2008

Government Data Mining: The Need for a Legal Framework

Fred H. Cate
Indiana University Maurer School of Law, fcate@indiana.edu

Government Data Mining: The Need for a Legal Framework

*Fred H. Cate**

I. Introduction

The United States government has long sought data about individuals for a wide variety of important public purposes. The process of collecting this information was often time-consuming and expensive and resulted in data that were difficult to use because of the form in which they were captured. The Supreme Court described the effect as "practical obscurity."[1] Much of the "privacy" Americans have enjoyed results from the fact that it was simply too expensive or laborious to find out intimate data about them.

In the twenty-first century, technology and law have combined to erode the protection for personal privacy previously afforded by practical obscurity. Advances in digital technologies have greatly expanded the volume of personal data created as individuals engage in everyday activities. "Today, our biographies are etched in the ones and zeros we leave behind in daily digital transactions,"[2] Professor Kathleen Sullivan has written. Moreover, technology has contributed to an explosion not only in the ubiquity of data, but also in the range of parties with physical access to those data and in the practical and economic ability of those parties to collect, store, share, and use those digital footprints.

At the same time, the Supreme Court has refused to extend the Fourth Amendment to restrict the government's access to data held by third parties. In the 1976 decision *United States v. Miller*, the Court held that because there can be no reasonable expectation of privacy in information held by a

* Distinguished Professor and Director of the Center for Applied Cybersecurity Research, Indiana University; Senior Policy Advisor, Center for Information Policy Leadership at Hunton & Williams LLP. The author has participated in drafting a number of documents cited in this article as the reporter for the American Law Institute's project on Principles of the Law on Government Access to and Use of Personal Digital Information, counsel to the Department of Defense Technology and Privacy Advisory Committee, reporter for the third report of the Markle Foundation Force on National Security in the Information Age, and a panelist at the Cantigny Conference on Counterterrorism Technology and Privacy, organized by the Standing Committee on Law and National Security of the American Bar Association. The author also serves as a member of the National Academy of Sciences Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals. I am grateful for the instruction of my colleagues in these settings; however, the views expressed herein should not be attributed to them or to the sponsoring organizations. I also appreciate the generous help of Professor Craig Bradley, Beth E. Cate, Benjamin Keele, Michael Riskin, Professor Paul Schwartz, and Stefaan Verh. I alone am responsible for any errors.

[1] U.S. Dep't of Justice v. Reporters Comm., 489 U.S. 749, 762 (1989).

[2] Kathleen M. Sullivan, *Under a Watchful Eye: Incursions on Personal Privacy*, *in* The War on Our Freedoms: Civil Liberties in an Age of Terrorism 128, 131 (Richard C. Leone & Greg Anrig, Jr., eds., 2003).

third party, even if the third party possesses it because of a legal obligation to do so, the Fourth Amendment does not apply to the government's seizure of such data.[3] The *Miller* third-party exception thus invites an end run around the Fourth Amendment. As serious a threat to privacy as this may have posed when *Miller* was decided, the danger is far greater today, when for commercial and regulatory reasons, individuals' everyday activities are routinely captured in digital records.

The government faces new and intense pressure to collect and use personal data. Much of that pressure reflects the conviction that greater reliance on digital data will reduce costs and enhance convenience, speed, efficiency, and accountability. Perhaps the greatest source of that pressure, however, is the fear of terrorist attacks and the widely shared view, as the National Commission on Terrorist Attacks Upon the United States (commonly referred to as the 9-11 Commission) Vice Chairman Lee Hamilton testified before Congress in November 2005, that the inability of federal agencies to marshal and share information about suspected terrorists and their activities "was the single greatest failure of our government in the lead-up to the 9/11 attacks."[4]

This indictment has led Congress and the President to expand the authority of the government to collect personal data through mandatory disclosure, seizure, independent creation, and purchase.[5] It has also helped to fuel an apparently insatiable government appetite for access to and retention of personal data, especially from the vast databases routinely maintained by the private sector. The government uses these data sets for a spectrum of data mining activities, ranging from inquiries on specific individuals and the people with whom they interact to broad searches for unusual or predetermined patterns of activities or relationships.[6]

Data mining poses significant legal and policy issues. Many of these concern the government's access to data, especially from the private sector. In the absence of either practical obscurity or effective legal privacy protections, the government has unprecedented and virtually unlimited access to an extraordinary volume and variety of personal data on the behaviors, attributes, resources, associates, and beliefs of individuals who have done nothing to warrant suspicion. In addition, the government's use of the data creates serious concerns. Although data mining can have real consequences for individuals identified, it occurs without legal guarantees for the accuracy or appropriateness of the data or the searches, redress for people injured by being falsely identified as posing a threat, or judicial or legislative oversight. In

---

[3] 425 U.S. 435, 436 (1976); *see also* Cal. Bankers Ass'n v. Shultz, 416 U.S. 21 (1974).

[4] *Federal Support for Homeland Security Information Sharing: Role of the Information Sharing Program Manager: Hearing Before the Subcomm. on Intelligence Information Sharing and Risk Assessment of the H. Comm. on Homeland Security*, 109th Cong. 23 (2005) (statement of Lee Hamilton, Vice Chairman, 9/11 Public Discourse Project).

[5] *See, e.g.*, Alexandra Markes, *Privacy Advocates Fight for Ground Lost After 9/11*, CHRISTIAN SCI. MONITOR, Apr. 3, 2007, at 2.

[6] *See* Newton N. Minow & Fred H. Cate, *Government Data Mining*, *in* MCGRAW-HILL HANDBOOK OF HOMELAND SECURITY 1063, 1065-66 (David G. Kamien ed., 2005).

fact, most government data mining today occurs in a legal vacuum outside the scope of the Fourth Amendment and without a statutory or regulatory framework.

The absence of a legal regime governing data mining not only fuels privacy concerns, but also runs the risk of compromising the very objectives that data mining is designed to serve by permitting the use of outdated, inaccurate, and inappropriate data. It denies government officials guidance as to what is and is not acceptable conduct. The lack of a modern, coherent legal regime interferes with the ability of businesses and other possessors of potentially relevant databases to know when they can legally share information with the government. It slows the development of new and promising data mining programs, undermining research into this potentially important tool and hampering appropriate data sharing. And it significantly undercuts the confidence of the public and of policymakers that data mining will be carried out effectively or with appropriate attention to protecting privacy and other civil liberties.

This Article examines some of the critical issues surrounding the government's collection and use of personal data for data mining, especially for law enforcement and national security purposes—the area of greatest growth and most recent controversy. In particular, this Article focuses on the failure of law and the legal system to respond to the proliferation of data mining and the dramatic technological changes that make it possible. Part II surveys some of the many recent data mining efforts initiated by the government, especially in the law enforcement and counter-terrorism areas, and the range of government authority to seize, require the disclosure of, or purchase third-party data. Part III examines the Supreme Court's exclusion in *Miller* and subsequent cases of third-party data from the privacy protection of the Fourth Amendment. Part IV addresses Congress's privacy legislation and its failure to fill the gap created by *Miller* or to respond to the proliferation in government data mining. Part V suggests the range of issues that these programs raise in the face of a legal vacuum. Part VI offers recommendations for marshalling the potential power of data mining for appropriate uses while protecting personal privacy. Although addressed specifically to national security and law enforcement data mining, these recommendations apply equally to government data mining for other purposes.[7]

---

[7] This Article does not address the regulation of data collection and use in the private sector. Clearly, these issues relate to government data mining, since, as this article argues, the private sector is a major source of personal data used by the government, and the controversy over public-sector data mining affects the debate over private-sector activities. However, the issues are simply too broad to address together effectively in a single article. Moreover, there are important conceptual distinctions: the Fourth Amendment applies only to government searches and seizures, and only the government's data collection is entirely free from the constraints of competitive markets because only the government has the power to compel the production of personal data. Finally, there are important practical distinctions, because, as described in greater detail below, even when Congress has enacted privacy protections applica-

II. Data Mining Initiatives

A. Data Mining

"Data mining" is defined in many different ways but is perhaps best understood as encompassing a wide spectrum of data-based activities ranging from "subject-based" searches for information on specified individuals to "pattern-based" searches for unusual or predetermined patterns of activities or relationships.[8] Between these two ends are "relational" searches, which start with an individual but then reach out to determine who communicates or otherwise interacts with whom, and "data matching," which involves combining two or more sets of data looking for matches or discrepancies.[9]

Government agencies have long made use of subject-based, relational, and data matching searches. For example, law enforcement officials often search for a specific suspect (e.g., the driver of the car, the person with the fingerprint at the scene of the crime). They also frequently rely on relational searches (e.g., who knew the murder victim, who frequented the drug house). Data matching is also widely used, for example, by tax officials who compare individual tax filings with the records of financial institutions and employers, or national security officials, who compare flight manifests and visa applications with lists of known and suspected terrorists.

Pattern-based searches are a more recent innovation for government agencies, although they have long been used by commercial entities for tar-

---

ble to the private sector, it has consistently exempted the provision of personal data to the government. *See infra* text accompanying notes 191-98.

[8] *See* Minow & Cate, *supra* note 6, at 1065-66.

[9] There is a broad spectrum of definitions of "data mining." At one end are the narrow definitions such as that identified in the General Accountability Office's May 2004 report on government data mining: "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." U.S. Gen. Accounting Office, GAO-04-548, Data Mining: Federal Efforts Cover a Wide Range of Uses 1 (2004), *available at* http://www.gao.gov/new.items/d04548.pdf. At the other end of the spectrum are far broader definitions. For example, the Department of Defense Technology and Privacy Advisory Committee in 2004 defined the term to include "searches of one or more electronic databases of information concerning U.S. persons, by or on behalf of an agency or employee of the government." Tech. and Privacy Advisory Comm., U.S. Dep't of Def., Safeguarding Privacy in the Fight Against Terrorism, at viii (2004) [hereinafter, TAPAC, Safeguarding Privacy].

This Article employs the more comprehensive definition of data mining described above, in part because of the changing nature and difficulty of distinguishing among different data analysis techniques. The broader the search criteria, and the more people who will be identified by them, the more pattern-like subject-based 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. In addition, I prefer the broader definition in the belief that rules that apply more broadly, albeit with appropriate sensitivity to the distinguishing characteristics of different types of data mining, are more useful than a large number of narrower rules.

get marketing and risk assessment.[10] Businesses develop a pattern of attributes or behaviors that their good customers have in common and then search databases to find people meeting those patterns. Over the past decade, government agencies have experimented with applying similar technologies to other activities, such as detecting fraud or tax evasion.[11] After the terrorist attacks of September 11, 2001, pattern-based data mining struck many observers as a promising tool for law enforcement and national security. If government officials could develop models of what criminal or terrorist behavior might look like and then search for those patterns across a sufficiently broad range of information, observers hoped it would be possible to detect criminals or terrorists, perhaps even before they executed their nefarious enterprises. In the Homeland Security Act of 2002, Congress required the new Department of Homeland Security ("DHS") to "establish and utilize . . . data-mining and other advanced analytical tools" to "access, receive, and analyze data to detect and identify threats of terrorism against the United States."[12]

A 2004 report by the then-General Accounting Office ("GAO") found that forty-two federal departments—including every cabinet-level agency that responded to the survey—engaged in, or were planning to engage in, 122 pattern-based data mining efforts involving personal information.[13] Thirty-six of those involve accessing data from the private sector; forty-six involve sharing data among federal agencies.[14] Fourteen data mining programs in the GAO report are concerned with "[a]nalyzing intelligence and detecting terrorist activities" and fifteen involve "[d]etecting criminal activities or patterns."[15] Of these twenty-nine national security and law enforcement programs, all but four use personal data.[16] The following section examines a cross-section of government data mining programs involving third-party data, many of which were not included in the GAO's 2004 survey even though they were ongoing at the time.

B. Government Data Mining Programs

1. Administrative and Regulatory Programs

The government today increasingly relies on personal data—obtained not only from third parties, but also directly from individuals—to administer social service programs such as Social Security, Medicare, and workers'

---

[10] *See* JEFFREY W. SEIFERT, CONG. RESEARCH SERV., DATA MINING AND HOMELAND SECURITY: AN OVERVIEW 4 (2007).

[11] *Id.*

[12] 6 U.S.C. § 201(d)(1), (d)(14) (2000).

[13] U.S. GEN. ACCOUNTING OFFICE, *supra* note 9, at 3, 27-64 tbls. 2-25.

[14] *Id.* at 3.

[15] *Id.* at 8 tbl. 1.

[16] *See id.* at 10 fig. 1.

compensation insurance; to administer tax programs and collect revenue; to issue licenses for many personal, business, and professional activities; to support hundreds of regulatory regimes ranging from voter registration and political campaign contributor disclosures to employee identity verification to child support obligation enforcement; to maintain vital records about major lifecycle events, including birth, marriage, divorce, adoption, and death; and to operate facilities such as toll roads and national parks. The role of personal information collected as part of these programs is striking and reflects what Professor Paul Schwartz has described as the "*data processing model* of administrative control."[17] In this model, government agencies become largely information processors, substituting information-based determinations for what previously might have involved subjective judgment by clerks, who feed data into computers and act on the result, for professionals who assess and evaluate independently. The increased reliance on personal data helps to provide services to a larger population, diminishes the perceived inequality of subjective determinations, reduces the costs of litigating decisions and maintaining more skilled personnel, and enhances accountability. "Compared to its historic role, the state today depends upon the availability of vast quantities of information, and much of the data it now collects relates to identifiable individuals."[18]

Combined with the twentieth-century expansion of government services and oversight into the market and the family, the net effect of the evolution towards a data processing model of administration is that "[b]ureaucracies now use data processing to manage information about every aspect of human existence. . . . Data are now gathered about every individual before birth, during life, and after death."[19] It is no exaggeration to say that "[i]nformation is the lifeblood of regulatory policy,"[20] and "[r]egulators depend on information for nearly everything they do."[21]

a. Government Benefits and Social Service Programs

Many of the government's data-based programs involve the delivery of social services, which would be "impossible without detailed information on the citizen as client, customer, or simply person to be controlled."[22] The largest of these programs, such as Social Security, Medicare, and workers' compensation, involve most Americans.[23] The personal information col-

---

[17] Paul Schwartz, *Data Processing and Government Administration: The Failure of the American Legal Response to the Computer*, 43 HASTINGS L.J. 1321, 1325 (1992) (emphasis in original).

[18] *Id.* at 1332.

[19] *Id.* at 1328-29.

[20] Cary Coglianese, Richard Zeckhauser & Edward Parson, *Seeking Truth for Power: Informational Strategy and Regulatory Policymaking*, 89 MINN. L. REV. 277, 285 (2004).

[21] *Id.*

[22] Schwartz, *supra* note 17, at 1332.

