Principles for Government Data Mining

Preserving Civil Liberties in the Information Age

THE CONSTITUTION PROJECT

Safeguarding Liberty, Justice & the Rule of Law

Principles for Government Data Mining:

Preserving Civil Liberties in the Information Age

THE CONSTITUTION PROJECT

Safeguarding Liberty, Justice & the Rule of Law

THE CONSTITUTION PROJECT

Established in 1997, The Constitution Project (TCP) is renowned for its ability to bring together *unlikely allies*—experts and practitioners from across the political spectrum—in order to promote and safeguard America's founding charter. TCP is working to reform the nation's broken criminal justice system and to strengthen the rule of law by undertaking scholarship, policy reform and public education initiatives. TCP was born out the belief that we must cast aside the labels that divide us, in order to keep our Constitution and our democracy strong.

The Constitution Project
1200 18th Street, NW
Suite 1000
Washington, DC 20036
Tel 202.580.6920
Fax 202.580.6929

Email: info@constitutionproject.org

www.constitutionproject.org

TABLE OF CONTENTS

The Constitution Project (TCP) is a national, bipartisan think tank that develops consensus-based solutions to some of the most difficult constitutional challenges of our time. Established in 1997, TCP is renowned for its ability to bring together *unlikely allies*—experts and practitioners from across the political spectrum—in order to promote and safeguard America's founding charter. TCP works on criminal justice and rule of law issues by undertaking scholarship, policy reform and public education initiatives. TCP's Rule of Law Program addresses threats to our constitutional system and our civil liberties. TCP's Criminal Justice Program seeks to counter a broad-based effort to deny fundamental day-in-court rights and due process protections to those accused of crimes.

The Rule of Law Program's Liberty and Security Committee comprises an ideologically diverse group of prominent Americans who work with the Constitution Project to recommend ways to preserve individual rights while also ensuring public safety. The Committee addresses a wide range of issues, including the ones discussed in this report: reform of data mining procedures to ensure the preservation of civil liberties. This is part of the Committee's ongoing efforts to examine areas in which technology is developing more quickly than the law, and to ensure that government programs incorporate proper safeguards to preserve constitutional rights in the digital age and beyond.

The amount of data available electronically has increased dramatically over the last decade, and will only grow more copious. Thus, the government is relying increasingly upon data mining programs, namely the use of computing technology to examine large amounts of data to reveal patterns and identify potential wrongdoing. Used properly, data mining can provide a valuable tool for the government to uncover fraud or criminal activity. However, uses in national security cases may prove more challenging, and it is important to ensure that the government's collection, acquisition, and use of data does not infringe upon individual privacy rights and respects the constitutional rights of freedom of expression, due process, and equal protection.

The report reflects the belief that effective data mining reform requires a commitment to the rule of law and respect for constitutional rights. This publication contains three parts. The first is a discussion of what data mining is, what its purposes are, and the current legal status of data mining. The second part outlines broad principles that should inform future government regulation of data mining. The final part makes specific recommendations to the government to reform data mining and ensure that such programs can both be effective and at the same time protect individual constitutional rights. In particular, the recommendations cover the need for transparency and notice; accountability, oversight, and redress; and data integrity and security when the government is conducting data mining operations.

The Constitution Project sincerely thanks Will DeVries, formerly of the law firm WilmerHale and now at Google, Inc., and Anne Sherwood, of WilmerHale, for sharing their expertise in law and technology and for their invaluable work in researching and drafting this report. We are also grateful to Robert Greffenius, a Fried Frank Legal Fellow at the Constitution Project, for his assistance in finalizing this report.

The Constitution Project also thanks Atlantic Philanthropies, Community Foundation for the National Capitol Region, CS Fund/Warsh-Mott Legacy, Foundation to Promote Open Society, Lawrence and Lillian Solomon Fund, Overbrook Foundation, Rockefeller Brothers Fund, and Wallace Global Fund for their support of this report.

Virginia E. Sloan, President, **Sharon Bradford Franklin**, Senior Counsel

December 2010

ENDORSERS OF THE CONSTITUTION PROJECT'S REPORT

Principles for Government Data Mining:
*Preserving Civil Liberties in the Information Age**

LIBERTY AND SECURITY COMMITTEE MEMBERS

CO-CHAIRS:

David Cole, Professor, Georgetown University Law Center

David Keene, Chairman, American Conservative Union

MEMBERS:

Bob Barr, Former Member of Congress (R-GA); CEO, Liberty Strategies, LLC; the 21st Century Liberties Chair for Freedom and Privacy, the American Conservative Union; Chairman, Patriots to Restore Checks and Balances; Practicing Attorney

Mickey Edwards, Lecturer, Woodrow Wilson School of Public and International Affairs, Princeton University; former Member of Congress (R-OK) and Chairman of the House Republican Policy Committee

Thomas B. Evans, Jr., Former Member of Congress (R-DE) and Co-Chairman, Republican National Committee; Founder, Florida Coalition for Preservation

Eugene R. Fidell, Florence Rogatz Lecturer in Law, Yale Law School

Michael German, Policy Council, American Civil Liberties Union; Adjunct Professor, National Defense University School for National Security Executive Education; Special Agent, Federal Bureau of Investigation, 1988-2004

Asa Hutchinson, Senior Partner, Asa Hutchinson Law Group; Undersecretary, Department of Homeland Security, 2003-2005; Administrator, Drug Enforcement Administration, 2001-2003; Member of Congress (R-AR), 1997-2001; United States Attorney, Western District of Arkansas, 1982-1985

Dave Lawrence Jr., President, Early Childhood Initiative Foundation; former Publisher, *Miami Herald* and *Detroit Free Press*

Rear Admiral James E. McPherson, USN, Retired, JAGC, former Judge Advocate General of the Navy

Paul R. Pillar, Visiting Professor and Director of Studies, Security Studies Program, Georgetown University; Deputy Chief, DCI Counterterrorist Center, 1997-1999; former National Intelligence Officer for the Near East and South Asia; former Executive Assistant to the Director of Central Intelligence; Intelligence Officer, Central Intelligence Agency and National Intelligence Council, 1977-2005

Peter Raven-Hansen, Glen Earl Weston Research Professor, George Washington University Law School

William S. Sessions, Partner, Holland and Knight LLP; Director, Federal Bureau of Investigation, 1987-1993; Judge, United States District Court for the Western District of Texas, 1974-1987, Chief Judge, 1980-1987; United States Attorney, Western District of Texas, 1971-1974

Neal R. Sonnett, Former Chair, American Bar Association Task Force on Treatment of Enemy Combatants and Task Force on Domestic Surveillance in the Fight Against Terrorism; Former Assistant United States Attorney and Chief, Criminal Division, Southern District of Florida; Past President, National Association of Criminal Defense Lawyers; Past Chair, ABA Criminal Justice Section

John W. Whitehead, President, The Rutherford Institute

Colonel Lawrence Wilkerson, USA, Retired; Visiting Pamela C. Harriman Professor of Government at the College of William and Mary; Professorial Lecturer in the University Honors Program at George Washington University; former Chief of Staff to Secretary of State Colin Powell

---

CONSTITUTION PROJECT POLICY ADVISORY COMMITTEE MEMBER

Christopher G. Caine, Mercator XXI, LLC

REPORTERS

Will DeVries, WilmerHale (now Policy Counsel, Google, Inc.)

Anne Hazlett Sherwood, WilmerHale

CONSTITUTION PROJECT STAFF

Sharon Bradford Franklin, Senior Counsel, Rule of Law Program

*Affiliations listed for identification purposes only

In the Information Age, enhancing information awareness is a critical objective for the federal government. Indeed, Lee Hamilton, the Vice Chairman of the 9/11 Commission, pointed to the intelligence agencies' inability to organize and share information as "the single greatest failure of our government in the lead-up to the 9/11 attacks."[1] In the wake of 9/11, and as the federal government tools to mine data have developed, the government has built or is building thousands of databases and is deploying hundreds of data mining applications to mine law enforcement, communications, and intelligence data for criminal, terrorist, or national security threats. It is clear that government data mining operations will only grow in the years to come.

Government data mining, which this report defines broadly,* can offer significant benefits, but without adequate processes and controls, it can encroach on constitutional rights and values—including privacy, freedom of expression, due process, and equal protection. Innocent people could mistakenly be added to terrorist "watch lists," leading to travel delays, reputational harms, or more serious consequences. Rogue government employees can abuse database access and look for information on the famous or infamous—as occurred with the 2008 presidential candidates.[2] Careless contractors can lose laptops with unencrypted personal data. In addition, not all types of data mining programs are equally valuable. Some, such as those that improve program efficiency and evaluate performance, are very worthwhile. Others, such as predictive pattern-based analysis for terrorism prevention, are evolving rapidly but, due to the particular difficulties of predicting terrorist activity and these analyses relative youth, their efficacy has yet to be demonstrated to the public.

Even where these programs are effective, despite the weightiness of the interests involved, the current legal regime fails to clearly or uniformly regulate government data mining activities: While the Constitution and several federal statutes are implicated by government data mining, the practice has substantially avoided direct regulation. This is a situation that can be and must be remedied. We can adopt rules that both allow the government to harness the vast seas of information for our collective benefit and simultaneously protect the delicate relationship our Constitution established between the government and the governed.

For these reasons, the Constitution Project's bipartisan Liberty and Security Committee urges Congress and the executive branch to incorporate critical protections for individual rights into all government data mining programs. We offer the principles below as a starting point for agency-specific data mining regulations, or for government-wide rules. These common sense principles are rooted in constitutional values and the Fair Information Practice Principles.** We recognize that not all of these principles can or should dictate government behavior when classified information is involved. But when applied to the greatest extent possible, our recommendations can enable the government to use data mining techniques without sacrificing constitutional rights and values.

---

* We define "data mining" to include *any use of computing technology to examine large amounts of data to reveal relationships, classifications, or patterns.* Our principles apply only to any data mining that is undertaken by a government entity, on behalf of the government, or where government personnel are permitted to access the data.

** The Fair Information Practice Principles ("FIPPs") underlie the Privacy Act of 1974 and many United States' agencies and international entities' approaches to privacy protection. *See* Federal Trade Commission, Fair Information Practice Principles, http://www.ftc.gov/reports/privacy3/fairinfo.shtm. The Department of Homeland Security describes the FIPPs as a set of eight principles: Transparency, Individual Participation, Purpose Specification, Data Minimization, Use Limitation, Data Quality and Integrity, Security, and Accountability and Auditing. Department of Homeland Security, *Privacy Policy Guidance Memorandum*, No. 2008-01, at 1 (Dec. 29, 2008), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf (memorializing DHS adoption of the FIPPs); *see also* The Constitution Project, *Guidelines for Public Video Surveillance: A Guide to Protecting Communities and Preserving Civil Liberties* (2006) (providing additional information on these principles).

Principles for the Development of Data Mining Programs

- *Prior to acquisition, clearly articulate, in writing, the purpose(s) of data acquisition and the intended use(s) of that data.*
- *Create a comprehensive data mining Plan ("Plan") covering data sources, data acquisition, system design and capabilities, and intended uses.*
- *Perform an internal evaluation of the program's expected effectiveness, costs, benefits, compliance with existing law, and impact on civil liberties and constitutional values.*
- *Submit the Plan and internal evaluation for review and comment in the Federal Register, as feasible; ensure congressional and high-level administrative review for non-public aspects.*
- *Respond to and incorporate administrative, public, and congressional commentary on the Plan.*

Principles for the Operation of Data Mining Programs

Transparency and Notice

- *To the greatest extent feasible and consistent with national security concerns, provide notice to an individual of any specific government action or classification of that individual pursuant to data mining. When immediate notice is not possible, establish a framework for delayed notification to the extent feasible.*

Accountability, Oversight, and Redress

- *Create administrative standards and procedures governing acquisition, use, and sharing of information for data mining.*
- *Establish and enforce penalties for misuse and abuse of data mining programs by operators or others.*
- *Establish a system of appeal and redress for individuals misclassified or harmed.*
- *Require private companies to have data correction procedures in place as a requirement of contracting with a government agency.*
- *To the extent feasible, permit individuals to review and correct data.*
- *Conduct and publish the results of regular audits, and report regularly to Congress.*

Authority and Choice

- *Coordinate uses, best practices, regulations and protections among agencies.*
- *Establish approval procedure for data acquisition and actions taken pursuant to data mining with decisionmakers at highest possible level and outside of the program's operational structure.*

Data Integrity and Security

- *Incorporate technical and administrative measures to limit access to or availability of personal data, particularly sensitive or personally identifiable data.*
- *Evaluate and improve data security, integrity, accuracy, and timeliness on a regular basis, including the use of audit trails.*
- *Conduct training and evaluation for employees with access to personal data or data mining systems.*
- *Take all reasonable steps to rectify and minimize harm from data breach, including prompt notification of affected individuals.*

Data Appropriateness/Minimization

- *Ensure that acquisition only includes data relevant to the purpose of the program and minimize the extent to which databases are aggregated.*
- *Use deidentified or aggregate data formats or other techniques with respect to personal information to minimize potential harm.*
- *Continue safeguards for personal data through technical measures, rule, or contract as data moves "downstream."*
- *Set limited retention periods and ensure complete destruction of expired data.*

In the Information Age, human output is increasingly measured by the terabyte rather than the kilo. The newest industrial giants manufacture products not out of steel but out of data. Google makes its billions from one single commodity—information—and from services that have one goal: the accessible organization of that information.[3] Successful governments must similarly harness the power of information to improve efficiency, deliver services, fight crime, and protect national security.

