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

EXECUTIVE SUMMARY

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 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[*] can offer significant benefits, but without adequate processes and controls, it can encroach on the constitutional rights and values that govern the citizen's relationship to the state—including privacy, freedom of expression, due process, and equal protection. Innocent people could be mistakenly 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.

Despite the weightiness of the interests involved, the current legal regime fails to clearly or uniformly regulate government data mining activities: While both the Constitution and several federal statutes implicate government data mining, it has substantially avoided direct regulation. This is a situation that can be and must be remedied. We can adopt a system that both encourages the government to harness the vast seas of information for our collective benefit and simultaneously protects the delicate relationship our Constitution established between the government and the governed.

For these reasons, we, the undersigned members of the Constitution Project's bipartisan Liberty and Security Committee, 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. Rooted in constitutional values and the Fair Information Practice Principles, these common sense principles, if applied and followed, can facilitate the government's use of data mining techniques without sacrificing constitutional rights and values.

Principles for the Development of Data Mining Programs

[*] We define "data mining" 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 is undertaken by a government entity, on behalf of the government, or where government personnel are permitted to access the data.

- *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 covering data sources, data acquisition, system design and capabilities, and intended uses.*

- *Perform an internal evaluation of the program's 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 Plan.*

Principles for the Operation of Data Mining Programs

Transparency and Notice

- *Provide notice to an individual of the acquisition of his or her personal data or any specific government action or classification of that individual pursuant to data mining.*

- *When individual notice is not applicable or permissible, provide for judicial or administrative notice, as feasible.*

Accountability, Oversight and Redress

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

- *Establish penalties for misuse and abuse by operators or others.*

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

- *To the extent feasible, permit individuals to review and correct data through FOIA or other procedure.*

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

Authority and Choice

- *Coordinate uses, best practices, regulations and protections with other 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.*

- *Where consistent with the goals of the program, allow individuals a meaningful opportunity to opt out of the acquisition and use of their data.*

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 regular basis.*
- *Conduct training and evaluation for employees with access to personal data or data mining system.*
- *Take all reasonable steps to rectify and minimize harm from data breach, including prompt notification of affected individuals.*

Data Minimization

- *Minimize data acquisition and aggregation of databases.*
- *Use deidentified or aggregate data formats or other techniques with respect to personal information to minimize potential harm.*
- *Limit "downstream" use of personal data through technical measures, rule, or contract.*
- *Set retention periods and ensure complete destruction of expired data.*

TABLE OF CONTENTS

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

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 data.[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 tools to mine data have developed, the government is improving its capabilities, although it is not yet as adept as Google or other private sector actors. It has built or is building thousands of databases and 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 the constitutional rights and values that govern the citizen's relationship to the state—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 undersigned members of the Constitution Project's bipartisan Liberty and Security Committee, 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.

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.[6] 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.[7] 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 is simply scanned for items or events meeting specified parameters.[8] 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.[9] Of course, this distinction is often blurry, and many data mining systems use both search-based and pattern-based techniques.[10]

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 the Federal Trade Commission (FTC) recently noted that "the traditional notion of what constitutes PII versus non-PII is becoming less and less meaningful."[11] Accordingly, the relevant question for this report is not whether some set of data identifies a particular individual, but whether it is or can be linked to a specific individual or device.[12]

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,[13] 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 becomes available. Old data is now digital—census records, meteorological data, stock prices—while new forms of data have been created, like social networking data, internet search queries, and data from radio-frequency-ID systems like E-Z-Pass and employee access cards. Observers estimate that the world's collected data volume doubles every year.[14]

Notably, advances in technology can also 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 mine data remotely without copying it to a local database—lessening the chances of accidental release or

[*] 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 is obtained directly from an individual. "Acquisition" refers to both direct collection and obtaining data from a third party or another government entity.

exposure.[15] 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 is undertaken by a government entity, on behalf of the government, 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 discuss some of the most relevant programs.[16]