[23] Other social service programs include Medicaid, Social Security disability insurance, Supplemental Security Income, WIC (Special Supplemental Nutrition Program for Women,

lected under these and similar programs includes identifying data (such as Social Security Number ("SSN")) about the applicants and family members, extensive financial information, health information, and data on what products or services are purchased or consumed. The information is collected not only from the data subject—the person to whom the information pertains—but also from third parties such as health care providers, employers, and service providers. The Social Security Administration's ("SSA's") Numerical Identification File ("NUMIDENT"), for example, maintains identifying information, including name, birth date, citizenship, and SSN, on more than 441 million individuals, and the information is accessible to other government agencies and private employers.[24]

b. Taxes

The government also collects and stores extensive personal information to administer tax programs and collect revenue. The Internal Revenue Service ("IRS") estimated in 2002 that it collected data on 116 million individual taxpayers, 45 million fully or partially self-employed individuals and small businesses, 210,000 larger corporations, and 2.4 million not-for-profit entities.[25] These data include not only self-reported information on identity, income, and activities, but also a vast array of third-party data that include both identifying information (such as SSN) and financial data.[26] As Professor Lillian BeVier has written, "[i]n part because the Internal Revenue Code has become such a complex maze of deductions, exemptions, surcharges, and credits, citizens cannot pay taxes without at the same time providing the government with quite detailed information about their families, jobs, investments, misfortunes, and favorite charities."[27] In addition, individuals are required to provide even more personal data to private-sector institutions so

Infants and Children), Nutrition Program for the Elderly, Child and Adult Care Food Program, School Lunch and Breakfast Program, Senior Farmers' Market Nutrition Program, Temporary Assistance for Needy Families, Aid to Families with Dependent Children, and unemployment insurance.

[24] *Employment Eligibility Verification Systems: Hearing Before the Subcomm. on Social Security of the H. Comm. on Ways and Means*, 110th Cong. (2007) (statement of Frederick G. Streckewald, Assistant Deputy Comm'r for Disability and Income Security Programs, Social Security Admin.), http://waysandmeans.house.gov/hearings.asp?formmode=view&id=6093.

[25] DEP'T OF THE TREASURY, A REPORT TO CONGRESS IN ACCORDANCE WITH § 357 OF THE UNITING AND STRENGTHENING AMERICA BY PROVIDING APPROPRIATE TOOLS REQUIRED TO INTERCEPT AND OBSTRUCT TERRORISM ACT OF 2001, at 12 n.11 (2002).

[26] *See* Press Release, Internal Revenue Serv., Dep't of the Treasury, IRS Updates National Research Program for Individuals IR-2007-113 (June 6, 2007), *available at* http://www.irs.gov/newsroom/article/0,,id=171023,00.html (last visited Mar. 13, 2008).

[27] Lillian R. BeVier, *Information About Individuals in the Hands of Government: Some Reflections on Mechanisms for Privacy Protection*, 4 WM. & MARY BILL RTS. J. 455, 456 (1995).

that those institutions have the data they need to comply with their tax reporting requirements.[28]

c. Employment

Under the U.S. Citizenship and Immigration Services' new E-Verify employment verification program, within three days of each new employee's hiring date many U.S. employers must enter basic identification information—including SSN and name, date of birth, citizen status claimed by employee, and other data—into an automated government database.[29] The database attempts to match the data against the SSA's NUMIDENT database or, for noncitizens, DHS databases.[30]

Employers are also required to report to their "State Directory of New Hires" the name, address, and SSN of all new hires within twenty days of their hiring date, and then to withhold from their paychecks any child support payments they may owe.[31] The states can determine whether any amount is owed by matching data with another federal database, the Federal Case Registry, which centralizes data from federally mandated State Case Registries about all child support orders established or modified on or after October 1, 1998.[32] Even parents who are not in the New Hires Directory might be located because Congress also has mandated the creation of the Federal Parent Locator Service ("FPLS").[33] This service accesses data from, and provides data to, not only the New Hires Directory but also federal tax authorities, the Department of State passport database, private-sector financial institutions, insurers and their agents, and other public- and private-sector sources.[34]

2. Law Enforcement

The Federal Bureau of Investigation ("FBI") maintains extensive databases in its Criminal Justice Information Services Division ("CJISD") that collect data from, and supply data to, a wide array of public- and pri-

---

[28] *See* INTERNAL REVENUE SERVICE, DEP'T OF THE TREASURY, EMPLOYMENT TAXES FOR BUSINESSES, http://www.irs.gov/businesses/small/article/0,,id=172179,00.html (last visited Mar. 13, 2008).

[29] *See* U.S. CITIZENSHIP AND IMMIGRATION SERVICES, DEP'T OF HOMELAND SEC., I AM AN EMPLOYER . . . HOW DO I . . . USE E-VERIFY?, M-655 (2007), *available at* http://www.uscis.gov/files/nativedocuments/E4_english.pdf (last visited Mar. 13, 2008).

[30] *See Employment Eligibility Verification Systems, supra* note 24, at 5 (testimony of Richard M. Stana, Director, Homeland Security and Justice Issues, Gen. Accountability Office), http://www.gao.gov/new.items/d07924t.pdf.

[31] Personal Responsibility and Work Opportunity Reconciliation Act of 1996, 42 U.S.C. § 653a (2000).

[32] *See id.* § 654a(e).

[33] 42 U.S.C.A. § 653 (West 2007).

[34] *Id.* § 653(e)(2).

vate-sector entities.[35] For example, the Integrated Automated Fingerprint Identification Service ("IAFIS") provides for automated data matching with three fingerprint databases containing 51 million records: the criminal history database; the civil file, containing records on individuals who have been required to submit fingerprints for employee background checks, security clearances, state licensure, and other non-criminal purposes; and the Unsolved Latent File, which includes fingerprints from crime scenes that could not be matched with either of the other databases.[36] The FBI's Next Generation Integrated Automated Fingerprint Identification System will not only allow for faster data mining, but also allow matching of other biometric identifiers.[37]

The FBI already collects data on one of those additional biometric identifiers—DNA. The Bureau's Combined DNA Index System ("CODIS") includes separate databases for DNA collected from: convicted criminals, arrestees, and parolees; forensic profiles from crime scenes; unidentified human remains; and missing persons and their relatives.[38] CODIS interconnects state and local databases to facilitate faster data matching. The Bureau has announced plans to spend $1 billion to build a more comprehensive biometric database.[39]

In addition, the FBI houses the national sex offender database, which aggregates information from the federally mandated state registries.[40] States are required to share information in their registries with the FBI, and vice versa, within three days of receiving it.[41] Registration is required of any person convicted of a "sexually violent offense" or a "criminal offense against a victim who is a minor."[42] The information that must be provided includes the offender's name, address, photograph, and fingerprints. Some state laws also require that the offender supply a biological specimen.[43]

The CJISD's National Crime Information Center ("NCIC") aggregates extensive data about missing persons, unidentified persons, criminal suspects wanted by law enforcement, sex offenders, federal prisoners, persons

---

[35] *See* Fed. Bureau of Investigation, Dep't of Justice, Fingerprint Identification Records System (1999), http://foia.fbi.gov/firs552.htm (last visited Mar. 13, 2008).

[36] *See* Jeff Carlyle, FBI Criminal Justice Information Services Division, at 2, http://fingerprint.nist.gov/standard/presentations/archives/IAFISoverview_Feb_2005.pdf (Feb. 2006).

[37] Jason Miller, *FBI Expanding Access to Fingerprint Database*, FCW.com, Aug. 22, 2007, http://www.fcw.com/online/news/103568-1.html.

[38] *See* Kimberly A. Polanco, *The Fourth Amendment Challenge to DNA Sampling of Arrestees Pursuant to the Justice for All Act of 2004*, 27 U. ARK. LITTLE ROCK L. REV. 483, 489 (2005).

[39] *See* Ellen Nakashima, *FBI Prepares Vast Database of Biometrics*, WASH. POST, Dec. 22, 2007, at A1.

[40] Pam Lychner Sexual Offender Tracking and Identification Act of 1996, 42 U.S.C. § 13701 (2000); Jacob Wetterling Crimes Against Children and Sexually Violent Offender Registration Act, 42 U.S.C. § 14071 (2000).

[41] 42 U.S.C.A. § 14071(b)(2) (West 2007).

[42] 42 U.S.C. § 14072(b) (2000).

[43] *See* Catherine L. Carpenter, *The Constitutionality of Strict Liability in Sex Offender Registration Laws*, 86 B.U. L. REV. 295, 332-33 (2006) (citations omitted).

on parole or probation, suspected terrorists, gangs, persons enrolled in the U.S. Marshal Service's Witness Security Program, victims of identity theft, foreign fugitives, and stolen vehicles and property.[44] In 2003, the NCIC contained 71 million state criminal history files.[45]

The FBI aggregates data from multiple databases into its Investigative Data Warehouse ("IDW").[46] According to press briefings given by the FBI in 2006, the IDW contains more than 659 million records, which come from 50 FBI and outside government agency sources.[47] The system's data mining tools are so sophisticated that they can handle many variations in names and other data, including up to twenty-nine variants of birth dates. The 13,000 agents and analysts who use the system average one million queries a month.[48]

3. National Security

a. Financial Institution Reporting

The government requires extensive data reporting by financial institutions as part of its counter-terrorism efforts. The Bank Secrecy Act, as amended in 2001 by the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001 ("USA PATRIOT Act"),[49] requires financial institutions and a wide range of other businesses to report to the government on certain transactions that are "determined to have a high degree of usefulness in criminal, tax, regulatory, intelligence, and counter-terrorism matters."[50] The nation's 24,000 banks and credit unions, as well as broker-dealers and commodity traders, must file Suspicious Activity Reports ("SARs") concerning suspicious financial transactions. More than 160,000 money service businesses (such as checking cashers and money transmitters) must register with the Department of Treasury. Currency Transaction Reports ("CTRs") for cash or coin transactions of $10,000 or more must be filed by financial institutions, the Post Office, casinos, travel agencies, pawnbrokers, real estate agents, automobile and boat retailers, jewelers, and anyone who accepts a check, travelers' check, or

---

[44] Fed. Bureau of Investigation, Dep't of Justice, National Crime Information Center, http://www.fbi.gov/hq/cjisd/ncic_brochure.htm (last visited Mar. 4, 2008).

[45] BUREAU OF JUSTICE STATISTICS, DEP'T OF JUSTICE, CRIMINAL RECORD SYSTEMS STATISTICS, http://www.ojp.usdoj.gov/bjs/crs.htm (last visited Mar. 4, 2008).

[46] *See Intelligence Reform – FBI and Homeland Security: Hearing Before the S. Select Comm. on Intelligence*, 110th Cong. 7 (2007) (testimony of John S. Pistole, Deputy Director, FBI), http://intelligence.senate.gov/070125/pistole.pdf.

[47] Ellen Nakashima, *FBI Shows Off Counterterrorism Database*, WASH. POST, Aug. 30, 2006, at A6.

[48] *Id.*

[49] Pub. L. No. 107-56, 115 Stat. 272 (2001) (codified in scattered sections in numerous titles of U.S.C.).

[50] DEP'T OF THE TREASURY, *supra* note 25, at 4.

money order.[51] The reports are received by the IRS and by the Treasury's Financial Crimes Enforcement Network ("FinCEN").

In 1996, these two agencies received 15,994,484 CTRs and 1,049,149 SARs.[52] FinCEN has collected and stored more than 75 million reports over the past decade.[53] According to a 2002 Treasury report, the agency combines these data "with other governmental and commercial information from a variety of data sources" and "link[s to] a variety of databases," to operate "one of the largest repositories of financial information available to law enforcement in the country."[54] State and local law enforcement "in every state" as well as federal law enforcement officials have online access to this information.[55]

The USA PATRIOT Act also mandates new rules requiring all 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 (e.g., a driver's license or passport); and (3) provide the information to the government for matching with terrorist watch lists.[56] This reporting and recordkeeping requirement is much broader than it might first appear because federal law defines "financial institutions" very broadly to include entities that "significantly engage" in activities as diverse as appraising real estate and personal property; leasing personal or real property; furnishing general economic information or statistical forecasting services; providing finance-related educational courses or instructional materials; providing tax-planning and tax-preparation services; providing ancillary services in or through a bank (such as notary public services, selling postage stamps or bus tickets, or providing vehicle registration services); and support services for any of these activities, including courier and data processing services.[57]

b. Aviation and Transportation Security

The Transportation Security Administration ("TSA") has struggled with how to screen airline and other passengers to determine whether they are on government terrorist watch lists or otherwise present a threat to aviation security. For example, the Federal Aviation Administration ("FAA") has required airlines to deny boarding or give "enhanced screening" to pas-

---

[51] *See id.* at 6; *See also* 31 U.S.C. § 5312(a)(2).

[52] *See Suspicious Activity and Currency Transaction Reports: Balancing Law Enforcement Utility and Regulatory Requirements: Hearing Before Subcomm. on Oversight and Investigation of the H. Comm. on Fin. Servs.*, 110th Cong. 46 (2007) (statement of William F. Baity, Deputy Director, Fin. Crimes Enforcement Network, Treasury Dep't).

[53] DEP'T OF THE TREASURY, *supra* note 25, at 9.

[54] *Id.*

[55] *See id.* at 10.

[56] *See, e.g.*, Customer Identification Programs for Banks, Savings Associations, Credit Unions, and Certain Non-Federally Regulated Banks, 68 Fed. Reg. 25,090 (June 9, 2003) (to be codified at 31 C.F.R. pts. 21, 103, 208, 211, 326, 563, and 748).

[57] *See* 12 C.F.R. § 225.86 (2007).

sengers identified by the FAA.[58] Furthermore, airlines for the past decade have been required to use a computer assisted passenger pre-screening system ("CAPPS"), which designated certain passengers as "Selectees" based on "customized, FAA-approved criteria."[59] Selectees undergo additional screening based on this crude form of subject- and pattern-based data mining. TSA has experimented unsuccessfully with Computer Assisted Passenger Pre-Screening System II ("CAPPS II") that would have reviewed passengers against both governmental and commercial databases to determine which of three levels of risk they posed.[60] This far more ambitious data mining was derailed by public and congressional controversy over privacy issues.[61]

DHS has published a notice of proposed rulemaking for "Secure Flight," a program that will require airlines to request each passenger's full name, gender, and birth date, and submit those data, along with the reservation record locator and other itinerary information,[62] to the TSA for matching against terrorist watch lists.[63] Passengers would only be required to provide their full name at the time of reservation to allow TSA to perform watch list matching. However, if the absence of the other requested information meant that the TSA had insufficient information to distinguish a passenger from a person on the watch list, the individual could "experience delays, be subject to additional screening, be denied transport, or be denied authorization to enter a sterile area."[64] In short, the TSA is trying to use the consequences of poor data matching to motivate passengers to provide more complete information necessary for more accurate matching.

The Automated Targeting System ("ATS") is designed to assess the risk of passengers, vehicles, and cargo entering or leaving the United States. Based on data from numerous sources, ATS compiles an assessment on the risks presented by each person (passenger or crew member) seeking to enter, exit, or transit through the United States by land, air, or sea; people who engage in any form of trade or other commercial transaction related to the importation or exportation of merchandise; and people who serve as booking

---

[58] National Commission on Terrorist Attacks Upon the United States, The Aviation Security System and the 9/11 Attacks, Staff Statement No. 3, at 6 (Jan. 27, 2004), *available at* http://www.9-11commission.gov/staff_statements/staff_statement_3.pdf.

[59] *Id.*

[60] *See* U.S. GEN. ACCOUNTING OFFICE, GAO-04-385, AVIATION SECURITY: COMPUTER-ASSISTED PASSENGER PRESCREENING SYSTEM FACES SIGNIFICANT IMPLEMENTATION CHALLENGES 6-7 (2004), *available at* http://www.gao.gov/new.items/d04385.pdf (last visited Mar. 13, 2008).

[61] Mimi Hall & Barbara DeLollis, *Plan to Collect Flier Data Cancelled*, USA TODAY, July 15, 2004, at 1A.

[62] Itinerary information is the following information about a flight: (1) departure airport code; (2) aircraft operator; (3) departure date; (4) departure time; (5) arrival date; (6) scheduled arrival time; (7) arrival airport code; (8) flight number; (9) operating carrier (if available). Secure Flight Program, 72 Fed. Reg. 48,356 (Aug. 23, 2007) (to be codified at 49 C.F.R. pts. 1540, 1544, and 1560).

