

LEGAL STANDARDS FOR DATA MINING

Fred H. Cate¹

for Robert Popp & John Yen, eds.,
21st Century Enabling Technologies and Policies for Counter-Terrorism

1. Introduction

Data mining is a promising tool in the fight against terrorism. It already plays a number of important roles in counter terrorism including locating known suspects, identifying and tracking suspicious financial and other transactions, and facilitating background checks. Rapid increases in the power and speed of computing technologies, the capacity of data storage, and the reach of networks have added exponentially to both the volume of data available for possible use and the ability of the government to meaningfully examine them. As a result, as discussed elsewhere in this volume, new data mining applications are likely to play increasingly important roles in fighting terrorism.

Government data mining also poses significant issues for individual privacy and other civil liberties. Proposals for enhanced government data mining have provoked serious controversy, beginning with the first large-scale computerized government benefits databases created by the then-Department of Health, Education and Welfare. More recently, public concern over proposals for Total Information Awareness and second-generation Computer Assisted Passenger Profiling was sufficient to block at least public development of these systems.

One of the major contributors to the controversies over government data mining is the absence of clear legal standards. Forty years ago the lack of relevant law was understandable: the technologies were new, their capacity was largely unknown, and the types of legal issues they might raise were novel. Today, it is inexplicable and threatens to undermine both privacy and security.

The situation is exacerbated by the Supreme Court's 1976 decision in *United States v. Miller* that there can be no reasonable expectation of privacy in information held by a third party, so the Fourth Amendment does not apply to the government's seizure of such data.² With the growth in digital technologies and networks and the resulting proliferation of digital data, the *Miller* decision today means that the government faces few, if any, constitutional limitations when it seeks the personal data maintained by banks, credit card issuers, brokers, airlines, rental car companies, hospitals, insurers, Internet service providers, real estate agents, telephone companies, publishers, libraries, educators, employers, and information brokers. According to

¹ Distinguished Professor and Director of the Center for Applied Cybersecurity Research, Indiana University; Senior Policy Advisor, Center for Information Policy Leadership at Hunton & Williams. Professor Cate served as Reporter for the Department of Defense Technology and Privacy Advisory Committee. The author gratefully acknowledges the generous aid of his Rosenzweig and K.A. Taipale and the excellent research assistance of David Scott Dickinson and Lindsey Ann Rodgers. The author alone is responsible for any errors or omissions.

² *United States v. Miller*, 425 U.S. 435 (1976).

the Supreme Court, it does not matter whether the information sought by the government was disclosed to a third party as a necessary part of a consumer transaction, provided to a third party “on the assumption that it will be used only for a limited purpose,” or conveyed pursuant to a “confidence” that it will not be shared at all.³ The government is still free to seize it and the Fourth Amendment imposes no limits on either its collection or use.

Congress reacted to *Miller* and subsequent cases by enacting statutory protections for customer financial records held by financial institutions and other sectoral statutes,⁴ but these laws offer limited protection and do not apply to the terabytes of other personal data maintained by third parties today. So, while most individuals believe that the government is constitutionally prohibited from seizing their personal information without a warrant, granted by a judge, based on probable cause, the reality is that the government is only restricted from obtaining that information directly from the individuals to which it pertains.

The massive volume of data about individuals and the ease with which they are collected, aggregated, and shared, means that the private sector often holds a treasure trove of information to which the government desires access as part of its counter-terrorism efforts. Often those data are already collected together by private-sector service providers and data aggregators. The exponential growth in detailed personal information that third parties possess, the ease of access to that information, and the absence of constitutional protections for such data have greatly reduced the government’s need to seek information directly from the individuals to which it pertains. Instead, the government can often obtain—whether by seizure or purchase—the same or an even wider and more revealing supply of data from third parties, thereby avoiding entirely the need to comply with the Fourth Amendment. As a result, the barrier that the Fourth Amendment historically interposed between the government and data about individuals has been greatly reduced and may soon be eliminated effectively. Congress’ failure to respond with broad legislation has resulted in a situation in which the vast majority of personal data about individuals is now accessible to the government without legal limit and contrary to the public’s expectations.

Moreover, the growing importance of new forms of data mining to fight terrorism is enhancing the government’s interest in third-party data. No longer does the government just want occasional access to a narrow range of records for specific searches for information about identified individuals. Instead, new data mining programs would require broad access to vast amounts of third-party records about people who have done nothing to warrant attention in an effort to detect relationships and patterns of behavior that might highlight terrorist activities. Even if the Supreme Court had not decided *Miller* or otherwise excluded personal information held by third parties from the protection of the Fourth Amendment, Congress would still need to create a legislative framework for these new and very different data mining activities. *Miller* only exacerbates the need for that framework which, to date, Congress has declined to provide.

The absence of a coherent legal regime applicable to data mining significantly undercuts the confidence of the public and of policymakers that it will be carried out with appropriate

³ Id. at 443 (citation omitted).

⁴ Right to Financial Privacy Act, 12 U.S.C. §§ 3401-3422. See notes 51-63 and accompany text.

attention to protecting privacy and other civil liberties. That absence denies government officials charged with fighting terrorism guidance as to what is and is not acceptable conduct. It interferes with businesses, universities, and other possessors of potentially relevant databases knowing when they can legally share information with the government. And it is threatening individual privacy while slowing the development of new and promising data mining programs, undermining research into this potentially important weapon in the war on terrorism, and hampering the data sharing that is key to national security. This is an untenable situation for both protecting privacy and fighting terrorism.

This chapter examines the Supreme Court's exclusion of government mining of third-party data from the privacy protection of the Fourth Amendment, Congress' failure to fill the gap left by the Court's jurisprudence or to otherwise respond to the growing use of data mining in counter-terrorism by specifying its appropriate roles and limits, and the resulting threat to privacy and security. Even as other technologies, such as anonymized data matching, help reduce the impact of government data mining on privacy and other civil liberties, legal rules will still be necessary to allow those technologies to reach their full potential, protect privacy in settings they do not reach, build public confidence in appropriate data mining, enhance national security, facilitate more rational and consistent policymaking, and foster further innovation.

2. What is Data Mining?

“Data mining” is defined in many different ways, but most have in common the elements of searches of one or more databases of personally identifiable information by or on behalf of the government. Data mining in the antiterrorism context usually involves third-party data that have been provided voluntarily, purchased or seized by the government, or reported to the government in compliance with routine reporting requirements. It may also involve the use of data previously collected by the government for other purposes.

Dramatic advances in information technology have greatly enhanced the government's ability to search vast quantities of data for the purpose of identifying people who meet specific criteria or otherwise present unusual patterns of activities. These technologies have exponentially increased the volume of data available about individuals and greatly reduced the financial and other obstacles to retaining, sharing, and exploiting those data in both the public and private sector. They also have eliminated the need to physically combine disparate data sets to be able to search them simultaneously.

Government data mining in general is widespread and expanding. A 2004 report by the Government Accountability Office found 42 federal departments—including every cabinet-level agency that responded to the GAO's survey—engaged in (88), or were planning to engage in (34), 122 data mining efforts involving personal information.⁵ Thirty-six of those involve accessing data from the private sector; 46 involve sharing data among federal agencies.⁶

⁵ U.S. General Accounting Office, *Data Mining: Federal Efforts Cover a Wide Range of Uses* (GAO-04-548), May 2004, at 3, 27-64, tables 2-25.

⁶ *Id.* at 3.

Data mining is increasingly being looked to as a tool to combat terrorism. For example, in 2002 the Defense Advanced Research Projects Agency in the Department of Defense launched “Total Information Awareness”—later renamed “Terrorism Information Awareness”—a research and development program that included technologies to search personally identifiable transaction records and recognize patterns across separate databases for the purpose of combating terrorism.⁷ The Advanced Research and Development Activity center, based in the National Security Agency in DOD, has a project—Novel Intelligence from Massive Data—to develop tools to examine large quantities of data to “reveal new indicators, issues, and/or threats that would not otherwise have been found due to the massiveness of the data.”⁸

Army defense contractor Torch Concepts, with the assistance of DOD and the Transportation Security Administration, obtained millions of passenger records from U.S. airlines in 2003 to study how data profiling can be used to identify high-risk passengers.⁹ The TSA also worked to develop the second generation of the Computer-Assisted Passenger Prescreening System to compare airline passenger names with private- and public-sector databases to assess the level of risk a passenger might pose.¹⁰ In the Homeland Security Act, signed into law in November 2002, Congress required the new Department of Homeland Security to “establish and utilize . . . data-mining and other advanced analytical tools,” to “access, receive, and analyze data detect and identify threats of terrorism against the United States.”¹¹

These are just a handful of the more prominent and controversial publicly disclosed projects since the terrorist attacks of September 11, 2001. These projects are far removed from traditional inquiries to locate information about a particular individual. Criminal investigators have long made use of “subject-based” data mining to look for information about a specific individual. These inquiries start with *known* suspects and search for information about them and the people with whom they interact. The law applicable to such searches is quite complex, but fairly well settled.

Many new government data mining programs, especially in the counter-terrorism arena, feature “pattern-based” searches. These involve developing models of what criminal or terrorist behavior might look like and then examining databases for similar patterns. The GAO, for example, defines “data mining” as “the application of database technology and techniques—such as statistical analysis and modeling—to uncover hidden patterns and subtle relationships in data and to infer rules that allow for the prediction of future results.”¹² The pending Federal Agency

⁷ U.S. Department of Defense, Technology and Privacy Advisory Committee, *Safeguarding Privacy in the Fight Against Terrorism* 15-20 (Mar. 2004).

⁸ See <ic-arda.org>.

⁹ Department of Homeland Security Privacy Office, *Report to the Public on Events Surrounding jetBlue Data Transfer—Findings and Recommendations* (Feb. 20, 2004).

¹⁰ Privacy Act; System of Records, 68 Fed. Reg. 45,265 (2003) (DHS, TSA) (interim final notice); U.S. General Accounting Office, *Computer-Assisted Passenger Prescreening System Faces Significant Implementation Challenges* (GAO-04-385), Feb. 2004.