Efficiency and program evaluation. The most common purpose for which government agencies mine data is to improve efficiency and evaluate performance.[17] 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 DOJ and Department of Veterans Affairs have successfully used operational data to more efficiently allocate agency resources.[18] 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 GAO estimated in 2004 that the federal government had employed or will employ data mining for these purposes in at least 30 separate programs.[19] The IRS makes extensive use of data mining to increase tax compliance and detect tax fraud.[20] A data mining program helped uncover millions of dollars in Medicare fraud.[21] As we discuss below, while these applications can pose serious risks to civil liberties, data mining is well-suited to situations

[*] The private sector has long used data mining for marketing and other purposes. Although private data mining is beyond the scope of this report, because it implicates similar privacy concerns, 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.

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.

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. The controversial "fusion centers" are an example of data mining for this purpose. These are 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 openness, and lack of oversight for the program create significant risks to civil liberties.[22]

Crime prevention and counter-terrorism. 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 is worth trying every conceivable tool to combat it. On the other hand, civil libertarians are concerned for several reasons: First, the value of data mining to help *prevent* bad acts is unclear due to the shortage of data on past terrorist and criminal acts. As explained below, it is difficult to develop 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.

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

- ***Investigative Data Warehouse.*** The FBI describes 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.[23] In addition to the IDW's value for investigative purposes, the Electronic Frontier Foundation concluded that the FBI was likely employing advanced, predictive "data exploitation" tools based on the IDW data.[24]

- ***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, counter-terrorism 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 might continue as part of classified operations.

- ***Secure Flight / CAPPS II.*** Also following 9/11, the FAA (later the 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 program is still in development, but the TSA has recently taken steps to prepare the program for deployment, including asking travel agents and airlines to begin collecting more information.[25]

2. Practical hurdles to effective data mining

Data mining—particularly for criminal prevention or counter-terrorism purposes—presents practical issues that may outweigh 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 is correctly attributed. Names change or are recorded differently, addresses and other identifiers change, or data is entered incorrectly.[26] This problem is 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.[27] Relatedly, underlying databases may be incompatible, and prevent or complicate data mining applications that use multiple sets of data.[28] 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.[29] In the credit card context, the harm from these errors may be generally small, and consumers generally feel that the benefits outweigh the inconvenience of false positives. In the criminal or terrorism context, however, these errors can be disastrous. Individuals who are erroneously added to the "Do Not Fly" list, for example, can be materially harmed in a variety of ways. Such errors also increase costs to the government and airlines and create public skepticism about the value of security measures.[30] In 2009, 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.[31] When there are few model events—as is particularly the case with terrorist acts[*]—the potential for false identification is increased.[32]

[*] Terrorists may attempt to consciously alter their methods to avoid mimicking past terrorist plots—undermining pattern-based data mining methodologies.

C. Legal Implications of Data Mining

1. Data Mining and Constitutional Rights and Values

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. While the Supreme Court has delineated the legal contours of our civil liberties, a broader consideration of the values that underlie those liberties is a necessary step in making sound policy. We briefly review those broader values here.

a) Privacy

In most discussions of the legal and ethical implications of data mining, privacy is the central concern. 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."[33] Related to privacy is the right of "confidentiality," which concerns how personal information is disseminated,[34] 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."[35] 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.[36] Even more than the Constitutional standards, the social conception of what is personal has altered with technological development. Mobile PCs, cloud computing, online social networks—all have further 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 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 ISP or a bank, can also implicate privacy and confidentiality, especially in the absence of notice. While the current constitutional standard generally does not limit the government's access to information that an individual has shared with a third party,[37] 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.[38] 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 and airline operations 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.

b) 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.[39]

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 counter-terrorism 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.[40]

c) Government accountability and procedural safeguards

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, but it more broadly requires 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.

d) Equal protection and anti-discrimination

The communal values expressed in our American society reject discrimination for many reasons. Most obviously, it degrades its victims and reinforces that most anti-American of social institutions: a class structure. Moreover, discrimination retards the very ability of any insular minority group—be it religious, cultural, political, or ethnic—to participate fully in civil society. Justice Stone famously wrote that "prejudice against discrete and insular minorities may be a special condition, which tends seriously to curtail the operation of those political processes ordinarily to be relied upon to protect minorities."[41] This idea paved the way for much of modern Equal Protection and Due Process jurisprudence.

