TOTAL INFORMATION AWARENESS AND BEYOND

The Dangers of Using Data Mining Technology to Prevent Terrorism

Shannon R. Anderson
Haywood Burns Fellow,
Bill of Rights Defense Committee
241 King St., Suite 216
Northampton, MA 01060
(413) 582-0110
info@bordc.org

Table of Contents

Introduction

In an enlightened 1928 opinion, Supreme Court Justice Brandeis wrote:

> The progress of science in furnishing the government with means of espionage is not likely to stop with wire tapping. Ways may some day be developed by which the government, without removing papers from secret drawers, can reproduce them in court, and by which it will be enabled to expose to a jury the most intimate occurrences of the home.[1]

Unfortunately for the American public, Justice Brandeis's prediction has largely come true. Under the guise of preventing terrorism, the federal government has dramatically increased the tools it uses to obtain private information. This upsurge of government surveillance has increased public concern about privacy and the protections of the Fourth Amendment. While these concerns led Congress to deny funding to Total Information Awareness (TIA)[2] in the Department of Defense Appropriations Act of 2004[3] last September, the federal government has continued to develop large-scale data mining and data surveillance technology systems. Additionally, the online publication Capitol Hill Blue recently reported that TIA has gone underground and is very much "alive and well."[4]

This article (1) surveys current government data mining and data surveillance programs used for counter-terrorism purposes, (2) explains some of the drawbacks of these programs and how they may threaten civil rights and liberties, and (3) suggests how citizens can take action in response to these programs.

I. Government Data Mining and Data Surveillance Initiatives

The Government Accountability Office (GAO), at the request of Senator Akaka, released a report in May 2004 that reviews the various data mining initiatives of 52 executive branch agencies.[5] Of the 199 data mining projects listed, 122 collect and store personal information, and 54 "mine" data from the private sector.[6] Fourteen projects aim at analyzing intelligence information and detecting suspected terrorists,[7] and seven of the fourteen mine personal information. This paper discusses those seven and seven other data mining and data surveillance projects that were not included in the GAO report, beginning with TIA itself.

---

[1] Olmstead v. U.S., 277 U.S. 438, 474 (J. Brandeis *dissenting*).
[2] After public objection, TIA was later renamed to Terrorism Information Awareness. However, the project format or objective did not change. Throughout this paper the acronym TIA refers to either of the project's names.
[3] Pub. L. No. 108-87 § 8131, 117 Stat. 1054, 1102 (2003).
[4] Theresa Hampton & Doug Thompson, *Where Big Brother Snoops on America 24/7*, CAPITOL HILL BLUE, June 7, 2004, *at* http://www.capitolhillblue.com/artman/publish/article_4648.shtml.
[5] GAO-04-548, *Data Mining: Federal Efforts Cover A Wide Range of Uses*, Report to the Ranking Minority Member, Subcommittee on Financial Management, the Budget, and International Security, Committee on Governmental Affairs, U.S. Senate, (May 2004), *available at* www.gao.gov/cgi-bin/getrpt?GAO-04-548. For Senator Akaka's public statement on the report, visit http://www.senate.gov/~akaka/releases/04/05/2004527449.html.
[6] *Id.* at 8.
[7] *Id.* at 13.

Data mining is defined as "the practice of automatically searching large stores of data for patterns…us[ing] computational techniques from statistics and pattern recognition." [8] Data mining differs from data surveillance, which largely involves records keeping and records searching. Data mining involves searching for patterns of behaviors that fit a defined algorithm for suspicious behavior, while ordinary data surveillance does not have that capability. Nevertheless, both types of technology are troubling from a privacy perspective. Therefore, this paper discusses both data mining and data surveillance projects.

Unfortunately, limited information is available on the full extent of these projects or how much they cost taxpayers. David Sobel, the General Counsel for the Electronic Privacy Information Center (EPIC), recently initiated a Freedom of Information Act (FOIA) request to obtain more information on some of the GAO-listed projects, so more information may be available to the public soon.

A. TIA

TIA, a project of the Defense Department's Defense Advanced Research Projects Agency (DARPA), aims[9] to build a centralized database containing private transactional data on all Americans, including "records on credit-card purchases, plane flights, e-mails, websites, [and] housing."[10] In the tradition of the Bush Administration's policy of preemptive action, TIA's goal is to create electronic tools to "better detect, classify, and identify potential foreign terrorists…to increase the probability that authorized agencies of the United States can preempt adverse actions."[11] The government establishes baseline patterns identifying what they see as suspicious behavior, such as buying one-way plane tickets or drastic changes in spending habits, and then conducts pattern-based electronic searches of huge amounts of information to find matches for their trends. The personal information is "mined" from private sector databases as well as government databases.

In September 2003, Congress gave into public objection by denying funding to TIA "or any successor program."[12] Nevertheless, the Pentagon's own Technology and Privacy Advisory Committee (TAPAC) reported that the Defense Appropriations Act that ended TIA funding also "explicitly permitted funding for 'processing analysis, and collaboration tools for counter-terrorism foreign intelligence' specified in a classified, and therefore non-public, annex to the Act."[13] Therefore, "TIA-like activities could be continued to be pursued outside the public's view."[14] Since that time, it has been reported that aspects of TIA went underground and are now funded directly through the Pentagon's budget.[15] In addition, Congress left alone an Advanced

---

[8] Data Mining, *at* http://en.wikipedia.org/wiki/Data_mining.

[9] Although the public does not know which aspects of TIA are still in place, for the sake of simplicity I use present tense to describe the project.

[10] John Allen Paulos, *Privacy, Terrorists, and Science Fiction*, ABC NEWS, Jan. 5, 2003, *at* http://www.abcnews.go.com/sections/scitech/WhosCounting/whoscounting030105.html.

[11] Report to Congress Regarding the Terrorism Information Awareness Program in Response to Consolidated Appropriations Resolution, 2003, Pub. L. No. 108-7, 6 (May 20, 2003), *available at* http://www.epic.org/privacy/profiling/tia/may03_report.pdf.

[12] Dep't of Defense Appropriations Act, *supra* note 3.

