U.S. Department of Justice
Office of Legislative Affairs

Office of the Assistant Attorney General

Washington, D.C. 20530

July 9, 2007

The Honorable Patrick J. Leahy
Chairman
Committee on the Judiciary
United States Senate
Washington, D.C. 20510

Dear Mr. Chairman:

We are pleased to transmit the report required by Section 126 of the USA PATRIOT Improvement and Reauthorization Act of 2005, Pub. L. No. 109-177 ("the Act"). Section 126 required the Attorney General to submit a report to Congress concerning "any initiative of the Department of Justice that uses or is intended to develop pattern-based data-mining technology," as defined by that section. The report is enclosed.

After a thorough review by the components and agencies of the Department of Justice, we have identified seven initiatives that meet the definition of "pattern-based data mining" set forth in section 126. In an effort to provide you with the majority of the report without further delay, the enclosed report covers six of those initiatives, and the seventh initiative will be covered in a supplemental report to follow. In addition, we have also provided information regarding additional initiatives and systems that do not technically meet the definition of "pattern-based data mining" as set forth by Congress but may be perceived by some as "data mining." As some of the programs, particularly the STAR program, involve sensitive law enforcement information, we would urge that you treat this material accordingly.

The seventh initiative that we believe meets the definition will be led by the FBI and going forward expects to receive some funding from the Office of the Director of National Intelligence ("ODNI"). Because the FBI and ODNI are continuing to discuss policy and procedures that will be used as part of the initiative, we are unable to include it in the enclosed report at this time. However, we would appreciate the opportunity to brief interested Members and staff, and once the parameters for the initiative are finalized, we will submit a supplemental report.

We apologize for the delay in transmitting this report. Please do not hesitate to contact this office if we may be of further assistance with this or any other matter.

Sincerely,

Brian A. Benczkowski
Principal Deputy Assistant Attorney General

Enclosure

cc: The Honorable Arlen Specter
Ranking Minority Member

UNITED STATES DEPARTMENT OF JUSTICE

REPORT ON "DATA-MINING" ACTIVITIES PURSUANT TO SECTION 126 OF THE USA PATRIOT IMPROVEMENT AND REAUTHORIZATION ACT OF 2005

Section 126 of the USA PATRIOT Improvement and Reauthorization Act of 2005, Pub. L. No. 109-177 ("the Act"), requires the Attorney General to submit a report to Congress on "any initiative of the Department of Justice that uses or is intended to develop pattern-based data-mining technology," as defined by that section. For each such initiative, the Attorney General must provide:

1) a thorough description of the pattern-based data-mining technology;
2) a thorough discussion of the plans for use of the technology, including target dates for deployment of the technology;
3) an assessment of the likely efficacy of the technology's quality assurance controls to ensure that the technology provides accurate and valuable information;
4) an assessment of the likely impact of the implementation of the technology on privacy and civil liberties;
5) a list and analysis of the laws and regulations applicable to the Department that govern the application of the data-mining technology to the information used with the data-mining technology; and
6) a thorough discussion of the Departmental procedures, policies and guidelines that are to be developed and applied in the use of the technology to protect privacy and due process rights and ensure that only accurate information is collected and used.[1]

As background, this report first discusses "data mining" as a conceptual matter, as well as the privacy concerns that may be implicated by advanced analysis of information obtained and retained by the government. Next, for qualifying initiatives, this report provides information responsive to each of the six categories set forth in section 126 and listed above.[2] Finally, this report provides information on certain advanced analytic activities conducted by the Department that do not meet the definition of "data mining" set forth in section 126, but may nonetheless be perceived as "data mining" as that term is commonly understood. These additional descriptions do not provide a comprehensive canvas of all advanced analysis undertaken at the Department, but are included to provide additional background on the types of investigative techniques currently in use. Similarly, the report contains descriptions of information technology (IT) systems that have been or are being developed by the Department that may at some point have

[1] USA PATRIOT Improvement and Reauthorization Act of 2005, Pub. L. No. 109-177, § 126(a), 120 Stat. 192, 227-28.

[2] Given the range of sophistication of the various initiatives, the information responsive to the statute varies as well. For example, the third category of information—efficacy of the technology's quality assurance controls—would be far more relevant with respect to initiatives that use specially designed software. By contrast, the responsive initiatives generally use widely available and time tested commercial off the shelf software (COTS).

the capacity to support advanced analytical initiatives. These initiatives and systems were identified through a query to the individual agencies and components of the Department, with the information about the initiatives and systems provided by the agencies and components.