¹¹ Homeland Security Act of 2002, Pub. L. No. 107-296, §§ 201(d)(1), (d)(14) (Nov. 25, 2002).

¹² GAO *Data Mining* Report, *supra* at 4.

Data-Mining Reporting Act would require that the search or analysis of the database(s) be intended “to find a predictive pattern indicating terrorist or criminal activity.”¹³

The power of data mining technology and the range of data to which the government has access have contributed to blurring the line between subject- and pattern-based searches. The broader the search criteria, and the more people other than actual criminals or terrorists who will be identified by those criteria, the more pattern-like these searches become. Even when a subject-based search starts with a known suspect, it can be transformed into a pattern-based search as investigators target individuals for investigation solely because of apparently innocent connections with the suspect. The more tenuous the connection, the more like a pattern-based search it becomes. Moreover, even searches that are not pattern-based raise similar significant issues if they use third-party data from the private sector or data previously collected by the government for other purposes. Whether pattern- or subject-based, “data mining,” as the term is used in this chapter, involves the government’s access to and use of personal information about people who have done nothing warrant suspicion in an effort to identify those individuals who do.

3. Constitutional Protection for Information Privacy

a. The Fourth Amendment

Historically, the primary constitutional limit on the government’s ability to obtain personal information about individuals is the Fourth Amendment:

The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.¹⁴

The Fourth Amendment does not purport to keep the government from conducting searches or seizing personal information; it only prohibits “unreasonable” searches and seizures. The Supreme Court interprets the Fourth Amendment also to require that searches be conducted only with a warrant issued by a court, even though this is not a requirement contained in the amendment itself.¹⁵ For a court to issue a warrant, the government must show “probable cause” that a crime has been or is likely to be committed and that the information sought is germane to that crime. The Supreme Court also generally requires that the government provide the subject of a search with contemporaneous notice of the search.¹⁶

¹³ S1169, Federal Data-Mining Reporting Act, 109th Cong., 1st Sess. (2005).

¹⁴ U.S. Constitution amend. IV.

¹⁵ Akihl Reed Amar, *The Constitution and Criminal Procedure* 3-4 (1997).

¹⁶ *Richards v. Wisconsin*, 520 U.S. 385 (1997).

The Fourth Amendment applies to searches and surveillance conducted for domestic law enforcement purposes within the United States, and those conducted outside of the United States if they involve U.S. citizens (although not necessarily permanent resident aliens). The Fourth Amendment also applies to searches and surveillance conducted for national security and intelligence purposes within the United States if they involve U.S. persons who do not have a connection to a foreign power.¹⁷ The Supreme Court has not yet addressed whether the Fourth Amendment applies to searches and surveillance for national security and intelligence purposes that involve U.S. persons who are connected to a foreign power or are conducted wholly outside of the United States.¹⁸ Lower courts have found, however, that there is an exception to the Fourth Amendment's warrant requirement for searches conducted for intelligence purposes within the United States that involve only non-U.S. persons or agents of foreign powers.¹⁹

Where it does apply, while the protection afforded by the Fourth Amendment can be considerable, it is not absolute. The Supreme Court has determined, for example, that warrants are not required to search or seize items in the "plain view" of a law enforcement officer,²⁰ for searches that are conducted incidental to valid arrests,²¹ and for searches specially authorized by the Attorney General or the President involving foreign threats of "immediate and grave peril" to national security.²² Moreover, the Fourth Amendment poses no limits on how the government may use information, provided that it has been obtained legally. So personal data seized by the government in compliance with the Fourth Amendment may later be used in a context for which the data could not have been obtained consistent with the Fourth Amendment.

The Fourth Amendment prohibits only "unreasonable" searches and seizures, but is silent about what makes a search or seizure "unreasonable." In his 1967 concurrence in *Katz v. United States*, Justice Harlan wrote that reasonableness was defined by both the individual's "actual," subjective expectation of privacy and by an objective expectation that was "one that society was prepared to recognize as 'reasonable.'"²³ The Court adopted that test for determining what was "private" within the meaning of the Fourth Amendment in 1968 and continues to apply it today.

b. The *Miller-Smith* Exclusion of Third-Party Records

The Supreme Court held in 1976 in *United States v. Miller*²⁴ that there can be no reasonable expectation of privacy in information held by a third party. The case involved

¹⁷ U.S. v. U.S. District Court for the Eastern District of Michigan, 407 U.S. 297 (1972) (commonly referred to as the *Keith* decision).

¹⁸ Jeffrey H. Smith & Elizabeth L. Howe, "Federal Legal Constraints on Electronic Surveillance," Markle Foundation Task Force on National Security in the Information Age, *Protecting America's Freedom in the Information Age* 133 (2002).

¹⁹ See *U.S. v. Bin Laden*, 126 F. Supp. 2d 264, 271-72 (S.D.N.Y. 2000).

²⁰ *Coolidge v. New Hampshire*, 403 U.S. 443 (1971).

²¹ *U.S. v. Edwards*, 415 U.S. 800 (1974).

²² 68 *American Jurisprudence 2d*, Searches and Seizures § 104 (1993); Smith & Howe, *supra* at 136, n.16.

²³ 389 U.S. 347, 361 (1967).

²⁴ *United States v. Miller*, 425 U.S. 435 (1976).

cancelled checks, to which, the Court noted, “respondent can assert neither ownership nor possession.”²⁵ Such documents “contain only information voluntarily conveyed to the banks and exposed to their employees in the ordinary course of business,”²⁶ and therefore the Court found that the Fourth Amendment is not implicated when the government sought access to them:

The depositor takes the risk, in revealing his affairs to another, that the information will be conveyed by that person to the Government. This Court has held repeatedly that the 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 the third party will not be betrayed.²⁷

Congress reacted to the decision by enacting modest statutory protection for customer financial records held by financial institutions,²⁸ but there is no constitutional protection for financial records or any other personal information that has been disclosed to third parties. As a result, the government can collect even the most sensitive information from a third party without a warrant and without risk that the search may be found unreasonable under the Fourth Amendment.

The Court reinforced its holding in *Miller* in the 1979 case of *Smith v. Maryland*, involving information about (as opposed to the content of) telephone calls.²⁹ The Supreme Court found that the Fourth Amendment is inapplicable to telecommunications “attributes” (e.g., the number dialed, the time the call was placed, the duration of the call, etc.), because that information is necessarily conveyed to, or observable by, third parties involved in connecting the call.³⁰ “[T]elephone users, in sum, typically know that they must convey numerical information to the phone company; that the phone company has facilities for recording this information; and that the phone company does in fact record this information for a variety of legitimate business purposes.”³¹

As a result, under the Fourth Amendment, the use of “pen registers” (to record out-going call information) and “trap and trace” devices (to record in-coming call information) does not require a warrant because they only collect information about the call that is necessarily disclosed to others. As with information disclosed to financial institutions, Congress reacted to the Supreme Court’s decision by creating a statutory warrant requirement for pen registers,³² but the Constitution does not apply.

²⁵ Id. at 440.

²⁶ Id. at 442.

²⁷ Id. at 443 (citation omitted).

²⁸ Right to Financial Privacy Act, 12 U.S.C. §§ 3401-3422. See notes 51-53 and accompanying text.

²⁹ 442 U.S. 735 (1979).

³⁰ 442 U.S. 735 (1979).

³¹ Id. at 743.

³² 18 U.S.C. §§ 3121, 1841. See notes 51-53 and accompanying text.

The third party exemption from the Fourth Amendment made little sense in the two cases in which it was created. Individuals who write checks and dial telephone calls do not “voluntarily” convey information to third parties. They have no choice but to convey the information if they wish to use what in the 1970s were the overwhelmingly dominant means of making large-value payments or communicating over physical distances.

Moreover, and more importantly, the information collected and stored by banks and telephone companies is subject to explicit or implicit promises that it will not be further disclosed. Most customers would be astonished to find their checks or telephone billing records printed in the newspaper. As a result of those promises and the experience of individuals, the expectation that such information would be private was objectively reasonable and widely shared. The Court’s decisions to the contrary, while they served important law enforcement objectives, made little logical or practical sense and did not reflect the expectations of either the public or policymakers, as demonstrated by the fact that Congress responded so quickly to both with gap-filling legislation.

Irrespective of whether *Miller* and *Smith* were correctly decided, however, excluding records held by third parties from the protection of the Fourth Amendment makes no sense today because of the extraordinary increase in both the volume and sensitivity of information about individuals necessarily held by third parties. Professor Daniel Solove writes: “We are becoming a society of records, and these records are not held by us, but by third parties.”³³ He offers these examples:

[L]ife in modern society demands that we enter into numerous relationships with professionals (doctors, lawyers, accountants), businesses (restaurants, video rental stores), merchants (bookstores, mail catalog companies), publishing companies (magazines, newspapers), organizations (charities), financial institutions (banks, investment firms, credit card companies), landlords, employers, and other entities (insurance companies, security companies, travel agencies, car rental companies, hotels). Our relationships with all of these entities generate records containing personal information necessary to establish an account and record of our transactions, preferences, purchases, and activities.³⁴

Thanks to the proliferation of digital technologies and networks such as the Internet, and tremendous advances in the capacity of storage devices and parallel decreases in their cost and physical size, those records are linked and shared more widely and stored far longer than ever before, often without the individual consumer’s knowledge or consent.

In addition, as more everyday activities move online, records contain more detailed information about individuals’ behavior. No longer do merchants record data only on what individuals buy and how they pay for their purchases. Instead, those data include every detail of

³³ Daniel J. Solove, “Digital Dossiers and the Dissipation of Fourth Amendment Privacy,” 75 *Southern California Law Review* 1083, 1089 (2002).

³⁴ *Id.*

what we look at, the books we read, the movies we watch, the music we listen to, the games we play, and the places we visit. Instead of comparatively barebones data about the checks individuals write and telephone calls we place, the government today has access unrestricted by the Fourth Amendment to private-sector records on every detail of how we live our lives.