Enhancing information awareness is a critical objective for the federal government, whether to reduce waste or to better organize intelligence.[4] Indeed, Lee Hamilton, the Vice Chairman of the 9/11 Commission, pointed to the intelligence agencies' inability to organize and share information as "the single greatest failure of our government in the lead-up to the 9/11 attacks."[5] In the wake of 9/11, and as the federal government's tools to mine data have developed, the government is improving its capabilities, although it is not yet as adept as many private sector actors. Indeed, flaws in these systems contributed to the failure to preempt the 2009 attempted terrorist attack by the "Christmas Day bomber."[6] Nevertheless, the government has built or is building thousands of databases and is deploying hundreds of data mining applications, and those are increasingly being applied to mine law enforcement, communications, and intelligence data for criminal, terrorist or national security threats. It is clear that government data mining operations will only grow in the years to come.

Increased government access to and use of information brings significant benefits, but also increases the risk of encroachment on constitutional rights and values—including privacy, freedom of expression, due process, and equal protection. With insufficient controls, innocent people could be mistakenly added to terrorist "watch lists" and potentially barred from air travel. Government employees can abuse database access and look for information on the famous or infamous. Careless contractors can lose laptops with unencrypted personal data.

These harms can be contained, but only with firm rules that ensure government actors undertake data mining with adequate safeguards and minimize the potential for mistake, misuse and abuse. For these reasons, we, the members of the Constitution Project's Liberty and Security Committee endorsing this report, urge Congress and the executive branch to incorporate critical protections for individual rights into any government data mining programs. We offer the principles below as a starting point for agency-specific data mining regulations, or for government-wide rules.

I. Background

To better understand and evaluate our recommended principles, we offer the following background discussion to explain what we mean by "data mining," how the government has mined data to this point, and the constitutional and legal implications of data mining.[7]

A. What is Data Mining?

Types of data mining. Data mining is the broad term used to refer to many types of activities involving data processing. Most agree that the core of data mining is the use of "pattern-based" searching to uncover novel patterns or relationships in large sources of data.[8] Such pattern-based systems learn over time by examining the data, comparing the data to a model, and then searching databases for patterns matching the revised model.[9] Federal money-laundering investigators, for example, might input information about financial crimes and criminals into a sophisticated data mining system, which would review banking

data for transactions or accounts that share suspicious attributes with the criminal data points. The system would continue to refine its model of suspect behavior over time.

Broader definitions might also include "subject-based" queries, where data are simply scanned for items or events meeting specified parameters.[10] For example, law enforcement officers might start with a known suspect or person of interest, and use a multi-jurisdictional law enforcement database to search for information about that person, such as prior arrests or known associates.[11] Of course, this distinction is often blurry, and many data mining systems use both search-based and pattern-based techniques.[12]

Although the term "data mining" does not technically refer to the process of *collecting* or *acquiring* the data,* we consider the acquisition process as part of this report. The acquisition stage can often have a great impact on individual rights and the ability to safeguard personal information, and therefore several of our recommendations below focus on the data acquisition process.

Types of data. While data mining techniques can be applied to any type of data, policy debates involving data mining are generally focused on mining of data about individuals. Privacy advocates often talk about "personally identifiable information" ("PII"), or information that can be used to uniquely identify a specific individual, but a broader swath of information is increasingly being treated as PII. The Office of Management and Budget and others now define PII as "information which can be used to distinguish or trace an individual's identity . . . alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual."[13] Accordingly, the relevant question for this report is not whether some set of data actually identifies a particular individual, but whether it is or can be linked to a specific individual or device.[14]

Technology of data mining. The ability to mine data has improved dramatically in recent years thanks to advances in computing technology and digitization of information. Computer processing speed continues to double every 18-24 months or sooner,[15] data storage has become cheaper and more compact,** and private networks and the internet have immensely increased the availability of information. At the same time, more and more data are becoming available. Old data are now digital—census records, meteorological data, stock prices—while new forms of data have been created, like social networking data, internet search queries, and systems like E-Z-Pass for highway tolls and employee access cards. Observers estimate that the world's collected data volume doubles every year.[16]

Notably, advances in technology also can be employed to reduce the potential civil liberties impact of data mining. For instance, anonymization engines, data masking, or data transformation can help shield sensitive data from human operators. Other tools can securely mine data remotely without copying it to a local database—lessening the chances

---

* We distinguish here between the "collection" and the "acquisition" of data. As used in this report, "collection" generally refers to the process by which data are obtained directly from an individual. "Acquisition" refers to both direct collection and obtaining data from a third party or another government entity.

** Recent advances in computer chip technology have led to great magnification of processing speeds and vast increases in data storage capabilities, "giving rise to a constant escalation of computing power at lower costs." John Markoff, *Intel Says Chips Will Run Faster, Using Less Power*, N.Y. Times, at B9, Jan. 27, 2007. Those developments are expected to continue in the near-term, potentially facilitating new, "astronomical" enhancements in data storage. *See* John Markoff, *Advances Offer Path to Shrink Computer Chip*, N.Y. Times, at A12, Aug. 31, 2010. Utilizing these improvements, the government will have the capability to store and access information on an unprecedented scale. Such technological advances have significant implications for both the practical challenges of data mining and the constitutional concerns raised by data mining programs, discussed in detail below.

of accidental release or exposure.[17] Use of encryption and security features, while not specific to data mining, could also reduce potential harms.

The purpose of this report is to offer a set of rules and procedures requiring that when government entities collect or acquire data for data mining purposes or use data mining tools, they do so in ways least invasive of civil liberties and constitutional values. This policy goal requires a broad conception of data mining, since a wide range of data tools can affect these values. We thus define the term to include *any use of computing technology to examine large amounts of data to reveal relationships, classifications, or patterns.* Furthermore, our principles apply only to any data mining involving information that is or can be linked to a specific individual or device and that is undertaken by a government entity, on behalf of the government (whether a government contractor or private party), or where government personnel are permitted to access the data.*

B. Uses and Purposes of Data Mining

1. Government Uses of Data Mining

In recent years the federal government has disclosed hundreds of data mining programs, varying widely in scope and purpose. Many of these programs have been abandoned, and others are still in planning stages. There are also—almost certainly—programs that have not been publicly disclosed. Below we discuss the principal purposes for which government actors engage in data mining, and describe some of the most relevant programs.[18]

Efficiency and program evaluation. The most common purpose for which government agencies mine data is to improve efficiency and evaluate performance.[19] This category would include human resources and/or internal operations management. Most observers believe that data mining can improve government performance if used appropriately. For example, the Department of Justice ("DOJ") and Department of Veterans Affairs have successfully used operational data to more efficiently allocate agency resources.[20] While these data mining applications would fall within our broad definition, they would not, as a general rule, raise significant constitutional concerns due to the non-personal nature of the data used. Nevertheless, even internal data mining programs like this one may pose risks to the privacy rights of employees and should be reviewed and evaluated following the principles described below.

Fraud detection and compliance. Data mining can also be effective at combating fraud and auditing for compliance. The Government Accountability Office ("GAO") estimated in 2004 that the federal government had employed or will employ data mining for these purposes in at least 30 separate programs.[21] The Internal Revenue Service ("IRS") makes extensive use of data mining to increase tax compliance and detect tax fraud.[22] A data mining program helped uncover millions of dollars in Medicare fraud.[23] As we discuss below, while these applications can pose serious risks to civil liberties, data mining is well-suited to situations like these where there are established patterns of misbehavior, many data points from which to draw inferences, and post-hoc enforcement of privacy safeguards can be effective.

---

* The private sector has long used data mining for marketing and other purposes. Although private data mining is beyond the scope of this report, it still implicates similar privacy concerns and therefore we recommend that federal, state, and local governments contemplate private-industry regulation or oversight to protect individual liberty interests. Yet, many of the relevant constitutional and legal restrictions only apply to government actors, and government data mining can have substantively different and farther-reaching impacts than private-sector mining, including restricting an individual's ability to travel by plane or flagging an individual for criminal investigation, as explained below. We therefore focus our principles on government data mining.

Criminal investigation. Law enforcement officials have employed data mining tools to help investigate crimes or enhance their understanding of criminal patterns and behavior. Data mining tools can assist investigators in matching crime scene evidence to other crimes or suspects or finding known associates or other information about persons of interest. An increasing amount of such data mining is occurring at "fusion centers," centers within each state that bring together federal, state, and local law enforcement personnel to share information and coordinate activities. Through these fusion centers, the federal government has acquired data from state and local law enforcement databases to improve information sharing and availability among law enforcement and intelligence agencies. While more efficient sharing of data can undoubtedly aid law enforcement efforts, the unlimited scope, lack of transparency, and lack of oversight for the program create significant risks to civil liberties.[24] Perhaps in recognition of these concerns, the Department of Homeland Security ("DHS") recently linked grant money to a requirement that the recipient fusion centers comply with certain privacy and civil liberties guidelines.[25]

Crime prevention and counterterrorism. In the last decade, government officials have increasingly sought to employ data mining tools to anticipate and prevent crime and terrorist acts. Use of data mining for this purpose has garnered the most attention and concern, despite the relatively small number of applications involved. On the one hand, the damage caused by terrorism and other serious crimes is so large that it seems worth trying every conceivable tool to combat it. On the other hand, such uses raise several civil liberties concerns: First, the value of data mining to help prevent bad acts is unclear due to the particular difficulties of developing a predictive model to identify plans for terrorist acts. Second, the consequences for an individual who is misidentified as a potential criminal or terrorist can be devastating. Third, as with any other form of data mining involving personal information, there is a risk of misuse or abuse of the data.[26]

While there are undoubtedly many classified projects in this category, some have been disclosed. For example:

- ***Investigative Data Warehouse ("IDW").*** The Federal Bureau of Investigation ("FBI") describes the IDW as its "single largest repository of operational and intelligence information;" it serves as a centralized data access point for FBI agents across the country.[27] In addition to the IDW's value for investigative purposes, the Electronic Frontier Foundation concluded that the FBI has likely been employing advanced, predictive "data exploitation" tools based on the IDW data.[28]

- ***Total Information Awareness.*** The Defense Department's Advanced Research Projects Agency ("DARPA") began a program after 9/11 to gather vast amounts of domestic and foreign data and develop tools to discern patterns and relationships in the data for the benefit of defense, counterterrorism, and law enforcement agencies. The program (later renamed Terrorism Information Awareness) was criticized by the public and Congress and eventually abandoned. It is possible that aspects of the program continue as part of classified operations.

- ***Secure Flight / CAPPS II.*** Also following 9/11, the Federal Aviation Administration (later the Transportation Security Administration ("TSA")) began work to develop a replacement for the existing air passenger screening system (Computer-Assisted Passenger Prescreening System, or CAPPS) to screen air passengers for inclusion on "watch lists" or for terrorist or criminal threat. The system was designed to use information acquired from government sources, airlines, and commercial data brokers. The CAPPS II proposal was scrapped for a new program, Secure

Flight, which was designed to focus exclusively on identifying and preventing terrorism threats. After privacy advocates and the GAO raised concerns, Congress mandated that any such program must satisfy a GAO evaluation to ensure compliance with specified privacy protections. The TSA put Secure Flight's requirements into effect in the fall of 2009, and all airlines are expected to fully implement the program by the end of 2010.[29]

2. Practical Hurdles to Effective Data Mining

Data mining—particularly for criminal prevention or counterterrorism purposes—presents practical issues that may countervail their potential effectiveness.

First, all data mining faces challenges related to *data integrity*. The outcome of data mining can only be as good as the underlying data. Duplicate records, incomplete records, timeliness of updates, and human error all create data integrity problems. For instance, both business and the government have struggled with how to ensure that data about individuals are correctly attributed. Names change or are recorded differently, addresses and other identifiers change, or data are entered incorrectly.[30] For instance, Umar Farouk Abdulmutallab—the alleged "Christmas Day bomber" —was permitted to board a plane to the United States in part because his name was misspelled, contributing to the failed integration of government databases.[31] These data inaccuracies can also result in false or improper flagging of individuals as discussed below. Indeed, data integrity problems are particularly acute in the criminal and terrorism context, as information is often sketchy or incomplete, or from sources that do not allow verification or follow-up.[32] Relatedly, underlying databases may be incompatible, and prevent or complicate data mining applications that use multiple sets of data.[33] At the same time, correcting for problems of data quality can dramatically increase the cost of a data mining program.