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.[42] Even more insidious, data mining could be pointed to as a supposedly objective validation for inappropriate stereotyping of minority groups. For example, a data mining program searching air passenger lists for potential security risks might 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. Investigators might see this result and assume it objectively "proves" that middle-eastern Muslim men are security risks. Such racial profiling would not only unfairly target certain minorities, but can also undermine the effectiveness of programs.

2. The Current Legal Regime

As explained above, at stake in the conduct of government data mining programs are both security interests and civil liberties, particularly individual privacy. Despite the weightiness of these two interests, the current legal regime fails to clearly or uniformly regulate government data mining activities: While both the Constitution and several federal statutes implicate government data mining, it has substantially avoided direct regulation. The below review discusses the applicable laws.

a) Constitutional protections

The Fourth Amendment is the primary constitutional provision that pertains to issues of information privacy.[43] It protects individuals against "unreasonable searches and seizures,"[44] but only in circumstances in which a person has a reasonable expectation of privacy.[45] 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 as yet unclear whether courts will construe those expectations as "reasonable" under the Fourth Amendment.[46] Nevertheless, the contours of a reasonable expectation of privacy have been defined in at least one significant respect. In the seminal case *United States v. Miller*, the Supreme Court held that a person does not have a reasonable expectation of privacy in information held by a third party.[47]

Miller's "third-party doctrine" has substantial implications for government data mining activities. 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.[48]

The Fourth Amendment also generally applies 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.[49] Indeed, 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."[50] Therefore, 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.[51]

Yet, government data mining may implicate Fourth Amendment interests more than other systems previously considered by the Court.[52] Government data mining raises concerns not yet directly decided by that body in large part because technology has outpaced the development of judicial doctrine.[53] The prevalence of record-keeping; the number of records maintained for daily transactions; the sensitivity of the information contained in those records; the breadth of information captured by, transferred for, and searched during data mining; and the use to which

the government will put those records all distinguish today's government data mining activities.[54] The privacy implications of data mining activities in today's world cannot be compared to the government data operations of thirty years ago.[55]

The Supreme Court has cracked open the door to broader constitutional protections for personal privacy. 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."[56] 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.[57] The courts may continue to develop this doctrine.

b) Legislative and administrative protections

Some of the gaps in constitutional protections have been filled by statute. However, because many of those laws have broad exceptions, and because different laws apply to different sectors, privacy 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")[58] regulates electronic surveillance for law enforcement purposes. The Foreign Intelligence Surveillance Act of 1978 ("FISA"),[59] 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. Moreover, these laws impose little or no limitation on how the information may be used by the government once it is obtained.[60]

Data Processing and Disclosure. A handful of statutes primarily regulate government treatment of individuals' data. The Privacy Act of 1974[61] 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.[62] Various data thus may not receive privacy protections in the data mining context,[63] 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.[64] 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.[65] In addition, certain information technologies used for national security are excepted from the PIA requirement,[66] and publication of a PIA may be waived for security reasons.[67]

Under the Federal Agency Data Mining Reporting Act of 2007 ("Data Mining Act"),[68] 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. The Department of Homeland Security's Privacy Office released reports on its data mining activities for 2006, 2007, and 2008, and the Office of the Director of National Intelligence released reports addressing its 2007 and 2008 activities.[69] As of the date of this report, both have yet to release their Data Mining Act reports for 2009 activities. The ODNI report for 2008 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.[70]

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,[71] and adoption of federal steps to support the development of state and local information fusion centers.[72]

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.[73] Laws such as the Right to Financial Privacy Act of 1978,[74] the Cable Act,[75] the Video Privacy Protection Act,[76] the Family Educational Rights and Privacy Act of 1974,[77] the Health Insurance Portability and Accountability Act of 1996 Privacy Rule,[78] the Communications Act,[79] and the Fair Credit Reporting Act[80] 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.

