions. In *Nader v. General Motors*, a tort case, Judge Breitel wrote that “extensive or exhaustive monitoring and cataloging of activities that are normally disconnected and anonymous” could be found to violate the right of privacy.⁸⁴

By contrast, in the more recent search and seizure case of *California v. Greenwood*, the U.S. Supreme Court ruled that a sealed trash bag was not entitled to protection against police surveillance, even though it

80. TAPAC REPORT, *supra* note 16, at viii.

81. *In re Sealed Case*, 310 F.3d 717, 727–28 (Foreign Intel. Surv. Ct. Rev. 2002). This firewall between these activities had been justified on the grounds that criminal investigators are reactive by nature, seeking to catch the perpetrators of crimes that have already happened and paying strict attention to the rules governing the introduction of evidence in court, so that the perpetrators can be tried and punished. *Id.* Intelligence agencies, in contrast, are proactive by nature, seeking to give advance warning of threats to national security. *See id.* at 738–39. However, the gathering of intelligence by intelligence agencies is not subject to the Federal Rules of Evidence, since the material gathered is not intended for use in court. For a critique of this decision, see Swire, *Foreign Surveillance*, *supra* note 26, at 1337–38.

82. *In re Sealed Case*, 310 F.3d at 733.

83. *Id.* at 721–36; *see also* Brief for United States at § II.C.1, *In re All Matters Submitted to Foreign Intelligence Surveillance Court*, 218 F. Supp. 2d 611 (Foreign Intel. Surv. Ct. 2002) (No. 02-001), at <http://fas.org/irp/agency/doj/fisa/082102appeal.html>. For a critique of the decision, see Swire, *Foreign Surveillance*, *supra* note 26, at 1334–35.

84. *Nader v. General Motors Corp.*, 255 N.E.2d 765, 772 (N.Y. 1970). The court held that “[s]urveillance may be so ‘overzealous’ as to render it actionable.” *Id.* at 771 (citation omitted).

provides an integrated picture of a person's life.⁸⁵ In a cogent dissent, Justice Brennan noted, "A trash bag . . . is a common repository for one's personal effects Almost every human activity ultimately manifests itself in waste products. . . . A single bag of trash testifies eloquently to the eating, reading, and recreational habits of the person who produced it."⁸⁶ In the opinion of the Court, however, the bag was sufficiently exposed to the public to defeat the defendant's claim of protection on grounds of privacy.⁸⁷

D. Deanonymization

There is little jurisprudence on the legal criteria governing the disclosure of a name in order to meet some social need. Perhaps the situation most nearly analogous to the deanonymization of a possible terrorist threat is the disclosure by public health authorities of the name of an individual with a communicable disease, for example, to alert those who came into contact with him or her or to initiate quarantine measures.⁸⁸ Gostin has proposed that such identification be minimally intrusive (i.e., carried out in the least intrusive way capable of accomplishing the public purpose), and that the procedures for doing so provide for due process and be both equitable and transparent.⁸⁹

Coincidentally, the issue of deanonymization appeared at the heart of a very recent U.S. Supreme Court case, *Hiibel v. Sixth Judicial District Court of Nevada*.⁹⁰ This case, on its face, has nothing to do with terrorism or surveillance. A Nevada rancher refused to give his name to a policeman carrying out a so-called "Terry stop,"⁹¹ defined as a "minimally intrusive" pat-down or "frisk" permitted when a police officer has a "reasonable, articulable suspicion" that a search is objectively necessary.⁹² The issues in *Hiibel* are of major importance to U.S. criminal law since disclosure of one's identity opens the door to an extensive data search that could, in principle, establish that the person being questioned is wanted for a serious crime. As Rotenberg points out, "A name is no longer a simple identifier: it is the key to a vast, cross-

85. *California v. Greenwood*, 486 U.S. 35, 37, 39 (1988) (internal quotations and citations omitted).

86. *Id.* at 50 (Brennan, J., dissenting). For analogous arguments about checkbook records in stinging dissents, see *Cal. Bankers Ass'n v. Shultz*, 416 U.S. 21, 79–99 (1974) (Douglas, Marshall, JJ., dissenting). "[Banking records] mirror not only one's finances but his interests, his debts, his family, and his civic commitments." *Id.* at 89 (Douglas, J., dissenting).

87. *Greenwood*, 486 U.S. at 40.