Second, any data mining application that automatically categorizes records based on model criteria raises issues about *"false positives"* and *"false negatives."* For example, credit card companies use data mining to flag potentially fraudulent purchases. If the company places a hold on the account of a customer who in fact was making legitimate purchases, it has made a "false positive." If it fails to put a hold on an account that has been stolen, it has made a "false negative." Because the models used by data mining applications cannot be perfect, false positives and false negatives are inevitable.[34] In the credit card context, the harm from these errors may be generally small, and consumers usually feel that the benefits outweigh the inconvenience of false positives. In the criminal or terrorism context, however, these errors can be disastrous. As discussed, despite government-held information about the suspect in the Christmas Day 2009 bombing attempt, the counterterrorism community failed to identify him as a threat and therefore to prevent what could have been a serious disaster aboard a United States-bound airplane.[35] Conversely, individuals who are *erroneously* added to a terrorist watch list or the "Do Not Fly" list can be materially harmed in a variety of ways. In 2009, for example, the DOJ revealed that nearly 24,000 individuals had been incorrectly listed on a terrorist watch list—and admitted that the errors caused harm to those individuals and posed risks to national security.[36] Both false positives and false negatives thus increase costs to the government and airlines and create public skepticism about the value of security measures.[37] Moreover, when there are few actual events from which to create a model—as, fortunately, is particularly the case with terrorist acts*—the potential for false identification is increased.[38]

---

* In addition, terrorists may attempt to consciously alter their methods to avoid mimicking past terrorist plots—undermining pattern-based data mining methodologies.

C. Legal Status of Data Mining

To fully understand the costs and benefits of a data mining program, one must understand not only the practical hurdles but also the potential impact on civil liberties and constitutional values. We briefly review here those values and the limits the Constitution imposes on such programs. We then discuss the additional protections afforded by legislation and administrative rules and policies. We conclude that, despite the weightiness of the values at stake, the current legal regime fails to clearly or uniformly regulate government data mining activities.

1. Constitutional Limits on Data Mining

Privacy. In most discussions of the legal and ethical implications of data mining, privacy is the central concern. We will first briefly discuss the general concept of a right to privacy and next summarize existing constitutional protections for data privacy. Privacy is a general term, covering concepts that are often very different from one another—from data protection to freedom from warrantless search—but all conceptions of privacy center on the right to personal autonomy, or what Justice Brandeis famously called "the right to be let alone."[39] Related to privacy is the right of "confidentiality," which concerns how personal information is disseminated,[40] and "anonymity," a form of privacy that "occurs when the individual is in public places or performing public acts but still seeks, and finds, freedom from identification and surveillance."[41] For instance, in the latter case, an individual does not expect to be recorded or identified as he or she enters a fertility clinic or an Alcoholics Anonymous meeting, even if the person must pass through a public space to enter. Nor do most individuals expect that information they disclose to family members while in a public park will be broadcast, despite the location of the personal revelations. The Supreme Court has validated the right to anonymity in certain circumstances by recognizing that people should be able to remain anonymous while exercising certain constitutionally protected rights.[42]

While the constitutional standards as interpreted by the courts have been slow to change, the popular understanding of what is personal has altered considerably with technological development. Whereas most people historically have safeguarded private documents and information within their homes, now use of laptops, mobile communications devices, and the internet have radically changed how we store our personal information. As a result, mobile PCs, cloud computing, and online social networks all have shifted our understanding of privacy toward an independent liberty interest removed from the traditional focus on the home and person.

Government mining of personal data can implicate the right of privacy, broadly considered, in many ways. First, to the extent the data used comes *directly* from individuals—via surveillance or other means—the practice obviously implicates privacy rights. Acquisition of data from third parties, such as an internet service provider or a bank, can also implicate privacy and confidentiality, especially in the absence of notice. Regardless of whether a court would currently view the Constitution as limiting the government's access to information that an individual has shared with a third party,[43] government access to or use of personal information in databases can still violate our privacy values—as demonstrated by, for instance, the passage of laws limiting law enforcement access to stored communications.[44] Second, even when the collection or acquisition of information is not objectionable, the use of it for data mining can be. Data willingly shared by airline passengers for security screening, airline operations, and meal preferences can implicate privacy and anonymity values if mined for political and

religious affiliations, for instance. Finally, privacy rights can be implicated by inappropriate sharing and downstream uses of information gleaned from data mining.

The Fourth Amendment is the primary constitutional provision that pertains to issues of information privacy.[45] It protects individuals against "unreasonable searches and seizures,"[46] but only in circumstances in which a person has a reasonable expectation of privacy.[47] Technology is increasingly blurring the lines between spheres in which people commonly do or do not expect privacy. In many instances, such as internet use, it is not yet clear whether courts will construe privacy expectations as "reasonable" under the Fourth Amendment.[48] Nevertheless, the contours of a reasonable expectation of privacy have been defined in at least one significant respect. In the seminal 1976 case *United States v. Miller*, the Supreme Court held that, as a general matter, a person does not have a reasonable expectation of privacy in information revealed to a third party.[49]

Miller's "third-party doctrine" has substantial implications for government data mining activities. Under present Supreme Court jurisprudence, data obtained by the government from private companies, or obtained by one government agency from another, very likely fall under its purview, and therefore outside the protection of the Fourth Amendment.[50]

However, a host of scholars have vociferously objected to the third-party doctrine, and there are signs that judges are beginning to reject the application of the doctrine in the personal data context.[51] Indeed, the drastic changes in technology in recent decades have put in sharp focus the limitations the doctrine imposes on Fourth Amendment protections in the digital era. As privacy scholar Daniel Solove has observed, "[i]n the Information Age, so much of what we do is recorded by third parties that the Court's third party doctrine increasingly renders the Fourth Amendment ineffective in protecting people's privacy against government information gathering."[52] Furthermore, in many instances, individuals have no choice but to disclose information to a third party in order to be able to participate in basic aspects of modern society, such as online banking, storing electronic business or financial records online, communicating by phone or email, or using a credit card to make purchases.[53] The internet service provider, telecommunications company, or credit card company therefore becomes a repository of a host of personal data.* Finally, the searching of aggregated data today is different in kind from the government data operations of thirty years ago, which often involved a search of one data source. Thus, even assuming application of the third-party doctrine was proper in *Miller* and its progeny, some argue that government data mining should be treated differently because of the potential to search and review vast quantities of data from a wide variety of sources. As a result, data mining today may implicate Fourth Amendment interests more than other systems previously considered by the Court.[54]

Courts are increasingly recognizing reasonable expectations of privacy related to new technologies, and some have limited the reach of the third-party doctrine. The highest courts in more than twelve states have rejected the doctrine under their state constitutions.[55] The Supreme Court also has expressed some willingness to narrow the third-party doctrine. In one recent case, it recognized a reasonable expectation of privacy in the results of diagnostic tests by a hospital, and it acknowledged that the hospital's collection of information for the purpose of law enforcement's review—and not simply for its purported

---

* Furthermore, consenting to release personal data to third parties for purposes of banking, storing records, communicating, or making purchases is not the equivalent of consenting to allow that data to be given to the government. When individuals consent to the former, they often do not expect the scope of that consent to encompass the latter.

non-law enforcement use by the hospital—violated the Fourth Amendment.[56] Significantly, recent decisions by some United States courts of appeal have addressed the issue more directly, holding that individuals have a reasonable expectation of privacy in emails stored by an internet service provider and in text messages.[57] The stored email case, *Warshak v. United States*, was later vacated on procedural grounds,[58] but a related criminal appeal is pending, and as this report goes to print, the Sixth Circuit is expected to rule on that issue soon.[59] The Eleventh Circuit, however, recently held that an individual does not have a reasonable expectation of privacy in emails voluntarily provided to a third party internet service provider.[60] The text message case, *City of Ontario v. Quon*, made its way to the Supreme Court, but in a June 2010 opinion, the Court explicitly chose not to decide whether an employee has a reasonable expectation of privacy in his text messages, recognizing that "[r]apid changes in the dynamics of communication and information transmission" are causing privacy expectations to evolve.[61] In an effort to avoid "error by elaborating too fully on the Fourth Amendment implications of emerging technology before its role in society has become clear," the Court instead assumed for purposes of the opinion that there was a privacy interest, but found that the search at issue was still reasonable.[62]

The Court has also cracked open the door to broader constitutional protections for downstream uses of data that impact personal privacy. Traditionally, the courts applied the Fourth Amendment only to the government's initial *collection* of data. Unless the government has collected the data in contravention of it, the Fourth Amendment typically has not been interpreted to restrict the government's *processing, use,* or *disclosure* of the collected data.[63] This limitation meant that it is in theory constitutionally permissible for one agency to release data to another for purposes of scrutiny through data mining, so long as the data was legally collected in the first instance.[64] In a 1977 case, however, the Supreme Court indicated a reluctance to completely ignore downstream use of data. Addressing government *disclosure* of personal matters, in *Whalen v. Roe* the Court stated in dicta that it was "not unaware of the threat to privacy implicit in the accumulation of vast amounts of personal information in computerized data banks or other massive government files."[65] Based on that case, several lower courts have explicitly recognized a constitutional interest in protection against government disclosure of personal information, although they have applied a less rigorous level of scrutiny than is used for fundamental constitutional rights.[66] The courts may continue to develop this doctrine.

Freedom of expression and association. As the First Amendment attests, the United States is deeply committed to preserving the right of individuals to freely express their ideas and to associate freely to share those ideas. To protect this freedom, even laws or policies that merely "chill" free expression or freedom of association may be struck down.[67] For instance, the Supreme Court has recognized First Amendment protection extends to government surveillance of political activities[68] and to the contents of expressive groups' membership lists.[69]

As of the writing of this report, no federal cases have directly addressed whether a government data mining program violates First Amendment rights to freedom of speech or association, but we believe that these rights are implicated by data mining. By enabling the government to learn more about individuals from the data they generate (data that may previously have been too diffuse or voluminous to use), data mining can chill individuals' freedom of association and expression. For example, a data mining program used to help assess the security risk of air passengers based on group affiliations and contacts

may cause people to limit their affiliations and associations to "safe" organizations and individuals. The risk of this result is far from hypothetical: reporters recently exposed overzealous and inappropriate counterterrorism investigations by the Maryland State Police. The state police chief acknowledged that his force entered the personal information of 53 nonviolent antiwar and anti-death penalty activists into federal databases that track terrorism suspects.[70]

Government accountability and due process. Equally central to the concept of a free society is the principle that laws—rather than people—govern us. We submit to society's laws knowing the authorities must do the same. Through representatives, the public enacts rules and procedures that dictate when the government can deprive any individual of life, liberty, or property. The Fifth and Fourteenth Amendments' specific guarantee of "due process" is one aspect of this right. Under well-established case law interpreting the Due Process Clause, individuals may be entitled to basic process or to more extensive procedures, depending on the government activity and the weightiness of the interests at stake.[71]

More broadly, due process values require that government remain accountable to the governed through procedural rules such as open government or "sunshine" laws, notice requirements, and regular elections. Of course, central to this principle is the public's ability to know if the government is adhering to its own rules and how it reaches its decisions.

Data mining by the government, if unregulated, could undermine these values. The public uproar and congressional reaction to the "Total Information Awareness" and air passenger screening programs (discussed above) demonstrate our collective desire to make data mining publicly accountable and procedurally limited. As with any government use of personal information, procedures must be in place to limit error, misuse, or abuse of the information. Those with access to the data must be accountable to the public, through oversight and public notice, wherever possible.