Using data mining initiatives to analyze lawfully acquired information, as is the case with each of the qualifying initiatives,[3] can be extremely valuable tools for investigators. These advanced analytical activities are grounded in traditional investigative techniques, but are designed to process information more efficiently and effectively. Such initiatives must also be undertaken with deep respect for the privacy and civil liberties of Americans. All of the pattern-based data mining initiatives undertaken by the Department meet both of these goals.

I. Introduction

There is no universally accepted definition of "data mining." In fact, the term means different things to different people, such as technologists, policy makers and privacy officials. Similarly, although often used in common parlance, there are many understandings as to what data mining actually encompasses in a given situation. As a general matter, however, the term "data mining" refers to either "subject-based" or "pattern-based" database queries. Subject-based queries search for information on a predetermined individual based on a specific identifier. A wide range of ordinary investigative techniques may easily come within this understanding of subject-based data mining—from checking fingerprints or names against a set of computer records to querying multiple databases to find a telephone number. Many investigations employ such subject-based queries. Pattern-based queries, on the other hand, "search for data elements that match or depart from a predetermined pattern."[4] Pattern-based data mining has been used in the public and private sectors for a number of years for a wide range of applications from conducting market research to detecting financial fraud.

Section 126, by its terms, focuses on pattern-based initiatives and provides the following definition for data mining:

> a query or search or other analysis of one or more electronic databases, where—
>
> (A) at least one of the databases was obtained from or remains under the control of a non-Federal entity, or the information was acquired initially by another department or agency of the Federal Government for purposes other than intelligence or law enforcement;
>
> (B) the search does not use personal identifiers of a specific individual or does not utilize inputs that appear on their face to identify or be associated with a specified individual to acquire information; and
>
> (C) a department or agency of the Federal Government is conducting the query or search or other analysis to find a pattern indicating terrorist or other criminal activity.[5]

[3] Specifically, the information involved in the qualifying initiatives has been lawfully acquired by the Government through voluntary submission, legal process, or contracts with commercial data aggregators.

[4] United States Government Accounting Office, GAO-05-866, *Data Mining*, pg. 5 (August 2005).

[5] USA PATRIOT Improvement and Reauthorization Act of 2005, Pub. L. No. 109-177, § 126(b)(1), 120 Stat. 192, 228.

In addition, section 126 defines a database specifically to *exclude* "telephone directories, information publicly available via the Internet or available by any other means to any member of the public, any databases maintained, operated, or controlled by a State, local, or tribal government (such as a State motor vehicle database), or databases of judicial and administrative opinions."[6]

Pattern-based data mining, when used properly, can be a critical tool that increases investigative efficiency, thus enhancing security and prevention efforts and reducing crime. Pattern-based data mining also promotes important privacy interests. Specifically, pattern-based data mining serves a valuable role in narrowing a set of individuals meriting additional, perhaps more-intrusive investigation. Such pattern-based data mining may, however, present privacy issues, including questions about the retention, analysis, and potential sharing by the government of lawfully obtained information.

Federal statutes and internal DOJ policies and procedures are designed to mitigate potential privacy concerns. For example, privacy impact assessments (PIAs) completed by the Department pursuant to the E-Government Act of 2002 address the issue of the existing authority for the collection of information and advanced analysis of such information. The goal of a PIA is three-fold: (1) to ensure handling of information conforms to applicable legal, regulatory, and policy requirements regarding privacy; (2) to determine the risks and effects of collecting, maintaining and disseminating information in identifiable form via an electronic information system; and (3) to evaluate protections and alternative processes for handling information to mitigate potential privacy risks. In guidance updated in August of 2006, the Department's Privacy and Civil Liberties Office (PCLO) indicated that PIAs should be conducted when an office is, *inter alia*, developing or procuring any IT systems or projects that collect, maintain, or disseminate information in identifiable form from or about members of the public, initiating, consistent with the Paperwork Reduction Act, a new electronic collection of information in identifiable form for 10 or more persons (excluding agencies, instrumentalities or employees of the federal government), or is changing an existing system in a manner that creates new privacy risks (such as when converting from paper-based records to electronic systems or when merging, centralizing, or matching databases that contain information in identifiable form with other databases).[7]