88. This is commonly found in state laws. See, e.g., DEL. CODE ANN. tit. 16, § 702 (2004).

89. Lawrence O. Gostin, *Health Information Privacy*, 80 CORNELL L. REV. 451, 494–513 (1995); Lawrence O. Gostin, *Public Health Law in an Age of Terrorism: Rethinking Individual Rights and Common Goods*, 21 HEALTH AFF. 6, 79, 88–91 (2002); Lawrence O. Gostin, *When Terrorism Threatens Health: How Far Are Limitations on Personal and Economic Liberties Justified?*, 55 FLA. L. REV. 1105, 1141–54 (2003).

90. *Hiibel v. Sixth Judicial Dist. Court of Nev.*, 124 S. Ct. 2451 (2004).

91. *Id.* at 2452–53.

92. *Terry v. Ohio*, 392 U.S. 1, 21–22 (1968).

referenced system of public and private databases, which lay bare the most intimate features of an individual's life."⁹³ In ruling that the state could require an individual to reveal his identity during a *Terry* stop, the Court was careful to sidestep Fifth Amendment issues of self-incrimination, specifically deferring them for consideration in possible future cases.⁹⁴

The decision of the Supreme Court in *Hibel* could be an important precedent in setting the standard of proof for deanonymization of the kind of pattern-based searches envisioned under TIA. Future courts may be inclined to apply the *Terry* standard of "reasonable suspicion" in cases dealing with the deanonymization of pattern-based searches. The next section will argue that anticipated advances in technology make it important that they avoid this temptation.

V. THE NEED FOR NEW LEGAL STANDARDS

This Article's brief review of the jurisprudence surrounding the four possible intervention points in anti-terrorism surveillance supports Bennett's conclusion that "[t]here may be a lot of laws, but there is not much protection."⁹⁵ As outlined earlier in this Article, the legal responses to advances in technology have so weakened the limits on the government's ability to gather and aggregate information that a systematic review of the safeguards to privacy is now necessary, even if this requires a reexamination of well-settled constitutional precedent.⁹⁶ In this regard, the critical precedent is *Miller*,⁹⁷ which holds, in Solove's pithy summary, that "[s]ince information maintained by third parties is exposed to others, it is not private, and therefore not protected by the Fourth Amendment."⁹⁸

There is ample precedent for such a reexamination of well-established precedent in the light of advances in technology. Justice Scalia, the conservative theorist on the present Supreme Court bench, wrote in the majority opinion for *Kyllo* that:

[i]t would be foolish to contend that the degree of privacy secured to citizens by the Fourth Amendment has been entirely unaffected by the advance of technology. For example, . . . the technology enabling human flight has exposed to public view (and hence, we

93. Brief of Amici Curiae Electronic Privacy Information Center (EPIC) as Legal Scholars and Technical Experts at 3, *Hibel v. Sixth Judicial Dist. Court of Nev.*, 124 S. Ct. 2451 (2004) (No. 03-5554), at http://www.epic.org/privacy/hibel/epic_amicus.pdf (last visited Dec. 28, 2004) (including a comprehensive list of the databases to which police have access once they have a person's name).

94. *Hibel*, 124 S. Ct. at 2454.

95. Colin J. Bennett, *Convergence Revisited: Toward a Global Policy for the Protection of Personal Data?*, in *TECHNOLOGY AND PRIVACY: THE NEW LANDSCAPE* 99, 113 (Philip E. Agre & Marc Rotenberg eds., 1997), quoted in Solove, *Privacy and Power*, *supra* note 15, at 1444.

96. For a broader account of the erosion of Fourth Amendment rights, see generally DASH, *supra* note 8.

97. *United States v. Miller*, 425 U.S. 435 (1976).

98. See Solove, *Digital Dossiers*, *supra* note 8, at 1134.

have said, to official observation) uncovered portions of [a] house and its curtilage that once were private. The question we confront today is what limits there are upon this power of technology to shrink the realm of guaranteed privacy.⁹⁹

The question posed by Justice Scalia has yet to be resolved by the Court, or indeed by the American people. American values on privacy were defined in a previous, less technological era. These values need to be reexamined and redefined for the modern era of data mining and knowledge discovery. In particular, there needs to be a public discussion regarding the large-scale, anonymized “fishing expeditions” contemplated by TIA and (very likely) by its successors, and whether the deanonymization technology that TIA sought to develop would be a sufficient safeguard against unwarranted intrusion into personal privacy. This public dialogue could begin with Congressional hearings, in which a broad range of views are aired.

