

February 7, 2007

The Honorable Patrick Leahy
Chairman
Senate Judiciary Committee
Washington, DC 20510

Dear Chairman Leahy,

Thank you again for the opportunity to testify before the Committee on January 10 about the implications of government data mining. We also thank you, Senator Kennedy and Senator Specter for submitting follow-up questions, asking us to elaborate on the important issues raised at the hearing.

All of our answers should be read in the context of the statement in our written testimony about the broad way in which the term “data mining” is used. As we stressed in our testimony, one cannot be either for or against data mining. It is a tool for data analysis. The important 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?

Answer to Chairman Leahy’s question #1.A:

Yes, Congress should consider legislation to place limits on Governmental access to third-party records.

As we stated in our prepared testimony, 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. This reform could be accomplished by amending Subsection (m) of the Act to apply to all PII acquired by the government from private sector information services providers. 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 order to improve the utility of PIAs, Congress should require, as a general rule, that they be publicly issued some period of time (such as 60 days) before a program is launched.)

In addition, Congress should require the government to perform an audit 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 when it acquires information from the private sector for use in making decisions about individuals.

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.

A number of these ideas are reflected in the Personal Data Privacy and Security Act, introduced in this Congress by Senators Leahy and Specter, which CDT strongly supports.

Additional legislative reforms are needed to address the very low standards for compulsory governmental access to third-party records. In particular, Congress should strengthen the standards for issuance of National Security Letters and orders under Section 215 of the PATRIOT Act. The bi-partisan SAFE Act, S. 737 in the 109th Congress, is an excellent starting point for those reforms; it should be reintroduced and given priority consideration.

Answer to Chairman Leahy’s question #1.B:

Yes, the wall between the government and the private sector has been eroded. CDT would not say that data brokers should be considered quasi-governmental, but we do agree that information services companies should be subject to a comprehensive baseline federal privacy law. Unfortunately, Congress has not acted since the Committee’s April 2005 hearing, “Securing Electronic Personal Data: Striking a Balance Between Privacy and Commercial and Governmental Use,” which examined the roles and responsibilities of information services companies. CDT testified at that hearing and offered five main recommendations:

1. As a first step towards preventing identity theft, entities, including government entities, holding personal data should be required to notify individuals in the event of a security breach.
1. Since notice only kicks in after a breach has occurred, Congress should require entities that electronically store personal information to implement security safeguards, similar to those required by California AB 1950 and the regulations under Gramm-Leach-Bliley.
1. Congress should impose tighter controls on the sale, disclosure and use of Social Security numbers and should seek to break the habit of using the SSN as an authenticator.
1. Congress should address the federal government’s growing use of commercial

- databases, especially in the law enforcement and national security contexts.
1. Finally, Congress should examine the “Fair Information Practices” that have helped define privacy in the credit and financial sectors and adapt them as appropriate to the data flows of this new technological and economic landscape.

These ideas are reflected in the Personal Data Privacy and Security Act, introduced in this Congress by Senators Leahy and Specter, which CDT strongly supports.

Answer to Chairman Leahy’s question #1.C:

Yes, it is possible to balance the government’s legitimate need for information and our most important freedoms. The proposals in the Personal Data Privacy and Security Act (110th Congress) and the SAFE Act (109th Congress) reflect this necessary balance. Those bills would protect privacy and strengthen the national security and law enforcement.

Answer to Chairman Leahy’s question #2:

Yes, it is possible to strike a meaningful balance between privacy and security in government data mining programs. In fact, it is necessary if we are to improve security. Privacy protection, checks and balances, accountability and redress are not incompatible with security. To the contrary, clear guidelines and oversight mechanisms are part of the solution. As the 9/11 Commission stated: "The choice between security and liberty is a false choice." The shift in government power and authority that is occurring in response to terrorism, the 9/11 Commission concluded, "calls for an enhanced system of checks and balances to protect the precious liberties that are vital to our way of life."