Moreover, the Department has long been subject to, and is diligent in its compliance with, the Privacy Act of 1974, 5 U.S.C. § 552a. The Privacy Act's requirements generally are applicable to records that identify and are about U.S. citizens and legal permanent resident aliens, and that are retrieved from a system by reference to an individual's name or other personal identifier. As a result, any information that is produced as a result of pattern-based data mining that meets these criteria would be subject to the Act's requirements. Among these

[6] USA PATRIOT Improvement and Reauthorization Act of 2005, Pub. L. No. 109-177, § 126(b)(2), 120 Stat. 192, 228. The Department of Justice understands "publicly available via the Internet or available by any other means to any member of the public" to include databases of information that are available to the public for a fee.

[7] August 7, 2006 "Privacy Impact Assessments: Official Guidance," Privacy and Civil Liberties Office; *see also* E-Government Act of 2002, Pub.L. 107-347, Title II, § 208, Dec. 17, 2002, 116 Stat 2899, 2921, codified at 44 U.S.C.A. § 3501 note; OMB Memorandum 03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*, dated Sept. 23, 2003.

requirements are the following: that the agency maintain only such information about an individual that is "relevant and necessary to accomplish a purpose of the agency required to be accomplished by statute or by Executive order of the President," 5 U.S.C. § 552a (e)(1); that the agency publish descriptive notices in the Federal Register of all records systems about individuals from which information is retrieved by reference to their name or personal identifier, 5 U.S.C. § 552a (e)(4); that the agency "maintain all records which are used by the agency in making any determination about any individual with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination," 5 U.S.C. § 552a (e)(5); and that the agency "establish appropriate administrative, technical and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained," 5 U.S.C. § 552a (e)(10). While law enforcement agencies are entitled to exempt their systems from subsections (e)(1) and (e)(5) pursuant to subsection (j)(2) of the Act , the Department's components nevertheless recognize that relevancy and accuracy of the information that they rely upon serves to further their law enforcement mission. Furthermore, exemption cannot be claimed from (e)(4) or (e)(10), nor from subsection (b) of the Act, the very core of the Act that prohibits disclosure of Privacy Act information except under certain circumstances. In addition, the Department is in the process of developing a working group and process designed to analyze proposed FBI initiatives, such as those intending to use advanced analytical tools, on the basis of considerations such as efficacy and privacy impacts. This working group will be comprised of senior officials from the FBI's General Counsel's office and operational components, and the Department of Justice's National Security Division, Privacy and Civil Liberties Office, and other components.

One potential privacy issue with respect to any pattern-based data mining initiative is whether the pattern-based data mining is undertaken for a legitimate purpose. In general, such initiatives have long been recognized as legitimate and permissible law enforcement techniques.[8] In fact, each of the initiatives described in more detail below is grounded in traditional law enforcement techniques designed to discern patterns of criminal activity and to focus resources appropriately. The initiatives are simply designed to accomplish these goals with greater efficiency and accuracy. In addition, in each of the substantive areas in which pattern-based data mining initiatives have or are being developed, the FBI has statutory authority to conduct criminal investigations, which can include the data mining initiatives described herein.

A second potential privacy issue relates to the security of the information and how it is retained. In this regard, agencies that administer a pattern-based data mining initiative must ensure that the information initially collected is secure and that users utilize the particular tools only for authorized purposes. The Privacy Act requires that agencies "establish appropriate administrative, technical and physical safeguards to insure the security and confidentiality of records and to protect against any anticipated threats or hazards to their security or integrity which could result in substantial harm, embarrassment, inconvenience, or unfairness to any individual on whom information is maintained." 5 U.S.C. § 552a (e)(10). In addition, the Federal Information Security Management Act of 2002 (FISMA), along with the Federal Information Processing Standards (FIPS) published by the National Institute of Standards and

[8] *See, e.g., Data Mining: An Overview*, Congressional Research Service, December 14, 2004.