Such a reexamination should start with this general question: What are the limits to individual privacy that can be expected by a citizen of a free and open society in light of terrorist danger?¹⁰⁰ Posing the issue in this way leads directly to a reexamination of several important, presently well-established judicial precedents affecting the steps in knowledge discovery.¹⁰¹

A. Data Acquisition

Most existing jurisprudence has focused on only one of the four steps of knowledge discovery: data acquisition and the conditions under which it is permissible. Even this jurisprudence needs to be reexamined. As this Article has shown, the accepted standard of “reasonable expectation of privacy” has been steadily whittled down by a series of decisions responding to advances in technology.¹⁰² A useful starting point for such a reexamination would be the argument by Justice Stevens in *Whalen v. Roe*.¹⁰³ Writing for the majority, he stated that the constitutionally protected “zone of privacy” included “the individual interest in avoiding disclosure of personal matters.”¹⁰⁴ Another protected zone of privacy might be the common notion that there should be a limit to the amount of information that the government can collect, by

99. *Kyllo v. United States*, 533 U.S. 27, 33–34 (2001) (citations omitted).

100. *See United States v. White*, 401 U.S. 745, 758, 763–65 (1971) (examining privacy under the Fourth Amendment and noting that the Executive Branch has been weakening the Fourth Amendment since the 1940s to prevent sabotage of the government).

101. *See, e.g., Kyllo*, 533 U.S. at 34; *Cal. Bankers Ass’n v. Schultz*, 416 U.S. 21, 64 (1974).

102. *See, e.g., California v. Greenwood*, 486 U.S. 35 (1988); *Katz v. United States*, 389 U.S. 347 (1967).

103. *See generally Whalen v. Roe*, 429 U.S. 589 (1977).

104. *Id.* at 599–600. Nevertheless, the Court did not apply “strict scrutiny” to this newfound privacy interest, rejecting the plaintiff’s claim that providing copies of certain doctor’s prescriptions for controlled drugs to the state infringed his right of nondisclosure. TAPAC REPORT, *supra* note 16, at 25 (citing *Whalen*, 429 U.S. at 603–04).

whatever means, in the absence of a specific justification vetted by a neutral third party. Such a limit might be subsumed within the existing doctrine of “reasonable expectation of privacy.”

B. Data Sharing

As we have seen, the legal limits on the sharing of information among companies, among government agencies, and from private databases to government are narrowly drawn. The author is inclined to agree with the remark in the TAPAC report that “*Miller* and its progeny clearly conflict with American values concerning privacy”¹⁰⁵ Once a piece of information is revealed to either a tool of government surveillance or a private database during a commercial transaction, personal information can be freely shared and aggregated into patterns by programs of increasing effectiveness, subject to minimal limitations.¹⁰⁶ This information may be shared and aggregated even in the absence of evidence of any criminal or terrorist activity.¹⁰⁷

Whatever the Court’s eventual definition of the “limits . . . of technology to shrink the realm of guaranteed privacy,” as phrased by Justice Scalia in *Kyllo*,¹⁰⁸ the knowledge-discovery technology described here, if successfully developed, would surely exceed these limits. As Justice Stevens, speaking for the Supreme Court in *United States Department of Justice v. Reporters Committee for Freedom of the Press*, put it, “Plainly there is a vast difference between the public records that might be found after a diligent search of courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information.”¹⁰⁹

Granted that the data in commercial databases are available to the public, the government acquisition of personally identifiable data still involves different considerations. To quote the second Markle Foundation report:

Although there are consequences associated with the data’s being available in the private sector (such as loss of job opportunities, credit worthiness, or public embarrassment), the consequences of government access to and use of data can be more far-reaching, and can include loss of liberty and encroachment on the constitutionally rooted right of privacy . . . which is designed to protect citizens from intrusions by government, not neighbors or credit bureaus.¹¹⁰

105. TAPAC REPORT, *supra* note 16, at 23.

106. *See id.*

107. *Id.*

108. *Kyllo v. United States*, 533 U.S. 27, 34 (2001).

109. *United States Dep’t of Justice v. Reporters Comm. for Freedom of Press*, 489 U.S. 749, 764 (1989).