The robustness of these records is difficult to overestimate and is not limited to settings involving commercial transactions. For example, computers track every moment of most employees' days. Digital time clocks and entry keys record physical movements. Computers store work product, e-mail, and Internet browsing records—often in keystroke-by-keystroke detail, as more and more employees use technologies to monitor employee behavior. E-mail and voice mail are stored digitally. Even the content of telephone conversations may be recorded.

Nor is the ubiquitous nature of data collection and observation limited to the workplace. Digital devices for paying tolls, computer diagnostic equipment in car engines, and global positioning services that are increasingly common on passenger vehicles record every mile driven. Cellular telephones and personal digital assistants record not only call and appointment information, but location as well, and transmit this information to service providers. ISPs record online activities, digital cable and satellite record what we watch and when, alarm systems record when we enter and leave our homes, and all of these data are held by third parties.

All indications are that this is just the beginning. Broadband Internet access into homes has not only increased the personal activities we now engage in online, but also created new and successful markets for remote computer back-up and online photo, e-mail, and music storage services. With Voice Over IP telephone service, digital phone calls are becoming indistinguishable from digital documents: both can be stored and accessed remotely. Global positioning technologies are appearing in more and more products, and Radio Frequency Identification Tags are beginning to be used to identify high-end consumer goods, pets, and even people.

Moreover, those records may be held by more private parties than ever before. Digital transactions are likely to be observed by more parties. Data about online browsing or purchases are accessible not only to the consumer and merchants directly involved in the transaction, but also to their ISPs, the provider of the payment mechanism (for example, a credit card company), and the company that delivers the merchandise. The every day use of a credit card or ATM card involves the disclosure of personal financial information to multiple entities.

In addition, digital networks have facilitated the growth of vigorous outsourcing markets, so information provided to one company is increasingly likely to be processed by a separate institution. Customer service may be provided by another. And all of those entities may store their data with still another. Personal information is available from all of these. For example, many employers contract with separate ISPs. Information on browsing habits of employees is available to both the employer and the ISP. If an employee buys an airline ticket through an online travel service, such as Travelocity or Expedia, the information concerning that transaction will be available to the employer, the ISP, the travel service, the airline, and the provider of the payment mechanism, at a minimum.

The government would hardly need to visit all of these businesses separately, however, to gather personal information. A handful of service providers already process, or have access to, the large majority of credit and debit card transactions, ATM withdrawals, airline and rental car reservations, and Internet access. As demonstrated by the 2005 security breach at Atlanta-based CardSystems that revealed sensitive information about as many as 40 million Visa, MasterCard, Discover, and American Express customers,³⁵ there is no need to go to each of these companies separately for information on their customers when one service provider can supply the same data.

Moreover, there are information aggregation businesses in the private sector that already combine personal data from thousands of private-sector sources and public records. ChoicePoint, Acxiom, LexisNexis, the three national credit bureaus, and dozens of other companies maintain rich repositories of information about virtually every adult in the country. These records are updated daily by a steady stream of incoming data. These businesses supply this information, for a fee, to private- and public-sector customers for a variety of valuable uses. One of the common threads among most of the antiterrorist data mining programs that have been made public to date is their reliance—intended or actual—on these aggregators. Why seize data from many separate entities or even service providers when much the same information can be bought from one?

The *Miller* exclusion from the Fourth Amendment of information disclosed to third parties means that all of this information, no matter how sensitive or how revealing of a person's health, finances, tastes, or convictions, is available to the government without constitutional limit. The government's demand need not be reasonable, no warrant is necessary, and no judicial authorization or oversight is required. Moreover, it appears not to matter how explicitly confidentiality was promised by the third party as a condition of providing the information. Those promises and contractual provisions may restrict the ability of the third party to volunteer the information, although as Professor Solove notes, most privacy policies today are written to permit voluntary disclosures to the government,³⁶ but privacy promises have no effect on the power of the government to obtain the information.

Finally, technological developments over the past 40 years have both ensured that the data are in digital form and therefore more likely to be of use to the government, and put increasingly powerful tools in the hands of the government to be able to use those data. Millions of records stored on index cards were not likely to be of much use to the government. The cost of duplicating, transporting, storing, and using them would have been in most cases prohibitive. In electronic format, however, those costs are comparatively negligible. So while the impact of *Miller* in 1976 was primarily limited to government requests for specific records about identified individuals, today *Miller* allows the government to obtain the raw material for broad-based data mining.

This is a significant difference. The 1970s searches involved demands for information about individuals who had already done something to warrant the government's attention.

³⁵ Eric Dash, "68,000 MasterCard Accounts Are at High Risk in Breach," *New York Times*, June 19, 2005, at A22.

³⁶ Solove, *supra* at 1098-1100.

Whether or not the suspicious activity amounted to “probable cause,” there was at least some reason to suspect a particular person. Today, because of major technological and related changes, the government not only has the power under the Fourth Amendment to ask for everything about everybody, but increasingly the practical ability to do something with that information. As a result, as part of its on-going fight against terrorism, the government increasingly desires access to broad swaths of information about people who have done nothing to warrant suspicion. This new practical power offers potentially valuable new weapons in the war against a nearly invisible terrorist foe. But that power, especially in the absence of constitutional oversight, also raises important legal and political questions.

Advances in technologies, and the development of new products and services in response to those changes, have significantly expanded the scope of the *Miller* exclusion of records held by third parties from the protection of the Fourth Amendment. Today there are vastly more personal data in the hands of third parties, they are far more revealing, and much more readily accessible than was the case in the 1970s. Moreover, for the first time, the government has the practical ability to exploit huge data sets. As a result, the scope of the *Miller* decision has been greatly expanded and the balance between the government’s power to obtain personal data and the privacy rights of individuals fundamentally altered.

4. Congressional Roles

While the Supreme Court identifies and interprets constitutional boundaries between the government and the citizenry, Congress establishes statutory boundaries and rules that the government must follow to cross them. The congressional roles are vital because of the breadth of Congress’ power and its ability to provide detailed, prospective guidance to the public and to government officials about the seizure of personal information.

That guidance is also necessary to address the consensual and regulatory collection, use, storage, and disclosure of personal data. While the government has constitutionally unlimited power to search and seize third-party records, it is more likely to seek those records through routine reporting requirements or purchase, or by reusing data already collected by the government through these means for other purposes. Historically, the Fourth Amendment has played no role in restricting these activities. Moreover, as we have seen, the Fourth Amendment plays no role once information has been lawfully collected in determining how it is to be used, stored, or disclosed. These are significant omissions that legislation is well suited to address.

As a result, even if the Supreme Court had not excluded third-party records from the protection of the Fourth Amendment in *Miller*, congressional action would still be critical because of the need to provide a legal structure for the government’s collection of information through means other than seizure, and for its use, storage, and dissemination of that information. This is especially true in the face of technological advances that have exponentially increased the volume of data available about individuals, greatly reduced the financial and other obstacles to sharing and exploiting those data, and significantly enhanced the government’s ability to search vast quantities of data for the purpose of identifying people who meet specific criteria or otherwise present unusual patterns of activities. These changes have resulted in a new

environment and new challenges that require new rules. It is the responsibility of Congress to provide them.

a. The Privacy Act

Congress first regulated how the government collects and uses personal information in the Privacy Act of 1974.³⁷ In the early 1970s, mounting concerns about computerized databases prompted the government to examine the issues they raised—technological and legal—by appointing an Advisory Committee on Automated Personal Data Systems in the Department of Health, Education and Welfare. In 1973, the Advisory Committee issued its report, *Records, Computers and the Rights of Citizens*.³⁸ Congress responded the following year with the Privacy Act.

The Privacy Act requires federal agencies to store only relevant and necessary personal information and only for purposes required to be accomplished by statute or executive order; collect information to the extent possible from the data subject; maintain records that are accurate, complete, timely, and relevant; and establish administrative, physical, and technical safeguards to protect the security of records.³⁹ The Privacy Act also prohibits disclosure, even to other government agencies, of personally identifiable information in any record contained in a “system of records,” except pursuant to a written request by or with the written consent of the data subject, or pursuant to a specific exception.⁴⁰ Agencies must log disclosures of records and, in some cases, inform the subjects of such disclosures when they occur. Under the Act, data subjects must be able to access and copy their records, each agency must establish a procedure for amendment of records, and refusals by agencies to amend their records are subject to judicial review. Agencies must publish a notice of the existence, character, and accessibility of their record systems.⁴¹ Finally, individuals may seek legal redress if an agency denies them access to their records.

The Privacy Act is less protective of privacy than may first appear, because of numerous broad exceptions.⁴² Twelve of these are expressly provided for in the Act itself. For example, information contained in an agency’s records can be disclosed for “civil or criminal law enforcement activity if the activity is authorized by law.”⁴³ An agency can disclose its records to officers and employees within the agency itself, the Bureau of the Census, the National Archives, Congress, the Comptroller General, and consumer reporting agencies.⁴⁴ Information subject to

³⁷ 5 U.S.C. § 552a.

³⁸ U.S. Department of Health, Education & Welfare, Report of the Secretary’s Advisory Committee on Automated Personal Data Systems, *Records, Computer, and the Rights of Citizens* (1973).

³⁹ *Id.*

⁴⁰ *Id.* § 552a(b).

⁴¹ *Id.* § 552a(e)(4).

⁴² Sean Fogarty & Daniel R. Ortiz, “Limitations Upon Interagency Information Sharing: The Privacy Act of 1974,” Markle Foundation Task Force, *Protecting America’s Freedom in the Information Age*, *supra* at 127, 128.

⁴³ *Id.* § 552a (b)(7).