[13] Technology & Privacy Advisory Comm., *Safeguarding Privacy in the Fight Against Terrorism*, 39, Mar. 2004, DEP'T OF DEFENSE, 2, *available at* http://www.cdt.org/security/usapatriot/20040300tapac.pdf. Although the report was written in March, it was not publicly released until May. The over 50-page report extensively details government data mining programs and the threat they present to privacy interests.

[14] *Id.* at 3.

[15] Hampton, *supra* note 4. *See also* Feb. 23, 2004 articles, *infra* notes 17, 19.

Research & Development Activity (ARDA) $64 million research grant. ARDA, which sponsors U.S. intelligence research by corporations and universities, administers a research program called Novel Intelligence from Massive Data.[16] ARDA is controversial because it uses some of the same researchers as TIA,[17] and like TIA, it carries out its activities well below the public's radar.

DARPA's TIA project encompassed many different aspects, and no public information exists as to the status of its various programs. For instance, technology systems called Genoa II, Evidence Extraction and Link Discovery (EELD),[18] and Scalable Social Network Analysis (SSNA) were all parts of DARPA's project. CBS News reported in February 2004 that some TIA programs, including EELD, "were transferred to U.S. intelligence offices" for further research and development.[19] However, recent status information, including financial expenditure, is unavailable. In fact, DARPA's Information Awareness Office (IAO), the office that previously administered TIA, is no longer listed as a technical office on the DARPA webpage,[20] and DARPA has not disclosed which office has taken over the transferred TIA programs. Moreover, none of the TIA-related programs were mentioned in the Department of Defense's Fiscal Year 2005 Budget Estimate.[21]

B. CAPPS II

The enhanced Computer Assisted Passenger Prescreening System (CAPPS II) is an automated prescreening system for airports, which "yield[s] a recommended screening level, based on the degree of risk assessed, or specific identifiable terrorist threat."[22] The system uses commercial information and passenger provided flight reservation information.[23] This information, called Passenger Name Record (PNR) data, includes full name, date of birth, home and business addresses and phone numbers, credit card information, and meal information (which can hint to religion). Although CAPPS II uses commercial and other private sector information, the Transportation Security Administration (TSA) assures travelers that all non-government information will be kept outside the government firewall and will be accessed without permanent storage.[24]

Nevertheless, civil rights groups are still concerned. So are members of Congress; they requested a GAO report, which was released in February 2004 and highlighted "significant implementation challenges" of the program.[25] The report warned that TSA has not yet "identified

---

[16] *See* http://www.ic-arda.org/Novel_Intelligence/.

[17] Associated Press, *Pentagon Continues Some 'Data Mining': Controversial Office Eliminated, but Not All Research*, MSNBC, Feb. 23, 2004, *at* http://www.msnbc.msn.com/id/4350463/.

[18] *See* http://www.rl.af.mil/tech/programs/eeld/ for information on what the project was. EELD was in developmental stages as early as 2000 as part of DARPA's Asymmetric Threat Initiative. *See* http://www.darpa.mil/DARPATech2000/Presentations/iso_pdf/4ArmourATB&WRev1.pdf. The Asymmetric Threat Initiative, which later became TIA, is no longer receiving funding according to the Department of Defense's Budget Estimate for Fiscal Year 2005. Dep't of Defense, FY 2005 Budget Estimate, 30, *available at* http://www.darpa.mil/body/pdf/DoDFY2005BdgetEstFeb04.pdf.

[19] Anon., *Fed Data-Mining Research Lives On*, CBS News, Feb. 23, 2004, *available at* http://www.cbsnews.com/stories/2004/02/23/tech/main601728.shtml.

[20] *See* http://www.darpa.mil/body/off_programs.html.

[21] Dep't of Defense, *supra* note 18.

[22] TSA Congressional testimony: http://www.tsa.gov/public/display?content=09000519800a6143.

[23] *Id.*

[24] TSA CAPPS II Fact Sheet: http://www.tsa.gov/public/display?content=09000519800917e7.

[25] GAO-04-385, *Aviation Security: Computer-Assisted Passenger Prescreening System Faces Significant Implementation Challenges* (Feb. 2004), *available at* http://www.gao.gov/new.items/d04385.pdf

and addressed all privacy concerns" or established a process that allows redress of violations resulting from the use of erroneous information.[26]

CAPPS II will not be fully funded until the program meets certification requirements, such as establishing redress and oversight procedures, adopting privacy protections, demonstrating program effectiveness, and implementing safeguards against abuse or unauthorized access.[27] Because CAPPS II keeps failing to meet the certification requirements, its implementation has been delayed time after time, further increasing its price tag. The development stages of CAPPS II have been expensive, with the TSA requesting $60 million dollars for the 2005 fiscal year, $24 million of which will be recurring costs.[28] With funding already allocated by Congress, the TSA awarded a five-year contract to Lockheed Martin Management and Data Systems to "develop, implement, deploy and operate" its risk assessment system, which includes CAPPS II.[29]

C. MATRIX

The Multistate Anti-terrorism Information Exchange System (MATRIX) is a database system that includes information provided by participating states such as "criminal history records, driver's license data, vehicle registration records, incarceration records, and digitized photographs."[30] Like TIA, MATRIX can run pattern-based queries, which "seek information about people, places, and things based on patterns of activity, none of the components of which might on its own arouse suspicion or be in any way improper."[31] While MATRIX officials "have repeatedly denied that the Matrix is used for data mining,"[32] the technology certainly has that capability.

MATRIX receives funding from the Departments of Justice and Homeland Security and distributes an $8 million federal grant to states that are participating in the program.[33] The program's goal is to operate in all fifty states to enable information sharing among state and federal law enforcement agencies. Although only five states are currently participating, [34] the MATRIX database "has driver's license information from 15 states, motor vehicle registration from 12 states, Department of Corrections information from 33 states and sexual offender information from 27 states."[35] As the federal government maintains "managerial oversight and control" over the database, the ACLU and other civil liberties organizations have questioned the

---

[26] *Id*. at 4.