110. MARKLE FOUNDATION, CREATING A TRUSTED INFORMATION NETWORK FOR HOMELAND SECURITY 33 (Dec. 2003) [hereinafter MARKLE FOUNDATION, TRUSTED NETWORK], *at*

A more defensible rule would hold that a person may disclose personal information—a mortgage, a credit card, medical records—for a specific purpose, without fear that this information will be freely shared with the government, if refusal to disclose this information would effectively exclude him or her from modern life, or from the exercise of his or her constitutional rights.¹¹¹ This proposal extends Cohen’s concept of “transactional identity,” which she argues takes into account the fact that “data about one’s own transactional history and preferences are . . . bound up with identity.”¹¹²

Nissenbaum proposes a definition of privacy based on the idea of “contextual integrity.”¹¹³ She proposes that a complaint that privacy has been violated should be considered sound if an “informational norm”—defined as a norm of appropriateness of flow or distribution—has been “transgressed.”¹¹⁴ In simpler terms, contextual integrity has been violated if information is disseminated beyond what could be reasonably inferred from the circumstances in which it was originally divulged. This concept is a common-sense interpretation of the doctrine of “reasonable expectation of privacy.”¹¹⁵ On the other hand, the proposed doctrine of contextual integrity lends itself to broad interpretations that could be used to justify very narrow constraints on the sharing of almost any personal data. For purposes of this Article, it is argued that the application of these norms should be restricted to financial, medical, or educational information of which disclosure would result in a clearly definable loss.¹¹⁶

http://www.markle.org/downloadable_assets/nstf_report2_full_report.pdf. The Markle Foundation’s purpose is to more quickly implement new information and communication technologies in order to meet public needs in the areas of healthcare and national security. Markle Foundation, Who We Are, at http://www.markle.org/about_markle/who_we_are/index.php (last visited May 4, 2005).

111. Solove, *Privacy and Power*, *supra* note 15, at 1426–27 (“People must relinquish personal data to gain employment, procure insurance, obtain a credit card, or otherwise participate like a normal citizen in today’s economy. Consent is virtually meaningless in many contexts. When people give consent, they must often consent to a total surrender of control over their information.”); *cf.* TAPAC REPORT, *supra* note 16, at 24 (“[O]ften [these] disclosures are not voluntary. We unavoidably disclose information about ourselves everyday, often without knowledge, and increasingly without choice. To fly, open a bank account, or obtain employment, for example, one is required to provide a name (and, in the latter two cases, Social Security Number) and an acceptable form of identification. Such disclosures are hardly ‘voluntary.’”).

112. Cohen, *supra* note 19, at 597–98. Cohen proposes this concept in the course of a discussion of the limits of the rights of a copyright holder to explore the reading and listening habits of people suspected of violating his or her intellectual property rights. *Id.* at 585. This leads her to propose to extend the tort of “unauthorized appropriation of name and likeness” to “transactional identity.” *Id.* at 597. See generally Julie E. Cohen, *Overcoming Property: Does Copyright Trump Privacy?*, 2002 U. ILL. J.L. TECH. & POL’Y 375.

113. Nissenbaum, *supra* note 34, at 136.

114. *Id.* at 138.

115. See *id.*

116. *Id.* at 137.

C. Pattern Identification

The disagreement between the previously discussed *Nader* and *Greenwood* cases illustrates the difficulty in defining limits to the identification of patterns by integrating legally available data.¹¹⁷ A citizen may instinctively feel that there should be limits to the ability of government to aggregate data for this purpose, but it is difficult to translate this feeling into a legally enforceable doctrine. At some time in the future, U.S. courts may decide to extend the ideas of “reasonable expectation of privacy” (*Katz*) or “zone of privacy” (*Whalen*) to set substantive limits on data aggregation and knowledge discovery, although the *Greenwood* precedent indicates a reluctance on their part to do so.¹¹⁸ Even the European protections referred to an earlier focus on the gathering and sharing of information by the private sector and are basically procedural in nature.¹¹⁹

D. Deanonimization

The TIA program proposed to safeguard the process of information sharing and pattern recognition by a process of anonymization (or, in Taipale’s terminology, escrowed identity or pseudonymity),¹²⁰ by which the names of individuals identified by pattern-based or pattern-matching searches would be revealed only if specific conditions were met.¹²¹ The criterion for such deanonimization may be expressed as a legal standard of proof, or, equivalently, as the degree of uncertainty associated with the possible connection between the identified pattern and the possible terrorist activity. As in other branches of the law, this standard of proof would be an expression of the desired balance between the risk of false positives (in this case, mistaking an innocent pattern for one indicative of terrorist activity) and that of false negatives (missing a pattern that really did indicate terrorist activity).¹²²