⁴⁴ 5 U.S.C. § 552a(b),

disclosure under the Freedom of Information Act is exempted from the Privacy Act.⁴⁵ And under the “routine use” exemption,⁴⁶ federal agencies are permitted to disclose personal information so long as the nature and scope of the routine use was previously published in the Federal Register and the disclosure of data was “for a purpose which is compatible with the purpose for which it was collected.” According to OMB, “compatibility” covers uses that are either (1) functionally equivalent or (2) necessary and proper.⁴⁷

Moreover, the Privacy Act applies only to information maintained in a “system of records.”⁴⁸ The Act defines “system of records” as a “group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual.”⁴⁹ The U.S. Court of Appeals for the District of Columbia Circuit held that “retrieval capability is not sufficient to create a system of records. . . . ‘To be in a system of records, a record must . . . in practice [be] retrieved by an individual’s name or other personal identifier.’”⁵⁰ This is unlikely to be the case with new antiterrorism databases and data mining programs. They are more likely to involve searches for people who fit within certain patterns, rather than inquiries by name or other personal identifier.

As a result, the Privacy Act plays little role in providing guidance for government data mining activities or limiting the government’s power to collect personal data from third parties. In fact, the framework created by the Privacy Act, which was designed more than 30 years ago primarily for personnel records and benefits files, would appear to be altogether ill-suited for regulating counter-terrorism data mining. Like many laws relating to information, it has become outdated and outmoded by the passage of time and dramatic technological change.

b. The Response to *Miller* and *Smith*

Congress responded to *United States v. Miller* and *Smith v. Maryland* with specific statutes designed to address the vacuum created by the Supreme Court’s decisions. The Right to Financial Privacy Act, enacted in 1978, two years after *Miller*, regulates how federal agencies may obtain financial records from financial institutions.⁵¹ The Electronic Communications Privacy Act, enacted in 1986, seven years after *Smith*, broadly regulates electronic surveillance, including the use of pen registers and trap and trace devices.⁵²

⁴⁵ Id. § 552a (b)(2)

⁴⁶ Id. § 552a (b)(3).

⁴⁷ Privacy Act of 1974; Guidance on the Privacy, Act Implications of “Call Detail” Programs to Manage Employees’ Use of the Government’s Telecommunications Systems, 52 Fed. Reg. 12,900, 12,993 (1987) (OMB) (publication of guidance in final form); see generally Fogarty & Ortiz, *supra* at 129-130.

⁴⁸ 5 U.S.C. § 552a(b).

⁴⁹ Id. § 552a(a)(5).

⁵⁰ *Henke v. United States DOC*, 83 F.3d 1453, 1461 (D.C. Cir. 1996) (quoting *Bartel v. F.A.A.*, 725 F.2d 1403, 1408 n.10 (D.C. Cir. 1984)).

⁵¹ 12 U.S.C. §§ 3401-3422.

⁵² 18 U.S.C. §§ 3121, 1841.

Neither statute provides the level of protection that would have been required under the Fourth Amendment and both contain a number of exceptions. The Right to Financial Privacy Act, for example, does not restrict disclosures to state or local governments or private entities, nor does it apply to the federal government obtaining financial information from other third parties. Even in the limited area where it does apply, the Act allows the federal government to seize personal financial information pursuant to an administrative subpoena, judicial subpoena, search warrant, or formal written request. The Electronic Communications Privacy Act allows the government to obtain a judicial order authorizing the use of a pen register or trap and trace device upon a mere certification that the “information likely to be obtained is relevant to an ongoing ‘criminal investigation.’”⁵³

Nevertheless, despite their weaknesses, both statutes do impose some substantive limits on the government’s power to seize financial and calling attribute information and they do impose discipline on the government by specifying procedures to be followed. In short, the statutes help guard against the “unreasonable” searches and seizures that the Fourth Amendment, had it applied, would have prohibited.

This sectoral approach is not limited to financial and communications records, although other sectoral protections are often weaker, especially following post-9/11 amendments. For example, the Cable Act of 1984 prohibits cable companies from providing the government with personally identifiable information about their customers, unless the government presents a court order.⁵⁴ The USA Patriot Act, adopted in the immediate aftermath of the September 11 attacks, amended this provision to apply only to records about cable television service and not other services—such as Internet or telephone—that a cable operator might provide.⁵⁵ The Video Privacy Protection Act prohibits video rental companies from disclosing personally identifiable information about their customers unless the government presents a search warrant, court order, or grand jury subpoena.⁵⁶ The Family Education Rights and Privacy Act of 1974 contains a similar provision applicable to educational records.⁵⁷

The Fair Credit Reporting Act, enacted in 1970, permits disclosure of credit information only for statutorily specified purposes.⁵⁸ One of those purposes is “in response to the order of a court having jurisdiction to issue such an order, or a subpoena issued in connection with proceedings before a Federal grand jury.”⁵⁹ In addition, consumer reporting agencies may freely furnish identifying information (e.g., “name, address, former addresses, places of employment,

⁵³ Id. §§ 3122-23.

⁵⁴ 47 U.S.C. § 551.

⁵⁵ Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001, Pub. L. No. 107-56, 115 Stat. 272, Title II, § 211 (2001).

⁵⁶ 18 U.S.C. ‘ 2710.

⁵⁷ 20 U.S.C. ‘ 1232g.

⁵⁸ 15 U.S.C. ‘ 1681b.

⁵⁹ Id. ‘ 1681b(a)(1).

or former places of employment”) to the government.⁶⁰ The Act was amended following the September 11 terrorist attacks to permit virtually unlimited disclosures to the government for counter-terrorism purposes. All that is required is a “written certification” that the request information is “necessary for the agency’s conduct or such investigation, activity or analysis.”⁶¹

In 2001, the Department of Health and Human Services adopted rules, specifically authorized by Congress, protecting the privacy of personal health information.⁶² Those rules, while restrictive on their face, in reality permit broad disclosure of personal health information to the government “in the course of any judicial or administrative proceeding,” “in response to an order of a court or administrative tribunal,” “in response to a subpoena, discovery request, or other lawful process,” “as required by law,” “in compliance with . . . a court order or court-ordered warrant, or a subpoena or summons issued by a judicial officer,” “in compliance with . . . a grand jury subpoena,” and “in compliance with . . . an administrative request, including an administrative subpoena or summons, a civil or an authorized investigative demand, or similar process.”⁶³

These statutes and rules apply in limited areas; where they do apply they impose few substantive limits, although some procedural discipline, on government access to third-party data.

c. The Response to Data Mining

In 1988, Congress passed the Computer Matching and Privacy Protection Act as an amendment to the Privacy Act.⁶⁴ The new law responded to the growth in early forms of data mining within the federal government and the reality that the broad exceptions to the Privacy Act, and particularly the growing view of agency officials, the Office of Management and Budget, and even courts that data matching constituted a “routine use of data” and therefore was exempt from the Privacy Act,⁶⁵ rendered the Privacy Act inadequate to respond to data mining.

The Computer Matching and Privacy Protection Act provides a series of procedural requirements, such as written agreements between agencies that share data for matching,⁶⁶ before an agency can disclose personal information for data mining. These requirements deal only with federal agencies supplying—not obtaining—records for data mining.⁶⁷ Moreover, they only apply to data mining for the purpose of “establishing or verifying the eligibility of, or

⁶⁰ Id. ¶ 1681f.

⁶¹ Id. ¶¶ 1681u, 1681v.

⁶² *Standards for Privacy of Individually Identifiable Health Information*, 65 Fed. Reg. 82,462 (2000) (HHS, final rule) (codified at 45 C.F.R. pt. 160, §§ 164.502, 164.506).

⁶³ 45 C.F.R. § 164.512.

⁶⁴ Pub L. No. 100-503 (1988) (codified at 5 U.S.C. §§ 552a(a)(8), 552a(o)-(r)).

⁶⁵ Office of Technology Assessment, Congress of the United States, *Electronic Record Systems and Individual Privacy* 57 (1986).

⁶⁶ 5 U.S.C. § 552a (o).

⁶⁷ Id. § 552a (o)(1).

continuing compliance with statutory and regulatory requirements by, applicants for, recipients or beneficiaries of, participants in, or providers of service with respect to, cash or in-kind assistance or payments under Federal benefit programs” or “recouping payment or delinquent debts under such Federal benefit programs” or “Federal personnel or payroll systems of records.”⁶⁸ Counter-terrorism data mining does not fit within the definition of activities covered by the statute. Moreover, the Act specifically excludes data mining for “law enforcement,” “foreign counterintelligence,” and “background checks.”⁶⁹

As of 1988, then, Congress had responded to the Supreme Court’s decisions in *Miller* and *Smith* and the growth of federal data mining with sectoral statutes imposing modest limits on the government’s ability to seize personal data from third parties and with a statute imposing procedural limits on the ability of the government to share data for data mining in connection with federal benefits or payroll programs. The 1988 law was effectively Congress’ last word on the subject prior to post-9/11 developments. Laws and regulations enacted since then have either ignored government data mining entirely or failed to provide any structure for when data mining is appropriate and how it should be conducted. Moreover, counter-terrorism laws and even so-called “privacy” laws have actually weakened the protections against government seizure of personal data held by third parties. As a result the government now has the technological capability, incentive, and authority to engage in data mining, ready access to a virtually unlimited store of personal data on which to work, and no legal or policy framework to guide its data mining activities.

5. The Current Policymaking Morass

a. The Immediate Post-9/11 Response

The terrorist attacks of September 11 focused the attention of national security officials and policymakers on the importance of effective data mining to combat terrorism. In the immediate aftermath of the attacks, government officials turned to private-sector data as never before in an effort to identify the perpetrators and track down their co-conspirators. The government sought information from credit card companies, banks, airlines, rental car agencies, flight training schools, and colleges. Law enforcement officials sought information on large or suspicious financial transactions, and on any accounts involving suspected terrorists, not just from U.S. banks, but from all U.S. financial institutions and even from foreign banks that do business in the United States. Private-sector data, it became clear, was a treasure trove of information that could and would be used to identify and trace the activities of the 19 hijackers and their accomplices.