Technology (NIST), define requirements for securing agency information systems. These requirements are implemented at the agency level with information security plans that include certain specific controls designed to ensure that only individuals with proper authorization can access the pattern-based data mining tools and that users have proper training on the use and sensitivity of the system. Controls such as audit logs also help ensure that authorized individuals are only using the data mining tools for official business. Again, where applicable, a PIA will include descriptions of such security controls. Moreover, the Department is required to comply with the Privacy Act's subsection (b) disclosure prohibition, which restricts disclosure within the agency to those officers and employees of the agency "who have a need for the record in the performance of their duties," 5 U.S.C. § 552a (b), as well as its subsection (c)(1) requirement to keep an accurate accounting of disclosures made outside of the agency.[9]

A subsequent potential privacy concern relates to the security of information once the analysis has been undertaken. Again, the protections required by FISMA, and implemented in Departmental security policies, ensure that such data is not accessed by unauthorized users through strict access controls and audit capabilities. If information from a data analysis initiative ends up in an investigative file, the data is retained in accordance with the retention schedule of the investigative file. If that investigative file and the underlying record of the pattern-based data mining initiative are subject to the Privacy Act, then that record will also be subject to the protections of the of the Privacy Act.

As described above, a PIA conducted for a system will require an agency to evaluate the potential privacy risks of a pattern-based data mining initiative and describe mitigation procedures that have been put in place to counter such potential risks. One of the required questions in the Department's standard PIA requires information about security features of the system. The Department's Privacy and Civil Liberties Office is fully engaged in the development and analysis of any PIA on a major information system or national security system done by any component within the Department, providing additional insight into the potential privacy concerns at stake and potential for mitigating those concerns. Furthermore, in several of the initiatives described herein, personal information is not forwarded to FBI investigators unless it is necessary for opening an investigation pursuant to the Attorney General Guidelines. By minimizing the access to personal information, the risk of a security breach of this data is lessened.

The final privacy issue relates to the accuracy of the data to be searched and the potential for misidentification of innocent persons by a pattern-based data mining initiative. Section 126 provides that for an initiative to qualify as pattern-based data mining, it must involve a query, search, or analysis of one or more electronic databases, where "at least one of the databases was obtained from or remains under the control of a non-Federal entity, or the information was acquired initially by the Federal Government for purposes other than intelligence or law enforcement."[10] Consequently, the initiatives below do not involve the analysis solely of information acquired in the first instance by the Department for law enforcement or intelligence

[9] Disclosures made under the Freedom of Information Act, 5 U.S.C. § 552, are excluded from this requirement. 5 U.S.C. § 552a (c)(1).

[10] USA PATRIOT Improvement and Reauthorization Act of 2005, Pub. L. No. 109-177, § 126(b)(1), 120 Stat. 192, 228.

purposes and retained in the Department's control. As such, initial responsibility for accuracy with respect to the information used in the initiatives below lies with the record owner. Where the source is another government agency and the data is in records covered by the Privacy Act of 1974, the attendant accuracy requirements of that statute apply to the agency that owns the records.[11] In some of the initiatives described in this report, the queried data is supplied by individuals who are likely to provide accurate data as they are voluntarily providing information as a victim. For example, in the FBI's Identity Theft Intelligence initiative, the data that is searched is derived from information that is voluntarily provided by individuals with the knowledge that the information may be forwarded to law enforcement. With respect to initiatives in which information is obtained from a commercial data aggregator, those private parties have strong business incentives to consistently provide accurate information. There are comparable measures in other initiatives designed to assure accuracy. Furthermore, in each initiative in which the data come from victims or other members of the public, an analyst will verify the data with basic analytical tools to correct misspellings and obvious errors before it is used.

As to the accuracy and completeness of data searched by pattern-based data mining initiatives, search results are routinely checked by the use of the following measures. Leads generated by pattern-based data mining initiatives are not automatically accepted and acted upon, thus reducing the risk of "false positives." Rather, query results from these initiatives are independently evaluated by highly skilled analysts. The results are then passed along to investigators who also closely review results before taking any investigative action. These results are only used for lead purposes and no action is taken based solely on the analytic products produced by such pattern-based data mining initiatives. Internal DOJ and FBI procedures, including the Attorney General's Guidelines On General Crimes, Racketeering Enterprise and Domestic Security/Terrorism Investigations and the Attorney General's Guidelines for FBI National Security Investigations and Foreign Intelligence Collection ("NSIG") (collectively "Attorney General Guidelines"), set forth the Department's general policy that investigations should be undertaken by the least-intrusive means such that, as a general matter, the investigation of lead information must be undertaken by non-intrusive means prior to the use of more intrusive investigative means.[12]

The Department of Justice realizes that there are privacy risks inherent in the use of pattern-based data mining initiatives, as there are with most law enforcement investigative techniques. As with all law enforcement techniques, the Department strives to mitigate such potential privacy risks through compliance with federal statutes and Departmental policies and regulations, so that the Department can carry out its law enforcement and prevention mission while protecting the privacy and civil liberties of our nation's citizens. In addition to this report,

[11] 5 U.S.C. § 552a(e)(5).