Equal protection and anti-discrimination. The Supreme Court has recognized that "the Constitution prohibits selective enforcement of the law based on considerations such as race."[72] That Court and lower federal courts have held that race, religion, and other protected bases may not be used to determine whom to subject to a search.[73] Those principles have been codified by the Civil Rights Division of the United Stated Department of Justice, which has issued guidance to prevent unconstitutional searches by federal law enforcement agencies.[74] That guidance applies only to law enforcement and does not apply to U.S. intelligence activities, nor does it "necessarily apply to classifications based on alienage."[75] Not surprisingly, "officials involved in homeland security may take into account specific, credible information about the descriptive characteristics of persons who are affiliated with identified organizations that are actively engaged in threatening the national security."[76] What is clearly prohibited by equal protection principles, however, is a policy or activity that has a discriminatory effect and is motivated by a discriminatory purpose.[77] Indeed, facts identified as part of a *particular*, intelligence-based fact pattern might be relevant for data mining purposes—for instance, a person's country of origin might be taken into account if relevant when considered in conjunction with the countries the individual recently has visited or whether the individual has specific ties to a radical group. However, using race, nationality, or religion alone as the basis for data mining is wholly inconsistent with equal protection principles, as was exemplified in the immediate aftermath of the 2009 Christmas Day bombing attempt. There, the federal government initially subjected persons from fourteen specific countries to enhanced screenings for air travel, but soon after it abandoned that criticized approach in favor of more rigorous screenings based on specific threat information.[78]

While we would not expect any government agency to use data mining to deliberately discriminate against or target a minority group, data mining can still have a disproportionate effect on certain groups—leading to harmful stigma and discriminatory effects.[79] Even more insidious, data mining could be used as a tool for racial profiling if investigators choose to use race or religion as a variable in attempting to create a predictive pattern—rather than simply as the known attribute of a particular suspect. For example, a data mining program searching air passenger lists for potential security risks would be designed to look for passengers matching the profile of known terrorists. Given the ethnic and religious makeup of the 9/11 perpetrators and other recent terrorists, the program might "flag" a high proportion of middle-eastern, Muslim men. However, would-be terrorists can come from any racial or religious groups or countries of origin, and thus such racial profiling would not only unfairly target certain minorities, but would also undermine the effectiveness of programs. If airline searches were based on ethnic or religious makeup, this would be both under- and over-inclusive of the actual pool of would-be terrorists. Although other variables that may be included as part of a data mining algorithm—such as a passenger travelling on a one-way ticket or carrying a large quantity of cash—may similarly generate under- or over-inclusive lists, we must be especially careful in the case of racial, ethnic, and religious classifications. The Equal Protection Clause demands that the government may only rely on racial classifications in very limited circumstances.[80]

In addition, the potential for false identification through data mining is increased in the case of terrorist activity, as explained above, and consequently, the discriminatory effect of any racial classifications would be amplified. Finally, as a practical matter, racial profiling reduces individuals' trust in the government. Innocent individuals marginalized due to racial profiling may be far less likely to participate in public affairs, or to cooperate with the government to combat true threats to national security in the future.

Government activities through contractors or private parties. The government cannot avoid its constitutional obligations by working through private contractors. Each of these constitutional values is implicated no less when, in lieu of government-conducted data mining activities, private contractors conduct those activities *on behalf of* the government. Likewise, when privately conducted data mining activities are made available to the government, the government's use of those activities also raises constitutional concerns. The government should not be permitted to avoid its constitutional and legal obligations by engaging or relying on other parties to perform prohibited activities.[81] Therefore in either scenario, data mining activities should be treated as government-run for purposes of these guidelines.

2. Laws, Statutes, and Regulations

Some of the gaps in constitutional protections, particularly those related to privacy, have been filled by statute. However, because many of those laws have broad exceptions, and because different laws apply to different sectors, that legislation provides no more than a patchwork of protections. What is more, only one federal statute explicitly contemplates data mining as it relates to privacy, and none provide direct guidance on implementing these activities.

Data collection. The Electronic Communications Privacy Act of 1986 ("ECPA")[82] regulates electronic surveillance for law enforcement purposes. The Foreign Intelligence Surveillance Act of 1978 ("FISA")[83] regulates electronic surveillance for foreign intelligence purposes. While both regulate government collection of data, the government showing required for some of this

surveillance is quite low. For example, under ECPA, the government can obtain most emails directly from internet service providers with a mere subpoena, and without having to demonstrate probable cause. Moreover, these laws impose little or no limitation on how the information may be used by the government once it is obtained.[84]

Data processing and disclosure. A handful of statutes primarily regulate government treatment of individuals' data. The Privacy Act of 1974[85] regulates the federal government's use, retention, and disclosure of individuals' personal data, but its rules are subject to many exceptions and limiting interpretations. In particular, records compiled by law enforcement agencies for purposes of criminal investigation may be exempt from the Act's requirements.[86] Various data thus may not receive privacy protections in the data mining context,[87] and because it only applies to federal entities, the Privacy Act does not impose any prohibitions on private companies disclosing data to the government.

The E-Government Act of 2002 requires federal agencies to conduct and publish privacy impact assessments ("PIAs") on data "collection[s]" conducted through information technology, and it requires federal agencies to post privacy policies on their websites.[88] The Act requires PIAs to include a description of the information collected, why it is being collected, the agency's intended use of the information, with whom the information will be shared, what security protections will be used for the information, and certain information related to the Privacy Act.[89] In addition, PIAs must describe the notice or opportunities for consent individuals will receive about the collection and sharing of their information.[90] These requirements apply to data "collections" by the government, and therefore searches of data already collected by third parties likely are not subject to the Act's obligations.[91] In addition, certain information technologies used for national security are excepted from the PIA requirement,[92] and publication of a PIA may be waived for security reasons.[93] However, if conducted properly and thoroughly, PIAs have substantial potential for adding transparency to many agencies' data mining activities, and some agencies to date have issued particularly useful PIAs. The DHS Privacy Office, for instance, recently released a PIA on a demonstration of technologies for possible use in its EINSTEIN program on network security and cyber threats.[94] This PIA explained the scope of internet traffic to be monitored, the type of review being conducted, and the extent to which personally identifiable information might be reviewed or stored.

Under the Federal Agency Data Mining Reporting Act of 2007 ("Data Mining Act"),[95] federal agencies must report to Congress annually on the data mining activities they are using or developing, and those reports must be made available to the public. However, classified and sensitive material must be attached as a non-public annex and made available, as appropriate, only to certain congressional committees consistent with the National Security Act of 1947. "Data mining" as defined for purposes of the Act only includes predictive, pattern-based analyses and does not include subject-based searches.

The report must include, among other information, a description of the activity, its goals, the technology used, and the basis for determining whether a particular pattern or anomaly is indicative of terrorist or criminal activity; the data sources used; assessments of the program's impact on privacy, including the actions that will be taken as a result of the implementation of the activity; and a description of the agency's privacy protection and data accuracy policies. DHS's Privacy Office released reports on its data mining activities for 2006, 2007, 2008, and 2009, and the Office of the Director of National Intelligence ("ODNI") released reports addressing its 2007, 2008, and 2009 activities.[96] The ODNI report for 2009 indicates that many of

its counterterrorism activities use "link analysis tools" that are subject based, and therefore, because that type of analysis is not included in the Act's definition of data mining, those activities are not reflected in the report.[97]

Both the E-Government Act and the Data Mining Act make some strides toward greater transparency of government programs, but they alone do not provide the affirmative privacy protections that data mining activities require.

Use, storage, retention, and disclosure are all implicated by recent government efforts to increase data sharing in the wake of the terrorist attacks of 9/11. Those efforts include creation of the Information Sharing Environment ("ISE") to facilitate sharing of terrorism information among federal agencies,[98] and adoption of federal steps to support the development of state and local information fusion centers.[99]

Assorted sector-specific laws protect certain types of information. Some apply to particular government agencies, while others regulate private entities' disclosure of customers' personal information to the government. For instance, certain statutory provisions generally prevent relevant agencies from disclosing Social Security, tax, and census records data to other agencies or the public, except in certain circumstances defined by law.[100] Laws such as the Right to Financial Privacy Act of 1978,[101] the Cable Act,[102] the Video Privacy Protection Act,[103] the Family Educational Rights and Privacy Act of 1974,[104] the Health Insurance Portability and Accountability Act of 1996 Privacy Rule,[105] the Communications Act,[106] and the Fair Credit Reporting Act[107] limit government access to third-party entities' relevant data, but the showing required by the government in order to gain access widely varies among them.

Administrative policies and procedures play a role in guiding agencies' conduct of activities that may implicate data mining. The most prominent of these is Executive Order 12333, which regulates the conduct of U.S. intelligence activities. Each intelligence agency promulgates procedures to implement the executive order, including data handling procedures. However, their implications for data mining remain unclear.[108]

Finally, the 9/11 Commission recommended a Privacy and Civil Liberties Oversight Board to help monitor the government's protection of civil liberties in the counterterrorism context. Congress established the Board in 2004, but the original legislation provided the Board with only limited authority, and it accomplished very little. In 2007, Congress enacted legislation strengthening the Board's ability to provide oversight by removing it from within the Executive Office of the President and making it an independent agency, vesting it with subpoena power, and instituting a bipartisan membership requirement.[109] These changes gave the Board enhanced authority and greater potential as a true oversight mechanism.* However, when the original Board lapsed, the newly authorized independent Board never came into existence. President Bush did not nominate a full slate of five members for the Board, and none of his nominees were confirmed. And as of the date of this report, President Obama has not nominated a single member for the Board. Thus, no Board has existed since early 2008, and the 2007 legislation remains an unfulfilled promise.[110]

---

* The Constitution Project advocated for the legislative changes that conferred greater authority on the Board and has long sought to make the Board an effective oversight mechanism. *See, e.g.,* Constitution Project, Press Release, *Constitution Project Joins Broad Coalition Urging President Obama to Nominate Members to Privacy and Civil Liberties Oversight Board*, Mar. 2, 2010; Ginny Sloan, *The Privacy and Civil Liberties Oversight Board: In Need of Attention*, The Huffington Post, Oct. 15, 2009, http://www.huffingtonpost.com/ginny-sloan/the-privacy-and-civil-lib_b_322527.html.

Today's data mining involves greater data sharing and more powerful technologies than ever before. Those enhancements must be balanced by increased oversight and more responsive privacy and civil liberties protections.[111] Agencies require a clear and uniform articulation of the principles they should follow in order to implement both lawful and effective data mining activities that serve our essential national values of security and individual rights.

3. Other Proposals

Government data mining has received substantial attention from nonprofit institutions, academics, and government groups alike, many of which have proposed guidelines for government entities engaged in data mining for national security or counterterrorism purposes.[112] These have included the Markle Foundation Task Force on National Security in the Information Age ("Markle Foundation"), the National Research Council, the Department of Defense Technology and Privacy Advisory Committee ("TAPAC"), the Cantigny Conference on Counterterrorism Technology and Privacy, and Professor Fred H. Cate.

Existing proposals are generally procedural in their approach to protecting civil liberties. In most cases, they track the Fair Information Practice Principles ("FIPPs"),[113] and at least one explicitly recommends that the Privacy Act be adapted to apply to government use of third-party data.[114] Key themes include authority and oversight, data characteristics, and data subjects' participation. Authority and oversight recommendations highlight authorization by proper officials; limiting access to those who are authorized, and educating and training those with authority; and auditing and ensuring meaningful program oversight. Recommendations focused on the data itself suggest a need for data minimization and anonymization, as well as mechanisms for ensuring data quality, accuracy, and security. Many explicitly outline means to empower individuals to protect their privacy interests, and they underscore transparency and public participation, as well as redress mechanisms for those aggrieved by a data mining activity.

One significant substantive consideration some groups have highlighted is the need for different procedures based on the type of data to be searched or the uses of the search. For instance, the Markle Foundation, the National Research Council, and TAPAC all suggest that certain factors, such as how closely linked data are to an individual's identity or whether the results of a data mining will be used for particularly sensitive purposes, should require the government to meet correspondingly higher standards when mining the data.[115]

Three other recommendations were made by multiple commentators. First, some emphasized the need for careful attention at the design and planning phase, to ensure that programs are effective, that they are properly aligned with their intended purposes, and that they are in accord with data mining guidelines.[116] Second, nearly every set of guidelines stressed the need for routine evaluation of program lawfulness and efficacy. Finally, several recommended that a government-wide set of policies is needed to consistently regulate government data mining programs.[117]

Our principles build on this well-laid foundation, and incorporate the most significant of these recommendations into our proposed framework to guide government entities and private government contractors engaged in data mining activities. In addition, our principles extend beyond the prior reports' largely procedural approach, and also recommend that agencies adopt substantive safeguards to protect constitutional rights and values.

II. Principles for Government Data Mining Programs

Below are our recommended principles for creation and management of government data mining programs. Rooted in constitutional values and the Fair Information Practice Principles,[118] these common sense principles, if applied and followed, can facilitate the government's use of data mining techniques without sacrificing constitutional rights and values.

In the case of classified programs or programs mining classified data, we recognize that the principles requiring public notice or participation cannot be implemented. To provide individual protection in those circumstances, we recommend substitute measures that utilize congressional oversight or independent review.*

A. Development of Data Mining Programs

- ***Prior to acquisition, clearly articulate, in writing, the purpose(s) of data acquisition and the intended use(s) of that data.***

As an initial step, any new acquisition of personally identifiable information,** or modification of or addition to existing data, by a government entity should be preceded by a clear articulation of the purpose(s) of the acquisition*** and the intended use(s) of the data, where not otherwise regulated or circumscribed by existing law. This principle would apply to the *initial* acquisition of data by a government entity, but not to simple re-use or intra-agency sharing.