As shock and recovery efforts gave way to inquiries into why the attacks had not been prevented, it became clear that U.S. counter-terrorism and law enforcement officials had failed to connect important pieces of information stored in disparate government agencies.⁷⁰ Moreover,

⁶⁸ Id. § 552a(a)(8)(A).

⁶⁹ Id. §§ 552a (a)(8)(B)(iii), (B)(v)(vi).

⁷⁰ See, e.g., *Joint Inquiry into Intelligence Community Activities before and after the Terrorist Attacks of September 11, 2001*, Report of the U.S. Senate Select Committee on Intelligence and U.S. House Permanent Select

with the clarity of 20-20 hindsight, newspaper and magazine articles showed the myriad connections among the 19 hijackers available from largely nonsensitive private-sector data. Two of the terrorists—Nawaf Alhazmi and Khalid Almihdhar, flying under their real names on September 11—were on a State Department watch list. Data analysis expert Jeff Jonas showed that a third hijacker used the same address as Alhazmi. Two others, including Muhammad Atta, shared a residence with Almihdhar. Five others had the same phone number as Atta. Another had the same frequent-flier number as Almihdhar.⁷¹ How could we have failed to spot such now-obvious connections? What could be done to ensure that never happened again? Government inquiries and private-sector task forces stressed the need to “connect the dots.”

The government’s authority to centralize or more effectively share data it collects through traditional intelligence and law enforcement information-gathering methods has been greatly enhanced. Section 203 of the USA Patriot Act amended federal law to allow intelligence information gathered in grand jury proceedings and from wiretaps to be shared with any federal law enforcement, intelligence, immigration, or national defense personnel.⁷² In the case of grand jury information, the government must notify the court after disclosure.⁷³

Legislative proposals emerged from many sources, some of which were ultimately adopted, to centralize the dozen or more terrorist watch lists maintained by separate federal agencies. Government intelligence data was also brought together initially under a Terrorist Threat Integration Center and then, following the recommendations of the 9/11 Commission, under the National Counterterrorism Center. Ultimately intelligence operations themselves were made subject to a new Director of National Intelligence.⁷⁴

Efforts to enhance the collection and use of private-sector data have been more fractured and controversial. The USA Patriot Act imposed significant new reporting requirements on financial institutions. The Act expanded the power of the Treasury Department’s Financial Crimes Enforcement Network—FinCEN—to require financial institutions to report suspected money laundering or terrorist activities by their customers.⁷⁵ The Act also mandated new “Know Your Customer” rules which require financial institutions to (1) verify the identity of any person seeking to open an account; (2) maintain records of the information used to verify the person’s identity; (3) determine whether the person appears on any list of known or suspected terrorists or terrorist organizations; and (4) report to the government if they do.⁷⁶

Committee on Intelligence, S. Rept. No. 107- 351, H. Rep. No. 107-792, 107th Congress, 2d Session (2002), at xv-xvi.

⁷¹ Steven Levy, “Geek War on Terror,” *Newsweek*, Mar. 22, 2004, at 40.

⁷² Pub. L. No. 107-56, Title II, § 203.

⁷³ *Id.*

⁷⁴ Intelligence Reform and Terrorism Prevention Act of 2004, Pub. L. No. 108-458, 118 Stat. 3638 (Dec. 17, 2004).

⁷⁵ Financial Crimes Enforcement Network; Special Information Sharing Procedures to Deter Money Laundering and Terrorist Activity, 67 Fed. Reg. 60,579 (2002) (Treasury) (final rule)

⁷⁶ Transactions and Customer Identification Programs, 68 Fed. Reg. 25,089 (2003) (Treasury, Comptroller of the Currency, Office of Thrift Supervision, Federal Reserve System, Federal Deposit Insurance Corporation,

Under section 215 of the USA Patriot Act, the Director of the FBI or a high-level designee of the Director may apply for an order from the Foreign Intelligence Surveillance Court requiring the production of any tangible things (including books, records, papers, documents, and other items) for an investigation to protect against international terrorism or clandestine intelligence activities.⁷⁷ The only substantive limit on obtaining section 215 orders is that the investigation of a U.S. citizen or permanent legal resident may not be “conducted solely upon the basis of activities protected by the first amendment to the Constitution.”⁷⁸ The orders are issued and executed in secret, and the statute prohibits the recipient of a section 215 order from disclosing its existence to anyone.⁷⁹

The following year, in November 2002, Congress enacted the Homeland Security Act establishing DHS.⁸⁰ Section 201 of the law requires DHS to:

- “access, receive, and analyze law enforcement information, intelligence information, and other information from agencies of the Federal Government, State and local government agencies (including law enforcement agencies), and private sector entities, and to integrate such information”;
- “request additional information from other agencies of the Federal Government, State and local government agencies, and the private sector relating to threats of terrorism in the United States, or relating to other areas of responsibility assigned by the Secretary, including the entry into cooperative agreements through the Secretary to obtain such information”; and
- “establish and utilize, in conjunction with the chief information officer of the Department, a secure communications and information technology infrastructure, including data-mining and other advanced analytical tools, in order to access, receive, and analyze data and information in furtherance of the responsibilities under this section, and to disseminate information acquired and analyzed by the Department, as appropriate.”⁸¹

This positive command from Congress to access private-sector data and engage in data mining in the fight against terrorism could hardly have been more explicit, but it was soon contradicted by Congress’ response to another counter-terrorism data mining initiative.

National Credit Union Administration, Commodity Futures Trading Commission, Securities and Exchange Commission) (final rules and proposed rule).

⁷⁷ Pub. L. No. 107-56, Title II, § 215.

⁷⁸ Id.

⁷⁹ Id. If not reauthorized by Congress, this section will sunset on December 31, 2005.

⁸⁰ Homeland Security Act of 2002, Pub. L. No. 107-296 (Nov. 25, 2002).

⁸¹ Id. §§ 201(d)(1), (d)(13), (d)(14).

b. Total Information Awareness

In April 2002, the Director of the Defense Advanced Research Projects Agency testified before the Senate Armed Services Committee about a new research program: Total Information Awareness.⁸² In August, Admiral John Poindexter, director of DARPA's Information Awareness Office, described TIA at the DARPATech 2002 Conference. He described the need to “become much more efficient and more clever in the ways we find new sources of data, mine information from the new and old, generate information, make it available for analysis, convert it to knowledge, and create actionable options.”⁸³ To accomplish these purposes, he articulated the need for a “much more systematic approach.”⁸⁴ “Total Information Awareness—a prototype system—is our answer.”⁸⁵

Admiral Poindexter went on to identify “one of the significant new data sources that needs to be mined to discover and track terrorists”—the “transaction space.”⁸⁶ “If terrorist organizations are going to plan and execute attacks against the United States, their people must engage in transactions and they will leave signatures in this information space.”⁸⁷ He then showed a slide of transaction data that included “Communications, Financial, Education, Travel, Medical, Veterinary, Country Entry, Place/Event Entry, Transportation, Housing, Critical Resources, and Government” records. He also noted the importance of protecting privacy.

In November 2002, at the height of the debate over enactment of the Homeland Security Act, public controversy erupted over TIA and its impact on privacy, sparked in large part by a *New York Times* column by William Safire.⁸⁸ In the seven months between the initial disclosure of TIA and Safire's column, only 12 press reports had appeared about the program. In the next 30 days, the press carried 285 stories.⁸⁹

In December 2002, the Assistant to the Secretary of Defense for Intelligence Oversight, conducted an internal review of TIA and related programs. The review concluded that no legal obligations or “rights of United States persons” had been violated.⁹⁰ Opposition to TIA, however, continued to mount. In late 2002 Senators Charles E. Grassley (R-Iowa), Chuck Hagel (R-Neb.), and Bill Nelson (D-Fla.) wrote to the DOD Inspector General asking him to review

⁸² *Fiscal 2003 Defense Request: Combating Terrorism*, Hearing before the Senate Armed Services Committee, April 10, 2002 (statement of Dr. Tony Tether).

⁸³ John Poindexter, Overview of the Information Awareness Office, prepared remarks for delivery at DARPATech 2002, Anaheim, CA, Aug. 2, 2002, at 1.

⁸⁴ *Id.*

⁸⁵ *Id.* at 2.

⁸⁶ *Id.*

⁸⁷ *Id.*

⁸⁸ William Safire, “You Are a Suspect,” *New York Times*, Nov. 14, 2002, at A35.

⁸⁹ *TAPAC Report*, *supra* at 16.

⁹⁰ Statement of George B. Lotz, II, Assistant to the Secretary of Defense (Intelligence Oversight), to TAPAC, July 22, 2003, at 2.

TIA. On January 10, 2003, the Inspector General announced an audit of TIA, including “an examination of safeguards regarding the protection of privacy and civil liberties.”⁹¹

Congress did not wait for the results of the audit. On February 13, 2003, Congress adopted the Consolidated Appropriations Resolution.⁹² The bill contained an amendment proposed by Senator Ron Wyden (D-Ore.) prohibiting the expenditure of funds on TIA unless the Secretary of Defense, the Director of the Central Intelligence Agency, and the Attorney General jointly reported to Congress within 90 days of the enactment of the law about the development of TIA, its likely efficacy, the laws applicable to it, and its likely impact on civil liberties. The amendment also prohibited deployment of TIA in connection with data about U.S. persons without specific congressional authorization.⁹³

Secretary of Defense Donald Rumsfeld sought to diffuse congressional tension by appointing two committees in February 2003. One was an internal oversight board to establish “policies and procedures for use within DOD of TIA-developed tools” and “protocols for transferring these capabilities to entities outside DOD. . . . in accordance with existing privacy protection laws and policies.”⁹⁴

The other committee was the Technology and Privacy Advisory Committee, the members of which were eight prominent lawyers, including four former senior officials from democratic administrations and two from republican administrations.⁹⁵ The Secretary charged TAPAC with examining “the use of advanced information technologies to help identify terrorists before they act.”⁹⁶

The report specified by the Wyden Amendment was delivered to Congress on May 20, 2003.⁹⁷ The report described an array of TIA and TIA-related programs. With regard to TIA itself, the report eschewed earlier descriptions of a “virtual, centralized grand database.” DARPA wrote that “the TIA Program is not attempting to create or access a centralized database that will store information gathered from various publicly or privately held databases.”⁹⁸ Instead, the report focused on technological tools. Some of those tools would help the government

imagine the types of terrorist attacks that might be carried out against the United States at home or abroad. They would develop scenarios for these attacks and

⁹¹ Letter from Joseph E. Schmitz, DOD Inspector General, to Senator Charles E. Grassley, Chairman, Committee on Finance, Jan. 17, 2003).