[63] *See id.*

[64] *Id.*

agents, brokers, or who otherwise provide information on behalf of persons seeking to enter, exit, or transit through the United States.[65] This same information must be provided to Customs and Border Protection under the "Advance Passenger Information System" electronic data system.[66] These data are retained by the government for fifteen years under an agreement with the European Commission.[67]

c. SWIFT Subpoenas

Another prominent example of data mining conducted by the U.S. government involves disclosures of data about international bank transfers. The Society for Worldwide International Financial Telecommunication ("SWIFT") is a cooperative of financial institutions established under Belgian law in 1973[68] that supplies secure, standardized messaging services to more than 8,100 financial institutions in 208 countries.[69] While it is neither a payment system nor a settlement system, it transfers more than 13.4 million messages a day about international financial transactions.[70]

Beginning shortly after September 11, 2001, the Treasury Office of Foreign Assets Control began issuing administrative subpoenas for the data held in SWIFT's U.S. operations center. As of December 2006, SWIFT had received sixty-five subpoenas, each of which required it to provide the government with data "relevant to terrorism investigations."[71] The Treasury Office of Foreign Assets Control and SWIFT failed to make transparent the negotiations regarding the release of the personal data, and neither SWIFT nor the U.S. government has confirmed the total number of records involved.[72] SWIFT maintains that it limited U.S. access to its full database.[73] Under an agreement reached in June 2007 between U.S. and European officials, per-

---

[65] *See* Privacy Act of 1974, 5 U.S.C. § 552(a) (2000); U.S. Customs and Border Protection, Automated Targeting System, System of Records, 72 Fed. Reg. 43,650 (Aug. 6, 2007) (DHS, system of records notice of clarification).

[66] Advance Electronic Transmission of Passenger and Crew Member Manifests for Commercial Aircraft and Vessels, 72 Fed. Reg. 48,320 (Aug. 23, 2007) (to be codified at 19 C.F.R. pts. 4 and 22).

[67] *See* Agreement Between the European Union and the United States of America on the Processing and Transfer of Passenger Name Record (PNR) Data by Air Carriers to the United States Department of Homeland Security (DHS), June 23, 2007, 2007 O.J. (L 204) 23, *available at* http://eur-lex.europa.eu/LexUriServ/site/en/oj/2007/l_204/l_20420070804en00180025.pdf [hereinafter 2007 PNR Agreement].

[68] *See* Soc'y for Worldwide Int'l Fin. Telecomm., SWIFT history, http://www.swift.com/index.cfm?item_id=1243 (last visited Mar. 17, 2008).

[69] *See* Soc'y for Worldwide Int'l Fin. Telecomm., SWIFT in Figures—SWIFTNet FIN Traffic July 2007 YTD (2007), http://www.swift.com/index.cfm?item_id=63134 (last visited Mar. 13, 2008).

[70] *Id.*

[71] JENNIFER STODDART, OFFICE OF THE PRIVACY COMM'R OF CANADA, COMM'R'S FINDINGS ¶ 30 (2007), *available at* http://www.privcom.gc.ca/cf-dc/2007/swift_rep_070402_e.asp.

[72] *See* BELGIAN DATA PROTECTION COMMISSION, SUMMARY OF THE OPINION ON THE TRANSFER OF PERSONAL DATA BY SCRL SWIFT FOLLOWING THE UST (OFAC) SUBPOENAS 1 (2006).

[73] STODDART, *supra* note 71 at ¶ 34.

sonal data obtained from SWIFT will not be retained for longer than five years.[74]

d. Terrorist Surveillance Program

On December 16, 2005, the *New York Times* revealed that the National Security Agency (NSA) was intercepting communications where at least one party was located inside the United States, without obtaining judicial authorization.[75] In the face of the ensuing controversy, the President acknowledged the existence of the surveillance program, which he and other administration officials described as involving only communications into and out of the United States where there is a "reasonable basis to conclude that one party to the communication is a member of al Qaeda, affiliated with al Qaeda, or a member of an organization affiliated with al Qaeda."[76] In addition, the Administration reported that surveillance activities were "reviewed approximately every 45 days" by the Attorney General to ensure they were being conducted "properly."[77] Administration officials have described this program as the "Terrorist Surveillance Program," and have acknowledged that it is only one of a "number of intelligence activities [that] were authorized in one order."[78]

The Administration has pursued and defended the Terrorist Surveillance Program with more sustained vigor than any other publicly acknowledged data mining program. It was the subject of the late-night visit to the hospital bedside of Attorney General John Ashcroft by White House Chief of Staff Andrew Card and then-Counsel Alberto Gonzalez in an effort to persuade the ailing Attorney General to overrule his deputy and reauthorize the program.[79] It was also at the heart of a successful Administration lobbying effort to persuade Congress to amend federal law to temporarily eliminate judicial oversight of surveillance "directed at a person reasonably believed to be located outside of the United States."[80]

---

[74] *See* James Risen, *U.S. Reaches Tentative Deal with Europe on Bank Data*, N.Y. TIMES, June 29, 2007, at A6.

[75] *See* James Risen & Eric Lichtblau, *Bush Lets U.S. Spy on Callers Without Courts*, N.Y. TIMES, Dec. 16, 2005, at A1.

[76] Press Briefing, Alberto Gonzalez, Att'y Gen. & General Michael Hayden, Principal Deputy Dir. of Nat'l Intelligence (Dec. 19, 2005), *available at* http://www.whitehouse.gov/news/releases/2005/12/20051219-1.html.

[77] Press Conference, President George W. Bush (Dec. 19, 2005), *available at* http://www.whitehouse.gov/news/releases/2005/12/20051219-2.html; *see also* U.S. DEP'T OF JUSTICE, LEGAL AUTHORITIES SUPPORTING THE ACTIVITIES OF THE NAT'L SEC. AGENCY DESCRIBED BY THE PRESIDENT (2006), http://www.usdoj.gov/opa/whitepaperonnsalegalauthorities.pdf.

[78] Letter from J.M. McConnell, Director of National Intelligence, to Senator Arlen Specter, Ranking Member of the Senate Judiciary Committee (July 31, 2007), *available at* http://www.washingtonpost.com/wp-srv/politics/documents/NID_Specter073107.pdf.

[79] *See* Jeffrey Rosen, *Conscience of a Conservative*, N.Y. TIMES, Sept. 9, 2007, at 40.

[80] Protect America Act of 2007, S. 1927, 110th Cong. § 105A (2007). *See generally* Joby Warrick & Walter Pincus, *How the Fight for Vast New Spying Powers Was Won*, WASH. POST, Aug. 12, 2007, at A1.

e. Domestic Surveillance Program

One program that may be among the other "intelligence activities" authorized by the President involves the installation by the NSA of sophisticated surveillance equipment in domestic switching facilities operated by AT&T and Verizon. According to class-action lawsuits brought by the Electronic Frontier Foundation, the equipment was used to "intercept[ ] and disclos[e] to the government the contents of its customers' communications as well as detailed communications records about millions of its customers."[81] *USA Today* reported in June 2006 that nineteen lawmakers who had been briefed on the program "verified that the NSA has built a database that includes records of Americans' domestic phone calls."[82] The data to which the NSA program has access include records, such as "the numbers dialed and the length of calls,"[83] about "most telephone calls in the United States,"[84] potentially "hundreds of billions of telephone calls each year."[85] This includes purely domestic communications.

Some telecommunications experts have asserted that the fiber optic connections to the NSA equipment are too large and are connected to the wrong part of the telephone network to be collecting only billing records. For example, investigative journalist Seymour Hersh, writing in the *New Yorker* in May 2006, quoted an unnamed "security consultant" as saying that the government had "direct access to the carrier's network core—the critical area of its system, where all of its data are stored. 'What the companies are doing is worse than turning over records,' the consultant said. 'They're providing total access to all the data.'"[86]

f. Total Information Awareness

The most visible and controversial data mining initiative to date has been the Defense Advanced Research Projects Agency ("DARPA") project ironically named "Total Information Awareness" ("TIA")—later renamed "Terrorism Information Awareness." TIA included technologies to search personally identifiable transaction records and recognize patterns across separate databases for the purpose of combating terrorism.[87] Speaking at the DARPATech 2002 Conference, John Poindexter, retired Admiral and director of DARPA's Information Awareness Office ("IAO"), described the need to

---

[81] Amended Complaint for Damages, Declaratory and Injunctive Relief at ¶ 6, Hepting v. AT&T Corp., No. C-06-0672-JCS (N.D. Cal. Feb. 22, 2006), *available at* http://www.eff.org/files/filenode/att/att_complaint_amended.pdf.

[82] Susan Page, *Lawmakers: NSA Database Incomplete*, USA TODAY, June 30, 2006, at 2A.

[83] *Id.*

[84] Eric Lichtblau & Scott Shane, *Bush is Pressed over New Report on Surveillance*, N.Y. TIMES, May 12, 2006, at A1.

[85] Barton Gellman & Arshad Mohammed, *Data on Phone Calls Monitored*, WASH. POST, May 12, 2006, at A1.

[86] Seymour M. Hersh, *Listening In*, NEW YORKER, May 29, 2006, at 25.

[87] *See* TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 15-20.

"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."[88]

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."[89] "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."[90] He then showed a slide of categories of transaction data that included "Communications, Financial, Education, Travel, Medical, Veterinary, Country Entry, Place/Event Entry, Transportation, Housing, Critical Resources, and Government" records.[91]

> According to a subsequent DARPA report, "Red Teams" would: [I]magine 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 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.[92]

This is the classic statement of pattern-based data mining: develop patterns of the targeted behavior and then search across databases to detect those patterns. But the DARPA assurance that the subsequent searches would be performed only on databases to which the government had "lawful access" did little to quell mounting opposition to the program, since the government has lawful access to virtually all private-sector databases.

On January 23, 2003, in response to a storm of protest about TIA's potential impact on privacy ignited by a column by William Safire,[93] the Senate adopted an amendment to the Omnibus Appropriations Act that prohibited deployment of TIA in connection with data about U.S. persons with-

---

[88] John Poindexter, Director, Info. Awareness Office, Overview of the Info. Awareness Office, Prepared Remarks for Delivery at DARPATech 2002 Conference (Aug. 2, 2002), at 1, *available at* http://www.fas.org/irp/agency/dod/poindexter.html.

[89] *Id.* at 2.

[90] *Id.*

[91] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 15.

[92] INFO. AWARENESS OFFICE, U.S. DEP'T OF DEF., REPORT TO CONGRESS REGARDING THE TERRORISM INFORMATION AWARENESS PROGRAM 15 (2003), *available at* http://usacm.acm.org/usacm/PDF/TIA_May_20_2003_report.pdf.

[93] William Safire, *You Are a Suspect*, N.Y. TIMES, Nov. 14, 2002, at A35.

out specific congressional authorization.[94] Eight months later, Congress terminated funding for TIA, with the exception of "[p]rocessing, analysis, and collaboration tools for counterterrorism foreign intelligence" specified in a classified annex to the Act.[95] This reference to the classified annex suggested that maybe research on data mining had merely been moved out of sight. According to press reports, the new home for the TIA successor is the Disruptive Technology Office under the Director of National Intelligence.[96]

C. Summary

These are just a sample of the disclosed government programs that collect and use personal data for data mining. They almost all have in common their reliance, in whole or in part, on data supplied—in most cases through some compulsory process—by the private sector. Most are part of some essential government service, whether administering social services, collecting revenue, enforcing the law, or protecting national security. It is in these latter two areas that we have seen the greatest growth in government data mining over the past seven years, the greatest reliance on third-party data, and the most heated controversy.

III. Constitutional Protection for Information Privacy: the Fourth Amendment

Historically, the primary constitutional limit on the government's ability to obtain personal information about individuals is the Fourth Amendment, which reflects the Framers' hostility to "general searches"—searches not based on specific suspicion.[97] Since such searches are at the heart of most government data mining programs, which involve collecting and analyzing vast swaths of data about individuals who have done nothing to warrant the government's suspicion, this section examines the Supreme Court's interpretation of the Fourth Amendment and its application to data obtained from third parties.

A. Framework

The Fourth Amendment does not purport to keep the government from conducting searches or seizing personal information. It only prohibits "unreasonable" searches and seizures but is silent about what makes a search or

---

[94] *See* S. Amend. 59 to H.R.J. Res. 2, 108th Cong. (Jan. 23, 2003); *see* Consolidated Appropriations Resolution of 2003, 10 U.S.C. § 2241 (Supp. III 2003).

[95] Department of Defense Appropriations Act, 2004, Pub. L. No. 108-87, § 8131, 117 Stat. 1054, 1102 (2003); *see also* H.R. Rep. No. 108-283 (2003) (Conf. Rep.) ("The conferees are concerned about the activities of the Information Awareness Office and direct that the Office be terminated immediately.").

[96] *See* Shane Harris, *TIA Lives On*, National Journal, Feb. 23, 2006.

[97] U.S. Const. amend. IV.

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.'"[98] 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.[99]

The Supreme Court interprets the Fourth Amendment also to require that certain searches be conducted only with a warrant issued by a court, even though this is not a requirement contained in the amendment itself.[100] 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.[101] The Supreme Court also generally requires that the government provide the subject of a search with contemporaneous notice of the search.[102]

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).[103] 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 (i.e., U.S. citizens and permanent resident aliens) who do not have a connection to a foreign power.[104] 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 those that are conducted wholly outside of the United States.[105]

Where it does apply, the Fourth Amendment's protection, while considerable, 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,[106] for searches that are conducted incidental to valid arrests,[107] or for searches specially authorized by the Attorney General or the

---

[98] Katz v. United States, 389 U.S. 347, 361 (1967).

[99] *See* Terry v. Ohio, 392 U.S. 1 (1968).

[100] *See* AKHIL REED AMAR, THE CONSTITUTION AND CRIMINAL PROCEDURE 3-4 (1997).

[101] 68 AM. JUR. 2D *Searches and Seizures* § 166 (1993).

[102] *See* Richards v. Wisconsin, 520 U.S. 385 (1997).

[103] *See* United States v. Verdugo-Urquidez, 494 U.S. 259 (1990).

[104] *See* United States v. U.S. Dist. Court (*Keith*), 407 U.S. 297 (1972).

[105] *See* Jeffrey H. Smith & Elizabeth L. Howe, *Federal Legal Constraints on Electronic Surveillance*, *in* PROTECTING AMERICA'S FREEDOM IN THE INFORMATION AGE: A REPORT OF THE MARKLE FOUNDATION TASK FORCE 133 (2002). 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. *See* United States v. Bin Laden, 126 F. Supp. 2d 264, 271-72 (S.D.N.Y. 2000).

[106] *See, e.g.*, Coolidge v. New Hampshire, 403 U.S. 443, 465 (1971).

[107] *See* United States v. Edwards, 415 U.S. 800 (1974).