[27] The Dep't of Homeland Security Appropriations Act for Fiscal Year 2005 specified that no funds, other than for testing, would be distributed to CAPPS II until certification. H.R. 4567, 108th Cong. § 523 (2004). H.R. 4567 has been passed by the House and is waiting on approval from the Senate. This requirement was also present in the Dep't of Homeland Security Appropriations Act for Fiscal Year 2004, which established the certification requirements. Pub. L. 108-90 § 519.

[28] U.S. Senate Comm. on Governmental Affairs Pre-hearing Questionnaire For the Nomination of Admiral David Stone to be Assistant Secretary of Homeland Security, Transportation Security Administration, 108th Cong. 15 (2004), *available at* http://www.epic.org/privacy/airtravel/stone_answers.pdf.

[29] TSA, *CAPPS II at a Glance*, *at* http://www.tsa.gov/public/display?theme=5&content=0900051980088d91.

[30] GAO Report, note 5, at 10.

[31] Mary DeRosa, *Data Mining & Data Analysis For Counterterrorism*, Center for Strategic & International Studies Report, (Mar. 2004) at 13, *available at* http://www.mafhoum.com/press7/189T42.pdf.

[32] Madeleine Baran, *Welcome to the MATRIX: Inside the Government's Secret, Corporate-Run Mega-Database*, THE NEW STANDARD, July 9, 2004, *at* http://newstandardnews.net/content/?action=show_item&itemid=662.

[33] Robert O'Harrow, Jr., *Anti-Terror Database Got Show at White House*, Washington Post, May 21, 2004, at A12, *available at* http://www.washingtonpost.com/wp-dyn/articles/A43608-2004May20.html.

[34] The participating states are Florida, Michigan, Ohio, Pennsylvania, and Connecticut.

[35] Baran, *supra* note 32.

MATRIX program as an effort to recreate TIA at the state level.[36] State governments are also concerned about the lack of privacy protection for information used by the pilot program,[37] and eleven states withdrew from the program for that reason.[38]

D. SEVIS

The Student and Exchange Visitor Information System (SEVIS) is an internet based system that "maintains and manages data about foreign students and exchange visitors during their stay in the United States."[39] The Department of Homeland Security's (DHS) Bureau of Immigration and Customs Enforcement (ICE) maintains the SEVIS database.

Congress authorized SEVIS in section 641 of the Illegal Immigration Reform and Immigrant Responsibility Act (IIRIRA) of 1996, and the program was initiated in July 2001. Later that year, the USA PATRIOT Act amended IIRIRA to require full implementation by January 1, 2003.[40] It also expanded the program to include air flight schools, language training schools, and vocational schools, in addition to colleges and universities.

The SEVIS program collects 230 different pieces of information for each student or exchange visitor, including name, address, date of birth, dependents' information, passport information, field of study, and employment information.[41] DHS officers, Department of State officials, and school administrators use SEVIS data for a variety of purposes.[42] Although the program aims more at catching immigration violations than at preventing terrorist threats, intelligence officers also use SEVIS information to "identify patterns of criminal activity, including terrorism."[43] ICE says the program helps the agency "to better identify trends and patterns to assist in planning and analyzing risks."[44] Additionally, the DHS requires schools to report when students who request educational visas fail to show up for classes, so that the DHS can track down the individuals.

SEVIS cost $28.2 million in Fiscal Years 2002 and 2003, and the DHS plans to spend an additional $9.6 million before September 30, 2004.[45] However, enabling statutes require the program to be funded by user fees, a process that will be implemented starting September 1, 2004. The SEVIS fee will be $100 per student (the maximum amount allowed by law), with a limited fee of $35 for those students participating in summer travel programs or working as summer camp counselors.[46] The fee includes not only the $54 it takes to operate the database, but

---

[36] The MATRIX: ACLU Issue Brief #2, *New Documents Obtained by ACLU Raise Troubling Questions About MATRIX Program*, May 20, 2004, *at* http://www.aclu.org/Files/OpenFile.cfm?id=14253.
[37] Michigan Reviews Continued Support of Federal ID Database, MICHIGAN TECH. NEWS, Mar. 19, 2004, *available at* http://www.mitechnews.com/technews/bydate.htm?id=10796724000005.
[38] O'Harrow, *supra* note 33.
[39] Department of Homeland Security, U.S. Immigration and Customs Enforcement, *at* http://www.ice.gov/graphics/enforce/imm/imm_sevis.htm.
[40] *Id.*
[41] GAO-04-690, *Homeland Security: Performance of Information System to Monitor Foreign Students & Exchange Visitors Has Improved, but Issues Remain*, (June 2004), 67, available at http://www.gao.gov/cgi-bin/getrpt?GAO-04-690.
[42] *Id.* at 66-68.
[43] *Id*. at 61-62.
[44] Dep't of Homeland Security, *supra* note 39.
[45] SEVIS GAO Report, *supra* note 41, at 31.
[46] Authorizing Collection of the Fee Levied on F, J, and M Nonimmigrant Classifications Under Public Law 104–208; SEVIS, 69 Fed. Reg. 39814, 39816 (July 1, 2004) (to be codified at 8 C.F.R. pt. 103, 214, & 299).

also additional money used to cover enforcement costs. For instance, the fee revenue will be used to fund SEVIS Liaison Officers, who will "conduct investigations to ensure compliance."[47]

The DHS received comments arguing that it is unfair for foreign students to have to pay for enforcement measures that unjustly target them; however, the DHS responded that this is a statutory mandate. Although educational officials have argued that the DHS fee plans financially prevent students, especially those from developing countries, from studying in the United States,[48] the DHS interpreted its Congressional mandate as preventing fee exemption or reduction for groups of SEVIS users besides those statutorily listed (the reduced $35 fee for summer programs).

