Statement of Leslie Harris
Executive Director
Center for Democracy & Technology*

before the
Senate Committee on the Judiciary

"Balancing Privacy and Security:
The Privacy Implications of Government Data Mining Programs"

January 10, 2007

Mr. Chairman, Senator Specter, and Members of the Committee:

Good morning, and thank you for the opportunity to testify at this important hearing. In our testimony this morning, we want to emphasize the following six points:

- Terrorism poses a grave threat to our nation. To prevent further terrorist attacks, the government should use information technology to better share and better analyze the ocean of information at its disposal in this digital age.

- Both national security and the protection of civil liberties require that the technology be used only when it is demonstrably effective and then only within a framework of accountability, oversight and protection of individual rights.

- "Data mining" broadly defined is the use of computer tools to extract useful knowledge from large sets of data. It is in the abstract neither good nor bad. Rather, the questions are: What kind of data mining should the government use, for what purposes, with what consequences for individuals, under what guidelines, and subject to what oversight, auditing and redress?

- There is a vast difference both in terms of proven effectiveness and in terms of risks to privacy and due process between pattern-based data mining, especially when based on hypotheticals, and subject-based data mining, such as where the government is starting with some particularized suspicion.

- The threshold question for the application of any technology is efficacy. So far, there has been no evidence of the effectiveness of the broad forms of predictive data mining that have been proposed and deployed by the government. Unless and until a particular application can be shown to be an effective tool for

* The Center for Democracy & Technology (CDT) is a non-profit public interest organization dedicated to promoting privacy and other democratic values for the new digital communications media. Among other activities, CDT coordinates the Digital Privacy and Security Working Group (DPSWG), a forum for computer, communications, and public interest organizations, companies and associations interested in information privacy and security issues.

counterterrorism, and appropriate safeguards to protect the privacy and due process rights of Americans are put in place, the government should not deploy pattern-based data mining as an antiterrorism tool.

- Finally, the technological and legal context for data collection and analysis has changed dramatically in recent years. Technology has far outstripped existing privacy protections at the very time that legal standards for government access to data have been lowered (or ignored by Executive fiat). Core laws like the Privacy Act are inadequate and almost irrelevant to data mining.

In light of these considerations, we offer below a series of recommendations to Congress, focused on oversight, accountability, and due process.

I. The Rights at Stake: Privacy and Due Process

It is very important to start by defining what we mean by "privacy." Information privacy is not merely about keeping personal information confidential. In the context of a function like data mining, privacy is equally about due process: how to make fair decisions about people.

It is well established by U.S. Supreme Court cases, the federal Privacy Act, and other privacy laws like the Fair Credit Reporting Act (FCRA) and the Health Insurance Portability and Accountability Act (HIPAA) that individuals retain a privacy (or due process) interest in information about themselves even after they have disclosed it in the course of a commercial or governmental transaction. Our interest in the fair use of information to make decisions about us extends even to data that is publicly available: if that information is used to make decisions that can have adverse consequences, then we should have a right to know about the use of that information and an opportunity to respond to information that is inaccurate or misleading.

The term "Fair Information Practices" (FIPs) best describes the values at stake with regard to data mining. First articulated in the 1970s, these principles govern not just the initial collection of information, but also its use. The "Fair Information Practices" have been embodied in varying degrees in the Privacy Act, FCRA, and the other "sectoral" federal privacy laws that govern commercial uses of information. The concept of FIPs has remained remarkably relevant despite the dramatic advancements in information technology that have occurred since these principles were first developed.

While applying these principles to the current data landscape and the context of counterterrorism poses challenges, FIPs provide a remarkably sound basis for analyzing the issues associated with data mining: what information is being collected, how long will it be kept, how accurate and reliable is the information, how will an individual be able to correct erroneous information, what are the redress and enforcement mechanisms?