Taipale and Rosenzweig have each suggested that this standard of proof should be a function of both the nature of the database from which the patterns were derived, and of the possible consequences of false negatives for national security.¹²³ Taipale proposes that the standard of

117. See *supra* Section IV.C.

118. Kerr and Swire, in separate articles, argue that the courts are moving away from the *Katz* doctrine. Orin S. Kerr, *The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution*, 102 MICH. L. REV. 801 (2004); Peter P. Swire, *Katz is Dead. Long Live Katz.*, 102 MICH. L. REV. 904 (2004) [hereinafter Swire, *Katz is Dead*].

119. See OECD GUIDELINES, *supra* note 25, at 13–18.

120. Taipale, *Technology, Security, and Privacy*, *supra* note 47, at 197–98.

121. *Id.*

122. Charles Weiss, *Expressing Scientific Uncertainty*, 2 LAW, PROBABILITY AND RISK 25, 29 (2003).

123. Taipale, *Technology, Security, and Privacy*, *supra* note 47, at 146–48; Rosenzweig & Scardaville, *The Need to Protect Civil Liberties*, *supra* note 48, at 17–18. The question of false positives

proof of the deanonymization process be either “probable cause” or “reasonable suspicion.”¹²⁴

This Article would propose to add more flexibility to this binary function by introducing a third, intermediate standard: “reasonable indication.”¹²⁵ Reasonable indication has not been used as a legal standard of proof, and therefore has the disadvantage of not being defined by case law. On the other hand, it is a well-established administrative standard that is used, for example, by the FBI as the criterion for beginning an investigation.¹²⁶ It is also the criterion for initiating inquiry into possible dumping violations under U.S. trade law.¹²⁷

In sum, this Article would propose that the standard of proof for deanonymization of patterns possibly indicative of terrorist activity have three tiers. The first standard should be reasonable, articulable suspicion—the Fourth Amendment standard for a *Terry* stop. It should apply to transaction patterns, thought to be associated with the most serious forms of terrorist activity, such as nuclear, biological, or large-scale chemical attacks. This standard is defined in criminal law as a suspicion based on “objective, articulable facts, leading an experienced, prudent officer to suspect that the individual is concealing something on his/her person contrary to law.”¹²⁸ The second standard should be reasonable indication, the criterion for initiating an FBI investigation. It should apply to those transaction patterns which do not point to the most serious forms of terrorist activity and are not derived from the most sensitive data or databases.¹²⁹ The third standard should be probable cause, the Fourth Amendment standard for search, seizure, and arrest. It would apply in cases of transaction patterns not associated with these most serious terrorist activities, but relying on the most sensitive data or databases—for example, those holding personal information on finances, medical conditions, and intellectual and political activities through an

involves the actions taken by the government as a result of the identification of a pattern it associates with terrorism, which is outside the scope of this Article.

124. Taipale, *Data Mining and Domestic Security*, *supra* note 4, at 67, 79.

125. The definition of these terms is reasonably consistent among different dictionaries. For example, the *Random House Dictionary* defines “indication” as “anything serving . . . as a sign [of something].” RANDOM HOUSE UNABRIDGED DICTIONARY 190 (2d ed. 1993). The term “suspicion” is defined as “imagination of anything to be the case . . . ; a vague notion” or “slight trace, hint, or suggestion.” *Id.* at 1917. A “belief” is defined as “confidence in the truth . . . of something not immediately susceptible to rigorous proof.” *Id.* at 972. Similarly, the *Oxford English Dictionary* defines “indication” as “a hint, suggestion, or piece of information from which more may be inferred.” 12 THE OXFORD ENGLISH DICTIONARY 861 (1989). “Suspicion” is defined as “apprehension . . . on slight grounds or without clear evidence.” 12 *id.* at 323. “Belief” is defined as “[m]ental acceptance of a proposition . . . on the ground of authority or evidence.” 12 *id.* at 86.