President involving foreign threats of "immediate and grave peril" to national security.[108]

Moreover, the Supreme Court interprets the Fourth Amendment to apply only to the collection of information, not the use of it. Even if information is obtained in violation of the Fourth Amendment, the Supreme Court has consistently found that the Fourth Amendment imposes no independent duty on the government to refrain from using it: "The Fourth Amendment contains no provision expressly precluding the use of evidence obtained in violation of its commands, and an examination of its origin and purposes makes clear that the use of fruits of a past unlawful search or seizure 'work[s] no new Fourth Amendment wrong.' "[109] Under the Court's "exclusionary rule," illegally seized data may still be used if the government agent acted in good faith,[110] to impeach a witness,[111] or in other settings in which the "officer committing the unconstitutional search or seizure" has "no responsibility or duty to, or agreement with, the sovereign seeking to use the evidence."[112] The Court suppresses the use of information obtained in violation of the Fourth Amendment only when doing so would have deterred the conduct of the government employee who acted unconstitutionally when collecting the information. So, for example, the Court has allowed records illegally seized by criminal investigators to be used by tax investigators on the basis that restricting the subsequent use would not deter the original unconstitutional conduct.[113] Protecting privacy is not a consideration. The Court wrote in 1974 that the exclusionary rule operates as "a judicially created remedy designed to safeguard Fourth Amendment rights generally through its deterrent effect, rather than a personal constitutional right of the party aggrieved."[114] If a court finds no independent Fourth Amendment basis for restricting the use of illegally obtained information, it goes without saying that the Court does not apply the Fourth Amendment to restrict the use of lawfully obtained information. Thus, the Fourth Amendment today sets no limit on the government's use of lawfully seized records, and in the case of unlawfully seized material, restricts its use only to the extent necessary to provide a deterrent for future illegal conduct.

---

[108] Smith & Howe, *supra* note 105, at 136 n.16; *see* 68 Am. Jur. 2d *Searches and Seizures* §§ 161, 347, 353 (1993).

[109] United States v. Leon, 468 U.S. 897, 906 (1984) (quoting United States v. Calandra, 414 U.S. 338, 354 (1974)).

[110] *See Leon*, 468 U.S. at 905-28.

[111] *See* Walder v. United States, 347 U.S. 62 (1954).

[112] United States v. Janis, 428 U.S. 433, 455 (1975).

[113] *Id.*

[114] *Calandra*, 414 U.S. at 354.

B. The Miller *Exclusion of Third-Party Records*

In 1976, the Supreme Court held in *United States v. Miller*[115] that there can be no reasonable expectation of privacy in information held by a third party. The case involved cancelled checks, to which, the Court noted, "respondent can assert neither ownership nor possession."[116] Such documents "contain only information voluntarily conveyed to the banks and exposed to their employees in the ordinary course of business."[117] 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.[118]

The Court's decision in *Miller* is remarkably sweeping. The bank did not just happen to be holding the records the government sought. Instead, the Bank Secrecy Act required (and continues to require) banks to maintain a copy of every customer check and deposit for six years or longer.[119] The government thus compelled the bank to store the information, and then sought the information from the bank on the basis that since the bank held the data, there could not be any reasonable expectation of privacy, and the Fourth Amendment therefore did not apply.[120] A majority of the Supreme Court was not troubled by this end run around the Fourth Amendment: "even if the banks could be said to have been acting solely as Government agents in transcribing the necessary information and complying without protest with the requirements of the subpoenas, there would be no intrusion upon the depositors' Fourth Amendment rights."[121]

Congress reacted to the decision by enacting modest statutory protection for customer financial records held by financial institutions,[122] 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 govern-

---

[115] 425 U.S. 435 (1976).
[116] *Id.* at 440.
[117] *Id.* at 442.
[118] *Id.* at 443 (citation omitted).
[119] 12 U.S.C. §§ 1829b(d), 1829b(g) (2000); *see Miller*, 425 U.S. at 436; Cal. Bankers Ass'n v. Shulz, 416 U.S. 21 (1974).
[120] *See Miller*, 425 U.S. at 443.
[121] *Id.* at 444.
[122] Right to Financial Privacy Act, 12 U.S.C. §§ 3401-3422 (2000); *see infra* text accompanying notes 146-153.

ment 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 *Smith v. Maryland*, involving information about (as opposed to the content of) telephone calls.[123] The Supreme Court held the Fourth Amendment 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.[124] "Telephone 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."[125]

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 the devices only collect information about the call that is necessarily disclosed to others.[126] As with information disclosed to financial institutions, Congress reacted to the Supreme Court's decision by creating modest statutory requirements applicable to pen registers,[127] but the Constitution does not apply.

C. The Miller *Exclusion of Third-Party Records Today*

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 place 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 and communicating over physical distances. Moreover, banks and telephone companies collect and store data not only because of business necessity, but also because the law requires them to. 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 individuals' general expectations of privacy, the assumption that such information would be private was objectively reasonable and widely shared. The Court's decisions to the contrary, while serving important law enforcement objectives, made little logical or practical sense and did not reflect the expectations of either

---

[123] Smith v. Maryland, 442 U.S. 735 (1979).
[124] *Id.* at 733-34, 745-46.
[125] *Id.* at 743.
[126] *Id.* at 742.
[127] 18 U.S.C. § 3121 (2000 & Supp. V 2005); *see also infra* text accompanying notes 154-156.

the public or policymakers, as demonstrated by the fact that Congress responded so quickly to both decisions 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, "[w]e are becoming a society of records, and these records are not held by us, but by third parties."[128] 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.[129] This is especially true as more activities move online, where merchants record data not only on what we buy and how we pay for our purchases, but also on every detail of what we look at, what we search for, how we navigate through web sites, and with whom we communicate.

These records are not only found in the Internet context. Computers track every moment of most employees' days. Digital time clocks and entry key cards record physical movements. Computers store work product, e-mail, and voice mail. Sensors monitor productivity—from check-out scanners at retail points-of-sale, which record how quickly cashiers process transactions, to key cards that monitor how long employees spend in the bathroom or break room each day. Digital devices for paying tolls, computer diagnostic equipment in car engines, and global positioning services—that are increasingly common on passenger vehicles—record how many miles we drive. Cellular telephones and personal digital assistants record not only call and appointment information, but location as well, and the devices transmit this information to service providers. Digital cable and satellite service providers record what we watch and when. Alarm systems record when we enter and leave our homes. ATMs and digital credit and debit card terminals record who and where we are, what we buy or how much money we withdraw, and where we bank.

Indications are that this is just the beginning. Broadband Internet access in homes has not only increased the personal activities in which we now engage online, but also created new and successful markets for remote computer back-up and online photo, e-mail, and music storage services. With

---

[128] Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1089 (2002).

[129] *See id.*; *see also* JEFFREY ROSEN, THE UNWANTED GAZE: THE DESTRUCTION OF PRIVACY IN AMERICA (2001); James X. Dempsey & Lara M. Flint, *Commercial Data and National Security*, 72 GEO. WASH. L. REV. 1459 (2004); Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137 (2002); Derek J. Somogy, *Information Brokers and Privacy*, 2 J.L. & POL'Y FOR INFO. SOC'Y 901 (2006).

Voice Over IP telephone service, digital phone calls are becoming indistinguishable from digital documents: both can be stored and accessed remotely.

Moreover, these technologies generate digital records that are available to many parties. For example, in a credit or debit card transaction, the data are collected by the retailer, the transaction processor, the card issuer, the cardholder's bank, and the merchant's bank.[130] Digital networks have also facilitated the growth of vigorous outsourcing markets, so information provided to one company is increasingly likely to be processed by a separate institution. Records containing personal data are linked and shared more widely and stored far longer than ever before, often without the individual consumer's knowledge or consent.

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. They provide a one-stop-shop for the government when it wants access to personal data, and most of the government's data mining initiatives depend on access to those data.[131]

New surveillance technologies are supplementing this already rich store of personal data and providing the government, primarily via the private sector, with ready access to increasingly revealing information about individuals:

- *Radio Frequency Identification* (RFID) tags are small computer chips used for tracking.[132] They are injected today into pets (and on occasion people) to facilitate identification and to provide medical or other information.[133] Tags are embedded in consumer goods to help prevent shoplifting and fraudulent returns. Electronic toll payment systems, such as EZ-Pass, I Pass, FastPass, and FasTrak, often rely on

---

[130] *See* National Federation of Independent Business, How a Basic Credit Card Transaction Works, http://www.nfib.com/object/2730732.html (June 6, 2003).

[131] *See generally Personal Information: Agencies and Resellers Vary in Providing Privacy Protections: Testimony Before the Subcomm. on Commercial and Admin. Law and the Subcomm. on the Constitution of the H. Comm. on the Judiciary*, 109th Cong. (2006) (statement of Linda D. Koontz, Director of Info. Mgmt. Issues, Gov't Accountability Office); Chris Jay Hoofnagle, *Big Brother's Little Helpers: How ChoicePoint and Other Commercial Data Brokers Collect and Package Your Data for Law Enforcement*, 29 N.C. J. INT'L L. & COM. REG. 595 (2004).

[132] RFID tags contain limited information, usually a unique identification number. A reader or scanner in the vicinity of a tag can read the information it contains. Passive tags can be almost microscopic and require close proximity to read. Active tags might be the size of a quarter, and can be read from several hundred feet. The data in the tag is often linked to a database, which provides additional information.

[133] In January 2008, the British government proposed inserting RFID tags under the skin of prisoners to make them easier to track. *See* Brian Brady, *Prisoners 'To Be Chipped Like Dogs'*, INDEPENDENT, Jan. 13, 2008, *available at* http://www.independent.co.uk/news/uk/politics/prisoners-to-be-chipped-like-dogs-769977.html.

RFID tags, and they are also found in U.S. passports, I-94 forms, and high-denomination Euro notes.

- *Global Positioning System* ("GPS") takes advantage of medium-orbit satellites to provide precise information about the location, speed, and direction of movement of a person or object, as well as the time. GPS is also routinely used in automobile navigational systems. Under federal law, all cell phones must now provide the cell phone service provider with precise information about the location of each cell phone.[134] This is designed to facilitate the dispatch of emergency services to a caller's location, but it also allows the government and other third parties to obtain location information on cell phone users.[135]
- Other *location sensors* are also used to determine an individual's location. For example, a laptop, PDA, or cell phone that connects to a Wireless Local Area Network necessarily provides information concerning the user's location. Similarly, cell phones that are not equipped with GPS can be located by "triangulating" the comparative strength with which the cell phone signal is received by three or more cell towers. It is difficult to imagine why government officials would ever resort to a "beeper" or physical surveillance when they can track the movement of a suspect through any number of other methods—such as GPS devices in her car or cell phone and RFID tags in her clothing, wallet, and car—accessible through the private sector.
- *Digital audio and video* have introduced significant new surveillance capabilities. Digital cameras offer ultra-high resolution images capable of identifying faces and license plate numbers from hundreds of feet away. They are increasingly wireless and are so small that they can be contained in a shirt button. Moreover, they are digital, which makes the data they collect easier and cheaper to store and share, and conducive to analysis with sophisticated voice, face, and threat recognition programs. Face recognition technologies that compare video images with databases of targeted individuals were used at the Super Bowl in Tampa, Florida, in 2001, and by numerous other authorities since then.[136]

---

[134] 47 U.S.C. § 222 (2000).

[135] In August 2007, New York City Public Schools terminated an employee because the location information generated by his employer-provided cell phone showed he was not at work when he claimed to be. *See* David Seifman, *'Track' Man Is Sacked—GPS Nails Ed. Guy*, N.Y. Post, Aug. 31, 2007, at 27. Trucking lines, rental car companies, and other businesses now routinely rely on GPS to locate their vehicles. *See* Anita Ramasastry, *Tracking Every Move You Make*, FindLaw, Aug. 23, 2005, http://writ.news.findlaw.com/ramasastry/20050823.html.

[136] *See U.S. Urged to Regulate Face-Scan Technology*, San Diego Union-Trib., Aug. 9, 2001, at A5. The San Francisco International Airport has deployed software to monitor images from its surveillance cameras and automatically classify objects and behaviors as "suspi-

- *Biometric identification* relies on behavioral and physiological characteristics of humans to verify identity and authorization to access protected facilities, funds, or data.[137] Modern security systems are increasingly relying on new biometric identification characteristics, for example, fingerprints. Many computers today come equipped with fingerprint scanners, and large organizations are increasingly moving to fingerprints to help verify identity (visitors to Disney World must now provide a fingerprint in an effort to prevent sharing of tickets).[138] Iris, retina, and voice recognition are also used in some settings today (the Clear Registered Traveler Program uses the distinct pattern of the individual's iris to verify identity).[139]
- *High resolution photography* has become an increasingly common way to collect personal data. Long a technique of national security agencies, high resolution satellite photography is used today by many businesses and available to individual users via internet services such as Google Earth, which provides high-resolution images of popular locations so that objects as small as six inches are recognizable, and Google Street View, which provides professional on-the ground images of major cities. Of course, not all surveillance technology has to be high-resolution. The cameras now universally included in cell phones form perhaps the largest sensor network in the world, especially as users increasingly post their pictures online and services such as Google Image Search make them easily accessible.

These are only a few of the most widely used surveillance technologies that add to the store of personal data that are available to the government via the private sector. These technologies are in addition to the "routine" data collection techniques that private- and public-sector institutions use every day as individuals work, play, shop, travel, invest, study, and communicate, and comply with the numerous government reporting requirements that attach to these activities. The *Miller* exclusion of information disclosed to third parties from the Fourth Amendment means that the government can access all of this information without constitutional limit, no matter how

---

cious," as part of a $30-million pilot program funded by the federal government. MARK SCHLOSBERG & NICOLE A. OZER, UNDER THE WATCHFUL EYE: THE PROLIFERATION OF VIDEO SURVEILLANCE SYSTEMS IN CALIFORNIA 4 (2007), *available at* http://www.aclunc.org/docs/criminal_justice/police_practices/Under_the_Watchful_Eye_The_Proliferation_of_Video_Surveillance_Systems_in_California.pdf. Law enforcement officials are also experimenting with a new technology that can "pick up aggressive tones on the basis of 12 factors including decibel level, pitch, and the speed at which words are spoken," via microphones from as far as 100 yards away. *Word on the Street . . . They're Listening*, SUNDAY TIMES, Nov. 26, 2006, at 1.

[137] Handwriting analysis (e.g., matching signatures) is a longstanding use of behavioral biometric identification; passport and driver's license photographs are common examples of physiological biometric identification.

[138] *See Talk of the Nation, High-Tech Spy Tools Aren't Just for James Bond* (NPR radio broadcast Aug. 8, 2007), *available at* http://www.npr.org/templates/story/story.php?storyId=12594656 (follow "Listen Now" hyperlink).

[139] *See* Rob Schneider, Fly by Those Lines: System Letting Registered Air Travelers Get Through Security Faster Takes Off, INDIANAPOLIS STAR, Jan. 19, 2007, at 1.

sensitive or how revealing of a person's health, finances, tastes, or convictions. The government's demand need not be reasonable; no warrant is necessary, and no judicial authorization or oversight is required.

D. Reversing Miller*?*

One response to the explosion in digital information that has transformed *Miller* into a broad exception to the Fourth Amendment would be for the Supreme Court to overturn the case. The Court could accomplish this simply by applying its current test for reasonableness to recognize that individuals do not, in fact, believe that information they provide in the course of ordinary activities is automatically available to the government and that their belief is "one that society was prepared to recognize as 'reasonable.'"[140] Alternatively, the Court could announce a new test for evaluating when warrants are required for the government to seize personal information held by third parties or could require that the government always obtain judicial authorization before accessing such records.

As important and desirable as such a judicial development would be from a civil-liberties perspective, it is highly unlikely. The Court has shown little willingness to extend the protection of the Fourth Amendment in any fashion, especially in response to new technologies. In only a handful of cases in the past twenty years has the Court responded positively to a Fourth Amendment challenge to the use of a new technology to capture information—and those cases involved intrusions into the home.[141] In fact, with the sole exception of physical searches inside the home, the Court has proven more likely to reduce, rather than preserve (much less expand), Fourth Amendment protections. The recent additions to the Court's membership seem unlikely to reverse this trend.