Finally, 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.[81]

Today's data mining involves greater data sharing and more powerful technologies than ever before. Those enhancements must be balanced by more responsive privacy protections.[82]

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. 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.[83] Our principles build on this well-laid foundation, and particularly on the recommendations of 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"), 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"),[84] and at least one explicitly recommends that the Privacy Act be adapted to apply to government use of third-party data.[85] 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.

Another procedural concern addressed by some commenters is the risks associated with data transfer. Groups such as the Markle Foundation,[86] the Center for Democracy and Technology, and members of the Cantigny Conference suggest that, because continual data transfers create a risk of unauthorized disclosures or breaches, third parties should not transfer data to the government. Rather, when the proper requirements are met, the government should be permitted to directly access the third-party entity's database or use some form of decentralized network.[87]

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 is 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.[88]

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.[89] 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.[90]

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,[91] 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, judicial, or administrative input or 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.***

[Tentative – for discussion] As an initial step, any acquisition of personal data by a government entity should be proceeded by a clear articulation of the purpose(s) of the collection or acquisition and the intended use(s) of the data. While we recommend this step for any acquisition of personal information, it is particularly important where the intended uses include data mining. Enabling internal decision-makers (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 cause harm to constitutional rights and values, and so should be carefully and publicly considered. 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 or 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 intra-branch 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:

  - *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, integrity, and minimization (as discussed below); and

- *Intended uses* of the program.

- ***Perform an internal evaluation of the program's 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. At a minimum, the evaluation should include:

- *Costs and benefits* of the program, including the economic benefits of automated data mining systems over other methods of accomplishing the same goals;

- *Compliance 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 Defense Department's TAPAC guidelines); and

- *Impact on civil liberties and constitutional values*, which may be accomplished through existing procedures, such as the Department of Homeland Security Privacy Office's "Privacy Impact Assessment" standard.[92]

- ***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 via 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 currently 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.

- ***Respond to and incorporate administrative, public, and congressional commentary on 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.[93]

1. Transparency and Notice

- ***Provide notice to an individual of the acquisition of his or her personal data or any specific government action or classification of that individual pursuant to data mining.***

[Tentative – for discussion] The majority of personal data used by government agencies for data mining does not require secrecy.[94] In such cases, the agency obtaining data should seek to provide notice to individuals of the acquisition[*] of personal data for data mining purposes or of any specific action or classification that is based on information gleaned from data mining operations. For data acquisition, this notification could be accomplished at the time of initial collection by the government directly from the individual, or, in the case of data acquired from third parties like credit reporting agencies, by subsequent notification. For actions or classifications that are made as a result of data mining, 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. 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.

- ***When individual notice is not applicable or permissible, provide for judicial or administrative notice, as feasible.***

[Tentative – for discussion] In those instances where notification of individuals would undermine public safety, national security, or other law enforcement goals, we recommend contemporaneous notification of an independent arbiter. As discussed in other principles, the body exercising such oversight should have independent authority from those running the data mining program. The independent arbiter can determine—based on the standards developed in the data mining Plan and in the administrative rules discussed below—whether the acquisition, action, or classification is justified.[95] With proper authorizing legislation, agencies could seek ex parte judicial review of such actions, similar to the process undertaken under the Foreign Intelligence Surveillance Act.[96] In the absence of legislation, administrative review can substitute.

2. Accountability, Oversight and Redress

[*] 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 is obtained directly from an individual. "Acquisition" refers to both direct collection and obtaining data from a third party or another government entity.