While we recommend this step for any acquisition of personally identifiable information, it is particularly important where the intended uses include data mining. Enabling internal decisionmakers (and, when feasible, the public) to evaluate the legitimacy and impact of the proposed collection or acquisition is important for two reasons: First, the very act of data collection or acquisition can intrude on individual privacy and thereby cause harm to constitutional rights and values, and so should be carefully and publicly considered beforehand. Second, the statement of purposes and intended uses will guide the creation of any data mining plans and allow meaningful review of the system by the public, Congress, and any other oversight entity.

In the situations in which the act of collection or acquisition cannot be made public, the written statement of purposes and intended uses can aid internal or congressional oversight.

- ***Create a comprehensive data mining Plan covering data sources, data acquisition, system design and capabilities, and intended uses.***

Fundamental to preserving accountability and limiting "scope creep" in data mining operations is creation of a comprehensive data mining plan ("Plan"). As discussed further below, the Plan enables the agency and outside reviewers to identify and evaluate the goals of the data mining program, its likelihood of success at achieving those goals, its economic costs, and the potential impact on civil liberties and constitutional values. The Plan should address:

---

* This Committee has separately addressed the problem of over-classification of information. *See* Constitution Project, *Reining in Excessive Secrecy: Recommendations for Reform of the Classification and Controlled Unclassified Information Systems* (July 2009), http://www.constitutionproject.org/manage/file/178.pdf. Because that activity presents serious constitutional concerns, we emphasize here that these principles should not create incentives to increase over-classification.

** See page 8 for a discussion of the definition of "personally identifiable information."

*** As noted above, we distinguish here between the "collection" and the "acquisition" of data. As used in this report, "collection" generally refers to the process by which data are obtained directly from an individual. "Acquisition" refers to both direct collection and obtaining data from a third party or another government entity.

- *Data sources* from which the program will draw data, including both direct collection and acquisition of data from outside entities or other programs;
- *Data acquisition*—how and under what conditions data will be acquired or accessed;
- *System design and capabilities*, including management of data integrity and minimization (as discussed below); and
- *Intended uses* of the program.

- ***Perform an internal evaluation of the program's expected effectiveness, costs, benefits, compliance with existing law, and impact on civil liberties and constitutional values.***

Using the draft Plan, the agency should evaluate the proposed program. This evaluation principle would also apply to any addition or material modification to existing programs. At a minimum, the evaluation should include:

- *Likely effectiveness* of the data mining program in identifying the desired targets;
- *Costs and benefits* of the program, including the economic benefits of automated data mining systems over other methods of accomplishing the same goals;
- C*ompliance with existing law*, including constitutional limits, statutory requirements, and any executive regulations or departmental standards that govern the processing of personal data (such as the Department of Defense's Policy on Safeguarding Personally Identifiable Information and Breach Notification[119]); and
- *Impact on civil liberties and constitutional values*, which may be accomplished through existing procedures, such as through preparation of a Privacy Impact Assessment. The DHS Privacy Office has developed a "Privacy Impact Assessment" standard which may be useful in this context.[120] As discussed above, such an assessment would contain detailed information about the program, including what information will be collected, why it will be collected, the agency's intended use of the information, with whom the information will be shared, the information-security protections to be used, and a description of the notice or opportunities for consent individuals will receive about the collection and sharing of their information. Where the data mining program will be newly developed, the assessment may be required by statute. Even where an impact assessment is not statutorily mandated, agencies should nonetheless perform an assessment in order to satisfy these principles and protect individual liberty interests.

Given skepticism about the usefulness of data mining in terrorism prevention, due to the particular difficulties of predicting patterns of terrorist activity and the ability of criminals and terrorists to revise or alter their patterns (see supra), the first two steps are particularly critical in such cases. Data mining operations aimed at detecting terrorism should particularly quantify the costs and benefits and the likelihood that the program can advance counterterrorism goals.

- ***Submit the Plan and internal evaluation for review and comment in the Federal Register, as feasible; ensure congressional and high-level administrative review for non-public aspects.***

Public oversight and accountability are fundamental to preventing misuse or abuse of personal data. To the extent publication does not conflict with operational goals of the Plan and other security requirements, the Plan should be made available for public review and comment through the Federal Register or other applicable means. Public review will compel the agency to

justify the Plan more thoroughly than may occur with internal review only, and will increase the likelihood that operation of the data mining program will conform to the Plan.

When a Plan or aspects of it cannot be made public due to security concerns or legal barriers, review by other government officials may substitute. To maximize the independence of such a review, we recommend review by Members of Congress through all relevant committees with jurisdiction. Certain Members and staff regularly review highly sensitive intelligence information, and thus have processes in place to protect non-public information, if necessary. In rare cases where intra-branch review is impractical, high-level administrative review can substitute. In such cases, the reviewers should have sufficient institutional independence and authority to exercise effective oversight.

The Privacy and Civil Liberties Oversight Board described above, which was revamped in 2007 through new legislation that granted the Board greater authority and independence, could serve ably in this capacity once the President and the Senate nominate and confirm its members. *See supra* Part I.C.2. The Board was created by Congress to review the civil liberties implications of national security programs. We urge the Administration and the Senate to re-constitute this Board and vest it with authority as described in these principles. Active oversight by an empowered board is necessary to ensure successful application of these recommendations.

- ***Respond to and incorporate administrative, public, and congressional commentary on the Plan***

Following a period of comment and review, the agency should address reviewers' concerns, update the Plan accordingly, and seek additional review as necessary. The final Plan, as revised, can then serve as a guide for operation of the data mining program and as a benchmark for evaluation.

B. Operation of Data Mining Programs

After the development and review of a data mining Plan, an agency can begin building and operating a data mining program pursuant to the Plan. Our principles for operation of a data mining program fall into several categories, based on the Fair Information Practice Principles familiar to many agencies.[121]

1. Transparency and Notice

- ***To the greatest extent feasible and consistent with national security concerns, provide notice to an individual of any specific government action or classification of that individual pursuant to data mining. When immediate notice is not possible, establish a framework for delayed notification to the extent feasible.***

In the majority of cases, the fact that government agencies are using personal data for data mining does not require secrecy.[122] In such cases, the agency obtaining data should seek to provide notice to individuals of any specific action or classification that is based on information gleaned from data mining operations. For actions or classifications that are made as a result of data mining, for instance where individuals are flagged for auditing for tax or Medicare fraud, individuals could be notified once the decision is made, and should then be informed about the individual's ability to challenge the action and/or update his or her information. In circumstances where immediate notification is not possible due to national security concerns, but the information is not classified, a framework should be created to provide delayed notification of a classification. Indefinite, secret classification of individuals should be avoided to the greatest extent possible.

To avoid unnecessary burdens on the government, notice should be undertaken only when an individual has been subject to a specific action or classification and it is feasible to locate and notify individuals. Electronic notice, such as via email, would be appropriate where available. While notification imposes costs on the agencies undertaking data mining, we believe that such upfront costs may reduce subsequent litigation expenses and should strengthen public trust in data mining operations.

2. Accountability, Oversight, and Redress

- ***Create administrative standards and procedures governing acquisition, use, and sharing of information for data mining.***

Federal agencies should collaborate to adopt government-wide, written, defined standards for the acquisition, sharing, and use of data. Those standards must clearly articulate what types of data may be shared, with whom, and under what circumstances. Written standards will guide all employees, in identical terms, about their obligations in handling and transmitting data, and thus will uniformly build privacy safeguards into all agencies' broader data mining programs. Where coordinated, multi-agency standards cannot be created or where circumstances require an agency to act independently, it should create its own standards.[123]

- ***Establish and enforce penalties for misuse and abuse of data mining programs by operators or others.***

As explained above, standards should be binding upon all federal employees and those operating under government contract. Operators who do not follow the standards, or who otherwise misuse or abuse personal data or data mining systems, should be subject to civil or criminal penalties. Reports of misuse of the 2008 presidential candidates' passport data, mentioned above, demonstrate that misuse or abuse by operators is difficult to prevent through security measures and training alone. Penalties will help ensure operators will vigilantly follow procedures. To ensure oversight, employee compliance also should be reviewed through regular audits, discussed below.

- ***Establish a system of appeal and redress for individuals misclassified or harmed.***

Adequate protection of the rights and liberties of individuals must include some mechanism for redressing the harm caused by errors, abuse, or misuse in data mining operations. As discussed above, when feasible, individuals should be provided notice whenever they have been subject to any specific action or classification based on information gleaned from data mining operations. In addition, individuals should be afforded the opportunity to challenge burdens imposed upon them as a result of data mining that result in harm or the denial of any right, privilege, or benefit—for instance, on the grounds that the data used was flawed or out-of-date. Individuals should be given the opportunity to challenge their classification and any resulting government action through an administrative proceeding, in which like claims would be grouped by the agency and reviewed in a single proceeding. Administrative decisions should be subject to appeal in court, on grounds defined by Congress.[124]

Where the involvement of classified information or public safety, national security, or law enforcement concerns preclude an individual's ability to bring a challenge, an independent arbiter should have the opportunity to review (on a collective basis and to the extent feasible) the classification or action through which the individual might be harmed.

The Data Privacy and Integrity Advisory Committee to the Secretary and the Chief Privacy Officer of DHS recently made similar recommendations to DHS to assist it in developing and enhancing its redress processes.[125] In addition, that report emphasized

the importance of assigning accountability for the development and management of redress processes, monitoring the process for transparency and fairness, and ensuring corrections to data made as a result of the process are propagated throughout the relevant data systems. We recommend that DHS adopt these proposals as it enhances its redress programs.

- ***Require private companies to have data correction procedures in place as a requirement of contracting with a government agency.***

The data used by government agencies in their data mining activities are often acquired from third-party sources such as private companies. Because private entities acting on the behalf of the government should be held to the same standards as the government agencies utilizing their services, and because the government should not be permitted to rely on private contractors' activities to avoid its constitutional obligations, private entities providing data to the government or performing data mining functions on its behalf should be required to abide by these principles. To provide accountability, private parties should be required, as a specific condition to contracting with the government, to have policies and procedures in place to ensure the integrity of the data they provide. Ensuring the quality of the data will strengthen the effectiveness of data mining activities and reduce potential error.

- ***To the extent feasible, permit individuals to review and correct data.***

Where feasible, individuals should have the opportunity to review the data the government has on file about them, and they should be able to correct any misinformation. This principle applies equally to any entity processing data on behalf of the government, and as with other principles should be incorporated through the contract or award. Because individuals have the most incentive to protect their personal information, this review—when combined with other practices recommended here—will provide a sound means of ensuring the data are accurate, reliable, timely, and complete. It will also preempt potential harm that may result from the use of inaccurate or unreliable data. We recommend federal agencies formulate a centralized means and procedure for individuals to request access to data to avoid the difficulty and delay of multiple Freedom of Information Act ("FOIA") requests or other multiple filings. Such procedures could be modeled on those now available to consumers for correction of credit reports. This principle is also compatible with a similar recommendation made by the National Broadband Plan, delivered by the Federal Communications Commission to Congress on March 16, 2010.[126] In addition, to accommodate any requests made through the FOIA process, agencies should work to respond promptly and reduce their backlogs by implementing the President's 2009 FOIA Memorandum.[127]

Any errors noted by individuals relating to their personal data should be promptly corrected, and the corrected data promulgated through all data systems.

- ***Conduct and publish the results of regular audits, and report regularly to Congress.***

Inspector General investigations or another oversight mechanism should be used to regularly audit government data mining activities to ensure that each agency complies with its Plan and with binding administrative standards. Through independent auditing, agencies may be able to subject to objective scrutiny even national security activities that are too sensitive to be made public. Regular reports on those audits should be made to committees of jurisdiction in Congress and should include reporting on classified activities.

The Data Mining Reporting Act already requires agencies to report annually to Congress on relevant information concerning their data mining activities, as defined in the Act. *See supra* Part I.C.2. Although the scope of the required reporting is fairly comprehensive, the Act's definition of data mining is relatively narrow. Under the Act, reports must include descriptions of the program and its goals, the technology used and the basis for determining whether a particular pattern is indicative of terrorist or criminal activity, and the data sources used; assessments of the efficacy of the data mining activity in providing accurate information, as well as of the impact of the implementation of the activity on the privacy and civil liberties of individuals; and a list and discussion of governing laws and policies.[128] The Act also requires agencies to include classified information as an annex to a report.[129] However, agencies have noted that the Act's narrow definition of data mining does not include "link analysis tools" because these are not "pattern-based," and consequently agencies have not reported on such programs.[130]

Accordingly, we recommend expanding the definition of data mining programs under the Act to reflect the broader definition applied in this report, and thereby require reporting on a greater number of programs. In addition, the categories of information required for all such reports should be expanded to include rates of false positives and false negatives generated by the data mining activity. All such information should be reviewed through the audit process and included in the annual reports to Congress.

Misuse, abuse, or systematic errors uncovered pursuant to the principles above should trigger an immediate review, regardless of the audit schedule.