Educational institutions also commented that the system might create unnecessary visa delay. In response, the DHS admits that it "will not be able to establish a workable arrangement for fee collection by DOS [the Department of State] prior to the effective date."[49] However, the agency will be conducting additional feasibility studies.[50]

E. US-VISIT

The U.S. Visitor and Immigrant Status Indicator Technology (US-VISIT) program requires non-immigrant visitors to the U.S. to be fingerprinted and photographed when they enter the U.S. at an airport or sea cruise terminal. The information is kept on file and checked against databases of known criminals.[51] In April, the program was extended to include citizens from countries of which the U.S. does not require travel visas (visa-waiver nations).[52]

The program started January 5, 2004, as part of the DHS's new entry-exit strategy. (See information in the NSEERS section below). US-VISIT cost $380 million in 2003, and $330 million was allocated for Fiscal Year 2004.[53] The DHS is seeking an additional $340 million for Fiscal Year 2005.[54] Accenture, whose parent corporation is based in Bermuda, recently contracted with the DHS to develop US-VISIT technology systems. The contract could be worth up to $10 billion dollars over the next decade.[55] The contract was awarded despite the concern of the House Appropriations Committee about the company's offshore ownership.[56]

EPIC and other civil liberties organizations are concerned about US-VISIT's accuracy, the lack of redress procedures for foreign visitors it harms, and its potential expansion for other uses.[57] Some members of Congress are also concerned about the system. In June, Representative Jim Turner (D-Texas), the ranking member on the House Homeland Security sub-committee,

---

[47] *Id.*
[48] SEVIS GAO Report, *supra* note 41, at 4.
[49] Authorizing Collection of SEVIS Fee, 69 Fed. Reg. at 39817.
[50] *Id.*
[51] *Foreign Visitors to USA get fingerprinted, photographed,* USA TODAY, Jan. 5, 2004 http://www.usatoday.com/news/nation/2004-01-05-security-side_x.htm.
[52] *Entering the United States FAQ: US-VISIT, NSEERS, and APIS*, CBC NEWS ONLINE, Apr. 2, 2004, *at* http://www.cbc.ca/news/background/airportsecurity/usvisit.html.
[53] Gitte Lassby, *International Visitors Face Stricter Immigration Controls*, Capitol News Service, Dec. 5, 2003, *available at* http://www.cns.jrn.msu.edu/articles/2003_1205/INTERNATIONAL.HTML.
[54] Dep't of Homeland Security Appropriations Act 2005, H.R. 4567.
[55] Allen Gibson, *US-VISIT Contracts Move Forward, Despite Politics*, MARKETWIRE.COM, June 23, 2004, *at* http://www.marketwire.com/mw/release_html_b1?release_id=69076. *See also* Greta Wodele, *Accenture Wins $10 Billion US-VISIT Contract*, NATIONAL JOURNAL, June 1, 2004, *available at* http://www.govexec.com/dailyfed/0604/060104tdpm1.htm.
[56] *Id.*
[57] EPIC, US-VISIT News, *at* http://www.epic.org/privacy/us-visit/.

questioned the lack of information on the project: "[W]e do not know how the system will work, who will be covered, what technologies will be deployed, and, how much the whole thing will cost." He encouraged the increased use of Congressional oversight as the program progresses.[58] Additionally, the GAO released a report on US-VISIT last September, which concluded, "the program is a very risky endeavor."[59] The GAO reported that the program's risk factors both "are inherent to the program" and "arise from the program's relatively immature state of governance and management."[60] In May, the GAO released another report on US-VISIT, which noted that while the DHS started implementing all of the GAO's previous suggestions, US-VISIT still does not have a security plan or a thorough privacy assessment.[61] The report also questioned the program's management and governance.

F. NSEERS

The National Security Entry-Exit Registration System (NSEERS) is a comprehensive entry and exit registration and prescreening system for non-immigrant visitors who are male, eighteen years and older, and from a country that the DHS identifies as having a high-risk for terrorism.[62] According to the DHS, "the program has collected detailed information about the background and purpose of an individual's visit to the United States, the periodic verification of their location and activities, and departure confirmation."[63] It requires initial registration interviews and may require, at DHS discretion, follow-up interviews and re-registration.[64]

Both US-VISIT and NSEERS are part of DHS efforts to fulfill a Congressional mandate that an entry-exit system be in place by 2005. This mandate was originally established in section 110 of the IIRIRA and amended by the Immigration and Naturalization Service Data Management Improvement Act of 2000.[65] Section 415 of the USA PATRIOT Act amended this Act by making the Director of Homeland Security a member of the Entry-Exit Task Force.

To date, no individuals in the NSEERS system have been charged with terrorism related offenses. Perhaps realizing that NSEERS is thus an expensive mistake, the DHS reports that once US-VISIT is fully operational, the entry and exit requirements of the program will subsume NSEERS special registration.[66]

---

[58] Wodele, *supra* note 55.
[59] GAO-03-1083, Homeland Security: Risks Facing Border and Transportation Security Program Needs to be Addressed, Report to Congressional Committees (Sept. 2003), *available at* www.gao.gov/cgi-bin/getrpt?GAO-03-1083.
[60] *Id*.
[61] GAO-04-586, Homeland Security: First Phase of Visitor and Immigration Status Program Operating, but Improvements Needed, Report to Congressional Committees (May 2004), 57-60, *available at* www.gao.gov/cgi-bin/getrpt?GAO-04-586.
[62] These countries are mostly Muslim countries and mostly located in the Middle East. For a complete list, visit http://www.cbc.ca/news/background/airportsecurity/usvisit.html.
[63] Department of Homeland Security, *Fact Sheet: Changes to National Security Entry/Exit Registration System*, Dec. 1, 2003, *at* http://www.ice.gov/graphics/news/newsrel/articles/NSEERSfactsheet120103.pdf
[64] *Id.* NSEERS used to require re-registration after 30 days in the U.S. and annually thereafter. This part of the program was changed by regulation at the end of last year.
[65] Pub. L. 106-215, 114 Stat. 337, June 15, 2000.
[66] Dep't of Homeland Security, US-VISIT FAQs: NSEERS & US-VISIT, *at* http://www.dhs.gov/dhspublic/display?theme=20&content=3569.