II. What Are the Kinds of Data Mining and What Risks Do They Pose?

Policy discussions about data mining often suffer from a lack of clarity about key terms and concepts. An informed policy discussion requires an understanding of the different ways the term "data mining" is used and the risks to privacy and due process associated with different applications of data analysis tools. For simplicity's sake, it may be useful to identify two different kinds of data mining: subject-based, which seeks information about a particular individual who is already under suspicion; and pattern-based (sometimes referred to as predictive) data mining, which seeks to find a pattern, anomaly or signature among oceans of personal transactional data.[1] As a general matter, the value of subject-based approaches is more readily apparent, and there are fewer privacy concerns associated with data searches that begin with particularized suspicion.[2] Throughout our testimony today, we focus on pattern-based data mining in the counterterrorism context.

Pattern-based data mining does not begin with any particularized suspicion. Rather, it searches large databases containing transactional information on the everyday activities of millions of people in an attempt to determine the level of risk associated with individuals or to find patterns that may indicate terrorist behavior. Some proponents of data mining have suggested that the searches may be based on no more than a hypothetical set of assumptions about how terrorists behave.

In the counterterrorism field, we must be careful in the adoption of any data analysis tool. The consequences to individuals of being mistakenly designated as a possible terrorist or an associate of terrorists can be devastating and can include arrest, deportation, loss of a job, more intrusive investigation, discrimination, damage to reputation and a lifetime of

---

[1] Strictly speaking, some would say that only the latter is data mining. The Government Accountability Office, after surveying the technical literature, focused specifically on what we call "pattern-based" data mining when it defined data mining as "the application of database technology and techniques—such as statistical analysis and modeling—to uncover hidden patterns and subtle relationships in data and to infer rules that allow for the prediction of future results." GAO, "Data Mining: Federal Efforts Cover a Wide Range of Uses," GAO-04-548 (May 2004). However, in public policy circles, the term data mining has been broadly used. For more on the varities of data mining from a policy perspective, see Mary DeRosa, "Data Mining and Data Analysis for Counterterrorism," CSIS (March 2004); James X. Dempsey and Lara M. Flint, "Commercial Data and National Security," *The George Washington Law Review*, Vol. 72, No. 6 (August 2004).

[2] As others have noted, "the power of data mining technology and the range of data to which the government has access have contributed to blurring the line between the subject- and pattern-based searches . . . [e]ven when a subject-based search starts with a known suspect, it can be transformed into a pattern-based search as investigators target individuals for investigation solely because of their connection with the suspect." U.S. Department of Defense, Report of the Technology and Privacy Advisory Committee (TAPAC), "Safeguarding Privacy in the Fight Against Terrorism," p. 45 (March 2004) <http://www.cdt.org/security/usapatriot/20040300tapac.pdf>.

suspicion, with little or no opportunity for redress or correction of errors. False leads also have serious consequences for national security, diverting resources from true threats.

Currently, there is little evidence of the efficacy of pattern-based data mining in the antiterrorism context. Indeed, there is substantial reason to believe that the technique will not prove useful in identifying terrorists, but will instead lead to significant violations of civil liberties. As experts have explained, the sample of known terrorists whose behavior can be studied is statistically insignificant to identify an unusual or unique pattern of behavior.[3] Any pattern-based search based on characteristics drawn from such a sample will make it difficult to separate the "noise" of innocent behavior from the "signal" of terrorist activities, leading innocent behavior to be viewed as suspicious.[4] When unproven pattern data mining algorithms are applied to the records of millions of people, the false positive rate can be higher than 99%, potentially subjecting large numbers of law abiding citizens to a range of consequences, often with little recourse. The danger of false positives is exacerbated by well-recognized problems with data quality, not only in government databases but also in data drawn from commercial sources. However, under current rules, once the data is collected and analyzed, there are few if any effective controls within the government to prevent inaccurate information from being widely disseminated and used for other purposes.

III. The Changed Legal and Technological Landscape

In the past, the government by and large collected data on one person at a time (i.e., with particularity), either in the course of administering a government program or where there was some suspicion that a person was engaged in criminal conduct, terrorism or intelligence activity. The government was authorized to keep this data for long periods of time, and to retrieve, share and analyze it for compatible purposes without serious controls. However, before it could take action based on that data, the government was bound by procedural due process principles of notice and an opportunity to respond. In the traditional data environment, the greater the consequences for the individual, the greater the due process requirements. For example, the criminal due process standards in the Bill of Rights place the burden of proof on the government and force it to disclose all of its evidence to the accused, for challenge.