126. See Memorandum from Larry D. Thompson, U.S. Department of Justice, Office of the Deputy Attorney General, to Michael Chertoff, Assistant Attorney General, U.S. Department of Justice (Aug. 6, 2001) [hereinafter Thompson Memorandum], at <http://www.fas.org/irp/agency/doj/fisa/dag080601.html>.

127. Antidumping and Countervailing Duties, 19 C.F.R. § 351.222(f)(2)(iii) (2004).

128. Weiss, *supra* note 122, at 33–34; see also W. R. LAFAYE, J. H. ISRAEL & N. J. KING, CRIMINAL PROCEDURE § 3.8, at 215 (2d ed. 1992).

129. Thompson Memorandum, *supra* note 126.

individual's library books, video rentals, magazine subscriptions, Internet surfing, and the like. This standard is defined in the law of search and seizure as evidence that "would warrant a belief by a reasonable man."¹³⁰ It is "more than a bare suspicion but less than evidence that would justify a conviction."¹³¹

If, as is likely, the legal limits on the process of deanonymization become the key protection of privacy rights in some future surveillance system, it will be particularly important to address the question of how intrusive pattern recognition may be in using anonymized data. The criteria set forth in the previous paragraph constitute the author's best effort both to reconcile the competing needs of national security with those of privacy and civil liberties, and to design an information technology that embodies social values in its programming code.¹³² They are intended to provoke further discussion. Limits along these lines might be enacted by Congress.¹³³ Alternatively, they might be codified in an Executive Order and published in the Federal Register, with only the provision for judicial review requiring an act of Congress.

VI. IMPLICATIONS FOR THE INTELLIGENCE AND LAW ENFORCEMENT COMMUNITIES

This Article concludes with a brief discussion of four policy issues that have not received much attention in the academic literature or the public press, but require further exploration.

The first of these issues arises from the proposed structure of the TIA knowledge-discovery system. TIA envisioned a system that would query a large number of databases, some belonging to government and some to private companies. TIA foresaw (correctly, in the author's opinion) that it would be extremely difficult for the government to maintain a centralized database, and that it would be much more effective and politically acceptable to query a large number of smaller databases, each of which was maintained by its individual owner in its own format.

This decentralized structure, however, presents a tricky question of liability and civil liberties. Who would be responsible for the consequences of "false positives"? What would the rights be of innocent people mislabeled as potential terrorists, and perhaps arrested or

130. LAFAVE, *supra* note 128, § 3.3(b), at 138.

131. BLACK'S LAW DICTIONARY 1219 (7th ed. 1999).

132. See generally Taipale, *Data Mining and Domestic Security*, *supra* note 4; LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE (1999) (arguing that software is often written for the convenience of its authors or in the commercial interest of its vendors, and may intentionally or unintentionally embody social values).

133. Rosen argues that Congress is the most likely source of proactive laws in this area, as the courts are inclined to follow public opinion only after a considerable delay. ROSEN, *supra* note 4, at 147. This Article suggests that the intelligence agencies themselves have an important role to play. See discussion *infra* at Section VI.

excluded from employment? The decentralized structure is likely to suffer from errors in one or more databases, from inaccuracies in the terrorism templates from which the transaction patterns are matched, or even from mistakes in computer programming.

Taipale has proposed to avoid this danger by stipulating that the results of pattern-based searches be used only for purposes of allocating resources for surveillance and investigation, not for evidentiary purposes.¹³⁴ In practice, however, there is likely to be great bureaucratic and political pressure for “mission creep,” especially in support of actions like placement on “no-fly” lists, which, while short of actual search and arrest, can cause considerable cost and inconvenience to the individual.¹³⁵ On the other hand, owners of private databases are likely to refuse government access unless they are relieved of this huge potential liability.

A second issue arises from the fact that information gleaned from pattern-based searches may be used not just to find terrorists, but also to identify leads for criminal investigation.¹³⁶ This is likely to result in “de facto profiling”: disproportionate investigation and prosecution of Arabs and other Muslims for crimes that have nothing to do with terrorism, and that are committed with comparable frequency by people of other religious or ethnic groups. To guard against this unintended consequence, it might be advisable that a prosecutor wishing to pursue a criminal charge derived from a pattern-based search be required to show that the pattern originally thought to link the suspect to terrorism met more than the usual minimum standard of proof prescribed earlier in this Article.¹³⁷ This higher standard might be “preponderance of the evidence” or at least “clear indication.”¹³⁸

The third of these issues concerns the locus within the federal government for the implementation of any future technology for knowledge discovery. TIA was a technology development program and did not devote much attention to the institutional means by which it was ultimately to be deployed.¹³⁹ For purposes of this discussion, assume that something like the technologies that were to be developed by the TIA program, as well as something like the knowledge-discovery technologies that are now in development, will in fact be ready for deployment at some time in the future.