[12] Traditional pattern-based data mining models that are applied to a large general population set to produce names that fit a pattern should be subjected to rigorous quality control particularly at the development stage to reduce the risk of producing false positives. As discussed in more detail below, the initiatives responsive to this reporting requirement either do not apply a pattern to produce identities, or apply traditional analytical tools to identify patterns and links among a relatively small, filtered data set of persons who are more likely than those outside of the data set to be engaged or have engaged in the criminal conduct of concern. Nevertheless, the danger of false positives is always a concern with any analytical tool and is mitigated in these initiatives as described throughout this report.

as set forth above, the Department is preparing a working group to further investigate potential privacy issues related to initiatives intending to use advanced analytical tools and determine whether further internal policies and procedures need to be implemented.

II. FBI'S Pattern-Based Data Mining Initiatives as Defined by Section 126

The Department of Justice has identified six initiatives that arguably meet the criteria established by the Act. The following six initiatives are described in the body of this report:

1) System-to-Assess-Risk (STAR) Initiative
2) Identity Theft Intelligence Initiative
3) Health Care Fraud Initiative
4) Internet Pharmacy Fraud Initiative
5) Housing Fraud Initiative
6) Automobile Accident Insurance Fraud Initiative

Brief descriptions of other initiatives conducted by the Drug Enforcement Administration (DEA), the Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF), and the FBI have been included in this report even though we do not believe that they qualify as pattern-based data mining as defined in section 126. We have also included brief descriptions of certain IT systems that are either operational or in development, even though none are being used in any initiative meeting the definition set forth in section 126.

As is clear from the detailed descriptions below, the initiatives vary widely in terms of sophistication and subject matter. At one end of the spectrum is STAR, an initiative currently in development that is focused on preventing terrorist activities. At the other end of the spectrum are initiatives that use widely available commercial off the shelf software to determine patterns in criminal activity relating to identity theft, mortgage fraud, and other types of financial crimes. The amount and type of information involved in the different initiatives also varies significantly, as do the privacy implications and protections. The descriptions below therefore provide varying degrees of detail.

A. System-to-Assess-Risk (STAR) Initiative

(1) *Description:* To address its growing data processing and analysis needs, the Foreign Terrorist Tracking Task Force (FTTTF) initiated the development of the System to Assess Risk, STAR. FTTTF analysts will use this risk assessment software system to help them prioritize persons of possible investigative interest that analysts are assessing in support of a specified terrorist threat. Analysts will begin the process with names and at least one other personal identifier of individuals already considered of interest, and this risk assessment software system will prioritize the analyst's search results in data sets already lawfully collected and available as part of the FTTTF Data Mart (defined below). STAR, itself, will not produce the identities of these persons of interest—it will only prioritize the risks associated with the persons of interest after they are identified. They will be identified based on either information from credible sources or, depending on the parameters of the threat, a preliminary search of existing FTTTF

government data bases. These individuals may be known or suspected terrorists who are watch listed in the Terrorist Screening Data Base (TSDB) but, again depending on the threat, the search is not limited to the TSDB and may include other identities, within FTTTF data bases, who fit the criteria. No commercial data base will be searched to produce the identities of persons of interest (although, as noted below, STAR will query a commercial data base with respect to individuals who have already been identified as of interest). In order to do this, STAR will evaluate and process voluminous FTTTF data in a timely and efficient manner by leveraging a data analysis system that is designed to assess the risk potential of possible terrorism threats. The intent of the system is to increase the efficiency and effectiveness of the FTTTF analysts as they track and detect known and suspected terrorists and their supporters. STAR is not yet an operational system, but a prototype will soon be tested. The FBI has prepared a PIA coordinated with the PCLO.