G. Verity K2 Enterprise

The May GAO report explained that Verity K2 Enterprise "mines data from the intelligence community and internet searches to identify foreign terrorists or U.S. citizens connected to foreign terrorism activities."[67] It uses personal information, private sector information, and information from other government agencies. Unfortunately, that is the extent of government-supplied information. While the report cited that Verity K2 Enterprise belongs to the Defense Intelligence Agency (DIA), it actually is a commercial product of the Verity corporation that the DIA is using. According to Verity, Enterprise 2 is a complex data systems manager, which has real-time analysis, monitoring, and altering capabilities.[68] Among other things, it allows users to conduct query-based or full-text searches.[69] The government has not disclosed why it is using the technology or how much it is costing.

H. Analyst Notebook 12 and Pathfinder

Analyst Notebook, a product of the British company i2, and Pathfinder, developed by the U.S. Army, are two common database software programs that are used in both the private and public sectors. In December of 2003, the Associated Press reported that Analyst Notebook 12 was even being used by coalition forces in Iraq. The news agency reports that the technology is "used by police across North America and Europe to map crime hotspots and track serial rapists or arsonists."[70] Similarly, as early as 1993, Pathfinder was "used by over 200 analysts in 18 government agencies."[71] Analyst 12 uses private sector information, and both programs use personal information.

The long-term use of these programs on a massive scale by a variety of private companies and government agencies well before September 11th make them appear to be relatively harmless database tools. Nevertheless, the May 2004 GAO report documented that these projects are using data mining to detect terrorist activities,[72] and not enough information is available with which to determine how this technology works to prevent terrorism. Clearly, one of the most troubling aspects about the use of these technology systems is the lack of government transparency.

I. Project Strikeback

The GAO reported that the Department of Education's Project Strikeback "compares Department of Education and Federal Bureau of Investigation (FBI) data for anomalies" and "verifies personal identifiers."[73] According to the report, the project uses personal information and information from other government agencies, but not private sector information. Unfortunately, this appears to be the extent of publicly available information on the project. The Office of Inspector General (OIG), an office that primarily investigates financial aid fraud, runs

---

[67] GAO report, *supra* note 5, at 9.
[68] Verity, K2 Enterprise *Technical Overview*, *available at* http://www.verity.com/products/k2_enterprise/pdf/MK0385d_K2E_Tech.pdf.
[69] *Id.* at 5.
[70] Jim Krane, *Computer-Sleuthing Aids Troops in Iraq*, ASSOCIATED PRESS, Dec. 23, 2003, *available at* http://www.siliconvalley.com/mld/siliconvalley/7556612.htm.
[71] 1993 National Performance Review of Intelligence Agency Activities, made available by the Loyola Dep't of Political Science *at* http://www.loyola.edu/dept/politics/intel/npr93act.html.
[72] GAO report, *supra* note 5, at 30, 44.
[73] *Id.* at 37.

the project. The OIG's website does not mention Project Strikeback or any other initiatives designed to prevent terrorism.

J. Case Management Data Mart

Another GAO-reported project that uses personal information to detect terrorist threats is Case Management Data Mart. A project of the DHS's Border and Transportation Security Directorate, Case Management Data Mart "reviews [law enforcement] case loads, status, and relationships among cases."[74] While this project appears to be a harmless tool used to manage law enforcement cases, it is unclear why the DHS needs to mine private sector databases or what judicial authorization DHS officers seek before obtaining private information on known suspects.

K. FBI Projects

The GAO reported that the FBI's Secure Collaborative Operational Prototype Environment (SCOPE) mines personal information and data from other government agencies, allowing "the FBI to search multiple data sources through one interface to uncover terrorist and criminal activities and relationships."[75] SCOPE uses the Convera Corporation's RetrievalWare technology, which can search through multiple media formats, "including surveillance videotapes, forensic reports, case files, credit card transactions, terrorist watch lists, wiretaps, bank records and local law enforcement arrest reports."[76]

According to the FBI, SCOPE has been replaced by the Investigative Data Warehouse (IDW), which launched in January 2004.[77] The IDW "provides analysts with full access to investigative information within FBI files, open source news feeds, and the files of other federal agencies such as DHS."[78]

Like SCOPE, the Foreign Terrorist Tracking Task Force (FTTTF) was initiated shortly after the September 11th attacks. In his Homeland Security Presidential Directive of October 29, 2001, President Bush called for the creation of the FTTTF to "1) deny entry into the United States of aliens associated with, suspected of being engaged in, or supporting terrorist activity; and 2) locate, detain, prosecute, or deport any such aliens already present in the United States."[79] As one way to achieve its objectives, the FTTTF mines data from the DHS, FBI, and the private sector.[80] The FBI reports that the FTTTF now has access to "over forty sources of data."[81]

L. JRIES

The DHS's Joint Regional Information Exchange System (JRIES) is an initiative to increase information sharing among federal, state, and local law enforcement. Although its mission seems innocent, JRIES has some troubling aspects. First, the system's creators specifically designed it

---

[74] *Id.* at 44.
[75] *Id.* at 47.
[76] *FBI Continues Its Search for Terrorists*, TRANSFORM MAGAZINE, *at* http://www.transformmag.com/techselections/showArticle.jhtml?articleID=16101056.
[77] FBI, *Report to the National Commission on Terrorist Attacks upon the United States*, April 14, 2004, 60, *available at* http://www.fbi.gov/publications/commission/9-11commissionrep.pdf.
[78] *Id.*
[79] Homeland Security Presidential Directive-2, Oct. 29, 2001, *available at* http://www.whitehouse.gov/news/releases/2001/10/20011030-2.html.
[80] GAO Report, *supra* note 5, at 47.
[81] FBI 9-11 Commission Report, *supra* note 77, at 16.

to avoid legal restraints on the DOD from gathering information on U.S. citizens.[82] Although the DIA transferred management of the system to DHS in February 2004, a Pentagon employee is still the project manager.[83] Additionally, the DOD (like any system member) can request information from other system members. System members do not have a mandatory duty to provide the information, but they can share information at their discretion.[84]

Second, the system information is not limited to international terrorism. The system's members can use the database information to monitor domestic political groups. Ed Manavian, the current JRIES executive board director, has stated that "information on political protests can be considered legitimate terror intelligence."[85] Like most of these programs, JRIES does not have sufficient privacy protections or redress procedures if law enforcement officers act on inaccurate information.