- ***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.[97]

- ***Establish penalties for misuse and abuse by operators or others.***

As explained above, standards should be binding upon all federal employees and those operating under government contract. Operators that do not follow the standards, or that otherwise misuse or abuse personal data or data mining systems, should be subject to civil or criminal penalties. Those penalties will ensure operators will vigilantly follow safeguards.

- ***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, individuals should be provided notice of the acquisition of personal data for data mining purposes or of any specific action or classification based on information gleaned from data mining operations. In addition, individuals also must 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 either an administrative or judicial proceeding that is subject to appeal in court.[98]

As with the individual notice principle, *supra*, 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.

- ***To the extent feasible, permit individuals to review and correct data through FOIA or other procedure.***

Where individual review is 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. 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 is 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 requests or other multiple filings.

- ***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 appropriate members or committees of Congress and should include reporting on classified activities.

3. Authority and Choice

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

Agencies should collaborate to define and adopt consistent data practices—whether through interagency effort, executive order, or legislation. Coordination will help ensure 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 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.

- ***Where consistent with the goals of the program, allow individuals a meaningful opportunity to opt out of the acquisition and use of their data.***

While there are many instances in which the goals of a data mining program would be undermined by individual notice and choice, many programs—such as those to identify usage patterns or improve efficiency—could allow such individual control without negative impact. When making decisions in the course of daily transactions, travel, use of technology, or other activities, individuals often do not contemplate or understand the implications for their personal privacy and other civil liberties. In particular, they do not consider that consenting to provide their personal information in one circumstance may mean they are ceding their control of that information for other purposes.

Thus, as described above, individuals should be notified at the initial point of data collection or acquisition what their consent entails, so that they may better understand the scope of their consent to that collection or acquisition. In addition, they should be given a meaningful opportunity to opt-out of that collection, acquisition, or downstream use where possible. If an

agency wishes to repurpose personal data for a new or expanded data mining use, it should permit additional opportunities for individuals to opt-out to the extent compatible with the program's purposes. The most effective means of facilitating informed individual choice may be to allow individuals to opt-out of private organizations' sale of their data to the government.

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.

With respect to technical protection measures, we recommend at a minimum implementing network access limits, using strong encryption for transmission and storage of data, and automated logging of system access. These measures are 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.

- ***Evaluate and improve data security, integrity, accuracy, and timeliness on regular basis.***

Operators should conduct regular evaluations of the effectiveness and adequacy of data integrity and security measures. 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 system.***

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 Minimization

By minimizing the amount and duration of data collected or acquired, stored, and shared, as well as by reducing the sensitivity of that data, agencies can substantially reduce the potential impact of data mining operations on civil liberties and constitutional values.

- ***Minimize data acquisition and aggregation of databases.***

Overinclusive data collection or acquisition heightens the risk of scope creep and data misuse. Even if that data is 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 be required to collect or acquire only data that has been 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.***

Personally identifiable information 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. 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.

- ***Limit "downstream" use of personal data through technical measures, rule, or contract.***

When an agency disseminates data to another agency or third party, including other federal agencies, state law enforcement personnel, 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.

- ***Set 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.

Endnotes

[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* § I.B.1, *infra*.

[5] Statement of Lee Hamilton, *supra*.

[6] 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").

[7] *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).

[8] 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.").

[9] *See* Minow & Cate at 3; National Research Council at 21.

[10] *See* National Research Council at 23; DHS Privacy Office at 32.

[11] *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.

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

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

[14] *See* CRS Data Mining Report at 2.

[15] *See* DHS Privacy Office at 34.

[16] 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").

[17] *Id.* at 7-12.

[18] CRS Data Mining Report at 4.

[19] GAO 2004 Report at 10.

[20] *Id.* at 52.

[21] CRS Data Mining Report at 4.

[22] 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.