The STAR program itself is more appropriately viewed as subject-based data analysis, rather than pattern-based data mining, in that STAR is only used to retrieve and analyze information with respect to a specified set of individuals determined by a given threat. However, due to the way the risk assessment scoring process is conducted and validated, it could be considered pattern-based data mining, which is why it is included in this report.

Purpose and Background

By way of background, the core mission of the FTTTF is to identify and track both known and suspected terrorists inside the United States or as they attempt to enter this country. The information derived from FTTTF assessments is then reported to U.S. intelligence agencies and to Federal law enforcement officials in order to prosecute, remove and/or deny these individuals entry into the United States. By identifying and locating known and suspected terrorists and their supporters, the daily efforts of this Task Force support the FBI and its Joint Terrorism Task Forces (JTTFs) in fulfilling the goal of protecting the American people from potential terrorist attacks.

As a component of the FBI, the FTTTF has developed a data mart ("FTTTF Data Mart") containing data from U.S. Government and proprietary sources (e.g., travel data from the Airlines Reporting Corporation) as well as access to publicly available data from commercial sources (such as ChoicePoint).[13] The need to process the ever increasing amounts of data as well as large quantities of data queries, combined with the requirement to provide more accurate and focused risk assessments of potential foreign terrorists, has necessitated the development of a new system to automate elements of the risk assessment process. This new system will help to provide actionable and time-sensitive intelligence.

[13] The FTTTF Data Mart is comprised of a classified and unclassified system, consisting of several ingested data sets as well as other data accessed through specific external queries. Some of the data sets are acquired on a one-time basis and other data sets are regularly updated. Data sets are acquired based upon specific mission needs, and a prioritized data ingest list is constantly updated by FTTTF to reflect the most current operational needs. The FTTTF Data Mart does include data from Choicepoint, a commercial data base collecting publicly available data, which is used to augment or verify existing identification information such as an individual's address or place of employment.

In sum, the proper design, implementation, and deployment of this system will provide the following:

- Analysis and processing of the data contained in the FTTTF Data Mart to focus analysts' resources on the most likely potential terrorist suspects.

- Enhancement of FTTTF's ability to assess the risk associated with known and suspected terrorists by repeatedly and rigorously applying a set of rules to generate a risk score for each subject processed by the system.

The objective of STAR is to help analysts determine whether an individual or group of interest may be associated with terrorism by producing a risk assessment score based on a series of indicators of potential terrorist behaviors. STAR processes the results of predicated database queries pertaining to persons who may merit further inquiry. STAR's terrorism risk assessment score helps prioritize and focus the analyst's attention on particular individuals, who might require more in depth individual analysis. STAR does not label anyone a terrorist. It only alerts the analyst that further assessment may be required. It is the analyst who decides if the person or group represents a significant terrorism threat. He or she then sends a lead to a JTTF or to an FBI field agent to assess the information to decide if further action is warranted. In effect, STAR runs a simultaneous series of database queries against a number of data sets, a process that the analysts currently run manually. Automating these queries makes the research and assessment processes more timely and accurate.

In short, STAR does no more than an analyst currently does now by searching through multiple databases and manually assessing the likelihood that a given person or group fits the parameters of a threat to an extent that justifies further investigation.