⁹² Consolidated Appropriations Resolution, Pub. L. No. 108-7, Division M, § 111(b) (Feb. 24, 2003).

⁹³ S. Amend. 59 to H.J. Res. 2 (Jan. 23, 2003).

⁹⁴ U.S. Department of Defense, Total Information Awareness (TIA) Update, News Release 060-03 (Feb. 7, 2003).

⁹⁵ Establishment of the Technology and Privacy Advisory Committee, 68 Fed. Reg. 11,384 (2003) (DOD, notice).

⁹⁶ U.S. Department of Defense, Technology and Privacy Advisory Committee Charter (Mar. 25, 2003).

⁹⁷ Report to Congress Regarding the Terrorism Information Awareness Program (May 20, 2003).

⁹⁸ *Id.*, Detailed Information, *supra* at 27.

determine what kind of planning and preparation activities would have to be carried out in order to conduct these attacks. . . . The red team would determine the types of transactions that would have to be carried out to perform these activities. . . . These transactions would form a pattern that may be discernable in certain databases to which the U.S. Government would have lawful access.⁹⁹

The report also addressed other tools for secure collaborative problem solving, creating more structured and automated ways of organizing and searching data, enhancing the ability to detect and understand links between different individuals and groups, presenting data in easier-to-understand ways that make important connections easier to visualize, and improving decision making and the ways in which decision-making processes draw on stored data.¹⁰⁰

The report also identified technological tools that were being developed as part of TIA to enhance privacy protection. For example, the Genisys Privacy Protection Program “aims to provide *security with privacy* by providing certain critical data to analysts while controlling access to unauthorized information, enforcing laws and policies through software mechanisms, and ensuring that any misuse of data can be quickly detected and addressed.”¹⁰¹ DARPA had begun funding research into privacy enhancing technologies from the start of the TIA program in March 2002.

With regard to the privacy issues posed by TIA and related programs, the report provided:

The Department of Defense’s TIA research and development efforts address both privacy and civil liberties in the following ways:

- The Department of Defense must fully comply with the laws and regulations governing intelligence activities and all other laws that protect the privacy and constitutional rights of U.S. persons.
- As an integral part of its research, TIA program itself is seeking to develop new technologies that will safeguard the privacy of U.S. persons.
- TIA’s research and testing activities are conducted using either real intelligence information that the federal government has already legally obtained, or artificial synthetic information that, ipso facto, does not implicate the privacy interests of U.S. persons.¹⁰²

Neither the report nor the appointment of TAPAC was sufficient to sway Congress. On September 25, 2003, Congress passed the Department of Defense Appropriations Act, 2004.¹⁰³

⁹⁹ Id. at 14.

¹⁰⁰ Id., Executive Summary, at 2-3.

¹⁰¹ Id., Detailed Information, at 6.

¹⁰² Id., Executive Summary, at 3.

¹⁰³ Department of Defense Appropriations Act, 2004, Pub. L. No. 108-84 (Sept. 25, 2003).

Section 8131 of the Act terminated funding for TIA, with the exception of “processing, analysis, and collaboration tools for counterterrorism foreign intelligence”¹⁰⁴ specified in a classified annex to the Act. Under the Act, those tools may be used by DOD only in connection with “lawful military operations of the United States conducted outside the United States” or “lawful foreign intelligence activities conducted wholly overseas, or wholly against non-United States citizens.”¹⁰⁵

In its report accompanying the Act, the Conference Committee directed that the IAO itself be terminated immediately.¹⁰⁶ The Act thus closed the IAO, further research on privacy enhancing technologies, and further publicly disclosed research on data mining, while keeping open the possibility of counter-terrorism data mining programs being developed outside of DARPA in secret.¹⁰⁷

On December 12, 2003, the DOD Inspector General released the results of his audit of TIA. The audit concluded that “although the DARPA development of TIA-type technologies could prove valuable in combating terrorism, DARPA could have better addressed the sensitivity of the technology to minimize the possibility for Governmental abuse of power and to help ensure the successful transition of the technology into an operational environment.”¹⁰⁸

With specific regard to privacy, the audit found that DARPA failed to perform any form of privacy impact assessment, did not involve appropriate privacy and legal experts, and “focused on development of new technology rather than on the policies, procedures, and legal implications associated with the operational use of technology.”¹⁰⁹ The report acknowledged that DARPA was sponsoring “research of privacy safeguards and options that would balance security and privacy issues,” but found that such measures “were not as comprehensive as a privacy impact assessment would have been in scrutinizing TIA technology.”¹¹⁰

On May 18, 2004, TAPAC released its report. The report described TIA as a “flawed effort to achieve worthwhile ends,”¹¹¹ but the report went on to “conclude that advanced information technology—including data mining—is a vital tool in the fight against terrorism.”¹¹²

¹⁰⁴ Id. § 8183(a).

¹⁰⁵ Id. § 8183(b). The President stated in his signing statement that the classified annex “accompanies but is not incorporated as a part of the Act” and therefore would be considered by the President as merely “advisory in effect.” Statement on Signing the Department of Defense Appropriations Act, 2004 (Oct. 6, 2003).

¹⁰⁶ Conference Report on Making Appropriations for the Department of Defense for the Fiscal Year Ending September 30, 2004, and for Other Purposes, House Rpt.108-283 (2003).

¹⁰⁷ See note 141 and accompanying text.

¹⁰⁸ Department of Defense, Office of the Inspector General, *Information Technology Management: Terrorism Information Awareness Program* (D-2004-033) 4 (2003).

¹⁰⁹ Id.

¹¹⁰ Id. at 9.

¹¹¹ *TAPAC Report*, supra at 43.

¹¹² Id. at 7.

“Technological tools to help analyze data and focus human analysts’ attention on critical relationships and patterns of conduct are clearly needed.”¹¹³

The TAPAC report stressed the inadequacy of law applicable to data mining. Describing the law as “disjointed,” “inconsistent,” and “outdated,” the Committee wrote: “Current laws are often inadequate to address the new and difficult challenges presented by dramatic developments in information technologies. And that inadequacy will only become more acute as the store of digital data and the ability to search it continue to expand dramatically in the future.”¹¹⁴ Enacting a new regulatory structure, the report continued, is necessary both to “protect civil liberties” and to “empower those responsible for defending our nation to use advanced information technologies—including data mining—appropriately and effectively.”¹¹⁵ “It is time to update the law to respond to new challenges.”¹¹⁶

TAPAC proposed the outline for that new legal structure applicable to anti-terrorist or law enforcement data mining conducted by the government. Under that framework, government data mining would require:

- written authorization by agency heads;
- compliance with minimum technical requirements for data mining systems (including data minimization, data anonymization, creation of an audit trail; security and access controls, and training for personnel involved in data mining);
- special protections for data mining involving databases from other government agencies or from private industry;
- programmatic authorization from the Foreign Intelligence Surveillance Court before engaging in data mining that involves personally identifiable information concerning U.S. persons that has not been anonymized, and case-by-case authorization from the Court before reidentifying previously anonymized information concerning U.S. persons; and
- regular audits to ensure compliance.¹¹⁷

Certain data would be excluded from these new requirements, such as data mining that is limited to foreign intelligence that does not involve U.S. persons; data mining concerning federal government employees in connection with their employment; data mining that is based on particularized suspicion; and searches to identify or locate a specific individual (e.g., a suspected terrorist) from airline or cruise ship passenger manifests or other lists of names or other

¹¹³ Id. at 48.

¹¹⁴ Id. at 6.

¹¹⁵ Id.

¹¹⁶ Id.

¹¹⁷ Id. at 49-52.

nonsensitive information about U.S. persons.¹¹⁸ The report also recommended that data mining that is limited to information that is routinely available without charge or subscription to the public—on the Internet, in telephone directories, or in public records to the extent authorized by law—should be subject to “only the requirements that it be conducted pursuant to the written authorization of the agency head and auditing for compliance.”¹¹⁹

The “special protections” for data mining involving third-party databases from private industry recommended by TAPAC included:

- The agency engaging in the data mining should take into account the purpose for which the data were collected, their age, and the conditions under which they have been stored and protected when determining whether the proposed data mining is likely to be effective.
- If data are to be used for purposes that are inconsistent with those for which the data were originally collected, the agency should specifically evaluate whether the inconsistent use is justified and whether the data are appropriate for such use.
- Data should be left in place whenever possible. If this is impossible, they should be returned or destroyed as soon as practicable.
- Government agencies should not encourage any person voluntarily to provide data in violation of the terms and conditions (usually reflected in a privacy policy) under which they were collected.
- Government agencies should seek data in the order provided by Executive Order 12333: from or with the consent of the data subject, from publicly available sources, from proprietary sources, through a method requiring authorization less than probable cause (e.g., a pen register or trap and trace device), through a method requiring a warrant, and finally through a method requiring a wiretap order.
- Private entities that provide data to the government upon request or subject to judicial process should be indemnified for any liability that results from the government’s acquisition or use of the data.
- Private entities that provide data to the government upon request or subject to judicial process should be reasonably compensated for the costs they incur in complying with the government’s request or order.¹²⁰

The TAPAC report met with modest support from both the political left and right, but to date neither the Administration nor the Congress has taken any action on the committee’s

¹¹⁸ Id. at 46-47.

¹¹⁹ Id. at 47.