Nonetheless, even if the Supreme Court unexpectedly reversed or narrowed its third-party doctrine, that would still be inadequate to address the range of issues presented by the government's use of third-party records for data mining. For example, there would still be a need to address the Court's historical unwillingness to apply the Fourth Amendment or other constitutional provisions to restrict the use or sharing of personal information obtained by the government, even when it has been illegally seized.[142] Many government data-mining programs involve data that were collected either directly from the individual or from a third party for a regulatory or administrative purpose. Extending the Fourth Amendment to restrict the reuse of these data would require a fundamental shift in the Court's jurisprudence. For more than thirty years the Court has focused its Fourth Amendment ju-

---

[140] Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).

[141] *See, e.g.*, Kyllo v. United States, 533 U.S. 27 (2001) (involving the use of a thermal imager to sense activities within a home); United States v. Karo, 468 U.S. 705 (1984) (involving the use of a beeper that tracked the defendant's movement inside his home).

[142] *See, e.g.*, United States v. Janis, 428 U.S. 433, 447 (1975).

risprudence concerning illegally obtained material exclusively on deterring illegal collection of data, rather than preventing subsequent use or protecting personal privacy.[143]

In addition, the abandonment or weakening of the Court's third-party doctrine would require the creation of new rules or principles to guide lower courts in deciding what conditions would justify granting warrants for seizing third-party records, how to deal with requests for entire data sets rather than targeted data, and how to reconcile the potential need for the government to obtain judicial authorization to seize data with the fact that in many cases data are available for purchase from the third party or an information aggregator. These and other fundamental policy issues are better addressed by Congress. Moreover, the Fourth Amendment could provide, at best, only broad limits on government data mining. While those limits are important, and their absence denies individual privacy its most potent protection, government officials require clearer guidance concerning the appropriate conduct of data mining. Given how unlikely it is that the Court will abandon its third-party doctrine in the first place, Congress is the only meaningful place for citizens and government officials to turn for modern, coherent rules for how data mining is to be conducted and privacy protected.

IV. Statutory Responses

The Supreme Court's decision to exempt third-party records from the protection of the Fourth Amendment does not necessarily mean that those records are freely available to the government. Congress has adopted a number of statutes—two in response to the Supreme Court's third-party doctrine—in an effort to provide some protection for the privacy of personal information. Congress's role is potentially vital because of the breadth of its power and its ability to provide detailed, prospective guidance to the public and to government officials about the government's access to personal information.

Unfortunately, while Congress's privacy enactments may be numerous, they provide only modest protection, limited to specific economic sectors and subject to broad exceptions. The result is a remarkably complex set of laws, yielding very limited protection for privacy and little clear guidance to government agencies or private-sector entities. Recent "privacy" laws have further complicated the situation by actually weakening the limits on government access to personal data held by third parties.

Finally, despite the proliferation of government data mining programs, Congress has enacted no legislation to provide a legal framework for how such programs are to be undertaken, to provide redress for innocent people harmed by them, or to specify how privacy is to be protected in the process. This is not to say that Congress has been silent on the subject of data mining.

---

[143] *See* United States v. Calandra, 414 U.S. 338, 354 (1974).

Indeed, Congress has simultaneously been an enthusiastic proponent and an active critic. For example, Congress directed the DHS to "establish and utilize . . . data-mining and other advanced analytical tools . . . to access, receive, and analyze data" in order to "detect and identify threats of terrorism against the United States,"[144] but then it acted to terminate specific data mining initiatives when confronted with them.[145]

A. 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.[146] The statute provides substantially less protection than would have been required under the Fourth Amendment and is subject to a number of exceptions. The Act provides that federal agencies may not access the financial records of customers of financial institutions without the customer's consent, an administrative subpoena, a search warrant, a judicial subpoena, or a "formal written request."[147] This is less protection than would be required under the Fourth Amendment, because administrative and judicial subpoenas can be issued without any showing of probable cause and often without any showing of suspicion regarding a particular matter.[148] For example, the Act specifies that subpoenas and formal written requests may issue upon the mere showing that "there is a reason to believe that the records sought are relevant to a legitimate law enforcement inquiry."[149]

In addition, the statute is subject to a number of exceptions, including disclosures required under any other federal statute or rule.[150] The Act does not restrict a financial institution from notifying federal authorities that it possesses information they should seek.[151] And while the Act requires contemporaneous notice to the customer, it allows for that notice to be delayed in a variety of circumstances.[152] Most importantly, the Act does not apply when the federal government obtains financial information from third parties

---

[144] Homeland Security Act of 2002, Pub. L. No. 107-296, §§ 201(d)(1), (d)(14), 116 Stat. 2135, 2146-47 (codified at 6 U.S.C. § 121 (Supp. V 2005)).

[145] *See* S. Amend. 59 to H.R.J. Res. 2, 108th Cong. (Jan. 23, 2003); *see also supra* notes 94-95 and accompanying text.

[146] Right to Financial Privacy Act, 12 U.S.C. §§ 3401-3422 (2000 & Supp. V 2005).

[147] *Id.* § 3402.

[148] *See* William J. Stuntz, *O.J. Simpson, Bill Clinton, and the Transsubstantive Fourth Amendment*, 114 HARV. L. REV. 842, 857-59 (2000).

[149] 12 U.S.C. §§ 3405(1), 3407(1), 3408(3) (2000).

[150] *See id.* § 3413(d).

[151] *See id.* § 3403(c).

[152] *See id.* § 3409.

other than financial institutions nor does it restrict disclosures to state or local governments or to private entities.[153]

The Electronic Communications Privacy Act of 1986, enacted seven years after *Smith*, broadly regulates electronic surveillance.[154] Title III—the Pen Register Act—applies to "pen registers" and "trap and trace" devices.[155] To obtain information similar to what is contained in a phone bill or revealed by Caller ID, or to capture e-mail header information (the "To," "From," "Re," and "Date" lines in an e-mail), or the IP address of a site visited on the Internet, the government need only obtain a court order. Courts, however, are required to issue the orders—there is no room for judicial discretion—if the government certifies that "the information likely to be obtained by such installation and use is relevant to an ongoing criminal investigation."[156] As a result, Title III poses no meaningful barrier to the government's use of pen registers and trap and trace devices. Moreover, the Act provides for no exclusionary rule for violations of Title III, so law enforcement may freely violate these provisions and still use the data in subsequent criminal prosecutions.

Title II—the Stored Communications Act—also adopted in 1986, deals with communications in electronic storage, such as e-mail and voice mail.[157] Traditional warrants are required to obtain access to communications stored 180 days or less.[158] To obtain material stored for more than 180 days, the government need only provide an administrative subpoena, a grand jury subpoena, a trial subpoena, or a court order, all of which are easier to obtain than a traditional warrant.[159] Information about a customer's account maintained by a communications provider can be obtained by the government merely by providing "specific and articulable facts showing that there are reasonable grounds to believe that . . . the records or other information sought[ ] are relevant and material to an ongoing criminal investigation."[160] Violations carry a minimum fine of $1,000 but no exclusionary rule applies.[161]

The weakness of the protections afforded by Titles II and III of the Electronic Communications Privacy Act are illustrated by a comparison with the protection provided by Title I—the Wiretap Act—which was originally adopted in 1968 and deals with the interception of the contents of communi-

---

[153] *See id.* §§ 3401(1)-(3).

[154] *See* Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended in scattered sections of 18 U.S.C.).

[155] Pen Register Act, Pub. L. No. 99-508, § 301(a), 100 Stat. 1868-72 (codified as amended at 18 U.S.C. §§ 3121-3127 (2000)).

[156] 18 U.S.C. § 3123(a) (2000).

[157] Stored Communications Act, Pub. L. No. 99-508, § 201, 100 Stat. 1860-68 (1986) (codified as amended at 18 U.S.C. §§ 2701-2711).

[158] *See* Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1264, 1283-84 (2004).

[159] *See id.*

[160] 18 U.S.C. § 2703(d) (2000).

[161] Solove, *supra* note 158, at 1285.

cations in transmission.[162] It applies to "wire communications," although not to video unaccompanied by sound.[163] To intercept wire communications in transit requires a "'super' search warrant,"[164] which can only be sought by designated federal officials and requires probable cause, details about the communication to be intercepted, minimization of any non-relevant communications inadvertently intercepted, and termination immediately upon completion.[165] Information obtained in violation of these requirements can subject the responsible agent to minimum damages of $10,000 per violation and (except for e-mail) is subject to the exclusionary rule so that it cannot be used in a subsequent criminal prosecution.[166]

Despite their weaknesses, both the Financial Right to Privacy Act and Title III of the Electronic Communications Privacy Act do impose some limits on the government's power to seize financial and calling attribute information. More importantly, they impose some discipline on the government by specifying procedures to be followed. But they are a far cry from the protection against "unreasonable" searches and seizures that the Fourth Amendment would provide.

B. The Privacy Act

The broadest federal privacy law, and Congress's earliest effort to regulate how the government collects and uses personal information, is the Privacy Act of 1974.[167] 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 then-Department of Health, Education and Welfare (HEW). In 1973, the Advisory Committee issued its report, *Records, Computers and the Rights of Citizens.*[168] Congress responded the following year with the Privacy Act.

The Privacy Act requires federal agencies to: (1) store only relevant and necessary personal information and only for purposes required to be accomplished by statute or executive order; (2) collect information to the extent possible from the data subject; (3) maintain records that are accurate, complete, timely, and relevant; and (4) establish administrative, physical, and

---

[162] *See* Wiretap Act, Pub. L. No. 90-351, 82 Stat. 197 (1968) (codified as amended at 18 U.S.C. §§ 2510-2522).

[163] 18 U.S.C. § 2510(12) (2000).

[164] *See* Orin S. Kerr, *Internet Surveillance Law After the USA Patriot Act: The Big Brother That Isn't*, 97 Nw. U. L. Rev. 607, 621 (2003).

[165] *See* Solove, *supra* note 158, at 1282.

[166] *See id.*

[167] 5 U.S.C. § 552a (2000 & Supp. IV 2004).

[168] U.S. Department of Health, Education & Welfare, Records, Computers, and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems, Records, Computer, and the Rights of Citizens (1973).

technical safeguards to protect the security of records.[169] 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.[170] Agencies must log disclosures of records and, in some cases, inform the subjects of such disclosures when they occur.[171] 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.[172] Agencies must also publish a notice of the existence, character, and accessibility of their record systems.[173] Finally, individuals may seek legal redress if an agency violates the Act with regard to data concerning them.[174]

The Privacy Act is less protective of privacy than may first appear because of numerous broad exceptions.[175] 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."[176] 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.[177] The Privacy Act also exempts information subject to disclosure under the Freedom of Information Act.[178] And under the "routine use" exemption,[179] federal agencies may 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."[180] According to the Office of Management and Budget, "compatibility" covers uses that are either (1) functionally equivalent or (2) necessary and proper.[181]

---

[169] 5 U.S.C. § 552a.
[170] *Id.* § 552a(b).
[171] *Id.* §§ 552a(c), 552a(e)(8).
[172] *Id.* §§ 552a(d), 552a(f)(4), 552a(g).
[173] *Id.* § 552a(e)(4).
[174] *Id.* § 552a(g)(1).
[175] *See* Sean Fogarty & Daniel R. Ortiz, *Limitations Upon Interagency Information Sharing: The Privacy Act of 1974*, *in* PROTECTING AMERICA'S FREEDOM IN THE INFORMATION AGE, *supra* note 105, at 127, 128.
[176] 5 U.S.C. § 552a(b)(7).
[177] *Id.* § 552a(b).
[178] *Id.* § 552a(b)(2).
[179] *Id.* § 552a(b)(3).
[180] *Id.* § 552a(a)(7).
[181] 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. 12,990, 12,993 (Apr. 20, 1987) (publication of guidance in final form); *see generally* Fogarty & Ortiz, *supra* note 175, at 129-130.

Moreover, the Privacy Act applies only to information maintained in a "system of records."[182] 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."[183] 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.' "[184] This is unlikely to be the case with new 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 does little to provide guidance for government data mining activities or to limit 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 thirty years ago primarily for personnel records and benefits files, appears increasingly ill-suited for regulating twenty-first century data mining.

C. The Response to Data Mining

Congress has enacted one law specifically targeting early data mining. In 1988, Congress passed the Computer Matching and Privacy Protection Act as an amendment to the Privacy Act.[185] The new law responded to both the growth in early forms of data mining within the federal government and perceived inadequacies within existing privacy law to respond to data mining. In particular, the Act was an effort to fill the gap created by the 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.[186]

The Computer Matching and Privacy Protection Act provides a series of procedural requirements, such as written agreements between agencies that share data for matching,[187] before an agency can disclose personal information for data mining. These requirements deal only with federal agencies supplying—not obtaining—records for data mining.[188] Moreover, they only apply to data mining for the purpose of "establishing or verifying the eligi-

---

[182] 5 U.S.C. § 552a(b).

[183] *Id.* § 552a(a)(5).

[184] Henke v. U.S. Dep't of Commerce, 83 F.3d 1453, 1460 (D.C. Cir. 1996) (quoting Bartel v. FAA, 725 F.2d 1403, 1408 n.10 (D.C. Cir. 1984)).

[185] Pub. L. No. 100-503, 102 Stat. 2507 (1988) (codified at 5 U.S.C. §§ 552a(a)(8), 552a(o)-(r) (2000)).

[186] *See* OFFICE OF TECH. ASSESSMENT, ELECTRONIC RECORD SYSTEMS AND INDIVIDUAL PRIVACY 57 (1986).

[187] 5 U.S.C. § 552a(o).

[188] *See id.* § 552a(o)(1).

bility of, or 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;" "recouping payment or delinquent debts under such Federal benefit programs;" or "Federal personnel or payroll systems of records."[189] Law enforcement, counter-terrorism, and many other purposes for which the government engages in data mining do 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."[190]

D. Sectoral Privacy Laws

The 1988 law was effectively Congress's last word on data mining. 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, how it should be conducted, and/or how privacy is to be protected. Furthermore, even so-called "privacy" laws have actually weakened the protections against government seizure of personal data held by third parties. 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.[191] 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.[192]

The Fair Credit Reporting Act, enacted in 1970, permits disclosure of credit information only for statutorily specified purposes.[193] 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."[194] In addition, consumer reporting agencies may freely furnish identifying information (e.g., "name, address, former addresses, places of employment, or former places of employment") to the government.[195] After the September 11 terrorist attacks, Congress amended the Act to permit virtually unlimited disclosures to the government for counter-terrorism purposes. All that is required is a "written certification" that the re-

---

[189] *Id.* § 552a(a)(8)(A).
[190] *Id.* §§ 552a(a)(8)(B)(iii), (B)(v)(vi).
[191] 47 U.S.C. § 551 (2000 & Supp. I 2001).
[192] Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act (USA PATRIOT Act) of 2001, Title II, § 211, Pub. L. No. 107-56, 115 Stat. 272 (2001).
[193] *See* 15 U.S.C.A. § 1681b (West 2007).
[194] *Id.* § 1681b(a)(1).
[195] *Id.* § 1681f.

quested information is "necessary for the agency's conduct or such investigation, activity or analysis."[196]

In 2001, the Department of Health and Human Services adopted rules, specifically authorized by Congress, protecting the privacy of personal health information.[197] While facially restrictive, in reality, those rules permit broad disclosure of personal health information to the government in response to a warrant, court order, subpoena, discovery request, administrative request, investigative demand, or even a law enforcement official's "request."[198]

These sectoral statutes and rules apply in limited areas. Where they do apply, they impose few substantive limits, despite some procedural discipline, on government access to third-party data. And they offer no guidance whatsoever as to the proper role or limits of government data mining.