Now, in contrast, Section 215 of the PATRIOT Act, the expanded National Security Letter authorities, the growing implications of the Supreme Court's "business records" decisions (which place most commercial data outside the protections of the Fourth Amendment), the President's claims of inherent power, and the nature of technology itself can result in the wholesale collection of data and databases by the government without particularized suspicion. Yet the traditional rules on storage and use remain in

---

[3] Jeff Jonas and Jim Harper, "Effective Counterterrorism and the Limited Role of Predictive Data Mining," Cato Institute (December 11, 2006).

[4] DeRosa, *supra* note 1, at 15.

place, permitting the government to keep that data forever and to go back to it for further analysis (e.g., data mining) with little legal constraint.

Meanwhile, many traditional limits on information sharing have been removed. The wall between intelligence and law enforcement is down. The Executive Branch is moving forward with development of the Information Sharing Environment.[5] State and local information sharing and analysis centers are proliferating. The Justice Department is developing its own information sharing system to make millions of law enforcement investigatory records available to state and local police.[6] The Administration has been expansive in exempting law enforcement and intelligence systems from the Privacy Act.[7]

We must stress that information sharing to prevent terrorism and for other governmental purposes is generally desirable. But, especially in the counterterrorism context, a major shift in the data collection and use landscape is taking place without a suitable privacy and due process framework. The detailed guidelines called for by the Markle Foundation Task Force on National Security in the Information Age, the Defense Secretary's Technology and Privacy Advisory Committee and others have not been issued yet, and existing privacy laws are not up to the task. Yet the government is moving ahead with screening and risk assessment programs. At the same time, the government is claiming the power to make highly consequential decisions about people, cut off from the normal checks and balances: for example, deporting immigrants on the basis of secret evidence, holding individuals for extended periods as "material witnesses," incarcerating hundreds of people at Guantanamo and elsewhere without fundamental due process, and even asserting the power to imprison citizens without the protections of the criminal justice system.

The impact of this "perfect storm" of technological innovation, increased government power and outdated legal protections is well illustrated by the government's recent acknowledgement that it is, through its "Automated Targeting System," collecting travel records on all American citizens entering and leaving the country, assigning risks scores to those citizens, and keeping the records for 40 years. In the current legal environment (we address the failure of the Customs and Border Patrol to comply with the Privacy Act later in our testimony), all those records (including the secret risk score) may be freely shared with other federal agencies engaged in a wide range of activities and also accessed by various state and local law enforcement agencies. In this context, a risk score developed for border screening purposes could easily migrate to other uses (years after

---

[5] The Information Sharing Environment (ISE) was mandated by Section 1016 of the Intelligence Reform and Terrorism Prevention Act of 2004. In November 2006, the ISE Program Manager issued an Implementation Plan for the ISE, marking an important milestone in its development.

[6] Dan Eggen, "Justice Dept. Database Stirs Privacy Fears," *Washington Post* (December 26, 2006).

[7] *See, e.g.*, Transportation Security Administration, Notice of Privacy Act System of Records, 68 Fed. Reg. 2101 (January 15, 2003); Department of Homeland Security, Notice of Privacy Act System of Records, 71 Fed. Reg. 64543 (November 2, 2006).

the citizen was determined not to be a threat) and result in a host of consequences where the individual would find it impossible to respond.

The Privacy Act of 1974 was intended to subject government agencies that collect personally identifiable information to the Fair Information Practices. It was intended to require notice to and consent from individuals when the government collects and shares information about them, give citizens the right to see whatever information the government has about them, and hold government databases to certain accuracy standards. While those practices remain highly relevant today, the Act is increasingly impotent to address the modern data sharing environment. For one, the Act's exemptions for law enforcement and intelligence data have been interpreted in a manner that neuters the Act. Second, the Act's protections only apply to federal "systems of records." That means that the government can bypass the Privacy Act by accessing existing private sector databases, rather than collecting the information itself. Currently, when it accesses commercial databases, the government need not ensure (or even evaluate) the accuracy of the data; it need not allow individuals to review and correct the data; and the government is not limited in how it interprets or characterizes the data.