II. Potential Drawbacks of Using Data Mining Technology to Prevent Terrorism

The government claims that these data mining programs are necessary to prevent terrorist threats in the United States. However, all the reports coming from experts and Congressional research agencies like the GAO highlight how untested these programs really are and how little we know about whether they will in fact help prevent terrorism. Anita Ramasastry, a law professor at the University of Washington, questions the technology by noting that "[t]errorists and other criminals often anticipate the factors that law enforcement will use to profile them and will circumvent them quickly enough."[86] In addition to being untested and possibly ineffective, these programs present dangers to civil liberties and personal privacy.

A. Impact on Constitutional Rights

Creating a system for surveilling innocent Americans dramatically changes the way law enforcement operates. No longer do criminal investigators need to have probable cause[87] or reasonable suspicion[88] in order to obtain private information about an individual. Now, everyone is a suspect and the government can obtain personal information on anyone using pattern-based investigations. It is also important to underscore that all these programs search through data without the authorization of a warrant or other protective judicial measure. Thus, large-scale electronic surveillance systems present serious concerns for the Fourth Amendment's guarantee against unreasonable searches.

---

[82] Justin Rood, *Pentagon Has Access to Local Police Intelligence through Office in Homeland Security Department*, CONGRESSIONAL QUARTERLY, July 6, 2004.
[83] *Id*
[84] *Id.*
[85] *Id.*
[86] Ramasastry, Why We Should Be Concerned About "Total Information Awareness" and Other Anti-terrorism Strategies for the Internet, FINDLAW'S WRIT, Dec. 31, 2002, *at* http://writ.news.findlaw.com/ramasastry/20021231.html.
[87] The Fourth Amendment requires that "no Warrants shall issue, but upon probable cause." U.S. CONST. amend. IV § 2.
[88] In Terry v. Ohio, the Supreme Court held that the Fourth Amendment can protect the privacy of citizens even if the police action is less than a full search or seizure and reaffirmed that a search must be limited in scope and duration. The Court held that in these cases "the police officer must be able to point to specific and articulable facts which, taken together with rational inferences from those facts, reasonably warrant [the] intrusion." 392 U.S. 1, 21 (1968). This standard is commonly called "reasonable suspicion."

While the text of the Fourth Amendment, like all Constitutional text, is far from clear, Supreme Court interpretation of this Amendment helps to clarify why these surveillance systems potentially threaten the Amendment's guarantees. The Court's paramount interpretation of the "unreasonable searches and seizures" clause of the Fourth Amendment is *Katz v. U.S.*[89] In *Katz*, the Court held that protections of the Fourth Amendment apply when an individual has a "reasonable expectation of privacy." Equally important, the Court held "that electronic as well as physical intrusion into a place that is in this sense private may constitute a violation of the Fourth Amendment."[90]

However, in the 1970s the Court held that consumers do not have a reasonable expectation of privacy in information they voluntary give to businesses.[91] Therefore, there is no Fourth Amendment protection for records of credit card transactions, travel information, or other transactional data. Similarly, individuals that "voluntarily"[92] participate in government programs such as US-VISIT or SEVIS would probably not have Fourth Amendment protection for the information they provide.

Although the law may seem clear, a lot has changed since the 1970s. Electronic data are available on a much wider scale, and new data mining tools give the government much greater access to much larger amounts of data. The 1970s cases seem outdated considering recent technological changes and the public's heightened expectation of privacy in electronic information. In light of these new tools, a re-visiting of Fourth Amendment law is necessary.

The second important aspect of the *Katz* decision, that judicial monitoring of law enforcement action is necessary to preserve Fourth Amendment rights, is critical to understanding why these electronic surveillance systems are such a threat. In *Katz,* the Supreme Court held, "searches conducted outside the judicial process, without prior approval by judge or magistrate, are per se unreasonable under the Fourth Amendment — subject only to a few specifically established and well-delineated exceptions."[93] Exceptions include search "incident" of arrest, exigent circumstances such as "hot pursuit," and consent. In referring to electronic surveillance, the Court remarked that "it is difficult to imagine how any of those exceptions could ever apply."[94] Judicial authorization of electronic surveillance has been an important aspect of Fourth Amendment law.

Unlike the financial records cases of the 1970s, TIA, MATRIX, and other data mining programs do not obtain records about a known suspect. Additionally, and most importantly, data mining initiatives operate without judicial oversight. Thus, data mining is a new step for the Fourth Amendment. Chances are there will be a court case deciding whether to suppress evidence obtained using data mining technology in the near future. Will the vague notion of "protecting national security" be enough to convince a court that these searches can constitutionally occur without judicial oversight? Hopefully a court would look back to our history and remember times when constitutional rights were denied on the basis of national

[89] 389 U.S. 347 (1967).
[90] *Id.* at 360, Harlan *concurring*.
[91] *See* U.S. v. Miller, 425 U.S. 435 (1976) (holding that bank records are outside Fourth Amendment protection) and Couch v. U.S., 409 U.S. 322 (1973) (holding that there is no reasonable expectation of privacy in tax records given to an accountant).
[92] There is a strong argument that individuals do not have a choice in providing the information because if they refuse they will either be prohibited from entering the United States or detained once they arrive.
[93] *Id*. at 357. Court footnotes omitted.
[94] *Id.*

security[95] and times when they were protected despite national security interests[96] and hold in the interest of protecting our constitutional guarantees of privacy and liberty.

B. The Limits of the Privacy Act

With a few minor exceptions, the Privacy Act of 1974 prevents federal government agencies from disclosing records "to any person, or to another agency, except pursuant to a written request by, or with the prior written consent of, the individual to whom the record pertains."[97] Of particular concern to this requirement is the GAO identification of 46 data mining projects that mine personal information from other agencies.[98] Citizens should worry whether information they provide to the government is being used for a completely different purpose than what it was intended for. As agencies are transferring information without consent from the citizen who provided the information, it raises a Privacy Act flag.

