

U.S. HOUSE OF REPRESENTATIVES
COMMITTEE ON SCIENCE AND TECHNOLOGY

SUITE 2320 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6301
(202) 225-6375
TTY: (202) 226-4410
<http://science.house.gov>

June 5, 2007

The Honorable David Walker
Comptroller General of the United States
Government Accountability Office
441 G Street, NW
Washington, D.C. 20548

Dear Mr. Walker:

In August 2005, the Government Accountability Office issued a report on data mining (GAO-05-866) that looked into the specific data mining initiatives of five federal agencies. The report concluded that none of the five programs examined, including the Federal Bureau of Investigation's (FBI) Foreign Terrorist Tracking Task Force (FTTTF), complied with all relevant federal laws and executive branch guidance. This included administrative, technical and physical safeguards as mandated by the Privacy Act of 1974, guidance from the Office of Management and Budget and federal information security standards set forth by the National Institute of Standards and Technology as detailed in the Federal Information Security Management Act of 2002. Further, the Computer Security Act of 1987, details requirements to establish security plans for Federal computer systems that contain sensitive information.

The Foreign Terrorist Tracking Task Force was established by the President in the immediate aftermath of the September 11, 2001, terrorist attacks as an interagency group under the auspices of the Department of Justice. Its original mission was to deny entry into the United States by aliens suspected of having ties to terrorist organizations and to locate, detain, prosecute, or deport such aliens already present in the United States.

But documents now indicate that the FTTTF is expanding its mission to encompass the "detection, identification, and tracking of individuals or entities that pose threats to the United States and its interests through the use of advanced analytical techniques, technologies, and data resources." This mission will be accomplished through the use of bulk data analysis, pattern analysis, trend analysis and other programs, according to Justice Department budget documents reviewed by the Subcommittee. "The FBI's efforts to define predictive models and patterns of behavior will improve efforts to identify "sleeper cells," the documents suggest. The centerpiece of this greatly enhanced effort will be a newly proposed National Security Branch Analysis Center (NSAC).

The FBI is seeking \$12 million for the center in FY2008, which will include 90,000 square feet of office space and a total of 59 staff, including 23 contractors and five FBI agents. Documents predict the NSAC will include six billion records by FY2012. This amounts to 20 separate "records" for each man, woman and child in the United States. The "universe of subjects will expand exponentially" with the expanded role of the NSAC, the Justice Department documents assert.

The expanded and sweeping scope of the NSAC bears a striking resemblance to the Defense Advanced Research Project Agency's Total Information Awareness program which Congress terminated funding for in 2003 because of privacy and other concerns. Sharing critical information that can help law enforcement officer's track down known terrorists is extraordinarily important and needs to be improved. But the NSAC proposes to do much more than simply track down known terrorist suspects. Eleven of its proposed 59 staff will constitute a Proactive Data Exploitation unit – tasked with ferreting out "patterns" of suspicious behavior in the data the center collects. "The NSAC will leverage existing data mining tools to help identify relationships between individuals, locations, and events that may be indicators of terrorist or other activities of interest," according to the Justice Department budget documents

Data mining experts outside of government see great potential for abuse in this sort of proposal. Jeff Jonas, a world renowned data mining expert and IBM Distinguished Engineer, recently co-authored a critical review of "predictive" counterterrorism data mining efforts for the Cato Institute. "It would be unfortunate if data mining for terrorism discovery had currency within national security, law enforcement, and technology circles," wrote Jonas, "because pursuing this use of data mining would waste taxpayer dollars, needlessly infringe on privacy and civil liberties, and misdirect the valuable time and energy of the men and women in the national security community." Jonas supports other non-predictive or "pattern analysis" data mining efforts that permit law enforcement agencies to "efficiently locate, access, and aggregate information about specific suspects," he writes. But he does not believe data mining is suited to discovering unknown terrorists as a result of culling through massive mounds of data that contain "patterns" of individual behavior. Jonas argues that with an extraordinarily limited pool of known terrorist patterns of behavior a hunt for terrorists in this way would inevitably "flood the national security system with false positives – suspects who are truly innocent." In addition, argues Jonas, collocating massive amounts of data in a central repository poses significant logistical and security challenges and may invite misuse of the information.