Finally, and most remarkably, unless the courts and Congress respond, the legal and technological changes of the past decade could spell the effective end of key protections associated with the Fourth Amendment. Traditionally, as we all know, to search the intimate details of one's life the government required a judicial warrant, issued on a finding of probable cause to believe that a specific crime was being committed and naming with particularity the person or place to be searched and the items to be seized. In the 1970s, before the digital revolution and all it has entailed for the creation of electronic databases about our daily lives, the Supreme Court held that the Fourth Amendment does not apply to personal information contained in records held by third parties, with the result that the government could acquire it without meeting Fourth Amendment requirements of probable cause, particularity and notice.[8]

CDT questions the continued viability of these business records cases, for they were decided, by and large, in the context where the government was still collecting information one person at a time, usually in the course of criminal investigative activity where the individual would eventually have a robust set of due process protections. And previously, although the government could keep that data and retrieve it and use it in subsequent investigations, its ability to do so was severely limited by practical realities of incompatible data formats and limited search technology. In today's data mining context, the government is accessing entire buckets of data without a warrant and without particularized suspicion – some by purchase or subscription, some from files generated in the course of other government activities, and some by the forced disclosure of datasets using NSLs or other instruments. If this information on presumptively innocent people, having been acquired without Fourth Amendment protections, can be kept forever and analyzed at will without probable cause or individualized suspicion, what would be left of the Fourth Amendment?

---

[8] *U.S. v. Miller*, 425 U.S. 435 (1976); *Smith v. Maryland*, 442 U.S. 735 (1979).

CDT believes that the business records cases are inapplicable to the modern data environment and we are at the beginning of a long project to urge the courts and legislatures to re-examine them. The Supreme Court itself has made it clear in a related context that the law must advance with the technology to ensure the continued viability of the Fourth Amendment.[9] We also believe that other trends in Supreme Court rulings indicate that the analysis of opaque datasets are "searches" for Fourth Amendment purposes, just the search of a computer lawfully in the hands of the government is itself a separate "search" under the Fourth Amendment. For all of these reasons, we believe that the automated, pattern-based analysis of massive databases should be recognized as a search within the definition of the Constitution. Even if, for the sake of argument, the initial collection of the information as part of a database is not considered a search for Fourth Amendment purposes, once the government applies analytic techniques to extract from it meaning not readily apparent on the surface, the law should, at a minimum, consider that analysis a new search that requires procedural protections.[10] For now, the law is not there. Hence, there are huge gaps in privacy law, and Congress needs to respond.

-- Comparison With the Commercial Sector

While there is no comprehensive privacy law controlling the collection and use of personally identifiable information by the private sector, the private sector is still subject to more robust privacy protections than the government. Sector-specific privacy rules such as the Fair Credit Reporting Act place comparatively strict controls on private entities engaging in profiling or risk assessment.[11] When commercial data analysis could have adverse implications for a person's credit, insurance or employment, the private sector is under a legal obligation to use only accurate data, individuals have a right to access and challenge data about them, and individuals must be given notice and an opportunity to respond before adverse action is taken.

Furthermore, while the private sector uses pattern-based data mining to detect fraud, it has a large baseline of known frauds that can be used to develop and constantly refine risk assessment models. In contrast, agencies searching for a terrorist signature have a very small sample set on which to base their predictions.

---

[9] *See Kyllo v United States,* 533 U.S. 27 (2001).

[10] Lee Tien, "Privacy, Technology and Data Mining," 30 Ohio N.U. L. Rev. 389 (2004).