Where might these technologies be implemented? The implementing system would need access to databases located in a wide

134. Taipale, *Data Mining and Domestic Security*, *supra* note 4, at 32; Kim Taipale, Technology, Security and Privacy: Designing Technical Features to Meet Policy Goals, Presentation at Rutgers University 74 (Apr. 15, 2004), at <http://www.taipale.org/presentations/CAS-DIMACS.pdf>.

135. Swire, *Foreign Surveillance*, *supra* note 26, at 1318.

136. *Id.* at 88–89.

137. See *supra* Section IV.C–D and accompanying notes.

138. Weiss, *supra* note 122, at 38 tbl.2.

139. See Rosenzweig, *Terrorism Information Awareness*, *supra* note 48.

variety of agencies in the defense, intelligence, and law enforcement communities, as well as in state and local governments. It would also gain access, presumably on commercial terms, to databases belonging to a large number of private firms in a wide range of industries. The patterns it would identify would be of interest to any of an even larger number of companies or branches of government.

In principle, the technologies might be implemented by a unified system in the sense that authority to initiate searches would be limited to a single office with a unified structure and a relatively homogeneous institutional culture. Alternatively, the technology might be operated as a service, analogous to wiretapping or remote sensing, in which case a variety of intelligence and law enforcement organizations would have such authority.

First, consider the organizational implications of a single, unified TIA office. Neither the intelligence nor the law enforcement community is likely to want such a powerful engine of knowledge discovery located in the other, in part for reasons of bureaucratic turf protection, and in part for fear of compromising either's core function. Intelligence would seek to use the technology to prevent acts of terrorism before they occur, whereas law enforcement would use it to catch and prosecute the evildoers after they have done their evil. Moreover, given the history of past abuse, the public is likely to object to such a concentration of information (and thereby power) in secretive agencies under relatively deferential Congressional supervision.¹⁴⁰ The White House, which in principle sits above both the intelligence and law enforcement communities, might seem to be the logical locus for this function, but as a practical matter is preoccupied with politics. Besides, past history shows that placing such a critical operational task in the White House is fraught with problems.¹⁴¹

The recent passage of the Intelligence Reform and Terrorism Prevention Act has reorganized the intelligence community and created a National Counter-Terrorism Center ("the Center") located within the office of the Director of National Intelligence ("DNI").¹⁴² The newly created DNI is to "oversee and direct the implementation of the National Intelligence Program."¹⁴³ Since the Center is designated as the "central and shared knowledge bank on known and suspected terrorists and international terror groups, as well as their goals, strategies,

140. Swire, *Foreign Surveillance*, *supra* note 26, at 1315–17.

141. *See id.* at 1325 (discussing the legislative response to Nixon's abuse of power in Watergate).

142. Intelligence Reform and Terrorism Prevention Act of 2004, Pub. L. No. 108-458, 118 Stat. 3638. For an account of government action leading up to this, see NATIONAL COMMISSION OF TERRORIST ATTACKS UPON THE UNITED STATES, THE 9/11 COMMISSION REPORT 411–14 (2004), available at http://www.9-11commission.gov/report/911Report_Ch13.pdf. Also, the Council on Foreign Relations has provided a short, clear guide to the various proposals for reorganizing the U.S. intelligence community. Council on Foreign Relations, *Intelligence: Major Reform Plans Compared*, at http://www.cfr.org/background/intel_compare.php (last updated Oct. 12, 2004).

143. Intelligence Reform and Terrorism Prevention Act, § 1011.

capacities and networks of contacts and support,” any future TIA-like function would be likely to be located in the Center.¹⁴⁴

The DNI is not to be under the control of the Executive Branch and the President.¹⁴⁵ However, since the DNI is required to report to the President, he or she might well be subject to political pressures comparable to those that would be felt under direct control of the White House. If the DNI is given or gradually develops independence comparable to that of the Director of the Federal Reserve Bank, the information provided by the agencies under the DNI’s jurisdiction could be used to persecute personal or political enemies—an abuse associated with the name of J. Edgar Hoover.¹⁴⁶

There is no simple organizational solution to this dilemma. One possible organizational solution would be a decentralized system open to all intelligence and law enforcement agencies (at the federal or, perhaps, also at state and local levels). This would resolve the issues created by centralized control, but would open the way to multiple localized abuses, especially given the difficulty of creating a uniform institutional culture within each user agency that would both stress sensitivity to privacy and civil liberties and resist efforts to politicize the agency’s function.