Most of the data mining projects discussed in this article are maintained by law enforcement agencies. Exemption 7 of the Act allows disclosure of personal records to a law enforcement agency only if the "law enforcement activity is authorized by law, and if the head of the agency or instrumentality has made a written request to the agency which maintains the record specifying the particular portion desired and the law enforcement activity for which the record is sought."[99] It is doubtful whether the DOJ and the DHS are complying with this mandate, as few of the projects are explicitly authorized by law. Nevertheless, various laws and regulations, such as the Homeland Security Act, have encouraged the use of data mining and could be used as evidence of legal authorization.

As for the "written request" requirement, clearly, lawmakers wrote the Privacy Act with individual requests about known suspects in mind, not the large-scale transfer of personal data that data mining projects facilitate. Thus, a court may weaken the requirements to allow blanket requests or hold that data mining projects are outside the scope of the Privacy Act. In other words, the Privacy Act probably provides enough loopholes for the DOJ and DHS to use.[100]

One of the largest concerns about data mining is federal government use of private sector data. For example, the TSA's CAPPS II uses information provided to airlines for a purpose completely unrelated to detecting terrorists. In addition to government-controlled systems, many private systems allow the government to obtain personal information, such as Ancestry.com, the White Pages, Choice Point, Orion Scientific Systems, LexisNexis, Locate PLUS, and Seisint (who developed the MATRIX).[101] Many of these systems can be just as intrusive and inaccurate as government-maintained databases. Unfortunately, the Privacy Act regulates records the U.S. government stores but does not regulate how agencies access records from the private sector. Similar to Fourth Amendment law, the Privacy Act gives government agencies a large loophole

---

[95] *See e.g.* Korematsu v. U.S., 323 U.S. 214 (1944) (holding that the internment of Japanese-American citizens during World War II was a valid use of Congressional and Executive war powers).

[96] *See e.g.* N.Y. Times Co. v. U.S., 403 U.S. 713 (1971) (holding that Pentagon refusal to allow newspaper agencies from reporting on certain documents about the Vietnam War violated the 1st Amendment).

[97] 5 U.S.C. § 552a(b).

[98] GAO report, *supra* note 5, at 3.

[99] 5 U.S.C. § 552a(b)(7).

[100] For an illuminating commentary on Privacy Act loopholes, *see* Julianne M. Sullivan, *Will the Privacy Act of 1974 Still Hold up in 2004? How Advancing Technology Has Created a Need for Change in the "System of Records" Analysis*, 39 CAL. W. L. REV. 395 (2003).

[101] London publishing company Reed Elsevier purchased Seisint in July 2004 and is expected to merge the data company into its own LexisNexis global data distribution and risk management unit. Stephen Pounds, *Boca's Seisint to be Sold for $775 million*, PALM BEACH POST, July 15, 2004.

by allowing them to have the private sector do something they could not do themselves without judicial authorization: collect and store scores of private information on persons in the United States without reasonable suspicion or probable cause.

Another cause of concern is that the Privacy Act only provides protection for U.S. citizens and permanent residents.[102] In other words, nonresident foreign nationals cannot use the Act's provisions to protect their privacy. This distinction becomes especially important at a time when the government is targeting non-citizens, especially those from Arab and Muslim countries, in its terrorism-related investigations. Additionally, many of the major data surveillance programs, such as US-VISIT and SEVIS, are specifically aimed at obtaining private information about noncitizens.

C. Lack of Legal Regulation

The Departments of Defense, Justice, State, and Homeland Security need to develop uniform administrative regulations for the implementation of their data mining projects in order to protect technological privacy and civil liberties. As the two previous sections showcase, current laws and regulations, and even the Constitution, do not adequately protect citizens against the dangers of these projects.[103] As TAPAC notes:

> Existing legal requirements applicable to the government's many data mining programs are numerous, but disjointed and often outdated, and as a result may compromise the protection of privacy, public confidence, and the nation's ability to craft effective and lawful responses to terrorism.[104]

TAPAC recommends that a number of items be included in new regulatory guidelines, including requiring "authorization from the Foreign Intelligence Surveillance Court before engaging in data mining with personally identifiable information."[105]

Like all agency regulations, these would have a dual purpose. As the national security task force of the Markle Foundation, a group that researches emerging technology, notes, "Policy guidelines are meant to empower government officials as well as limit them, and Congress and the Executive Branch should share a common commitment to both objectives."[106] Without clear guidelines for agency action, many of these programs pose a real threat to our privacy and liberty.

D. Creating "False Positives"

Many innocent citizens' spending or travel habits could fall into the selected high risk patterns. It is therefore possible that these data mining systems will create numerous false positives.[107] Evidence of this comes from tests of the MATRIX system in Florida, which "flagged" 120,000 people "who had a statistical likelihood of being terrorists."[108] Almost

---

[102] 5 U.S.C. § 552a(a)(2).

[103] *See e.g.* Jim Dempsey & Lara Flint, *Privacy's Gap: The Largely Non-Existent Legal Framework for Government Mining of Commercial Data*, CENTER FOR DEMOCRACY AND TECHNOLOGY, May 28, 2003, *available at* http://www.cdt.org/security/usapatriot/030528cdt.pdf.

[104] Technology & Privacy Advisory Committee, *supra* note 13, at ix.

[105] *Id*. at x.

[106] Zoë Baird, et al. *Creating a Trusted Network for Homeland Security*, 2d Report of the Markle Foundation Task Force on National Security in the Information Age 6 (Dec.2003), *at* http://www.markletaskforce.org/Report2_Full_Report.pdf.

[107] *See e.g.* TAPAC report, *supra* note 13.