[11] Sectoral privacy laws that apply to personal data held by the private sector include, *inter alia*, the Fair Credit Reporting Act of 1970, 15 U.S.C. § 1681 (sets out rights for consumers with respect to credit information), the Family Educational Rights and Privacy Act of 1974, 20 U.S.C §1232g (governs access to personally identifiable information in educational records held by federally funded educational institutions); the Health Insurance Portability and Accountability Act of 1996, P.L. 104-191§ 264 (requires issuance of a privacy rule for individually identifiable health information); and the Right to Financial Privacy Act of 1978, 12 U.S.C.§ 3414 (sets out procedures for the federal government's access to financial institutions customer records).

The contrast is striking. While the private sector is subject to strict rules for at least some of its data mining activities, we have not yet devised a suitable set of rules for government data mining, where constitutional liberties are at stake and the consequences of error are much higher.

IV. What We Know So Far About Government Data Mining Illustrates the Need for Closer Oversight and Control

Little is known about the full extent of pattern-based data mining for counterterrorism and homeland security. While Congress has broadly authorized collection of data under extraordinarily low standards in the USA Patriot Act and authorized data sharing among law enforcement and intelligence agencies, our understanding of the data mining activities that these changes in the law have encouraged remains limited. Since the disclosure of the existence of the Total Information Awareness Program ("TIA") in 2002, there has been a steady stream of revelations about other data mining programs that raised concerns about privacy and efficacy, including CAPPS II.[12] With each new revelation, Congress has scrambled to respond, ultimately de-funding some elements of the TIA program[13] and postponing the deployment of the CAPPS II or Secure Flight airline passenger screening program until the GAO reported that the system had met certain reliability and privacy requirements.[14]

Yet, notwithstanding public and Congressional discomfort with these programs, they continue to proliferate without any apparent controls. Just last month, the Customs and Border Patrol ("CBP") acknowledged that, without notice and in violation of the Privacy Act, it has been using the Automatic Targeting System (ATS), which was designed to screen shipping cargo, to conduct "risk assessments" on tens of millions of travelers,

---

[12] In 2004, the GAO reported 199 data mining efforts, of which 68 were planned and 131 were operational.[12] The programs spanned 52 agencies and departments. Out of all 199 data mining efforts identified, 122 used personal information. The uses of data mining included improving service or performance, detecting fraud, waste, and abuse, analyzing scientific and research information, managing human resources, detecting criminal activities or patterns, and analyzing intelligence and detecting terrorist activities. GAO, "Data Mining: Federal Efforts Cover a Wide Range of Uses," GAO-04-548 (May 2004). For detailed descriptions and analysis of current antiterrorism and homeland security data mining programs, see "Data Mining and Homeland Security: An Overview." Congressional Research Service (January 27, 2006); "Survey of DHS Data Mining Activities," Department of Homeland Security Office of Inspector General (August 2006); "Privacy: Total Information Awareness Programs and Related Information Access Collection and Protection Laws" Congressional Research Service (February 14, 2003).

[13] Congressional action did not actually end many of the program's components; they moved into classified environments. Shane Harris, "TIA Lives On," National Journal (February 26, 2006) p. 66.

[14] Section 514, Department of Homeland Security Appropriations Act, 2007, Pub. L. 109-295; Section 522, Department of Homeland Security Appropriations Act 2005, Pub. L. 108-334.

including U. S. citizens. These "risk assessments" determine whether individuals will be subject to more invasive searches. No Privacy Act notice was issued before the focus of the massive data program was turned towards individuals nor was a Privacy Impact Assessment conducted before initiating the program as required by the E-Government Act of 2002.[15] An "after the fact" privacy assessment fails utterly to address the risks posed by the system.

But for the recent ATS Privacy Act notice, which boldly asserts unprecedented uses of the "routine use" exception, sweeping exemptions for law enforcement and intelligence investigations, and wide sharing of the data for a wide variety of uses wholly unrelated to border security, it is unclear whether or when Congress would have been made aware of the program. The danger to the rights of Americans under this program is self-evident. It is not hyperbolic to assume that the data---including the secret risk scores--- will find its way through the government and down to the state and local where it can easily be abused.