3. Authority and Choice

- ***Coordinate uses, best practices, regulations and protections among agencies.***

Congress should enact legislation requiring agencies to define and adopt consistent data practices. Even if Congress does not pass such legislation, the President should issue an executive order directing a specific agency to coordinate the various agencies and set forth best practices and procedures to ensure consistent practices. Coordination will help ensure that all agencies are communicating in the same terms and understand the same restrictions with respect to data activities, and therefore will facilitate authorized sharing of data across federal government bodies. Uniform uses, best practices, regulations, and protections will also help ensure that data mining operations will meet substantive standards.

- ***Establish approval procedure for data acquisition and actions taken pursuant to data mining with decisionmakers at highest possible level and outside of the program's operational structure.***

An approval procedure will help reduce excessive subjectivity in data acquisition, use, and other data mining activities and will reduce the risk that program operators will act contrary to procedure or otherwise improperly. High-level approval will also help create internal support for compliance by establishing clear expectations for the treatment of personal data and data mining program operations. Approval of a neutral decisionmaker outside the agency or at the least above the operational structure will bring additional objectivity to the process by de-linking approval from operational politics or improper criteria. As mentioned above, a re-constituted Privacy and Civil Liberties Oversight Board could be ideally suited for this role.

4. Data Integrity and Security

- ***Incorporate technical and administrative measures to limit access to or availability of personal data, particularly sensitive or personally identifiable data.***

Administrative and technical measures should be employed together to reduce the potential for abuse or misuse of personal data. As in the private sector, government entities should seek out and adopt "best practices" to maintain security. As with all the above principles, contractors working on behalf of the government should be bound to incorporate measures at least as strong as those described here.

With respect to technical protection measures, we recommend at a minimum implementing network access limits, using strong encryption for transmission and storage of data, using audit trails, and automated logging of system access. These measures would rely upon inexpensive, widely-available technologies that can substantially reduce the possibility of accidental or malicious misuse of data. We also recommend administrative measures, such as physically restricting access to data mining and data storage systems, requiring operators to undergo training (see below) and security screening, and independent audit of system access logs. These administrative and technical measures will ensure preservation of an audit trail for use in regular audits and to investigate allegations of misuse or abuse.

- ***Evaluate and improve data security, integrity, accuracy, and timeliness on a regular basis, including the use of audit trails.***

Operators should conduct regular evaluations of the effectiveness and adequacy of data integrity and security measures, including the use of audit trails. To the extent that evaluations reveal weaknesses in existing systems or the development of new tools or processes to improve data integrity and security, the data mining system should be updated.

- ***Conduct training and evaluation for employees with access to personal data or data mining systems.***

Government employees and contractors should undergo thorough training prior to gaining access to personal data or data mining systems. Operators of data mining systems should be evaluated for compliance with procedures.

- ***Take all reasonable steps to rectify and minimize harm from data breach, including prompt notification of affected individuals.***

When a data breach occurs, the agency should immediately implement procedures designed to minimize disclosure and harm, such as shutting down network access, investigating involved operators and other personnel, and working with law enforcement to recover missing data or equipment. To the extent feasible, individuals at risk of harm from disclosed data should be promptly notified and told of steps they can take to reduce harm, such as identity theft protections.

5. Data Appropriateness/Minimization

By only including the appropriate data in collection or acquisition efforts, storage, and sharing, agencies can substantially reduce the potential impact of data mining operations on civil liberties and constitutional values.

- ***Ensure that acquisition only includes data relevant to the purpose of the program and minimize the extent to which databases are aggregate.***

Overinclusive data collection or acquisition heightens the risk of scope creep and data misuse. Even if those data are not improperly mined or misused, the maintenance of more personal data than necessary heightens the risk to individuals in the event of a data breach. Therefore, agencies should carefully design programs to minimize acquisition of extraneous data unrelated to the program's purpose. The scope of acquisitions and the information to be included should be specifically approved by the proper administrative authority and within the parameters of the Plan. In addition, unless necessary for those objectives, operators should avoid combining datasets or databases from different programs. For instance, if an agency contracts with credit reporting agencies to obtain personally identifiable mortgage information for a data mining program aimed at uncovering mortgage fraud, it may find that it has access to other credit history information—including former addresses, other lines of credit, FICO scores, and demographic information. Unless such data are required for the mortgage fraud program, this other data should not be acquired.

- ***Use deidentified or aggregate data formats or other techniques with respect to personal information to minimize potential harm.***

Personal information (i.e., information sufficient to uniquely identify a person or personal device) should be used for data mining programs only when necessary. In many cases, program goals require only aggregate data. In other cases, unique data may be needed, but the names and other identifying information about individuals may be stripped away or replaced by anonymous unique identifiers. Such efforts must ensure that data are truly anonymized, and that sufficient details do not remain to permit identification of a particular individual. Even if a program requires personally identifiable data, the agency and reviewing parties should evaluate whether any request for sharing or other dissemination of the data can be accomplished using aggregate or deidentified data.

- ***Continue safeguards for personal data through technical measures, rule, or contract as data moves "downstream."***

When an agency disseminates data to another agency or third party, including other federal agencies, state or local government agencies, or private contractors, it should use all available means to limit the potential misuse or downstream disclosure of the data. It can employ technical measures, such as encryption, to maintain control over the data. Binding federal data mining standards, appropriate legislation, and contractual terms can provide legal guarantees, as well. In addition, each of these recommendations, including the notification principles, should apply to the receiving agency or third party.

- ***Set limited retention periods and ensure complete destruction of expired data.***

Long-term retention increases the threat of breach, misuse, or abuse, and data stored for long periods becomes out-of-date and unreliable. Retention periods should be set at the minimum duration required to accomplish the operational purposes. At the expiration of the retention period (or before), data should be deleted from all storage locations, logs, buffers, and other locations. Existing regulations or internal standards relating to destruction of classified data may be incorporated in this context. Ensuring these measures are followed should be a specific focus of the audits discussed in the above recommendations.

III. Action Plan

Recommendations

The above principles should help guide the federal government as it develops and administers data mining programs. In order to translate these principles into action, we recommend that the government take several steps.

1. Action by the President and Congress

First, as indicated above, the President and the Senate should give effect to Congress's efforts to reform and strengthen the Privacy and Civil Liberties Oversight Board by nominating and confirming its members. A reconstituted, empowered Board would have the independence and authority to effectively review data mining Plans—particularly highly sensitive Plans—where Congress cannot do so, and to oversee their execution, for instance by reviewing and approving data acquisitions and data mining activities.

2. Congressional Action

Congress should revisit existing legislation, and consider adopting new legislation, to effectuate the principles. For example:

- Congress should amend the Data Mining Reporting Act to expand the definition of data mining programs to include link analysis tools, and to reflect the broad definition applied in this report so that reporting would be required for a greater number of programs. It also should revise the Act to require agencies to report on rates of false positives and false negatives generated by each data mining activity.
- Congress should consider amending the Privacy Act of 1974—for instance to expand its scope to reach private companies disclosing data to the government—and it should consider requiring agencies to adopt regulations to implement the principles.
- In their oversight capacity, congressional committees of jurisdiction should review each data mining Plan that cannot be subjected to public review due to national security concerns.

3. Executive Orders and Action

The President should issue executive orders, or undertake other action, to facilitate administrative agencies' implementation of the principles. For example:

- Most importantly, the President should issue an executive order directing a specific agency to define consistent data mining practices to be adopted by the various agencies, set forth best practices and procedures, and coordinate implementation by the agencies. The best practices would include limiting collection to include only data relevant to the purpose of the program. The order should also require regular audits of government data mining activities.
- The President should adopt an executive order that requires agencies to develop a meaningful system of appeal and redress for individuals who are misclassified or harmed by data mining programs.

4. Agency Action

Because many of the principles necessitate agency-level action, administrative agencies have especially great responsibility to execute the principles. Where their authority allows, agencies should adopt regulations to implement the principles. As a secondary option, they should rely on less formal mechanisms such as administrative guidelines and internal policies and procedures to implement the principles. These steps should include the following:

- Before implementing a new data mining program, agencies should perform an internal evaluation of the program's expected effectiveness, costs, benefits, and compliance with existing law.
- In order to ensure data integrity, agencies should require through contract that each contractor adopt policies and procedures to effectuate the principles.
- Agencies also should adopt a centralized procedure for individuals to request access to and correct their data held by the agency. In anticipation that some individuals may make requests through FOIA, or that the agency may adopt a similar procedure to handle requests, agencies should immediately implement President Obama's January 2009 guidance to promptly review and resolve all FOIA requests.
- Agencies should be responsible for adopting administrative and technical measures to reduce the potential for abuse and misuse of personal data, including the specific technological practices, and the employee screening and training measures, described above.

IV. Conclusion

Data mining capabilities, like many other rapidly developing technological tools, may offer new abilities to government to protect us from fraud, abuse, and national security threats. But as greater amounts of personally identifiable information are captured by increasingly powerful technological tools, those tools can also threaten our most cherished constitutional rights and liberties. Thus, as technology develops, it is critical that constitutional safeguards continue to apply to protect individuals' privacy, freedom of expression and association, due process, and equal protection interests. The Constitution Project's Liberty and Security Committee offers the principles outlined above to help ensure that government data mining programs are developed in a manner that preserves and enforces these safeguards. We urge Congress and the executive branch to incorporate these critical protections for individual rights into all government data mining activities.

1. *Federal Support for Homeland Security Information Sharing: Role of the Information Sharing Program Manager: Hearing Before the Subcomm. on Intelligence, Information Sharing and Terrorism Risk Assessment of the House Comm. on Homeland Security,* 109th Cong. 23 (2005) (statement of Lee Hamilton).

2. *Passport files of candidates breached,* Associated Press (Mar. 21, 2008), http://www.msnbc.msn.com/id/23736254/.

3. Google Corporate Information, Company Overview, http://www.google.com/intl/en/corporate/.

4. *See* Section I.B.1, *infra.*

5. Statement of Lee Hamilton, *supra.*

6. Noah Shachtman, *Obama: Software Flaws Let Christmas Bomber Get Through*, Wired.com, Jan. 7, 2010, http://www.wired.com/dangerroom/2010/01/obama-software-flaws-let-christmas-bomber-get-through/.

7. This report does not address the collection and examination of data about non-citizens located abroad.

8. The Congressional Research Service ("CRS"), for example, defines data mining as "the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets." Jeffrey W. Seifert, *Data Mining and Homeland Security: an Overview*, CRS, at 1 (Aug. 27, 2008) ("CRS Data Mining Report") (citing Two Crows Corporation, *Introduction to Data Mining and Knowledge Discovery, Third Edition* (Potomac, MD: Two Crows Corporation, 1999) and Pieter Adriaans & Dolf Zantinge, Data Mining (New York: Addison Wesley, 1996)). *See also* The *Data Mining* Reporting Act of 2007, Pub. L. No. 110-53, 121 Stat. 266, Section 804(b)(1) (defining "data mining" as "a program involving pattern-based queries, searches, or other analyses of one or more electronic databases, where . . . the queries, searches, or other analyses are not subject-based and do not use personal identifiers of a specific individual, or inputs associated with a specific individual or group of individuals, to retrieve information from the database or databases").

9. *See* Newton N. Minow & Fred. H. Cate, Government Data Mining, at 4, http://ssrn.com/abstract=1156989, in McGraw Handbook of Homeland Security (2008); National Research Council, Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment, at 22 (National Academies Press 2008).

10. DHS Privacy Office, *Data Mining: Technology and Policy*: *2008 Report to Congress*, at 31-32 (Dec. 2008) ("Data mining uses mathematical algorithms to construct statistical models that estimate the value of an unobserved variable—for example, the probability that an individual will engage in illegal activity. Data mining is best understood as an iterative process consisting of two separate stages: machine learning, where algorithms are applied against known data; and probabilistic inference, where the models built from algorithms are applied against unknown data to make predictions.").

11. *See* Minow & Cate at 3; National Research Council at 21.

12. *See* National Research Council at 23; DHS Privacy Office at 32.

13. Office of Management and Budget, *Memorandum for the Heads of Executive Departments and Agencies: Safeguarding Against and Responding to the Breach of Personally Identifiable Information*, at 1 (May 2007); *see FTC Staff Report: Self-Regulatory Principles for Online Behavioral Advertising*, at 21 (Feb. 2009). For example, America Online (AOL) provided an accidental example of this possibility in 2006 when it temporarily released three-months-worth of search query data from over 650,000 AOL users. Although the individual users were identified only by a serial number, a few reporters and bloggers claimed that they had surmised the actual identity of a handful of the users by looking for searches with geographic terms and proper names. See, e.g., Michael Barbaro & Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. Times, Aug. 9, 2006, http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=1&scp=1&sq=aol%20queries&st=cse&oref=slogin; Ellen Nakashima, *AOL Takes Down Site With Users' Search Data*, Wash. Post, Aug. 8, 2006, http://www.washingtonpost.com/wp-dyn/content/article/2006/08/07/AR2006080701150.html.