This conclusion - that privacy protection and accountability must be built into the design and implementation of counterterrorism information sharing systems -- is central to the recommendations of the Markle Task Force on National Security in the Information Age and other bipartisan expert bodies that have carefully studied information technology and its role in fighting terrorism. "We must not sacrifice liberty for security," concluded the Technology and Privacy Advisory Committee (TAPAC) appointed by Secretary of Defense Rumsfeld to study the Total Information Awareness program and related activities. Likewise, the Advisory Panel to Assess Domestic Response Capabilities for Terrorism Involving Weapons of Mass Destruction, chaired by former Virginia Governor James Gilmore, repeatedly stressed that personal freedoms must be at the foundation of the nation's efforts to counter terrorist threats.

Answer to Senator Kennedy’s question #1:

We do not believe that the line between “punishing” and merely developing leads for further investigation is as clear as Mr. Taipale suggests. There are many ways in which a government can “punish” a person. Indeed, it is accepted as a matter of First Amendment law that being targeted for investigation can itself have a chilling effect on fundamental

freedoms. And wiretapping is clearly an intrusion on Fourth Amendment rights, so the use of data mining as the trigger for wiretapping would clearly impose a harm on an individual. Moreover, the Executive Branch has been increasing the ways in which it seriously disrupts persons' lives without inflicting punishment in the context of criminal prosecution. Would the use of data mining to generate investigative "leads" which were then pursued by coercive interrogation be on the punishment side of the line or the investigative side of the line? How about something as common as being repeatedly stopped at the airport for secondary screening?

See also our answer to Senator Specter's question #3, where we challenge Mr. Taipale's assumption that the use of data mining results in the courtroom is more objectionable than the use of data mining results for investigative or screening purposes. The use of the results of data mining to punish, assuming such punishment is not extrajudicial, would, in many ways, be more subject to checks and balances than would the use of data mining for screening or investigative purposes.

Answer to Senator Kennedy's question #2.a:

We don't have a clear picture of how the Administration is using data mining, so it is hard to cite concrete examples of demonstrable harm that has resulted from data mining.

We do know, however, that the Administration has relied on seriously erroneous data and faulty analytic tools in some of its key security programs, resulting in demonstrable harm to individuals. Some of the most notorious cases involve the watch lists maintained by the government and their use in screening passengers at airports. Senator Kennedy himself has been incorrectly associated with someone on the watchlist. While the Senator brushed off the inconvenience, others, such as the famous David Nelson of Alaska, has suffered genuine harm in terms of disrupted travel and business plans. One mistake recently revealed involved Sen. Ted Stevens, R-Alaska, whose wife, Catherine, was being identified as "Cat" Stevens and frequently stopped due to confusion with the former name of the folk singer now known as Yusuf Islam, whose name is on the list. The GAO found last year that about half of the tens of thousands of potential matches sent to the Terrorist Screening Center between December 2003 and January 2006 for further research turned out to be misidentifications. Most of these people experienced at least the inconvenience of secondary search.

At the other end of the spectrum is Maher Arar, a Canadian citizen detained in the US on the basis of faulty information and removed to Syria, where he was tortured. After Arar returned to Canada, an investigation was conducted. Last month, Canadian Prime Minister Stephen Harper called on the U.S. government to remove Arar from any of its no-fly or terrorist watchlists, saying "We think the evidence is absolutely clear and that the United States should in good faith remove Mr. Arar from the list."

Answer to Senator Kennedy's question #2.b:

Traditional police stops are subject to a series of protections lacking in the data mining context: The person subjected to the traditional police stop receives immediate notice - the police officer comes up to him and tells him he has been singled out. The scope of the policeman's search is limited: he cannot, for example, look inside a person's luggage - he has to ask for consent or get a warrant. Moreover, the innocent person subject to a traditional police stop has immediate recourse to conclusively clear his name - he opens his luggage and empties his pockets and proves he has no drugs, in which case he is free to go and no adverse record is kept. In the data mining context, the government provides no notice, it denies access to the risk score, so there is no opportunity for a person to clear himself, and the adverse inference may linger for a very long time (40 years in the case of ATS). The protections available in the traditional police stop make it a "reasonable" search, while their absence in the data mining context makes the search unreasonable.