V. Data Mining Issues

Government data mining, especially of personal information obtained from third parties, presents many issues. The most important of those issues align roughly around two main themes. First, efficacy: does data mining work and work well enough to warrant the financial and human resources that it requires? Second, impact: will data mining, or the aggregation of private sector data in government hands, deter lawful behavior or otherwise harm individuals? These two broad categories of issues are interrelated. Questions about efficacy will always affect the assessment of the impact of data mining on individuals. After all, if data mining does not work, it does not justify any negative impact on individuals. Conversely, if its harmful impact is very low, even marginally successful data mining might be appropriate if used as an additional layer of protection against a particularly grave threat.

A. Efficacy

The first set of issues concerns the efficacy of government data mining: how well does it work to achieve its intended objectives? Mounting evidence suggests that data mining is not likely to be effective for many of the purposes for which the government seeks to use it, especially in the national security and law enforcement arenas. Not only have government officials failed to identify any successful efforts to detect or prevent terrorist activity based on analysis of databases, there are significant obstacles to such efforts succeeding. These include the impediments presented by data quality issues,

---

[196] *Id.* §§ 1681u, 1681v.

[197] *See* Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82,462 (2000) (codified at 45 C.F.R. pt. 160, §§ 164.502, 164.506).

[198] 45 C.F.R. § 164.512 (2005).

difficulties with data matching, and limits in data mining tools, especially when data mining in the national security setting is contrasted with data mining for commercial target marketing.

1. Data Quality

In its examination of data mining for national security, the Congressional Research Service ("CRS") noted that "[d]ata quality is a multifaceted issue that represents one of the biggest challenges for data mining."[199] The CRS went on to note that the "presence of duplicate records, the lack of data standards, the timeliness of updates, and human error can significantly impact the effectiveness of the more complex data mining techniques, which are sensitive to subtle differences that may exist in data."[200] In 1997 and again in 2002, the Inspector General of the Department of Justice ("DOJ") found that data from the Immigration and Naturalization Service (the predecessor of U.S. Citizenship and Immigration Services) was "seriously flawed in content and accuracy."[201] A December 2006 report by the SSA's Inspector General found that 4.1% of the records it surveyed (or an estimated 17.8 million total records) in the SSA's NUMIDENT database—the backbone identification verification tool for social service and other federal programs—contained "discrepancies in the name, date of birth or citizenship status of the numberholder" or concerned deceased individuals.[202]

The fact that government data mining almost always involves "repurposing" data—i.e., using data for a purpose different from that for which they were originally collected and stored—further exacerbates concerns about the accuracy of the underlying data. For example, for its CAPPS II program, TSA proposed accessing credit report information and other private-sector data to help determine what level risk a potential passenger posed.[203] Current aviation and border security data mining initiatives include

---

[199] SEIFERT, *supra* note 10, at CRS-21.

[200] *Id.*

[201] *See* OFFICE OF INSPECTOR GEN., U.S. DEP'T OF JUSTICE, IMMIGRATION AND NATURALIZATION SERVICE'S ABILITY TO PROVIDE TIMELY AND ACCURATE ALIEN INFORMATION TO THE SOCIAL SECURITY ADMINISTRATION (NO. I-2003-001) at 25 (2002), *available at* http://www.usdoj.gov/oig/reports/INS/e0301/final.pdf; OFFICE OF INSPECTOR GEN., U.S. DEPARTMENT OF JUSTICE, FOLLOW-UP REPORT ON INS EFFORTS TO IMPROVE THE CONTROL OF NONIMMIGRANT OVERSTAYS (NO. I-2002-006) (2002), *available at* http://www.usdoj.gov/oig/reports/INS/e0206/index.htm; OFFICE OF INSPECTOR GEN., U.S. DEP'T OF JUSTICE, IMMIGRATION AND NATURALIZATION SERVICE MONITORING OF NONIMMIGRANT OVERSTAYS (NO. I-97-08) (1997), *available at* http://www.usdoj.gov/oig/reports/INS/e9708/index.htm. *See generally* Electronic Privacy Information Center, Spotlight on Surveillance, E-Verify System: DHS Changes Name, But Problems Remain for U.S. Workers (July 2007), http://epic.org/privacy/surveillance/spotlight/0707/.

[202] OFFICE OF INSPECTOR GEN., SOC. SEC. ADMIN., CONGRESSIONAL RESPONSE REPORT: ACCURACY OF THE SOCIAL SECURITY ADMINISTRATION'S NUMIDENT FILE (A-08-06-26100), at ii (2006), *available at* http://www.ssa.gov/oig/ADOBEPDF/A-08-06-26100.pdf.

[203] SIEFERT, *supra* note 11, at CRS-9.

data from passenger records such as frequent traveler numbers.[204] One might reasonably ask whether these data were collected and stored with the degree of attention to accuracy appropriate for making security-related determinations.

Questions about the provenance of the data are especially acute in the national security context because the stakes of errors are so high for individuals and society. Many records contain errors, especially records maintained for uses where accuracy is not a paramount concern or the subject of significant resources. As noted in *Computerworld* magazine in 2003, "[a] single piece of dirty data might seem like a trivial problem, but if you multiply that 'trivial' problem by thousands or millions of pieces of erroneous, duplicated or inconsistent data, it becomes a prescription for chaos."[205] The problem of inaccurate data is multiplied, not diminished, when records in databases of varying accuracy are combined. The accuracy of records raises important practical concerns about the value of national security analyses performed on potentially bad data as well.

2. Data Matching

Errors in linking data are a major contributor to inaccuracies in data mining. Many factors contribute to the difficulty of integrating data accurately:

- Names may be recorded in a variety of different ways in different records (e.g., J. Smith, J.Q. Smith, John Q. Smith).
- Individuals, especially women, change their names. There are approximately 2.3 million marriages and 1.1 million divorces every year in the United States, often resulting in changed last names (and also changed addresses).[206]
- Many people have the same name.
- Many individuals have more than one address (e.g., home, office, vacation home, post office box), and are likely to change addresses. As of 1998 there were 6 million vacation or second homes in the United States, many of which were used as temporary or second addresses.[207] And, according to the U.S. Postal Service, about 43 million Americans—approximately seventeen percent of the U.S. population—change addresses every year.[208]

---

[204] *See* 2007 PNR Agreement, *supra* note 67, at 22-23.

[205] Tommy Peterson, *Data Scrubbing*, COMPUTERWORLD, Feb. 10, 2003, at 32.

[206] *See* NAT'L CTR. FOR HEALTH STAT., NATIONAL VITAL STATISTICS REPORTS, VOL. 51, NO. 8, at 1 tbl.A (2003).

[207] *Identity Theft: Hearing on H.R. 4311 Before the H. Comm. on Banking and Fin. Servs.*, 106th Cong., (2000) (statement of Stuart K. Pratt, V.P., Associated Credit Bureaus).

[208] UNITED STATES POSTAL SERVICE DEPARTMENT OF PUBLIC AFFAIRS AND COMMUNICATIONS, LATEST FACTS UPDATE (Jun. 24, 2002), *available at* http://usps.com/news/facts/lfu_062402.htm.

- The systems in which different data are stored may be incompatible and the process of overcoming interoperability issues may introduce additional errors.[209]

Inclusion of Social Security Numbers ("SSNs") improves the likelihood of a correct match to the accountholder, but even when accounts include SSNs, identification may be difficult because accounts for the same household may reflect different primary SSNs (e.g., husband, wife, minor beneficiary) and because of the presence of transcription errors in recording strings of numbers. Moreover, data about potential terrorists are unlikely to include SSNs.

A 2002 study by the Consumer Federation of America and the National Credit Reporting Association concluded that "almost one in ten consumers runs the risk of being excluded from the credit marketplace altogether because of incomplete records, duplicate files, and mixed files,"[210] despite the fact that credit report files are among the most heavily regulated business databases. Their report goes on to note that "[u]se of nicknames, misspellings, transposed social security numbers, and mixed files that report information under one person's name, but match that name to a spouse's social security number, are all examples of variations that can result from an automated interpretation of complex and sometimes contradictory personal identifying data."[211]

The problem is by no means limited to businesses or not-for-profit organizations. As discussed in greater detail below, the government faces a similar challenge in accurately matching data and people, especially in its anti-terrorism and law enforcement efforts. Post-September 11 programs for enhanced border, critical infrastructure, and passenger facility security all depend on being able to identify individuals and assess the risk they present by quickly connecting to accurate information about them. This is a substantial challenge, as stressed in the 2004 final report of Technology and Privacy Advisory Committee ("TAPAC"), the "blue ribbon"[212] bipartisan independent committee appointed by then-Secretary of Defense Donald Rumsfeld in 2003 to examine privacy and security issues following the controversy over TIA:

---

[209] *See* SIEFERT, *supra* note 10, at CRS-22.

[210] CONSUMER FED'N OF AM. & NAT'L CREDIT REPORTING ASS'N, CREDIT SCORE ACCURACY AND IMPLICATIONS FOR CONSUMERS 39 (2002).

[211] *Id.* at 35.

[212] Ronald D. Lee & Paul M. Schwartz, *Beyond the "War" on Terrorism: Towards the New Intelligence Network*, 103 MICH. L. REV. 1446, 1467 (2005). The TAPAC comprised eight senior statespeople with expansive government and corporate experience. *See* Fred H. Cate, *Terrorism, Technology, and Information Privacy: Finding the Balance*, BILL OF PARTICULARS, Fall 2004 at 5-6, *available at* http://www.law.indiana.edu/publications/particulars/2004fall.pdf ("The eight members (of TAPAC) read like a who's who of government, law, industry, and higher education . . . They represent all three branches of government, including one federal appellate court judge, one member of Congress, two cabinet secretaries, an attorney general, three White House lawyers . . . and one chair of the FCC.").

> Integrating and analyzing a large volume of data such as credit card transactions or airline ticket bookings raise many practical issues, even before considering the potential privacy threat. One of the most significant of these issues concerns the significant difficulties of integrating data accurately. Business and government have long struggled with how to ensure that information about one person is correctly attributed to that individual and only to that individual.[213]

Overcoming the many obstacles to linking data accurately is a major challenge for all organizations. International Data Corporation, a worldwide market research, analysis, and consulting firm, estimates that the process of accurately and rapidly integrating new data is the most critical part of managing and using customer databases, consuming up to seventy percent of an organization's total information technology resources.[214] Even in well-designed data-based studies such as those developed by the Census Bureau, automated matching is only seventy-five percent accurate and hand-matching of records is required to reduce the error rates substantially.[215]

The task of integrating data accurately is especially difficult in the counter-terrorism arena, which often involves matching data from disparate systems over which the intelligence community has no control, from intercepts and other sources where little or no identifying information is provided, and in ways that prevent seeking or verifying additional identifying information. The fact that many government data mining applications involve searches across incompatible datasets and unstructured data (e.g., audio and video surveillance records) exacerbates the aforementioned concerns.

In addition, even when data are accurately aggregated, the file or data mining result must then be linked to the right person. A number of reasons make this significantly challenging in the national security arena. The problems associated with misidentifying people, including well-known figures such as Senator Edward Kennedy, on the current "do not fly" lists are well documented.[216] These problems are further exacerbated by the poor quality of most identity documents and the ease with which fraudulent documents may be obtained. Some of the September 11 hijackers had false identification documents, either forgeries or legitimate driver's licenses issued by

---

[213] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 37.

[214] Emily Kay, *Coordinating Supply Chain Data: To Deliver Timely Information, Companies Must Overcome Data Synchronization Hurdles*, FRONTLINE SOLUTIONS, May 1, 2003, at 21.

[215] MARGO ANDERSON & STEPHEN E. FEINBERG, WHO COUNTS? THE POLITICS OF CENSUS-TAKING IN CONTEMPORARY AMERICA 117-18 (Russell Sage Found. 1999).

[216] *See* Editorial, *Glitches Repeatedly Delay Innocent Air Travelers*, USA TODAY, June 25, 2003, at 11A.

states to the wrong person.[217] Moreover, photographs on driver's licenses and passports, which are issued for terms of between four and ten years, often provide poor verification of identity. Better forms of identification, such as biometric identifiers (e.g., fingerprints or retinal scans), are not widely used today and raise significant questions about their cost, reliability, and impact on privacy.

The critical issues surrounding data mining efficacy, therefore, include concerns about the provenance of the data and ensuring that they were matched accurately before coming into the government's control, during the process of data mining, and when the results are linked to specific individuals.

3. Data Mining Tools in Context

The third set of issues affecting the effectiveness of data mining is the quality of the analytical tools—the search algorithms and target patterns—being used and how useful they are in the contexts in which they are increasingly being deployed. We have limited data about the experience of government, especially in the national security setting, because so much of the data mining is both new and classified. The experience of industry, however, which is generally acknowledged to be ahead of the government in developing and deploying data mining technologies, is not encouraging on the success of data mining. For example, even sophisticated target marketing, which relies heavily on data mining, recorded an average response rate of 2.24 percent for catalog promotions and 2.15 percent for direct mail in the year 2007.[218] Those figures suggest a high false positive rate—the proportion of people or activities wrongly identified by data mining.

Data mining for national security and law enforcement presents far greater challenges than data mining for target marketing for many reasons. For example, government data mining is often searching for a far smaller population of targets than is the case in the private sector. Without knowing the precise number of potential terrorists in the United States, the figure is certain to be far smaller than the population of potential customers most marketers wish to target. Moreover, terrorists and other criminals are working hard to blend in. Government data mining often is searching for a needle not in a haystack, but among millions of other needles.

Further, the government has a much harder time knowing the patterns it is looking for. Most marketers have thousands or even millions of customers upon whose actual behavior they can base patterns for data mining. This was a key point in the 2007 CRS report on *Data Mining and Homeland Security*:

---

[217] *See* JIM HARPER, IDENTITY CRISIS 207 n.10 (2006); NAT'L COMM'N ON TERRORIST ATTACKS ON THE U.S., 9/11 COMMISSION REPORT 390 (2004).

[218] Press Release, Direct Marketing Association, DMA Releases 5th Annual "Response Rate Trends Report" (Oct. 13, 2007), *available at* http://www.the-dma.org/cgi/disppressrelease?article=1008 (last visited Mar. 13, 2008).

"Successful 'predictive data mining' requires a significant number of known instances of a particular behavior in order to develop valid predictive models. For example, data mining used to predict types of consumer behavior . . . may be based on as many as millions of previous instances of the same particular behavior."[219] Government agencies, fortunately, have limited experience with terrorists on U.S. soil. Moreover, domestic terrorist attacks actually rarely follow a pattern: each one is new and different. As a result, intelligence officials can imagine attack strategies and they can learn from past terrorists activities, but they have comparatively few opportunities to test the accuracy of their analysis and little reason to think that analyses based on past attacks will be useful in anticipating future ones. "With a relatively small number of attempts every year and only one or two major terrorist incidents every few years—each one distinct in terms of planning and execution—there are no meaningful patterns that show what behavior indicates planning or preparation for terrorism."[220]

One corollary to limited frequency and individuality of terrorist acts within the United States is that national security data mining efforts, like other aspects of homeland security, tend to be backwards focused. Consider the U.S. approach to aviation security: after the 9/11 attacks, in which terrorists used box cutters to take over passenger airplanes, the government banned not only box cutters but anything that resembled them—nail clippers, nail files, pocket knives.[221] Only after Richard Reid attempted to blow up an airplane by detonating explosives hidden in his shoes, did TSA officials begin screening shoes.[222] It was only after British officials discovered a plot to blow up airplanes with liquid explosives—a threat anti-terrorism officials had known about for more than a decade—that the TSA began restricting liquids allowed to be carried on planes.[223] In each case, the government's action was wholly reactive to the most recently demonstrated threat, rather than proactive in responding to known threats whether or not they had been attempted. Government data mining seems similarly likely to be fighting yesterday's battles—a problem that commercial data miners face to a far less extent, since the characteristics of desirable consumers are likely to change far less rapidly than those of terrorists.