14. *Cf. FTC Staff Report* at 25 (applying this standard in the context of behavioral advertising).

15. Jo Twist, *Law that has driven digital life*, BBC News, Apr. 18, 2005, http://news.bbc.co.uk/2/hi/science/nature/4449711.stm.

16. *See* CRS Data Mining Report at 2.

17. *See* DHS Privacy Office at 34.

18. For a more comprehensive, though somewhat dated list, see GAO, *Data Mining: Federal Efforts Cover a Wide Range of Uses*, GAO 04-548 (May 2004) ("GAO 2004 Report").

19. *Id*. at 7-12.

20. CRS Data Mining Report at 4.
21. GAO 2004 Report at 10.
22. *Id*. at 52.
23. CRS Data Mining Report at 4.
24. For more information and recommendations on fusion centers, please see the Liberty and Security Transition Coalition, *Fusion Centers and the Expansion of Domestic Surveillance, Recommendations for the Next Administration and Congress* (2009), http://2009transition.org/liberty-security/index.php?option=com_content&view=article&id=10:10-fusion-centers-and-the-expansion-of-domestic-surveilance&catid=5:secrecy-surveillance-and-privacy&Itemid=20.
25. Harley Geiger, *Fusion Centers Get New Privacy Orders Via DHS Grants*, Center for Democracy and Technology, Dec. 15, 2009, http://www.cdt.org/blogs/harley-geiger/fusion-centers-get-new-privacy-orders-dhs-grants. The grant program guidance requires certification of compliance with the Information Sharing Environment privacy guidelines, although those guidelines have been criticized for lacking specificity. *See id.* See also *infra* for a discussion of the Information Sharing Environment.
26. *See* Shane Harris, *The Watchers* (The Penguin Press 2010), for a detailed depiction of the potential uses for, and the serious consequences of, data mining in the counterterrorism context.
27. Electronic Frontier Foundation (EFF), *Report on the Investigative Data Warehouse* (Apr. 2009), http://www.eff.org/issues/foia/investigative-data-warehouse-report.
28. *Id.*
29. See TSA, Secure Flight – Frequently Asked Questions, http://www.tsa.gov/what_we_do/layers/secureflight/faqs.shtm; Michael Fabey, *TSA introducing Secure Flight program in fits and starts*, Travel Weekly (May 22, 2009), http://www.travelweekly.com/article3_ektid195004.aspx; The Identity Project, *'Secure Flight' data formats added to the AIRIMP*, Papers, Please! Blog (May 1, 2009), http://www.papersplease.org/wp/2009/05/01/secure-flight-data-formats-added-to-the-airimp/.
30. Minow & Cate at 19.
31. Shachtman, *Obama: Software Flaws Let Christmas Bomber Get Through, supra*.
32. *Id*.
33. CRS Data Mining Report at 27.
34. *See* National Research Council at 39.
35. Shachtman, *Obama: Software Flaws Let Christmas Bomber Get Through, supra*.
36. *See* Eric Lichtblau, *Justice Dept. Finds Flaws in F.B.I. Terror List,* N.Y. Times, May 6, 2009, http://www.nytimes.com/2009/05/07/us/07terror.html.
37 *See* Minow & Cate at 20.
38. CRS Data Mining Report at 3.
39. *Olmstead v. United States,* 277 U.S. 438, 478 (1928) (Brandeis, J., dissenting). *See also* Samuel D. Warren & Louis D. Brandeis*, The Right to Privacy,* 4 Harv. L. Rev. 193, 205 (1890).
40. If privacy "relates to the ability to withhold personal data," then confidentiality "relates to the activities of an agency that has collected such data from others." National Research Council at 28.
41. Alan F. Westin, Privacy and Freedom 31 (New York: Athenaeum 1967) (adding that, because of this anonymity, "he does not expect to be personally identified and held to the full rules of behavior and role that would operate if he were known to those observing him").
42. For example, the Supreme Court has recognized that political or religious expression is not "free" if speakers are obliged to disclose their identity. *See McIntyre v. Ohio Elections Commission*, 514 U.S. 334, 343 (1995) (striking down an Ohio law prohibiting the distribution of anonymous campaign literature and taking note of "a respected tradition of anonymity in the advocacy of political causes") (citing *Talley v. California*, 362 U.S. 60 (1960)); *Watchtower Bible & Tract Soc'y of N.Y., Inc. v. Village of Stratton*, 536 U.S. 150, 166–69 (2002) (declaring unconstitutional a town law requiring those who wish to canvass door-to-door to first identify themselves in a permit application filed with the mayor's office and made available for public inspection). Similar rules apply to free expression rights, *see*

Lamont v. Postmaster General, 381 U.S. 301 (1965) (striking down government measure that required individual to notify post office of interest in certain political materials before receiving them in mail), and freedom of association, *see NAACP v. Alabama*, 357 U.S. 449, 462 (1958) (forbidding the state of Alabama from compelling the NAACP to disclose its membership lists).

43. This is known as the "third-party doctrine." *See infra.*

44. *See* 18 U.S.C. § 2701 *et seq. See also infra*, at 15.

45. Minow & Cate at 5-6.

46. U.S. Const. amend. IV.

47. *Katz v. United States*, 389 U.S. 347, 351 (1967).

48. *See Kyllo v. United States,* 533 U.S. 27, 33-34 (2001) (recognizing that Fourth Amendment protections have been affected by changes in technology and holding use of thermal imaging technology to gather information about interior of a home was a violation of Fourth Amendment where that technology was not in general public use); Orin S. Kerr, *Applying the Fourth Amendment to the Internet: A General Approach*, 62 Stan. L. Rev. 1005, 1006 (2010) (explaining that very few courts have addressed how the Fourth Amendment applies to the internet). As the Ninth Circuit has explained, "the extent to which the Fourth Amendment provides protection for the contents of electronic communications in the internet age is an open question. The recently minted standard of electronic communication via e-mails, text messages, and other means opens a new frontier in Fourth Amendment jurisprudence that has been little explored." *Id.* (quoting *Quon v. Arch Wireless Operating Co.*, 529 F.3d 892, 904 (9th Cir. 2008)). Under Section 215 of the Patriot Act, for instance, the federal government may seek an order requiring a third party to produce "any tangible things" related to a terrorism investigation, including the records of library computer use. 50 U.S.C. § 1861.

49. 425 U.S. 435, 443 (1976); see Minow & Cate at 7-8 (discussing *Miller*); National Research Council at 31-32 (same). .

50. *See* Daniel J. Solove, *Data Mining and the Security-Liberty Debate*, 75 U. Chi. L. Rev. 343, 357 (2008) ("As so much of our personal information is in the hands of various companies, the third-party doctrine severely limits Fourth Amendment protection."); Minow & Cate at 8-9 (recognizing similar limitation).

51. The nearly "unanimous" position of many observers is that the third-party doctrine is "horribly wrong." Orin S. Kerr, *The Case for the Third-Party Doctrine*, 107 Mich. L. Rev. 561, 564 (2009) (disagreeing with that widely held position); see, e.g., Daniel J. Solove, *Fourth Amendment Codification and Professor Kerr's Misguided Call for Judicial Deference*, 74 Fordham L. Rev. 747, 753 (2005) ("The third party doctrine presents one of the most serious threats to privacy in the digital age.").

52. Solove, *Fourth Amendment Codification*, at 753; *see* Cate at 456 (discussing changes in technology and reach of data mining).

53. Justice Brennan recognized this flaw of the third-party doctrine in his dissent in *Miller,* and it is no less salient today. *Miller*, 425 U.S. at 451 (Brennan, J., dissenting) ("For all practical purposes, the disclosure by individuals or business firms of their financial affairs to a bank is not entirely volitional, since it is impossible to participate in the economic life of contemporary society without maintaining a bank account.").

54. According to one commentator, the Supreme Court "backed off from *Miller* in two recent cases . . . . signal[ling] that the Court is willing to consider at least minor exceptions to *Miller*'s dictate that the government does not effect a constitutionally regulated search when it accesses information the subject shared with a third party." Christopher Slobogin, *Government Data Mining and the Fourth Amendment*, 75 U. Chi. L. Rev. 317, 330-31 (2008) (citing *Ferguson v. City of Charleston*, 532 U.S. 67 (2001), and *Georgia v. Randolph*, 547 U.S. 103 (2006)). Some have argued that government data mining raises concerns not yet directly decided by the Supreme Court in large part because technology has outpaced the development of judicial doctrine. *See, e.g., id.* at 317 ("Since at least the mid-1990s, the quantity of the world's recorded data has doubled every year. At the same time, the computing power necessary to store, access, and analyze these data has increased geometrically, at increasingly cheaper cost." (citations omitted)); Cate at 456-62 (discussing technological developments). As the National Research Council has observed, "a search-enabled digital world . . . [has] chang[ed] the technological milieu that surrounds privacy jurisprudence." National Research Council at 98.

55. *See* Kerr, *The Case for the Third-Party Doctrine,* at 3 n.12 (citing Stephen E. Henderson, *Learning From All Fifty States: How To Apply The Fourth Amendment And Its State Analogs To Protect Third Party Information From Unreasonable Search,* 55 Cath. U. L. Rev. 373 (2006)).

56. *See Ferguson*, 532 U.S. at 83-84.

57. *Warshak v. United States,* 490 F.3d 455, 482 (6th Cir. 2007) (email); *Quon v. Arch Wireless Operating Co., Inc.,* 529 F.3d 892 (9th Cir. 2008) (text messages), *rev'd, City of Ontario v. Quon,* 560 U.S. --, No. 08-1332 (June 17, 2010).

58. *See Warshak v. United States*, 532 F.3d 521 (6th Cir. 2008).

59. *See* Reply Brief of Appellants Steven Warshak, Harriet Warshak, and TCI Media, Inc., *United States v. Warshak,* No. 08-4085 (6th Cir.) (filed Nov. 18, 2009).

60. *Rehberg v. Pauk*, No. 09-11897 (11th Cir. Mar. 11, 2010) ("A person . . . loses a reasonable expectation of privacy in emails, at least after the email is sent to and received by a third party.").

61. *City of Ontario v. Quon*, 560 U.S. --, No. 08-1332, slip. op. at 11 (June 17, 2010).

62. *Id*. at 10.

63. Fred H. Cate, *Government Data Mining: The Need for a Legal Framework,* 43 Harv. C.R.-C.L. L. Rev. 435, 453 (2008). The exception for use is the exclusionary rule, which limits the government's use of data for evidentiary purposes in a criminal trial if it has been collected illegally. *See id.;* National Research Council at 151. However, the Supreme Court has left the door open to protections against government disclosure of personal data under constitutional privacy considerations. *See* discussion of *Whalen v. Roe*, 429 U.S. 589 (1977), *infra*.

    The *Miller* Court indicated that consent to provide information for one purpose effectively constitutes consent for use for other downstream purposes, noting that Fourth Amendment protections do not extend to information that 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." *Miller*, 425 U.S. at 443.

64. Assuming airport screening programs such as Secure Flight are constitutional, it also is conceivably constitutionally permissible for the government to use data collected through those programs for data mining purposes. However, thus far the TSA, which recently took control of the Secure Flight program, has indicated that it will limit its use of passenger data collected as part of Secure Flight. *See* TSA, *Secure Flight Q&A* (June 2, 2009), http://www.tsa.gov/blog/2009/06/secure-flight-q.html; *Secure Flight Program Privacy Impact Assessment* (Oct. 21, 2008), http://www.tsa.gov/assets/pdf/nprm_pia.pdf. Because passengers effectively consent to the government's terms of data use when they consent to the government's terms of travel, the government should ensure that travelers are adequately notified of the scope of their consent.

65. 429 U.S. 589, 605 (1977). Although the Court held that under the facts of the case—where the government employed adequate data security provisions—there was no Fourth Amendment violation, it suggested a constitutional interest exists in protecting personal information against government disclosure.

66. *See* Minow & Cate at 10 (citing cases in the District of Columbia, Second, Third, Fifth, and Ninth Circuits). The Sixth Circuit has narrowed *Whalen* by holding it will recognize the nondisclosure interest only "where the individual privacy interest is of constitutional dimension." *See Kallstrom v. City of Columbus*, 136 F.3d 1055, 1061 (6th Cir. 1998). That circuit nonetheless has applied *Whalen* to strike down a government disclosure. Id. at 1065.

67. *See Lamont v. Postmaster General,* 381 U.S. 301, 303 (1965) (invalidating a Federal law requiring recipients of "communist political propaganda" to specifically authorize the delivery of each such piece of mail).

68. *See* Daniel J. Solove, *The First Amendment as Criminal Procedure,* 82 NYU L. Rev. 112, 143-44 (2007) (*citing Laird v. Tatum,* 408 U.S. 1 (1972)).