¹²⁰ Id. at 50-51.

recommendations concerning a new legal framework for government data mining. The gap created by *Miller* and *Smith* remains unaddressed more than 25 years later.

6. The Need for Standards

The controversy surrounding TIA and other counter-terrorism data mining projects illustrates the need for Congress and the Administration to establish legal standards for when personal information may be obtained from third parties and how it may be used. Although such standards serve many valuable purposes, six warrant special attention.

a. Protect Privacy and Other Civil Liberties

Government data mining, and especially of personal information obtained from third parties, threatens the privacy that is at the core of the relationship between the government and the citizenry. The Court's failure to extend the protections of the Fourth Amendment to personal data maintained by third parties, combined with the technological changes that result in more and increasingly revealing information being necessarily disclosed to and stored by third parties, threaten to vitiate those protections entirely. Moreover, the government's new practical ability to analyze vast amounts of disparate data rapidly and affordably threaten to extend government surveillance to every aspect of daily life. TAPAC wrote: "Government data mining presents special risks to informational privacy. If conducted without an adequate predicate, it has the potential to be a 21st-century equivalent of general searches, which the authors of the Bill of Rights were so concerned to protect against."¹²¹

Updating the law to respond to these new challenges is a daunting, but urgent, challenge. On one side is the risk of failing to identify and deter terrorist attacks. On the other are the civil liberties put at risk by data mining. The impact of data mining on civil liberties may not be immediately obvious, but awareness that the government may, without probable cause or other specific authorization, obtain access to myriad, distributed stores of information about individuals is likely to alter their behavior. The original motto of the TIA program—*Scientia Est Potentia*—is certainly correct: "knowledge is power." Knowledge that the government is observing data we generate through thousands of ordinary activities can alter the way people live their lives and interact with others. This is not always a bad outcome.

However, knowledge of that power can cause people to change their behavior to be more consistent with a perceived social norm, to mask their behavior, and to reduce their activities or participation in society to avoid the surveillance. Vice President Hubert Humphrey observed almost 40 years ago: "we act differently if we believe we are being observed. If we can never be sure whether or not we are being watched and listened to, all our actions will be altered and our very character will change."¹²² The threats posed by government data mining in a democracy are not merely to information privacy, but to other civil liberties, including freedom of expression, association, and religion.

¹²¹ Id. at 49.

¹²² Hubert H. Humphrey, Foreword to Edward V. Long, *The Intruders* at viii (1967).

Alexander Hamilton wrote in Federalist Paper 8 in 1787, exhorting the people of New York to ratify the Constitution, that “safety from external danger is the most powerful director of national conduct. Even the ardent love of liberty will, after a time, give way to its dictates.”¹²³ “The violent destruction of life and property incident to war, the continual effort and alarm attendant on a state of continual danger,” Hamilton warned, “will compel nations the most attached to liberty to resort for repose and security to institutions which have a tendency to destroy their civil and political rights. To be more safe, they at length become willing to run the risk of being less free.”¹²⁴

b. Enhance Public, Policymaker, Press, and Private-Sector Confidence

Privacy and national security are also inherently linked because there are limits as to how much of the former the public is willing to trade in pursuit of the latter. The clear lesson of the series of controversies over data mining programs is that the American people will rebel and policymakers will change direction in an instant if they believe that privacy is being threatened too much or unnecessarily.

With TIA, as we have seen, Congress restricted development and then terminated funding entirely, at least from the public budget.¹²⁵ But other programs have been similarly retarded by a privacy backlash. In response to public and political pressure, CAPPS II was scaled back and the data mining aspects limited merely to verifying identify and determining if a passenger is on a government terrorist watch list. Delta Air Lines withdrew from a pilot program after it was threatened with a boycott.

“MATRIX” (Multistate Anti-Terrorism Information Exchange)—designed “to link law enforcement records across states with other government and private-sector databases” and to “find patterns and links among people and events faster than ever before”—has been hard hit by privacy concerns.¹²⁶ At its height, 16 states were participating in MATRIX, which is funded by the Justice Department and DHS. Partly in response to privacy issues, all but five states have withdrawn.¹²⁷

The experience of companies who participated voluntarily in a test of how data profiling can be used to identify high-risk passengers has been particularly illuminating. With the assistance of DOD and TSA, Army defense contractor Torch Concepts, obtained millions of

¹²³ Alexander Hamilton, “The Consequences of Hostilities Between the States” (Federalist Paper 8), *New York Packet*, Nov. 20, 1787.

¹²⁴ *Id.*

¹²⁵ Department of Defense Appropriations Act, 2004, Pub. L. No. 108-84, § 8183 (Sept. 25, 2003).

¹²⁶ Thomas C. Greene, “A Back Door to Poindexter’s Orwellian Dream,” *The Register*, Sept. 24, 2003; Robert O’Harrow, Jr., “U.S. Backs Florida’s New Counterterrorism Database,” *Washington Post*, Aug. 6, 2003, at A1; see also <http://www.matrix-at.org/>.

¹²⁷ Chris Maag, “The Matrix: An Expensive Government Program Was Doomed from the start,” *Cleveland Scene* (Ohio), Mar. 9, 2005.

passenger records from U.S. airlines to help test the system it was designing.¹²⁸ For many of the passengers Torch Concepts was able to buy demographic information including data on gender, occupation, income, Social Security Number, home ownership, years at current residence, number of children and adults in the household, and vehicles.¹²⁹ Now, JetBlue, Northwest, and American, all of whom provided passenger data for the test, face multiple class-action lawsuits under a variety of federal and state laws, as does Acxiom, the supplier of the third-party demographic data.

Section 215 of the USA Patriot Act is under renewed attack for fear that it may be used to seize broad collections of data about the reading habits of people who have done nothing to warrant the government's attention. The *New York Times* editorialized on June 21, 2005, that "law enforcement should be able to get information, including library records, about specific individuals it reasonably suspects of a crime," but that the law as currently written "allows requests for library records for large numbers of people without any reason to believe they are involved in illegal activity."¹³⁰ According to the *Times*, this goes too far: "Fishing expeditions of this kind invade people's privacy and threaten to bring people under suspicion based on what they read."¹³¹ Some members of Congress appear to agree. On June 15, the House of Representatives voted to prohibit the use of section 215 to obtain "circulation records, library patron lists, book sales records or book customer lists" altogether.¹³²

While the retreat from each of these programs may have been justified in the circumstances, collectively they raise the specter that valuable tools for enhancing security may have been compromised. Moreover, the public outcry over these programs has made the government wary of security programs that involve data matching and industry hesitant to share personal data with the government.

Promises by proponents of all of these data mining projects that they were "adhering to the law" did little to quell the controversies, because the law is so limited and uncertain. Inadequate or unclear privacy laws are slowing the development of new and promising data mining programs, they are undermining research into this important weapon in the war on terrorism, and they are hampering the very data sharing that the 9/11 Commission recommended. Clear rules would facilitate accountability, public and policymaker confidence, and the willingness of the private-sector to provide data for lawful counter-terrorism uses. The absence of those rules undermines efforts to protect privacy and security.

¹²⁸ Sara Kehaulani Goo, "Airlines Confirm Giving Passenger Data to FBI After 9/11," *Washington Post*, May 2, 2004, at A14.

¹²⁹ Department of Homeland Security Privacy Office, *Report to the Public on Events Surrounding jetBlue Data Transfer—Findings and Recommendations* (Feb. 20, 2004).

¹³⁰ "Fishing in the Card Catalogs," *New York Times*, June 21, 2005, at A20.

¹³¹ *Id.*

¹³² Richard B. Schmitt, "House Weakens Patriot Act's 'Library Provision'," *Los Angeles Times*, June 16, 2005, at A1.

c. Enhance Security

Good privacy protection not only can help build support for data mining and other tools to enhance security, it can also contribute to making those tools more effective. For example, data integrity—ensuring that data are accurate, complete, up-to-date, and appropriately stored and linked—is a key privacy principle. But it clearly enhances security as well. Legal obligations requiring data integrity inevitably make those data more useful for security application as well.

In March 2003, the Justice Department exempted the FBI’s National Crime Information Center from the Privacy Act’s requirements that data be “accurate, relevant, timely and complete,”¹³³ and in August 2003, the DHS exempted the TSA’s passenger screening database from the Privacy Act’s requirements that government records include only “relevant and necessary” personal information.¹³⁴ These efforts to avoid privacy obligations raise important security issues as well. Mismatched data and misidentified individuals pose serious risks for both privacy and security.

Similarly, the DOD Inspector General’s December 2003 audit of TIA concluded that DOD’s failure to consider privacy adequacy during the early development of TIA led the Department to “risk spending funds to develop systems that may not be either deployable or used to their fullest potential without costly revision.”¹³⁵ The report noted that this was particularly true with regard to the potential deployment of TIA for law enforcement: “DARPA need[ed] to consider how TIA will be used in terms of law enforcement to ensure that privacy is built into the developmental process.”¹³⁶ Greater consideration of how the technology might be used would not only have served privacy, but also likely contributed to making TIA more useful as well.

As this example suggests, privacy protections often build discipline into counter-terrorism efforts that serves other laudatory purposes. By making the government stop and justify its effort to a senior official, a congressional committee, or a federal judge, warrant requirements and other privacy protections often help bring focus and precision to law enforcement and national security efforts. In point of fact, courts rarely refuse requests for judicial authorization to conduct surveillance. For example, between 1968 and 2003, courts approved a total of 30,692 wiretap orders (10,506 federal and 20,186 state)—all but 32 sought by the government.¹³⁷ Between 1979 and 2003, Foreign Intelligence Surveillance Court judges approved 16,971 FISA warrants—all but five that the Attorney General had sought.¹³⁸ As government officials often

¹³³ *Privacy Act of 1974; Implementation*, 68 Federal Register 14140 (2003) (DOJ, final rule).

¹³⁴ *Privacy Act of 1974: Implementation of Exemption*, 68 Federal Register 49410 (2003) (DHS, final rule).

¹³⁵ OIG Terrorism Information Awareness Program Report, *supra* at 4.