[108] *Early Database Project Yielded 120,000 Suspects*, CNN ONLINE LAW CENTER, May 21, 2004, *at* http://www.cnn.com/2004/LAW/05/20/terror.database.ap/.

assuredly there are not 120,000 terrorists in the United States. Nevertheless, the government touts the results, claiming that five of the suspected September 11th hijackers were among the top 80 people named.[109] The New Standard reported that the 120,000 names were given to law enforcement which "led to 'several arrests within one week,' and 'scores of other arrests.' Who was arrested, and whether they were convicted of or even charged with any actual crime remains unknown."[110]

If a mistake is made, the ramifications for personal liberty and privacy may be enormous, as the FBI's recent debacle with fingerprints in the Brandon Mayfield case demonstrates.[111] Not only will a "flagging" result in greater surveillance, but it could also result in detention, interrogation, or otherwise intrusive investigation. For a person who is innocent, these events will be negatively life-altering. Additionally, as mentioned above, very few of these projects have adequate administrative procedures for redressing harm resulting from use of erroneous information.

E. Lack of Transparency and Democratic Decision-Making

As the debate surrounding TIA last year demonstrates, there is a significant lack of transparency about important decisions that affect the very heart of our personal privacy. As this article notes, no public information about the status of many of these projects exists. Citizens and members of Congress are left in the dark regarding project scope, financial cost, and potential for privacy infringement.This lack of transparency only increases public concern. As Mary DeRosa, of the Center for Strategic and International Studies, observes:

> There is significant public unease about whether protections for privacy are adequate to address the negative consequences of increased government use of private data. These concerns are heightened because there is so little understanding of how the government might use these data analysis tools. Nor is there typically much public debate or discussion before these tools are adopted. This lack of transparency not only can make the government's decisions less informed, but it increases public fear and misunderstanding about uses of these techniques.[112]

A lack of transparency is not only a threat to the project's effectiveness, but also a threat to democratic decision-making and to a viable system of checks and balances.

III. A Way Forward: Steps to Protecting Your Civil Liberties in the Information Age

Despite the seemingly overwhelming ability of the federal government to create data mining systems aimed at terrorism prevention, there are steps that citizens can take to combat further erosion of their personal privacy.

First, like the public opposition regarding TIA funding, citizens should oppose all attempts to revive Congressional funding of TIA or to fund other terrorist prevention data mining systems. Not nearly enough information exists as to how these system affect privacy and Constitutional rights to justify financial support of them.

---

[109] *Id.*
[110] Baran, *supra* note 32.
[111] *See* Jennifer L. Mnookin, *The Achilles' Heel of Fingerprints*, WASHINGTON POST, May 29, 2004, at A27, *available at* http://www.washingtonpost.com/wp-dyn/articles/A64711-2004May28.html.
[112] DeRosa, *supra* note 31, at 6.

Second, the public should continue to demand Congressional oversight. Two extremely positive examples of Congressional oversight underscore the importance of maintaining an effective system of checks and balances. The May 2004 GAO report on data mining use cited above is an important first step towards better understanding these data mining systems. Additionally, the Intelligence Authorization Act for Fiscal Year 2004 approved the use of data mining for foreign intelligence but requires the Administration to compile reports on the privacy implications of their projects. [113] It is important for citizens to watch Congress and the Attorney General to make sure these reports are produced and released to the public.

Next, even if legal guidelines are developed, the public needs to encourage the Administration to build privacy protection technology into the systems themselves. Kim Taipale, the Executive Director of the Center for Advanced Studies in Science and Technology Policy, contends that "[a] strategy for insuring the protection of privacy and civil liberties interests is to build features that support those values into the technologies in the first place."[114] Procedures that TAPAC encourages government to utilize include data anonymization, creating an audit trail, and security and access controls.[115]

To their credit, some governmental agencies are developing privacy protection technology. For instance, David Stone of the TSA said that CAPPS II would use software called Radiant Trust, which "maintains audit trails of who accessed the system and the time/date."[116] Stone said, "CAPPS II testing will not begin until security systems to ensure protection of the data are fully in place."[117] It is important for the public to hold Stone to his word. As Taipale recognizes, the public needs to participate in this process: "[I]t is only through involvement in and oversight of government sponsored research projects that public interest concerns can be incorporated in to the development process."[118]

Finally, and most importantly, the American public must keep discourse alive. Decisions about these systems cannot happen behind closed doors. Citizens must continue to question their government officials about the development and implementation of data mining systems aimed at preventing terrorism.

---

[113] Pub. L. No. 108-177 §§ 359(a)(9), 360(b)(12) (2003), *available at* http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ177.108.
[114] K.A. Taipale, *Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data*, 5 COLUM. SCI. & TECH. L. REV. 1, 13 (2003), *available at* http://www.stlr.org/cite.cgi?volume=5&article=2.
[115] Technology & Privacy Advisory Committee, *supra* note 13, at x.
[116] Stone Pre-hearing Questionnaire, *supra* note 28, at 20.
[117] *Id.*
[118] *Id.*

Appendix 1: List of Acronyms

ACLU	American Civil Liberties Union
ARDA	Advanced Research and Development Activity
CAPPS	Computer Assisted Passenger Prescreening System
DARPA	Defense Advanced Research Projects Agency
DIA	Defense Intelligence Agency
DHS	Department of Homeland Security
DOD	Department of Defense
DOJ	Department of Justice
EELD	Evidence Extraction and Link Discovery
EPIC	Electronic Privacy Information Center
FBI	Federal Bureau of Investigation
FOIA	Freedom of Information Act
FTTTF	Foreign Terrorist Tracking Task Force
GAO	Government Accountability Office (Government Accounting Office)
IAO	Information Awareness Office
ICE	DHS Bureau of Immigration and Customs Enforcement
IIRIRA	Illegal Immigration Reform and Immigration Responsibility Act of 1996
JRIES	Joint Regional Information Exchange System
MATRIX	Multistate Anti-terrorism Information Exchange System
NSEERS	National Security Entry-Exit Registration System
SCOPE	Secure Collaborative Operational Prototype Environment
SEVIS	Student and Exchange Visitor Information System
SSNA	Scalable Social Network Analysis
TAPAC	Technology and Privacy Advisory Committee
TIA	Total Information Awareness
TSA	Transportation Security Administration
US-VISIT	U.S. Visitor and Immigrant Status Indicator Technology