Answer to Senator Kennedy's question #3:

At one level, data mining might be seen as "color blind" or blind to ethnicity and religion. One would hope that government agents would not use overtly ethnic parameters for data analysis in the absence of a specific lead. (The FBI instituted a census of mosques in 2003, and it was reported in December 2005 that FBI agents had been secretly monitoring radiation levels at Islamic mosques, businesses and homes for several years in large cities to determine whether nuclear or chemical bombs were being assembled - no suspicious radiation levels were found.) But it is easy to see how a pattern-based analysis could use factors that are a substitute for ethnicity or religion. For example, a traffic analysis program targeting between the US and an Arab country will inevitably target the calls of Arab-Americans with relatives and legitimate business connections in that country.

CDT has proposed safeguards that could help prevent discriminatory profiling. One approach is what we call "section 215 with teeth." As you know, Section 215 is a provision in the PATRIOT Act giving the government access to commercial data under a very weak standard. An amended section 215 could require a judicial finding, based on facts shown by the government, that there is a reason to believe that terrorist activity is afoot fitting a certain pattern, and that reliable information relevant to the interdiction of that activity would likely be obtained from the search of one or more commercial databases. Under this approach, an agency that had intelligence information about a possible future attack and that wanted to run a pattern-based search to identify potential planners would be required to demonstrate to the court: (1) facts giving reason to believe that a threat existed displaying certain characteristics; (2) a description of the databases that the government wants to search, including an assessment of the sensitivity of the data involved and its accuracy and reliability; (3) an explanation of why other methods of investigation were inadequate; and (4) a statement indicating whether the commercial databases would remain under the control of the commercial source or whether they would be acquired by the government. Among other things, this approach would give the

court the opportunity to determine whether ethnic or religious profiling was an impermissible part of the government's proposed search.

Answer to Senator Specter's question #1:

Mr. Taipale says, "However, in counterterrorism applications patterns can be inferred from lower-level precursor activity—for example, illegal immigration, identity theft, money transfers, front businesses, weapons acquisition, attendance at training camps, targeting and surveillance activity, and recruiting activity, among others." If the government has a list of people who attended training camps, that alone gives it the basis for collecting pretty much whatever data it wants about those people and to collect a fair amount of data about those who are closely associated with them. This is not the kind of precursor activity from which one needs to discover some obscure pattern. The same is true of those engaged in "recruiting activity." On the other hand, if the government tries to compile a list of all illegal aliens who transfer money overseas, it is likely to get an undigestable number of leads. However, if the government could run an analysis for all illegal aliens (we're not sure such a list exists) engaged in identity theft who run "front businesses," make money transfers to Pakistan, and possess a lot of weapons, that might in fact be a justifiable "data mining" program. So far, as far as we know, the government has not shown that it has the kind of data that would support such an analysis. Like much of the discussion of data mining, Mr. Taipale's example seems highly speculative.

Answer to Senator Specter's question #2:

In CDT's view, the standard is not perfection. Rather the standard is: does the program materially assist in the pursuit of a mission (keeping terrorists off airplanes, keeping terrorists from entering the country), without high levels of collateral damage to civil liberties, to the extent that in a world of limited resources, the program deserves to made a priority over other efforts that would serve the same mission. That's not a mathematical formula, but we believe it is better than anything the government is applying today to decide which data mining programs to launch. With such a standard, the government would not be precluded from deploying data mining technology. Rather, it would be empowered to deploy data mining technology that meaningfully advances the national security

Answer to Senator Specter's question #3:

Yes, recognizing that "data mining" is a very broad term that may include intuitively uncontroversial data analysis techniques, one should distinguish between using data mining as an evidentiary tool in a court of law as opposed to an investigative or screening tool. Some "data mining" techniques might be perfectly suited to the analysis of evidence for presentation in the courtroom. For example, it might be appropriate to apply data analysis techniques to the large amount of data collected with a court authorized pen register, and to introduce those results in a courtroom to illustrate a chain of events of circumstantial significance.