VI. Recommendations: What Congress and the Executive Can Do to Create a More Balanced Framework for Data Mining

A. Transparency and Congressional Oversight

Non-partisan congressional oversight is one of the pillars of a system of checks and balances. Congress has a critical role to play in ensuring that pattern-based data mining programs are effective and protect civil liberties. The first step, of course, is for Congress to get a comprehensive and accurate picture of the data mining activities of the federal government. While Congress, on its own and through the GAO, has conducted some oversight of data mining and has used that oversight to impose some constraints on particular programs through the budget process, Congress' response has largely been reactive, driven by revelations about excesses in particular programs rather than by facts developed during comprehensive and consistent oversight. Congress and – to the extent possible – the American people need to know what programs are being developed and deployed, whether those programs are likely to be effective, and what risks those programs pose to the rights of the American people. Congress should hold public oversight hearings with testimony from the Executive Branch, and should conduct annual reviews of data mining programs and issue public reports on the effectiveness of data mining in counterterrorism programs and its impact on privacy and other civil liberties.

As a first step toward developing a more balanced framework for data mining, CDT believes that the relevant agencies should be required by law to report in detail on pattern-based data mining programs that are being developed or deployed, and to provide assessments of each program's efficacy and impact on civil liberties.

---

[15] E-Government Act of 2002 [H.R. 2458] Pub. L. 107-347 (December 17, 2002).

B. Prior Congressional Authorization for Data Mining Programs

CDT believes that Congress should go further and expressly limit deployment of pattern-based data mining in law enforcement and antiterrorism contexts, by requiring an authorization based on a showing of effectiveness before a program is launched against U.S. citizens. In essence, we are proposing that the language Congress, on a bi-partisan basis, has adopted and annually renewed since FY 2005 for Secure Flight be applied government-wide. Under the approach we are proposing, research and development would be permitted without express prior authorization, under the careful oversight of Congress. In addition, as Congress mandated for Secure Flight, pattern-based programs should not be authorized until and unless there is in place a set of guidelines for data sharing and mining that protect privacy and ensure due process. While it is the job of the Executive Branch in the first instance to adopt adequate government-wide guidelines, that job has not yet been accomplished. In the absence of detailed and comprehensive Executive Branch guidelines, Congress may need to step in and legislate guidelines.

C. The Elements of Effective Guidelines

The elements of a set of robust and workable guidelines for information sharing and analysis have already been outlined in specific laws adopted by Congress and in leading studies, notably the three reports of the Markle Foundation Task Force on National Security in the Information Age.[16]

Congress has already legislated on some of the elements of a sound framework for data analysis in the limitations it placed on implementation of the Secure Flight passenger screening system[17] and in the rules it established for improvements in the government's terrorist "watch lists."[18] Drawing upon these laws, the Markle Task Force reports, and experiences in the commercial sector, one can develop a detailed set of guidelines that include the following elements:

- A concept of sharing that leaves information with the originator, using directories and search techniques that permit discovery and sharing of relevant information but minimize unnecessary transfers of data to central repositories.
- Strong data quality standards, including minimum standards for watchlists, and other procedures to ensure that the databases the government uses to establish the identity of individuals or make assessments about individuals are sufficiently

---

[16] "Mobilizing Information to Prevent Terrorism: Accelerating Development of a Trusted Information Sharing Environment" (July 13, 2006); "Creating a Trusted Network for Homeland Security (December 2, 2003); "Protecting America's Freedom in the Information Age (October 7, 2002), available at http://www.markletaskforce.org/.

[17] Section 514, Department of Homeland Security Appropriations Act, 2007, Pub. L. 109-295; Section 522, Department of Homeland Security Appropriations Act 2005, Pub. L. 108-334.