[23] Electronic Frontier Foundation (EFF), *Report on the Investigative Data Warehouse* (Apr. 2009), http://www.eff.org/issues/foia/investigative-data-warehouse-report.

[24] *Id.*

[25] *See* 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/.

[26] Minow & Cate at 19.

[27] *Id.*

[28] CRS Data Mining Report at 27.

[29] *See* National Research Council at 39.

[30] *See* Minow & Cate at 20.

[31] *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.

[32] CRS Data Mining Report at 3.

[33] *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).

[34] 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.

[35] 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").

[36] 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).

[37] This is known as the "third-party doctrine." *See infra*.

[38] *See* 18 U.S.C. § 2701 *et seq*. *See also infra*, at 9.

[39] *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).

[40] 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.

[41] *United States v. Carolene Products*, 304 U.S. 144, 153 n.4 (1938).

[42] *See* DHS Privacy Office at 33.

[43] Minow & Cate at 5-6.

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

[45] *Katz v. United States*, 389 U.S. 347, 351 (1967).

[46] *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*, Stan. L. Rev., at 3 (forthcoming 2009), http://ssrn.com/abstract=1348322 (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.

[47] 425 U.S. 435, 443 (1976); *see* Minow & Cate at 7-8 (discussing *Miller*); National Research Council at 31-32 (same).

[48] *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).

[49] 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*.

[50] *Miller*, 425 U.S. at 443.

[51] 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 Transportation Security Administration ("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.

[52] This idea has been explored by others. 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)).

[53] *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)); Cateat 456-62 (discussing technological developments).

[54] *See* Cate at 456 (discussing changes in technology and reach of data mining).

[55] 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.

[56] 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.

[57] *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.

[58] 18 U.S.C. §§ 2510-2522, 2701-2711, 3121-3127.

[59] 50 U.S.C. §§ 1801-1885c.

[60] *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).

[61] 5 U.S.C. § 552a.

[62] *Id.* § 552a(b)(7).

[63] *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)).

[64] 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.

[65] *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).

[66] *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 . . . .").

[67] E-Government Act of 2002 § 208(b)(1)(C).

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

[69] *See, e.g.*, DHS Privacy Office(2008 report); 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).

[70] ODNI at 1-2.

[71] 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.

[72] Section 511 of the Implementing Recommendations of the 9/11 Commission Act of 2007 establishes the Department of Homeland Security ("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.

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

[74] 12 U.S.C. § 3402.

[75] 47 U.S.C. § 551.

[76] 18 U.S.C. § 2710.

[77] 20 U.S.C. § 1232g; 34 C.F.R. pt. 99.

[78] *See* 45 C.F.R. § 164.512.

[79] 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).

[80] 15 U.S.C. § 1681b(b)(4).

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

[82] 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").

[83] *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").

[84] 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. 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 The Constitution Project's *Guidelines for Public Video Surveillance: A Guide to Protecting Communities and Preserving Civil Liberties* (2006), for additional information on FIPPs.

[85] *E.g.*, Harris statement at 11.

[86] The Markle Foundation Task Force guidelines informed the development of ISE legislation. Markle Foundation, http://www.markle.org/markle_programs/policy_for_a_networked_society/national_security/projects/taskforce_national_security.php.

[87] *See* Markle Foundation, *Creating a Trusted Network for Homeland Security* at 72; Harris statement at 10; Cantigny Principles at 15.

[88] 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.

[89] *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.

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

[91] *See supra*, note 84.

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

[93] *See, e.g., supra*, note 84.

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

[95] The Constitution Project, *Promoting Accuracy and Fairness in the Use of Government Watch Lists*, at 6 (Mar. 2007) ("*Watch List Guidelines*").

[96] *See* 50 U.S.C. § 1804.

[97] *See* DHS Privacy Office at 37-39; *TAPAC Report* at 45-60.

[98] *See Watch List Guidelines* at 5-8.