Another challenge faced by national security data mining is the desire of terrorists not to be found. Commercial data mining is generally searching for potential customers who either want to be discovered or do not care if they are found. Government data mining, by contrast, is often looking for terrorists or criminals who do not want to be located and therefore may be

---

[219] SEIFERT, *supra* note 10, at CRS-3.

[220] JEFF JONAS & JIM HARPER, CATO INSTITUTE, EFFECTIVE COUNTERTERRORISM AND THE LIMITED ROLE OF PREDICTIVE DATA MINING 7-8 (2006).

[221] Steven Greenhouse, *The New Property*, N.Y. TIMES, Oct. 22, 2001, at A16.

[222] *See* Hector Becerra, Jennifer Oldham & Mitchell Landsberg, *Airline Terrorism Alert; Winging It Once Again*, L.A. TIMES, Aug. 11, 2006, at A1.

[223] *See id.*

assumed to be trying to hide their identities and behaviors from government sight.

"False positives" are a much bigger concern when searching for terrorists than for customers. According to Paul Rosenzweig, Deputy Assistant Secretary for Policy at DHS, "[t]he only certainty [in data mining] is that there will be false positives."[224] In the commercial setting, false positives do not matter much because people erroneously targeted can simply discard the mail or e-mail solicitation, and the marginal costs associated with those wasted communications are comparatively small. False negatives—failing to target appropriate individuals—while thought to be high are also not particularly problematic because there are other means of communicating with those people, and they can always seek out the solicitation if they desire it (e.g., by visiting a store, calling an 800- number, etc.).

The situation with government data mining is wholly different. Even if falsity rates are very low, the consequences in the national security settings are difficult to exaggerate. For example, if a data mining system intended to keep potential terrorists off of airplanes yielded a false positive rate of only one percent—a far better rate than that achieved by publicly disclosed government or commercial data mining—that would still mean that 7.4 million travelers (one percent of the 739 million passengers that the TSA screened in 2005)[225] would have been wrongly identified as terrorist suspects. These are not speculative issues. The TSA operated its data-based passenger screening programs for more than two years with no system in place to report or correct errors, despite the fact that innocent passengers were routinely denied or delayed in boarding aircraft.[226] And DHS continued to use and expand its automated employment verification system even though as many as forty-two percent of employees who received "final nonconfirmation" notices were in fact eligible to work.[227] False positives which result in innocent people being detained, denied boarding on airplanes, denied employment, or subject to additional investigation not only inconvenience individuals and threaten constitutionally protected rights, they also consume significant resources and may undermine security by diverting attention from real threats. The consequences of false negatives may be even greater, by failing to detect potential terrorists or criminals or to prevent their nefarious activities. Moreover, valid targets overlooked by government data mining are unlikely to seek means to self-identify.

---

[224] PAUL ROSENZWEIG, HERITAGE LEGAL FOUNDATION, PROPOSALS FOR IMPLEMENTING THE TERRORISM INFORMATION AWARENESS SYSTEM (2003).

[225] U.S. CENSUS BUREAU, STATISTICAL ABSTRACT OF THE UNITED STATES 672 (2007).

[226] *See* Del Quentin Wilber, *Fliers' Data Left Exposed, Report Says*, WASH. POST, Jan. 12, 2008, at D1.

[227] INSTITUTE FOR SURVEY RESEARCH, TEMPLE UNIVERSITY & WESTAT, INS BASIC PILOT EVALUATION SUMMARY REPORT 25 (2002), *available at* http://www.uscis.gov/files/nativedocuments/INSBASICpilot_summ_jan292002.pdf (last visited Mar. 13, 2008).

In light of these significant issues, many experts argue that using data mining to detect and prevent terrorist attacks is far more difficult than using it for commercial application. One of the bluntest assessments comes from Jeff Jonas, chief scientist of IBM's Entity Analytic Solutions Group, and Jim Harper, director of information policy studies at the Cato Institute: "Data mining is not an effective way to discover incipient terrorism. Though data mining has many valuable uses, it is not well suited to the terrorist discovery problem."[228] This reveals the need for the government to examine carefully its current and planned data mining programs to determine whether they in fact work, and if so, work well enough to justify their costs—financial and otherwise.

4. Assessing Efficacy

There is nearly universal agreement about the need to assess the efficacy of data mining systems. Many of the committees created to examine various aspects of government data mining and information use have proposed ways of doing so. One of the earliest proposals came from TAPAC, which recommended to the Secretary of Defense that any DOD data mining program should require a "written finding" by the "agency head" specifying, among other things:

i. the purposes for which the system may be used;
ii. the need for the data to accomplish that purpose;
iii. the specific uses to which the data will be put;
iv. that the data are appropriate for that use, taking into account the purpose(s) for which the data were collected, their age, and the conditions under which they have been stored and protected;
v. that other equally effective but less intrusive means of achieving the same purpose are either not practically available or are already being used;
vi. the effect(s) on individuals identified through the data mining . . .
vii. that the system has been demonstrated to his or her satisfaction to be effective and appropriate for that purpose; . . .
ix. that the system yields a rate of false positives that is acceptable in view of the purpose of the search, the severity of the effect of being identified, and the likelihood of further investigation.[229]

These recommendations are designed to ensure that data mining programs are not deployed without being shown to be effective for specific purposes, and to rely on data that are appropriate to those purposes. Moreover, they are intended to ensure that those purposes are served without unnecessarily burdening individuals or intruding into protected rights. The recommendations require an explicit balancing between the goals to be

[228] JONAS & HARPER, *supra* note 220, at 2.
[229] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 49-50.

achieved and the likelihood of achieving them on the one hand, and the impact on individuals on the other hand. Finally, by requiring written authorization by a senior government official, the TAPAC recommendations help to ensure that decision to engage in data mining are taken seriously, that the required determinations are undertaken explicitly, and that an individual is identified to be held accountable if they are not. As John O. Marsh, Jr., a TAPAC member, former member of Congress, and the longest-serving Secretary of the Army, testified before the House Judiciary Committee, "[W]e believed that accountability was absolutely critical to . . . ensuring that data mining was conducted efficiently and effectively, . . . [and that it] would be enhanced, we believed, first by ensuring that no agency engage in data mining involving personal information without making a conscious, thoughtful decision to do so."[230]

Despite the burden that the process of assessing efficacy clearly will present, it is essential. The argument that the perceived danger is too great to allow time for meaningful assessment is exactly backwards. The perceived severity of the terrorist threat only enhances the importance of ensuring that we invest our efforts in measures calculated to work. Investing in ineffective tools can seriously undermine security, divert scarce resources, and compromise public confidence, as well as endanger privacy. Assessment is critical at all times to ensure that the government is doing not merely "something," but the best thing in light of the available resources.

B. Impact

1. Data Mining and Privacy

Data mining that involves personal data necessarily affects personal privacy. This is the consistent conclusion of every inquiry into government data mining, from the 1973 report of the HEW Advisory Committee on Automated Personal Data Systems[231] to the 2004 TAPAC report.[232] "[D]ata mining concerning U.S. persons inevitably raises privacy issues."[233]

Perhaps the greatest impact of data mining on individual privacy is that individuals will change their behavior as a result of their awareness that the government may, without probable cause or other specific authorization, obtain access to myriad distributed stores of information about them. The original motto of the TIA program—*Scientia Est Potentia*—is certainly correct:

---

[230] *Privacy and Civil Liberties in the Hands of Government Post-September 22, 2001: Recommendations of the 9/11 Commission and the U.S. Department of Defense Technology and Privacy Advisory Committee: Hearing Before the Subcomm. on Commercial and Administrative Law and Subcomm. on the Constitution of the H. Comm. on the Judiciary*, 108th Cong. 5 (2004) (statement of John O. Marsh, Jr., TAPAC).

[231] *See* U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE, *supra* note 168.

[232] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9.

[233] *Id.* at 48.

"knowledge is power." Knowledge that the government is, or even may be, observing data we generate through thousands of ordinary activities can alter the way people live their lives and interact with others.

It is this principle that was at the heart of Jeremy Bentham's concept of the Panopticon—a model prison consisting of a central tower surrounded by a ring of prison cells.[234] One-way windows would allow a person in the tower to see into the prison cells but would prevent the prisoners from seeing into the tower. Bentham posited that a single inspector in the tower could control the behavior of all of the prisoners through making each prisoner "*conceive* himself to be . . . constantly . . . inspected."[235] Applying the analysis of philosopher and historian Michel Foucault, Professor Slobogin argues that, "modern society increasingly functions like a super Panopticon," in which government constrains individual behavior by the threat of surveillance.[236]

This may not always be a bad outcome, but knowledge of the government's surveillance 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. More than forty years ago, Vice President Hubert Humphrey observed, "[w]e 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."[237] Therefore, government data mining in a democracy threatens not merely information privacy but other civil liberties, including freedom of expression, association, and religion. In the words of professor and former Deputy Attorney General Philip Heymann, "[n]o matter how honest the government was in restricting its uses of the data, many citizens would become more cautious in their activities, including being less outspoken in their dissent to government policies. For two hundred years Americans have proudly distrusted their government."[238]

The impact on individual behavior is far more direct when individuals are identified through data mining for additional scrutiny, denied boarding, or detained. When that identification is in error, as the prior discussion suggests it frequently is, the injury becomes one that the government must take

---

[234] *See* 1 JEREMY BENTHAM, PANOPTICON OR, THE INSPECTION-HOUSE 5-8 (T. Payne 1791) (1787).

[235] *Id.* at 3; *see also* Jeffrey H. Reiman, *Driving to the Panopticon: A Philosophical Exploration of the Risks to Privacy Posed by the Highway Technology of the Future*, 11 SANTA CLARA COMPUTER & HIGH TECH. L.J. 27 (1995).

[236] Christopher Slobogin, *Public Privacy: Camera Surveillance of Public Places and the Right to Anonymity*, 72 MISS. L.J. 213, 241 (2002) (citing Michel Foucault, Discipline & Punish 187 (Alan Sheridan trans., Vintage Books 2d ed. 1995) (1977).

[237] Hubert H. Humphrey, *Foreword* to EDWARD V. LONG, THE INTRUDERS, at viii (1967).

[238] Philip B. Heymann, Investigative Uses of Files of Data About Many People Collected for Other Purposes 9 (2003) (unpublished manuscript).

into account when deploying the data mining system and must be prepared to address.

2. Assessing Impact

To minimize the harmful impact of government data mining on individuals and assess the magnitude of that impact in light of the value that the nation has historically placed on privacy and other civil liberties, every group to consider the issue has recommended some form of legal process in addition to the little required under current law.

For example, TAPAC recommended five new legal requirements to address the impact of government data mining on individuals and minimize it to the extent possible. The first would condition most data mining on a "[w]ritten finding by agency head authorizing data mining."[239] The finding would have to address, in addition to the points already discussed, "that other equally effective but less intrusive means of achieving the same purpose are either not practically available or are already being used;" "the effect(s) on individuals identified through the data mining (e.g., they will be the subject of further investigation for which a warrant will be sought, they will be subject to additional scrutiny before being allowed to board an aircraft, etc.);" and "that there is a system in place for dealing with false positives (e.g., reporting false positives to developers to improve the system, correcting incorrect information if possible, remedying the effects of false positives as quickly as practicable, etc.), including identifying the frequency and effects of false positives."[240]

TAPAC's second recommendation would require "[d]ata mining of databases known or reasonably likely to include personally identifiable information about U.S. persons"[241] to employ a series of "technical require-

[239] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 49. TAPAC recommended applying its new legal framework to "all DOD programs involving data mining concerning U.S. persons" except for "data mining (1) based on particularized suspicion (including searches of passenger manifests and similar lists); (2) that is limited to foreign intelligence that does not involve U.S. persons; or (3) that concerns federal government employees in connection with their employment." *Id.* The committee noted that "these three areas are already subject to extensive regulation, which we do not propose expanding." *Id.* The committee 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 written authorization and compliance audit requirements. *Id.*

[240] *Id.* at 50.

[241] *Id.*

ments,"[242] including "[d]ata minimization,"[243] "[d]ata anonymization,"[244] "[a]udit trail,"[245] "[s]ecurity and access,"[246] and "[t]raining."[247]

The fourth[248] TAPAC recommendation for protecting personal privacy in government data mining would require judicial authorization from the Foreign Intelligence Surveillance Court (FISC) for searches—or entire data mining programs—that would involve the use of "personally identifiable information" about U.S. persons.[249] That authorization would depend on "specific and articulable facts" that:

i. The search will be conducted in a manner that otherwise complies with the requirements of these recommendations however enacted;
ii. The use of personally identifiable information is reasonably related to identifying or apprehending terrorists, preventing terrorist attacks, or locating or preventing the use of weapons of mass destruction;
iii. The search is likely to yield information reasonably related to identifying or apprehending terrorists, preventing terrorist attacks, or locating or preventing the use of weapons of mass destruction; and
iv. The search is not practicable with anonymized data in light of all of the circumstances . . . .[250]

---

[242] *Id.*

[243] *Id.* ("[T]he least data consistent with the purpose of the data mining should be accessed, disseminated, and retained.").

[244] *Id.* ("[W]henever practicable data mining should be performed on databases from which information by which specific individuals can be commonly identified (e.g., name, address, telephone number, SSN, unique title, etc.) has been removed, encrypted, or otherwise obscured.").

[245] *Id.* ("[D]ata mining systems should be designed to create a permanent, tamper-resistant record of when data have been accessed and by whom.").

[246] *Id.* ("[D]ata mining systems should be secured against accidental or deliberate unauthorized access, use, alteration, or destruction, and access to such systems should be restricted to persons with a legitimate need and protected by appropriate access controls taking into account the sensitivity of the data.").

[247] *Id.* ("[A]ll persons engaged in developing or using data mining systems should be trained in their appropriate use and the laws and regulations applicable to their use.").

[248] The third TAPAC recommendation is not directly relevant to this discussion and is omitted here. *See id.* at 52.

[249] *Id.* at 51.