However, the key point to recognize is that, in the courtroom, use of data mining for evidentiary purposes would be subject to vigorous cross-examination, presentation of contrary evidence and the other due process protections afforded in the trial setting. Among other protections, there is full notice of the use of the technique. If the matter were criminal in nature, the burden of proof would be on the government. The data mining technique itself might be subject to scrutiny under Federal Rule of Evidence 702, which charges the federal courts with the responsibility of acting as gatekeepers for all scientific and expert testimony. The threshold question for introduction of the evidence is reliability (or, as we stressed in our testimony, “efficacy”).

None of these protections are available in the screening or investigative contexts, so in some ways data mining is riskier in those contexts. As we said in our written testimony, application of data mining in the investigative or screening contexts must be preceded by an independent assessment of the reliability or effectiveness of the technique. There should also be notice, beginning with the kind of generic notice that would be provided by Senator Feingold’s bill. Redress procedures must be adopted so that individuals can challenge false inferences drawn about them and correct faulty information.

The differences between the protections that would be available when data mining is used as an evidentiary tool and the current lack of those protections when it is used for investigative or screening purposes argues for the position CDT took at the hearing: Congress should use the power of the purse to prohibit the use of unauthorized data mining (defined as predictive or pattern-based scans of large sets of data, where the goal is to assign risk scores or find individuals whose behavior matches some pattern believed to be associated with terrorist or criminal behavior). If the Executive Branch thinks it has an effective program, it should come forward and tell Congress, explain the program and get the money for it. Congress has already put that limit on implementation of the risk assessment program “Secure Flight.” CDT urges Congress to do the same across the board.

Answer to Senator Specter’s question #4(a):

Your question asks whether requiring the government to demonstrate an application’s absolute effectiveness before permitting its use would interfere with or prohibit innovation. It might, but we know of no one who has proposed “absolute effectiveness” as the standard for deployment of any technique. That is certainly not CDT’s position. The current posture of the Executive Branch is that it need offer no showing of effectiveness before deploying a technique. That approach is dangerous to civil liberties and national security. As we stated above in answer to your question #2, we believe that a workable standard would be whether the program materially assists in the pursuit of a mission (keeping terrorists off airplanes, keeping terrorists from entering the country), without high levels of collateral damage to civil liberties, to the extent that in a world of limited resources, the program deserves to be made a priority over other efforts that would serve the same mission. Ultimately, it would be a judgment call. We believe

Congress, as the appropriator, should have a role in that judgment. Right now, as far as we can tell, that judgment is not made on a systematic basis by the Executive branch and is certainly made without Congressional input.

Answer to Senator Specter's question # 4(b):

The development and publication of authorized procedures or prohibitions for data mining could and should be done without enabling countermeasures and evasion. (Evasion isn't necessarily a bad thing. A lot of our national counterterrorism program is intended to induce evasion, in the sense that airline screening is intended to induce terrorists to avoid airports, and physical protection measures around important sites are intended to compel terrorists to go elsewhere.)

In our testimony, we outlined several elements of guidelines that could be developed without disclosing anything of use to the enemy:

1. 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 accurate and reliable that they will not produce a large number of false positives or unjustified adverse consequences.
2. 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.
3. 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.
4. Automated and tamper-proof audit trails that can protect against misuse of data, improve security, and facilitate oversight.
5. 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.
6. Effective oversight of the use and operation of the system, including privacy officers with sufficient powers and resources to enforce the guidelines.

CDT has prepared a detailed analysis of guidelines for information sharing issued by the Administration in December 2006. The analysis describes in further detail some of the issues that should be addressed in guidelines, none of which would jeopardize operational effectiveness. <http://www.cdt.org/security/20070205iseanalysis.pdf>.

The Center for Democracy and Technology appreciates this opportunity to discuss in greater detail the important questions surrounding the privacy implications of government data mining. We look forward to working with the Committee as you continue your oversight and legislative work in this area, seeking to develop a more balanced approach to the government's use of information.

Sincerely,

Leslie Harris
Executive Director