TIA was to address these complex issues in two ways. First, it was to develop technology for an immutable audit trail.¹⁴⁷ This would have provided a safeguard against abuse by making it possible to trace the origin and outcome of all past inquiries.¹⁴⁸ Second, it assumed that the deanonymization of any pattern thought to indicate terrorist activity would have to be authorized by a neutral third party.¹⁴⁹

Now that TIA has been cancelled, it is important to ensure that political and organizational implications remain important considerations in specifying and designing research and development on any knowledge-discovery technology intended for implementation by the government. A robust requirement for anonymization of pattern-based searches and for their deanonymization (for example, only upon the authorization of an independent external body, such as an FISA Court) is an essential minimum protection for privacy and civil liberties.¹⁵⁰

The fourth and last of these issues concerns the need for extensive consultations with the public on privacy issues by intelligence and law enforcement agencies which normally (and for good reason) shun such public discussion.¹⁵¹ The war on terrorism, combined with the

144. *Id.*

145. *Id.*

146. *See, e.g.,* MORTON H. HALPERIN ET AL., *THE LAWLESS STATE: THE CRIMES OF THE U.S. INTELLIGENCE AGENCIES* 61–89 (1976).

147. Poindexter, *supra* note 51.

148. DEROSA, *supra* note 63, at 19; Poindexter, *supra* note 51.

149. DEROSA, *supra* note 63, at 19–20.

150. *See* Rosenzweig, *Terrorism Information Awareness*, *supra* note 48.

151. *See generally* MARKLE FOUNDATION, *PROTECTING AMERICA’S FREEDOM IN THE INFORMATION AGE* (Oct. 2002), available at http://www.markle.org/downloadable_assets/nstf_full.pdf;

inadequacy of existing legal protections of individual privacy, puts American intelligence agencies in an awkward bind. If another incident of serious terrorism occurs in the United States, Congress, the media, and the American public will ask whether the intelligence community used all the legal powers at its command to interpret the clues that were, in retrospect, available to it. They are also likely to forgive any past transgressions against privacy rights that would have helped stop such attacks. On the other hand, if there are no more major acts of terrorism on U.S. soil within the next several years, the public is likely to protest against what will by then be perceived as infringements on civil liberties imposed by the “war on terrorism,” regardless of the letter of the law. This could well lead to a public scandal resulting in the enactment of undesirable limits on intelligence gathering. Paradoxically, then, the intelligence and law enforcement communities would likely be forgiven their trespasses against civil liberties if they fail to prevent acts of terrorism, but are likely to be blamed for them if they succeed in preventing such attacks.

This Article would therefore suggest that the intelligence community ask itself how much access it really needs to sensitive personal information of the types discussed in this Article, and what kinds of restrictions on the use of this material it could accept without seriously impairing its function. The intelligence community should carry out this review in consultation with Congress and with members of the general public. The public will likely demand limits on the government’s access as it becomes more aware of the practical effect of intrusions under the likely conditions of a prolonged “war on terrorism” and increased efficiency of government surveillance.

VII. CONCLUSION

Advances in the technology of knowledge discovery are likely soon to overwhelm existing piecemeal protections to individual privacy vis-à-vis government surveillance. It is important to reexamine the legal standards covering the acquisition and sharing of information, and the identification and deanonymization of patterns in that information. Given the progressive weakening of privacy protection in the face of changing law and advancing technology, any implementation by government of data-mining and knowledge-discovery technology should be carried out under privacy safeguards that go well beyond present constitutional and statutory requirements. One may hope that Congress and the courts will take the lead in establishing new law to protect civil liberties even as the United States fights the terrorist threat. Paradoxically, it may be in the institutional self-interest of the

intelligence and law enforcement agencies to be *ahead* of Congress and the courts in protecting civil liberties. However useful it might prove in the war on terrorism, this technology is like a genie that needs to be put back in its bottle until privacy and civil liberties can be adequately protected.