[18] Section 4012(c), Intelligence Reform and Terrorism Prevention Act of 2004, Pub. L. 108-548, 118 Stat. 3638, 3718.

accurate and reliable that they will not produce a large number of false positives or unjustified adverse consequences.
- Corrective mechanisms, including assessments of the reliability of commercial databases and automated mechanisms that can identify and correct errors in shared data, with responsibility on both the originator and the recipient of data.
- Access controls, security measures and permissioning technologies that can protect against improper access to personal information, including the ability to restrict access privileges so that data can be used only for a particular purpose, for a finite period of time, and by people with the necessary permissions.
- Automated and tamper-proof audit trails that can protect against misuse of data, improve security, and facilitate oversight.
- Redress mechanisms that allow individuals to respond when they are about to face adverse consequences based on information. This includes the right to challenge inaccurate information.
- Effective oversight of the use and operation of the system, including privacy officers with sufficient powers and resources to enforce the guidelines.

While technology is no substitute for policy, various commercially-available technologies can help implement and enforce these policies. Auditing technology can provide built-in recordation and documentation capabilities to track how information is used and shared. Technologies can help assure that information is up-to-date. Software can ensure that information is updated regularly and that it is unusable after a certain date if not refreshed. Other technology can permit users to track where information came from and who received it and alert users if the original data is subsequently disproved or corrected. Anonymization technologies can minimize unnecessary disclosure of personal information when not needed.

D. Apply Fair Information Practices to Commercial Databases Accessed For Pattern-Based Data Mining

Congress should legislate to ensure that commercial databases accessed for data mining are subject to strong privacy rules. Congress should make clear that the Privacy Act applies whether the government is creating its own database or acquiring access to a database from a commercial entity. In addition, Congress should require Privacy Impact Assessments for the acquisition of commercial databases. Section 208 of the E-Government Act of 2002 already requires a PIA if the government initiates a new "collection" of information. The same process should apply when the government acquires access to a commercial database containing the same type of information that would be covered if the government itself were collecting it.

In addition, Congress should require the government to perform an accounting of private sector databases before using them and to publish in the Federal Register a description of the database, the name of the entity from which the agency obtained the database and the amount of the contract for use of the database. Agencies should further be required to adopt regulations that establish fair information practices including a process for redress. Finally, Congress should require agencies to incorporate provisions into their contracts

with commercial entities provisions that provide for penalties when the commercial entity sells information to the agency that the commercial entity knows or should know is inaccurate or when the commercial entity fails to inform the agency of corrections or changes to data in the database.[19]

These approaches that have been proposed strike a balance between the government's need for information and the privacy interests of individuals. Adapting the Privacy Act and Fair Information Principles to government uses of commercial databases would go a long way toward closing the unintended gap in privacy protection that exists under the current law.

E. Strong Internal Mechanisms for Accountability and Oversight

Congress has created Chief Privacy Officers for the Departments of Homeland Security and Justice and for the office of the Director of National Intelligence. The independence and authority of these officers should be improved. If taken seriously, Privacy Act notices and Privacy Impact Assessments can help in raising and mitigating privacy concerns surrounding the government's use of personal information. Inspectors General should also have a role to play. Inspectors General, in particular, provide a critical internal ability to identify civil liberties violations, and should regularly review agency actions to assess their privacy implications.

VI. Conclusion

The Center for Democracy and Technology appreciates the opportunity to present its views on government data mining. Our nation is at a critical moment on this issue. As the ATS revelations indicate, pattern-based data mining is moving forward in the Executive Branch without a legal framework that will protect the privacy and due process rights of Americans. Congress needs to ensure that the proper legal and policy framework is in place before these programs move forward, and limit their deployment to those with proven effectiveness. Oversight and accountability, done right, will benefit both national security and civil liberties. Checks and balances result in clear lines of responsibility, well-allocated resources, protection against abuse, and the ability to evaluate and correct past mistakes. Appropriate, well-implemented accountability mechanisms will help to ensure that systems are effective as well as protective of due process.

---

[19] A number of bills were proposed in the 108th and 109th Congresses that incorporate many of these concepts. For example, S.1484, the "Citizens Protection in Federal Databases Act," sponsored by Sen. Wyden in the 108th Congress; S.1789, the "Personal Data Privacy and Security Act of 2005," sponsored by Sens. Leahy and Specter; and S.1169, "The Federal Agency Data Mining Reporting Act of 2005," sponsored by Sens. Feingold, Sununu, Leahy, Akaka, Jeffords and Wyden.