[250] *Id.* FISC authorization meeting similar requirements would be required to reidentify search results conducted with anonymized or pseudonymized personal data. *Id.* at 52. The recommendations also include a provision dealing with "exigent circumstances," which would allow the government to engage in data mining without FISC authorization. According to that provision:

> Without obtaining a court order, the government may, in exigent circumstances, search personally identifiable information or reidentify anonymized information obtained through data mining if:
>
> i. The agency head or his or her single designee certifies that it is impracticable to obtain a written order in light of all of the circumstances (e.g., the type of data, type of search, the need for the personally identifiable information, and other issues affecting the timing of the search), and provides a copy of that certification to the privacy officer;

Fifth, TAPAC recommended that data mining "known or reasonably likely to include personally identifiable information about U.S. persons should be audited not less than annually to ensure compliance" with these requirements.[251]

Finally, TAPAC recommended administrative and reporting changes to enhance accountability and transparency concerning data mining. These included training, which was included under the "Technical Requirements" already discussed, the appointment of a "policy-level privacy officer,"[252] the creation of "a panel of external advisors,"[253] and the establishment of other "meaningful oversight mechanisms,"[254] including an annual report to Congress and, "[t]o the extent consistent with national security and applicable classification laws and regulations," the public.[255]

These recommendations thus create a significant incentive for using anonymized or pseudonymized data whenever possible and providing for systemic privacy protections and judicial oversight when not possible. Despite their far-reaching scope, they were accepted by the Department of Defense ("DOD") in August 2006.[256]

The TAPAC recommendations reflect the deep-seated view that privacy is a value that matters in its own right, and it is inevitably affected by government data mining.[257] As a result, TAPAC recommended the adoption of laws that would subject government data mining with personally identifiable information to external authorization from a court and external oversight from the judiciary and Congress.[258] TAPAC thus sought to fill the gap created by judicial and legislative inaction and to ensure that government data mining would be subject to the checks and balances of constitutionally divided government.

In the quest to protect individuals from the undesirable impact of government data mining programs, most of the TAPAC recommendations are primarily procedural. They do not purport to determine whether the government should engage in data mining or even how it should be conducted.

---

ii. DOD subsequently applies to the court for a written order within 48 hours or, in the event of a catastrophic attack against the United States, as soon as practicable; and

iii. The agency terminates any on-going searches of personally identifiable information or use of reidentified information obtained through data mining if the court does not grant the order. *Id.*

[251] *Id.* at 50.

[252] *Id.*

[253] *Id.* at 53.

[254] *Id.*

[255] *Id.* at 55.

[256] *See* Letter from William J. Haynes II, Gen. Counsel, Dep't of Def., to Carol E. Dinkins, Chair, Privacy and Civil Liberties Oversight Bd. (Sep. 22, 2006) (on file with the Harvard Civil Rights-Civil Liberties Law Review) (attaching a list of TAPAC's recommendations with each of those applicable to the DOD initialed by the Deputy Secretary as "approved").

[257] *See supra* text accompanying notes 232-250.

[258] *See supra* text accompanying notes 239-255, 274-281.

Instead, most of the recommendations set forth how those decisions should be made and who should make them. The committees' recommendations appear to be designed to facilitate discipline and rationality by those who develop and deploy data mining programs; meaningful oversight by policymakers, legislators, and judges; and accountability throughout the process.

Even when the committee's recommendations include substantive terms, most would leave it to individual agencies, subject to judicial and legislative oversight, to define the specific content of the substantive requirements. For example, while the recommendations instruct agencies to employ appropriate access controls on personal data, explicitly determine whether a system provides an acceptable rate of false positives, and engage in data mining whenever practicable with anonymized or pseudonymized personal data, they leave to those agencies the determination of which access controls are "appropriate," how many false positives are "acceptable," and when it is "practicable" to use anonymized or pseudonymized personal data.

This deference undoubtedly reflects many factors: the fact that technology and threats are constantly changing, that much government data mining takes place in secret and so is hard to set specific standards in public documents, and that many determinations concerning the impact of data mining inevitably are context-specific. In fact, TAPAC and most of the other initiatives concerning the proper conduct of government data mining recommend only one substantive and absolute requirement: government data mining should comply with applicable law.[259]

C. The Link Between Privacy and Security

Concerns about data quality and the impact of data mining, although often described in the literature in the context of national security as raising privacy concerns, also inevitably raise significant security concerns. For example, if data mining does not work for its intended purpose, whether or not it invades privacy, it may compromise security. In short, while the discussion of data mining issues might be thought useful in helping to balance privacy with security, it is really more focused on ensuring that data mining is conducted in such a way as to enhance both privacy and 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.

---

[259] *See* TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 46.

In March 2003, the DOJ exempted the FBI's NCIC from the Privacy Act's requirements that data be "accurate, relevant, timely and complete,"[260] 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.[261] 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 the 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."[262] 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."[263] 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 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. 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.[264]

---

[260] Privacy Act of 1974: Implementation, 68 Fed. Reg. 14140, 14140 (Mar. 14, 2003) (codified at 28 C.F.R. pt. 16).

[261] Privacy Act of 1974: Implementation of Exemption, 68 Fed. Reg. 49410, 49412 (Aug. 8, 2003) (codified at 49 C.F.R. pt. 1507).

[262] OFFICE OF THE INSPECTOR GEN., DEP'T OF DEF., INFORMATION TECHNOLOGY MANAGEMENT: TERRORISM INFORMATION AWARENESS PROGRAM (D-2004-033) at 4 (2003).

[263] *Id.* at 7.

[264] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 48.

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.[265] The originator of the concept, Admiral John Poindexter, was forced to resign in the wake of the controversy.[266] Other programs have been similarly retarded by a privacy backlash. In response to public and political pressure, Delta Air Lines withdrew from a CAPPS II pilot program after the airline was threatened with a boycott,[267] and the Secretary of Homeland Security ultimately terminated CAPPS II.[268]

The experience of companies who have participated in supplying data to the government for data mining is illuminating. When JetBlue, Northwest, and American, at the urging of DOD and TSA, provided millions of passenger records to defense contractor Torch Concepts to help test a security system it was designing, they were rewarded with multiple class-action lawsuits by outraged customers.[269] Financial network SWIFT endured multiple investigations from European data protection commissioners and a suit in federal court in Chicago for its role in supplying the Treasury with access to consumer records.[270] AT&T and Verizon face more than three dozen lawsuits for their alleged role in providing the federal government with bulk access to billing records and potentially telephone traffic.[271] The debate over whether they and other firms should be provided immunity for their role in supplying data for data mining has occupied the U.S. Congress for months.[272]

In short, the lack of legal clarity over the role of the private sector in supplying massive data sets to the government, and the resulting backlash when that role is disclosed, raise the specter that valuable tools for enhancing security may be compromised as industry grows hesitant to share personal data with the government. In addition, government officials may grow wary of data mining programs that threaten to embroil them in controversy and may even cost them their jobs.

---

[265] Department of Defense Appropriations Act, 2004, Pub. L. No. 108-87, § 8131, 117 Stat. 1054, 1102 (2003).

[266] *See* Stephen J. Hedges, *Poindexter to Quit over Terror Futures Plan*, CHICAGO TRIBUNE, Aug. 1, 2003, at C1.

[267] *See* Sara Kehaulani Goo, *Agency Got More Airline Records*, WASH. POST, Jun. 24, 2004, at A16.

[268] *See* Hall & DeLollis, *supra* note 61, at 1A.

[269] *See* Sara Kehaulani Goo, *Airlines Confirm Giving Passenger Data to FBI After 9/11*, WASH. POST, May 2, 2004, at A14; Sara Kehaulani Goo, *American Airlines Revealed Passenger Data*, WASH. POST, Apr. 10, 2004, at D12.

[270] *See* Risen, *supra* note 74, at A6.

[271] *See* James Oliphant, *Phone Firms Want Shield if Spy Suits Come Calling*, CHICAGO TRIBUNE, Nov. 15, 2007, at C1.

[272] *See id.*

Moreover, as demonstrated in the controversy over TIA, promises by government officials that data mining is limited to "lawfully obtained" data may carry little weight with lawmakers or with the public in the absence of meaningful legal constraints on accessing personal data, especially from the private sector. Similarly, even though a particular data mining project might be focused solely on a serious concern—for example, keeping terrorists off of airplanes—that may warrant incursions into personal privacy, lawmakers or journalists may nevertheless be skeptical because of the absence of legal constraints that limit the data mining to that particularly important purpose. As the Congressional Research Services has noted, "[m]ission creep is one of the leading risks of data mining cited by civil libertarians."[273] Clear legal standards applicable to data mining would facilitate not only privacy, but also accountability, public, and policymaker confidence, and could increase 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.

VI. 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 its scope. 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 thirty-two years of technological developments since then, it means today 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, they 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 data sets 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

[273] Seifert, *supra* note 10, at CRS-22.

the terrorist attacks of September 11 mean that the government now has both the ability and the motivation to explore huge arrays of private-sector data about individuals who have done nothing to warrant government attention. 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. There have been many efforts to articulate some or all of the content of those standards, including the work of TAPAC, the Markle Foundation Task Force on National Security in the Information Age,[274] the Cantigny Conference on Counterterrorism Technology and Privacy organized by the Standing Committee on Law and National Security of the American Bar Association,[275] as well as think tanks, advocacy groups, academic institutions, and individuals.[276]

While proposals differ in their details, there is broad consensus on many key points. Viewed together, they provide a clear case for why congressional action is needed and the broad roadmap for what that action should include. There is sweeping agreement about the critical need for Congress to establish a legal framework for the appropriate use of data mining to enhance both privacy and security and that the current law is wholly inadequate to that task. In the words of the TAPAC final report, "[l]aws regulating the collection and use of information about U.S. persons are often not merely disjointed, but outdated."[277] They "fail to address extraordinary developments in digital technologies, including the Internet," even though those technologies have "greatly increased the government's ability to access data from diverse sources, including commercial and transactional databases."[278] As a result, "[c]urrent laws are often inadequate to address

---

[274] *See* TASK FORCE ON NATIONAL SECURITY IN THE INFORMATION AGE, MARKLE FOUND., CREATING A TRUSTED NETWORK FOR HOMELAND SECURITY (2003); TASK FORCE ON NATIONAL SECURITY IN THE INFORMATION AGE, MARKLE FOUND., MOBILIZING INFORMATION TO PREVENT TERRORISM (2006); PROTECTING AMERICA'S FREEDOM IN THE INFORMATION AGE, *supra* note 105.

[275] *The Cantigny Principles on Technology, Terrorism, and Privacy*, NATIONAL SECURITY LAW REPORT, Feb. 2005, at 14.

[276] *See, e.g.*, JONAS & HARPER, *supra* note 220; Francesca Bignami, *European Versus American Liberty: A Comparative Policy Analysis of Antiterrorism Data Mining*, 48 B.C. L. REV. 609 (2007); Anita Ramasastry, *Lost in Translation? Data Mining, National Security, and the "Adverse Inference" Problem*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 757 (2006); Ira S. Rubinstein, Ronald D. Lee & Paul M. Schwartz, *Data Mining and Internet Profiling: Emerging Regulatory and Technological Approaches*, 75 U. CHI. L. REV. (forthcoming 2008); David J. Steinbock, *Data Matching, Data Mining, and Due Process*, 40 GA. L. REV. 1 (2005); K. A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 2 (2002-2003); Joseph T. Thai, *Is Data Mining Ever a Search Under Justice Stevens's Fourth Amendment?*, 74 FORDHAM L. REV. 1731 (2006); Michael Isaac, Note, *Privatizing Surveillance: The Use of Data Mining in Federal Law Enforcement*, 58 RUTGERS L. REV. 1057 (2006).

[277] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 6.

[278] *Id.*

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."[279] "It is time to update the law to respond to new challenges."[280]

That framework for regulating government data mining should include, at a minimum, requirements for:

- Authorization by Congress or a senior administration official of new data mining programs that seeks to ensure their efficacy, compliance with legal requirements, and a high level of oversight and accountability within each federal agency.
- Compliance with the law both when accessing data and engaging in data mining. The government should neither encourage nor press third parties to violate their legal obligation when providing data to the government.
- Evaluation of effectiveness in accomplishing specified objectives prior to being deployed and regularly thereafter. Those assessments should take into account practical experience with the system, technological advances, changing needs, and the impact on individuals.
- Limits on who can use data mining systems, have access to large data sets, and purposes for that use, as well as tools to enforce those limits. Rules can be built into data mining systems so, for example, "the analyst might be asked to specify whether she has a search warrant, and if she does not, the system might not allow her to retrieve certain kinds of information."[281]
- Some form of judicial authorization before data mining systems are deployed or used where personally identifiable information will be used in a way that affects U.S. persons, whether by denying or delaying access to a facility or benefit, subjecting them to intrusive investigation, or in some other way. The specific court matters less than that the authorization be external to the agency engaging in the data mining and specified by Congress.
- The use of data minimization and anonymization and other tools to limit the amount of information revealed to only what is necessary and authorized. This has been a major focus of the Markle Foundation Task Force on National Security in the Information Age, which has proposed that "anonymizing technologies . . . allow analysts to perform link analysis among data sets without disclosing personally identifiable information. By employing techniques such as one-way hashing, masking, and blind matching, analysts can perform their

[279] *Id.*
[280] *Id.*
[281] Rubinstein, Lee & Schwartz, *supra* note 276, at 6.

jobs . . . without the need to gain access to personal data until they make the requisite showing for disclosure."[282]

- The use of audit tools to ensure that the rules surrounding data mining are being followed.
- A system of redress to ensure that innocent individuals harmed by data mining are aware of the role of data mining, given the opportunity consistent with the nature of the data mining to dispute and seek correction of erroneous data, and compensated for their injury. The system must also ensure that data mining programs "learn" from their mistakes and that those errors are logged. If false positives are "inevitable," they must be addressed from the perspective of both affected individuals and the systems that are generating them.
- Serious oversight of data mining operations that delivers a high degree of accountability that data mining systems are used appropriately, lawfully, and effectively. Components of such oversight include review by an agency privacy officer, routine audits, agency Inspector General investigations, and regular reports to Congress and, to the extent consistent with the data mining context, the public. It is important that the oversight be—and be seen to be—both rigorous and independent.

It is striking that the justifications for these and other measures include both privacy *and* security. There is widespread agreement that privacy is an important right that is necessary to a productive life. Congress, and perhaps the Supreme Court, need to act to restore the protection for privacy that technology, legislative and judicial inaction, and the past missteps of these two branches have eroded. The TAPAC final report could not have been more explicit in its conclusion: "[i]nformational privacy is critical to participation in democracy and society."[283]

These undertakings, however, are also necessary to ensure that data mining serves its intended purpose—whether detecting fraud or protecting homeland security. With its seemingly insatiable quest for more data, the government threatens to exacerbate what may be its greatest challenge in the national security context: making sense of the data it already has. This point has been highlighted by every group to consider government data mining.[284] In addition to the government's many domestic data sources, it also receives more than 650 *million* intelligence intercepts every day.[285] The problem is not having too little data; it is having too much data and not being able to make sense out of the data it has. It is "separating out the 'signal' of useful information from the 'noise' of all of those data."[286] Jonas and Harper write that the "key goal—and challenge—is to produce not just more information,

---

[282] CREATING A TRUSTED NETWORK FOR HOMELAND SECURITY, *supra* note 274, at 34.
[283] TAPAC, SAVEGUARDING PRIVACY, *supra* note 9, at 34.
[284] *See, e.g.*, *id.*; Smith & Howe, *supra* note 105, at 1.
[285] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 48.
[286] *Id.*

but more *useful* information."[287] Data mining is essential to that objective, but poor tools or inappropriate data do not merely fail to advance security, they actively threaten it. In the words of the TAPAC report, a new regulatory structure is necessary to "help protect civil liberties and national security, and to help empower those responsible for defending our nation to use advanced information technologies—including data mining—appropriately and effectively. It is time to update the law to respond to new challenges."[288] The need could not be greater, and thanks to the work of so many organizations which have addressed this issue, the broad outlines of a path forward are clear. It only remains for Congress to act.

[287] JONAS & HARPER, *supra* note 220, at 5.
[288] TAPAC, SAFEGUARDING PRIVACY, *supra* note 9, at 6.