Given the scope of the NSAC endeavor, Congress has a duty to understand fully what information will be contained in the "records" it collects, whether the "records" of U.S. citizens will be included in its database, how this data will be employed and how the FBI plans to ensure that the data is not misused or abused in any way. A critical question is how the FBI will ensure that the records it obtains from other agencies is accurate, valid and complies with federal legal guidelines and policies. The FTTTF, for instance, shares "innovative technology" with the Defense Department's Counterintelligence Field Activity (CIFA) and the proposed NSAC will presumably maintain or expand on this relationship. This is of particular concern given the fact that the Defense Department has acknowledged that CIFA was compiling data in one of its databases on non-violent war protestors and civil rights activists in violation of DOD's own

policies. The Bureau needs to beware that it does not repeat the mistakes of other agencies. Even with those assurances the agency may have difficulty developing and operating the NSAC.

The FBI has historically been unable to develop information systems in a reliable, cost effective and technically proficient manner. In 2005, after investing \$170 million, the agency cancelled its Virtual Case File computerized records management system because of technical troubles. Sentinel, the replacement for this system, is now reportedly running behind schedule. Most troubling, last year it was revealed that a FBI-computer consultant managed to hack into the FBI's classified computer system, gaining access to records on counterespionage and the Witness Protection Program, as well as the passwords of 38,000 employees, including FBI Director Robert S. Mueller III.

In March 2007, the Department of Justice Office of the Inspector General issued a report on the FBI's use of National Security Letters. That report found that the Bureau had demanded personal data without proper authorization, improperly obtained personal telephone and banking records and underreported to Congress how often it used national security letters to obtain information on thousands of U.S. citizens and legal residents. Inspector General Glen Fine said that he found 48 separate violations of law in the use of national security letters that resulted in as many as 3,000 violations among more than 143,000 requests for information between 2003 and 2005.

These examples lead the Subcommittee to question whether the NSAC design, development and implementation is incorporating the lessons learned by the Bureau from previous systems. Are the safeguards required for such systems in place within the NSAC's database? We request a review of the NSAC to address the following questions:

1. What is the specific role and purpose of the NSAC and what requirements in the center's mission explain the size and scope of this planned database?
2. What types of "records" will be incorporated into the database, from which agencies or commercial enterprises will they be obtained and will any other entities be granted access to the database and under what restrictions?
3. Will the NSAC include any records on U.S. citizens and what provisions are in place to guarantee that any records collected or accessed are consistent with existing law, regulation, policy or other agency guidance?
4. How does the center intend to exploit the data it collects by utilizing specific analytical tools – including "pattern recognition," "predictive data mining," "social network analysis," and related software programs?

Mr. Walker
Page 4
June 4, 2007

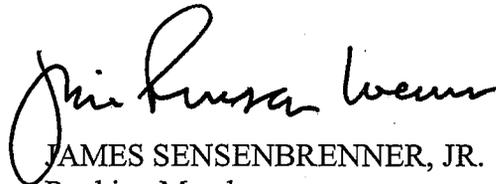
Please have your staff contact Douglas Pasternak, Subcommittee professional staff member at (202) 226-8892, Bart Forsyth, Counsel to Rep. Sensenbrenner at (202) 225-6371 or Dan Pearson, Subcommittee staff director at (202) 225-4494 to discuss this request further.

Your assistance in this matter is greatly appreciated.

Sincerely,



BRAD MILLER
Chairman
Subcommittee on
Investigations & Oversight



JAMES SENSENBRENNER, JR.
Ranking Member
Subcommittee on
Investigations & Oversight