69. *See* Katherine J. Strandburg*, Surveillance of Emergent Associations: Freedom of Association in a Network Society,* Digital Privacy: Theory, Technologies, and Practices, at 15-16 (Alessandro Acquisti et al. eds., Auerbach Publications 2007), *available at* http://works.bepress.com/katherine_strandburg/11.

70. Lisa Rein, *Md. Police Put Activists' Names on Terror Lists*, Wash. Post, Oct. 8, 2008, http://www.washingtonpost.com/wp-dyn/content/article/2008/10/07/AR2008100703245.html.

71. Daniel J. Steinbock, *Data Matching, Data Mining, and Due Process,* 40 Ga. L. Rev. 1, 46 (Fall 2005) (discussing *Mathews v. Eldrige*, 424 U.S. 319 (1976)).

72. *Whren v. United States,* 517 U.S. 806, 813 (1996).

73. *See United States v. Armstrong,* 517 U.S. 456, 464 (1996); *Chavez v. Illinois State Police,* 251 F.3d 612, 635 (7th Cir. 2001).

74. *See* U.S. Dep't of Justice, Civil Rights Division, *Guidance Regarding the Use of Race by Federal Law Enforcement Agencies* (June 2003), *available at* http://www.justice.gov/crt/split/documents/guidance_on_race.php.

75. *Id*. at n.3 & 4.

76. *Id*. at n.6.

77. *See Armstrong*, 517 U.S. at 465.

78. *See* Mike Ahlers, *U.S. Announces New Airport Security Measures,* CNN.com, Apr. 2, 2010, http://www.cnn.com/2010/TRAVEL/04/02/airline.security/index.html.

79. *See* DHS Privacy Office at 33.

80. *See Parents Involved in Cmty. Sch. v. Seattle Sch. Dist. No. 1,* 551 U.S. 701, 720 (2007) ("It is well established that when the government distributes burdens or benefits on the basis of individual racial classifications, that action is reviewed under strict scrutiny."); *Adarand Constructors, Inc. v. Peña,* 515 U.S. 200, 227 (1995) ("[A]ll racial classifications, imposed by whatever federal, state, or local governmental actor, must be analyzed by a reviewing court under strict scrutiny.").

81. *See, e.g.,* Note, Joshua L. Simmons*, Buying You: The Government's Use of Fourth-Parties to Launder Data about 'The People,'* 2009 Colum. Bus. L. Rev. 950, 956-57 (Sept. 19, 2009).

82. 18 U.S.C. §§ 2510-2522, 2701-2711, 3121-3127.

83. 50 U.S.C. §§ 1801-1885c.

84. *See* discussion of collection versus use and discussion of exclusionary rule, *supra; see* Cate at 463-64 (explaining deficiencies of ECPA). Whereas uses of information were at one time limited based on whether the information was collected for foreign intelligence versus criminal law enforcement purposes, that distinction has been narrowed by legislation and judicial interpretation. After FISA's amendment by the USA PATRIOT Act, *see* 50 U.S.C. § 1804(a)(7)(B), and as interpreted by the Foreign Intelligence Surveillance Court of Review in *In re Sealed Case*, collection of foreign intelligence must be a *significant* purpose of a surveillance sought under the Act, but it need not be the only purpose. See *In re Sealed Case*, 310 F.3d 717, 735 (Foreign Int. Surv. Ct. Rev. 2002) (recognizing that foreign intelligence and criminal law enforcement purposes are not entirely distinct).

85. 5 U.S.C. § 552a.

86. *Id*. § 552a(b)(7).

87. *See* National Research Council at 157-58 (discussing exceptions); Minow & Cate at 11-12 (same). An early amendment to the Privacy Act, the Computer Matching and Privacy Protection Act, implicates some data mining, but it covers a very narrow set activities, excluding data mining for law enforcement, foreign counterintelligence, and background checks. See Pub. L. No. 100-503, 102 Stat. 2507 (1988) (codified at 5 U.S.C. §§ 552a(a)(8), 552a(o)-(r) (2000)).

88. Pub. L. No. 107-347, § 208(b), 116 Stat. 2899, 2921-22 (2002) (codified at 44 U.S.C. § 3501 note); Minow & Cate at 18.

89. E-Government Act of 2002 § 208(b)(2)(B)(ii).

90. *Id*.

91. *See Balancing Privacy and Security: The Privacy Implications of Government Data Mining Programs: Hearing Before the S. Comm. on the Judiciary,* 110th Cong. 11 (2007) (statement of Leslie Harris, Executive Dir., Center for Democracy & Technology) ("Harris statement") (stating that E-Government Act requirements should apply to government access to third-party databases).

92. *See* Minow & Cate at 18; E-Government Act of 2002 § 202(i); OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002, M-03-22, § II.B.3(c) (Sept. 26, 2003) "[N]o PIA is required for national security systems defined at 40 U.S.C. § 11103 as exempt from the definition of information technology . . . .").

93. E-Government Act of 2002 § 208(b)(1)(C).

94. *See* DHS, *Privacy Impact Assessment for the Initiative Three Exercise (*Mar. 18, 2010), http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_nppd_initiative3.pdf. The DHS Privacy Office conducts PIAs for information technology, rulemaking, human resources, national security systems, programs involving personally identifiable information, privacy-sensitive technology, pilot testing, and social media. DHS Privacy Office, *Guide to Implementing Privacy,* at 14 (June 2010*).*

95. Section 804 of the Implementing the Recommendations of the 9/11 Commission Act of 2007, 42 U.S.C. § 2000ee-3(c).

96. *See, e.g.,* DHS Privacy Office (2008 report); DHS Privacy Office, *2009 Data Mining Report to Congress* (Dec. 2009); Office of the Director of National Intelligence (ODNI), *Data Mining Report* (Mar. 2009) (covering data mining activities for January 31, 2008 to January 31, 2009); ODNI, *2009 Data Mining Report* (Mar. 2010) (covering data mining activities for February 1, 2009 through December 31, 2009).

97. ODNI, *2009 Data Mining Report* at 2.

98. Section 1016 of the Intelligence Reform and Terrorism Prevention Act of 2004 directs the President to create the ISE. 6 U.S.C. § 485. The law requires that ISE "incorporate[] protections for individuals' privacy and civil liberties," *id.* § 485(b)(2)(H), and guidelines for privacy protection in the ISE have been promulgated.

99. Section 511 of the Implementing Recommendations of the 9/11 Commission Act of 2007 establishes the DHS State, Local, and Regional Fusion Center Initiative, designed to support information sharing among these levels of government. It also requires DHS to conduct, within 90 days of the Act's enactment and again within one year after, a privacy and civil liberties impact assessment of the initiative. *See* 6 U.S.C. § 124h.

100. *See* 42 U.S.C. § 1306(a)(1); 26 U.S.C. § 6103 & 7431; 13 U.S.C. §§ 8-9.

101. 12 U.S.C. § 3402.

102. 47 U.S.C. § 551.

103. 18 U.S.C. § 2710.

104. 20 U.S.C. § 1232g; 34 C.F.R. pt. 99.

105. *See* 45 C.F.R. § 164.512.

106. 47 U.S.C. § 222; 47 C.F.R. § 64.2001-2011; see Congressional Research Service, CRS Report for Congress, *Government Access to Phone Calling Activity and Related Records: Legal Authorities* 15 n.43 (Aug. 20, 2007).

107. 15 U.S.C. § 1681b(b)(4).

108. "These procedures . . . . provide little direct guidance concerning data mining." Minow & Cate at 16.

109. *See* Section 801 of the Implementing the Recommendations of the 9/11 Commission Act of 2007 (amending Section 1061 of the National Security Intelligence Reform Act of 2004 (5 U.S.C. § 601 note)).

110. Alan Charles Raul, *Privacy and Civil Liberties: Where's the Watchdog?*, Wash. Post, Oct. 23, 2008, available at http://www.washingtonpost.com/wp-dyn/content/article/2009/10/22/AR2009102203802.html.

111. As the Markle Foundation Task Force on National Security in the Information Age has noted, with more powerful data mining activities must come corresponding privacy safeguards. Markle Foundation Task Force on National Security in the Information Age (Markle Foundation), *Nation at Risk: Policy Makers Need Better Information to Protect the Country*, App. A, at 23 (Mar. 2009) (urging the President and Congress to "[e]nhance security and privacy protections to match the increased power of shared information").

112. *See, e.g.*, Harris statement at 9-12; Office of the Director of National Intelligence, *Guidelines to Ensure that the Information Privacy and Other Legal Rights of Americans are Protected in the Development and Use of the Information Sharing Environment* (Dec. 2006) (*"ISE Privacy Guidelines"*); *DHS Report* at 37-39; Markle Foundation, *Nation at Risk* at App. A; Markle Foundation, *Mobilizing Information to Prevent Terrorism: Accelerating Development of a Trusted Information Sharing Environment*, at 29-32 (July 2006); Markle Foundation, *Creating a Trusted Network for Homeland Security* 69-74 (Dec. 2003); Markle Foundation, *Protecting America's Freedom in the Information Age*, at 31-35 (Oct. 2002); National Research Council at 86-101; Technology & Privacy Advisory Committee, Dep't of Defense, *Safeguarding Privacy in the Fight Against Terrorism*, at 45-60 (Mar. 2004) (*"TAPAC Report"*); *id.* at 31-32 (citing European Union's data protection directive); Cate at 487-88; Minow & Cate at 23 (citing *TAPAC Report*); *The Cantigny Principles on Technology, Terrorism, and Privacy*, National Security Law Report, at 14-16 (Feb. 2005) ("Cantigny Principles").

113. DHS describes the FIPPs as a set of eight principles: Transparency, Individual Participation, Purpose Specification, Data Minimization, Use Limitation, Data Quality and Integrity, Security, and Accountability and Auditing. DHS, *Privacy Policy Guidance Memorandum*, No. 2008-01, at 1 (Dec. 29, 2008), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf (memorializing DHS adoption of the FIPPs). See The Constitution Project's *Guidelines for Public Video Surveillance: A Guide to Protecting Communities and Preserving Civil Liberties* (2006), for additional information on FIPPs.

114. *E.g.*, Harris statement at 11. Others have also called for Privacy Act reform. For instance, the National Broadband Plan, released by the Federal Communications Commission in March 2010, recommends that Congress "consider re-examining the Privacy Act . . . to account for changes in technology." Federal Communications Commission, *Connecting America: The National Broadband Plan,* at 290-91 (Mar. 16, 2010) (*"National Broadband Plan"*).

115. Markle Foundation, *Creating a Trusted Network for Homeland Security* at 71-72; National Research Council at 98; *TAPAC Report* at 49-52; *see also* Slobogin at 321-22.

116. *See, e.g.*, National Research Council at 86-87; Markle Foundation, *Creating a Trusted Network for Homeland Security* at 71; Cantigny Principles at 16; *DHS Report* at 38.

117. *See, e.g.*, Markle Foundation, *Mobilizing Information to Prevent Terrorism* at 33; *TAPAC Report* at xi.

118. *See* supra, note 113.

119. *See* Office of the Secretary of Defense, *Policy on Safeguarding Personally Identifiable Information and Breach Notification* (June 5, 2009), *available at* http://privacy.defense.gov/files/PII_Memo_Safeguard.pdf.

120. *See* DHS Privacy Office, *Annual Report to Congress,* at 33 (Sept. 2009), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_rpt_annual_2009.pdf.

121. *See, e.g., supra,* note 113.

122. *See generally* GAO 2004 Report (discussing use of data mining for efficiency and other administrative purposes).

123. *See* DHS Privacy Office at 37-39; *TAPAC Report* at 45-60.

124. *See Watch List Guidelines* at 5-8.

125. DHS Data Privacy and Integrity Advisory Committee, *The Elements of Effective Redress Programs*, No. 2010-01, at 6-7 (Mar. 25, 2010), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_dpiac_report2010_01.pdf.

126. The National Broadband Plan calls for the executive branch to establish a website that allows citizens to request their personal data held by government agencies, so that they can correct it. *See National Broadband Plan* at 290.

127. *See* The President, *Memorandum for the Heads of Executive Departments and Agencies: Freedom of Information Act* (Jan. 2009), *available at* http://edocket.access.gpo.gov/2009/pdf/E9-1773.pdf.

128. 42 U.S.C. § 2000ee-3(c)(2).

129. *Id.* § 2000ee-3(c)(3).

130. *See, e.g.*, ODNI, *Data Mining Report* (Feb. 2008), *available at* http://www.fas.org/irp/dni/datamining.pdf.

THE CONSTITUTION PROJECT

Safeguarding Liberty, Justice & the Rule of Law

1200 18th Street, NW
Suite 1000
Washington, DC 20036
Tel 202.580.6920
Fax 202.580.6929

Email: info@constitutionproject.org

www.constitutionproject.org