¹³⁶ *Id.* at 7.

¹³⁷ Administrative Office of the United States Courts, *2003 Wiretap Report*, tab. 7; Electronic Privacy Information Center, *Title III Electronic Surveillance 1968-2002*.

¹³⁸ Center for Democracy and Technology, *The Nature and Scope of Governmental Electronic Surveillance Activity* (June 2004); Electronic Privacy Information Center, *Foreign Intelligence Surveillance Act Orders 1979-2002*.

note, one reason for these high success rates is the quality of internal decision-making that the requirement to obtain judicial authorization requires.

As TAPAC noted in the introduction to its recommendations for new privacy protections:

Our conclusion, therefore, that data mining concerning U.S. persons inevitably raises privacy issues, does not in any way suggest that the government should not have the power to engage in data mining, subject to appropriate legal and technological protections. Quite the contrary, we believe that those protections are essential *so that* the government can engage in appropriate data mining when necessary to fight terrorism and defend our nation. And we believe that those protections are needed to provide clear guidance to DOD personnel engaged in anti-terrorism activities.¹³⁹

d. Improve Policymaking

One of the most striking lessons from Congress' response to TIA and other data mining programs is that the absence of a clear regulatory regime for data mining contributed to erratic and inconsistent behavior by policymakers. Clear standards are necessary not only to help guide the actions of counter-terrorism personnel, but also to help guide policymakers as well.

After all, it was only three months after Congress required DHS to engage in data mining with private-sector data that it prohibited DOD from deploying data mining tools within the United States, collecting or using data about U.S. persons, or developing other elements of TIA, including translation software, networks to link the intelligence community, and other tools that had few if any privacy implications.¹⁴⁰ Seven months later, Congress blocked the development of TIA entirely, but then established rules for how classified funding for TIA might be used. TIA's opponents in Congress and the privacy advocacy community proudly claimed that they had "killed" TIA, but the statutory language suggests that they had merely driven it from public view.¹⁴¹ Moreover, President Bush stated in his signing statement that the classified annex "accompanies but is not incorporated as a part of the Act" and therefore would be considered by the President as merely "advisory in effect."¹⁴² Ironically, the part of TIA that Congress did eliminate entirely was the funding for the development of privacy enhancing technologies.

The immediate result, therefore, of congressional intervention was to drive the development and deployment of data mining at DOD from public view, relieve it of the statutory restrictions that had previously applied to it, block funding for research into privacy enhancing

¹³⁹ *TAPAC Report*, supra at 48.

¹⁴⁰ S. Amend. 59 to H.J. Res. 2 (Jan. 23, 2003).

¹⁴¹ See K. A. Taipale, "Data Mining and Domestic Security: Connecting The Dots to Make Sense of Data," 5 *Columbia Science and Technology Law Review* 1, 48, n.96 (2003) ("The former TIA projects Genisys and Genoa II are believed to be included in the classified annex to the Defense Appropriations Bill," citing to the statement of Major General Paul Nielson before TAPAC, Nov. 20, 2003).

¹⁴² Statement on Signing the Department of Defense Appropriations Act, 2004 (Oct. 6, 2003).

technologies, and undermine the policy debate over the appropriate roles for and limits of data mining. Law and technology scholar K.A. Taipale writes:

At first hailed as a “victory” for civil liberties, it has become increasingly apparent that the defunding [of TIA] is likely to be a pyrrhic victory. . . . [N]ot proceeding with a focused government research and development project (in which Congressional oversight and a public debate could determine appropriate rules and procedures for use of these technologies and, importantly, ensure the development of privacy protecting technical features to support such policies) is likely to result in little security and, ultimately, brittle privacy protection.

Indeed, following the demise of IAO and TIA, it has become clear that similar data aggregation and automated analysis projects exist throughout various agencies and departments not subject to easy review.¹⁴³

Congress’ inconsistent treatment of similar technologies confuses the public and government officials charged with following these widely varying statutes. It runs the risk of compromising the protection of both national security and information privacy. And it is the inevitable result of the absence of clear legal structure concerning data mining and access to third-party data.

e. Facilitate Innovation and Research

The inconsistency that results from the absence of a legal framework may have its longest term effect on the innovation and research that is necessary to improve the accuracy and effectiveness of data mining, enhance privacy protections, and develop next generation tools for fighting terrorism.

TAPAC explicitly recognized the importance of research into technological and other tools for making data mining more precise and accurate and for protecting privacy. One unfortunate consequence of Congress blocking further public development of TIA was to prohibit further research by DARPA into both data mining and privacy.

Congress’ inconsistency and the controversy that data mining projects have provoked in the absence of strong legal protections for privacy are likely to undermine forward-looking research elsewhere as well. After all, what federal funding agency would invest seriously in an area where Congress had already acted to ban research once, and what investigator would invest her career in research on such a politically sensitive subject? It is instructive to remember that DARPA funded the development of the precursor of the Internet as a secure tool for connecting defense researchers. Where would the World Wide Web be today if Congress, at the infancy of ARPANet in the 1960s, had prohibited further research because the emerging technologically posed a clear threat to privacy?

¹⁴³ Taipale, *supra* at 4 (citations omitted).

Clear standards are necessary to support the investment of financial, institutional, and human resources in often risky research that may not pay dividends for decades. But that type of research is essential to counter-terrorism efforts and to finding better ways of protecting privacy.

f. Make New Technologies Work

Some observers suggest that the issues presented by data mining will be resolved by technologies, not by law or policy. There are indeed technologies emerging, some of which, such as anonymous entity resolution and immutable audit trails, are both very promising and described elsewhere in this volume. But even the best technological solutions will still require a legal framework in which to operate, and the absence of that framework may not only slow their development and deployment, as described above, but make them entirely unworkable.

Anonymous entity resolution is a perfect example. This technology makes it possible to standardize and match data that is completely anonymized through a one-way hash function. Only when there is a match between data sets, for example, a terrorist watch list and a list of airline passengers, would the government be entitled to seek the underlying, personally identifiable information from the data source. The technology protects privacy, enhances the accuracy of matches, and promises to facilitate the sharing of information likely to enhance national security.

However, it will work only if the private sector is willing to share its data with the government, and to anonymize it appropriately before doing so. After the experiences of JetBlue, Northwest, and American, companies might understandably require some legal comfort before they are going to share even anonymized data.

There is going to be a need for rules about when and through what process the government may seek the underlying data. This is the key question that the Court's *Miller* decision and Congress' inaction have left unanswered. That void will have to be filled before the public will have confidence in the system. There will also need to be rules to help protect the system. While anonymous entity resolution systems are very secure, they can still be challenged by relentless attacks (for example, through so-called "dictionary attacks," where one party runs thousands of queries against another party's anonymized data in an effort to pierce the anonymization). We will need laws that stop users of the system from engaging in conduct designed to defeat the privacy protection it provides.

Similarly, technologies that create immutable audit trails hold great promise for monitoring access to data and ensuring that rules are followed, but there will need to be legal standards for when immutable audit trail technologies are used, who holds the audit trail data, and who can obtain access to them.

Information technologies, far from eliminating the need for law, actually exacerbate it. The failure of Congress and the Administration to adopt a coherent legal framework applicable to data mining threatens not only to eliminate the useful role of law in protection privacy and fighting terrorism, but to reduce the effectiveness of technologies as well. Moreover, there will always be gaps left by technological protections that law will be essential to fill.

7. Conclusion

In *Miller v. United States* and subsequent cases the Supreme Court created a broad gap in the privacy protection provided by the Fourth Amendment by finding that the government's seizure of personal information from third parties is outside of the scope of the Fourth Amendment. As a result, the government's behavior need not be reasonable nor is any judicial authorization required when the government searches or seizes personal information held by third parties.

As striking as the Court's decision was in 1976, in the face of 29 years of technological developments since then, it today means that the government has at its disposal an extraordinary array of personal data that individuals necessarily deposit in the hands of third parties as we live our daily lives. As we rely more and more on technologies, that situation will only increase, until the Fourth Amendment is entirely swallowed up by the *Miller* exclusion. Although Congress has responded with specific, sectoral statutes, these are limited in their scope and in the protections they create. As a result, the government's ability to seize data from third parties is effectively unregulated.

Until recently, the government has had little practical use for massive datasets from the private sector. Significant advances in data mining technologies, however, now make it possible for the government to conduct sophisticated analysis, rapidly and affordably, of disparate databases without ever physically bringing the data together. These technologies allow the government to move beyond looking for data on specific people to search data about millions of Americans in the search for patterns of activity, subtle relationships, and inferences about future behavior. These technologies and the terrorist attacks of September 11 mean that the government now has both the ability and the motivation to use huge arrays of private-sector data about individuals who have done nothing to warrant government attention.

Even if *Miller* had not excluded these records from the protection of the Fourth Amendment, there would still be a critical need for Congress to establish a legal framework for the appropriate use of data mining. To date, Congress has failed to respond to this challenge. In fact, Congress has behaved erratically toward data mining, requiring and encouraging it in some settings and prohibiting it in others.

There is an urgent need for Congress and the Administration to address this situation by creating clear legal standards for government data mining, especially when it involves access to third-party data. It is beyond the scope of this chapter to try to articulate the content of those standards. There have been many efforts to do so, including the work of TAPAC, the Markle Foundation Task Force on National Security in the Information Age, the Cantigny Conference on Counterterrorism Technology and Privacy organized by the Standing Committee on Law and National Security of the American Bar Association,¹⁴⁴ think tanks and advocacy groups concerned with national security and civil liberties issues, and individuals, including other contributors to this volume.

¹⁴⁴ "The Cantigny Principles on Technology, Terrorism, and Privacy," *National Security Law Report*, Feb. 2005, at 14.

Standards for government data mining and access to third-party data are essential to protect privacy, build public confidence in appropriate data mining, enhance national security, facilitate more rational and consistent policymaking, foster innovation, and help new technologies for protecting privacy and security reach their full potential.