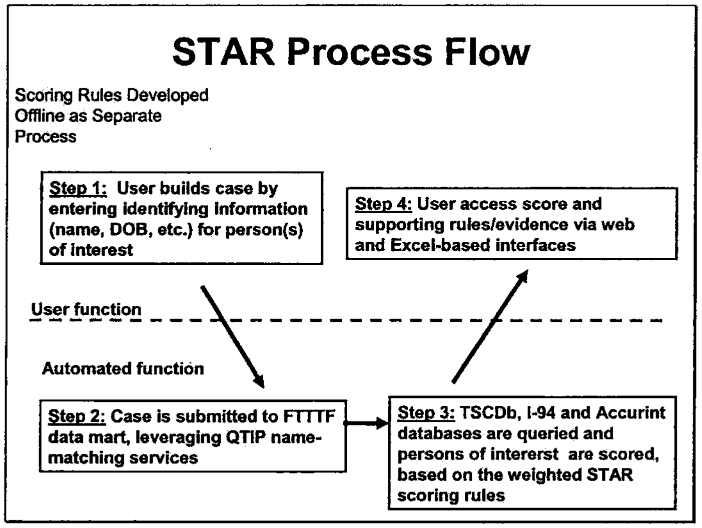
How the Program Works

Experienced FTTTF counterterrorism analysts developed criteria they determined to be indicative of behavior that has historically been associated with terrorism. The criteria were based on the particular experiences of individual analysts, recent terrorism case information from FBI case files, research reports written by government agencies and academic researchers, and on knowledge of available data elements within the FTTTF databases. The criteria were transformed into established scoring "rules" that are applied to volumes of information in order to assess the potential risk of individuals who fall within the parameters of threats received by the FBI. An example of such a rule would be whether a person is on a terrorism watch list. If the name being queried matches a name on the watch list, it would represent a "hit" by STAR, which could increase the person's score.

STAR creates a quantitative risk score of the risk posed by a given person or group of people by following the rules written by the analysts. STAR uses scoring to separate and prioritize those individuals who exhibit characteristics associated with terrorism or have links to terrorist activities. A terrorist scoring model, similar to that of a credit scoring model, has a list of rules that are applied to an individual's data. The significance of each scoring rule is measured by its associated weights. Scores associated with an individual are calculated by summing the weight associated with rules that are triggered by the data. A high score provides the analyst a good place to begin a more in-depth analysis by identifying potential persons of interest for further inquiry without extensive manual processing.

STAR uses Thomas Saaty's Analytical Hierarchy Process (AHP) as the quantitative method for ranking decision alternatives. AHP develops a numerical score relative to how well each rule compares to another rule given the objective of determining match with prioritization criteria. The ranking information is turned into scoring weights, which capture the relative importance of the different rules. STAR compares the data collected about an individual and determines whether each rule is true or false for that individual.

The rules are composed of single statements, a format required by the AHP process; however, most rules, such as age, have little or no weight by themselves. A person has to "trip" a number of rules or "trip" a heavily weighted rule, such as being on a terrorist watch list, before he or she will receive a high STAR risk assessment score. As noted above, the score is the sum of the weights for all of the triggered rules.



STAR Process Flow

Scoring Rules Developed Offline as Separate Process

Step 1: User builds case by entering identifying information (name, DOB, etc.) for person(s) of interest

Step 4: User access score and supporting rules/evidence via web and Excel-based interfaces

Step 2: Case is submitted to FTTTF data mart, leveraging QTIP name-matching services

Step 3: TSCDb, I-94 and Accurint databases are queried and persons of intererst are scored, based on the weighted STAR scoring rules

The STAR Process Flow diagram illustrates how the program will operate. The rules used to score persons of interest have been developed first as a separate process, as described above. The analyst begins the process of using STAR by building a STAR data set that is composed of the names of persons of interest being assessed in support of a predicated threat. These names are often provided to FTTTF as part of the threat information; if the threat information does not identify individuals, the analyst will develop a list of names from the FTTTF Data Mart based on the nature and the specificity of the terrorist threat. Only individuals considered emergent foreign terrorist threats (as opposed to other criminal activity such as U.S. bank robbery threats) will be analyzed. A "terrorist threat" could originate from human intelligence or signals intelligence sources, or from a citizen-originated tip. The analyst must have the individuals' names and at least one other personal identifier before using STAR. STAR does not identify people—it scores people given to it by the STAR user. In this regard, STAR is clearly a subject-based advanced analytical tool.

Once the analyst has submitted the data set of names of the persons of interest into STAR, STAR submits the names to the FTTTF Data Mart by using the Query Tracking Initiation Program (QTIP), a data query tool developed by the FTTTF. STAR, through QTIP, will initially search only three databases for assessment: the Terrorist Screening Center database (TSCDB) which contains consolidated watch lists; the I-94 database which is provided by DHS; and Accurint, which contains public records data which includes information such as addresses, phone numbers, employer information, drivers licenses and pilot licenses (the availability of this information being dependent on state public records policies and laws). It is anticipated that future spirals of STAR will add additional databases. All of these databases are contained within the pre-existing FTTTF Data Mart, and there are formal memoranda of agreement in place with the issuing agencies that govern their use.

Consistent with the FTTTF mission, the focus of the query is foreign terrorists—not U.S. Persons—who fit the threat parameters. Two of the databases searched may nonetheless have U.S. Person information. If the subject of the query has U.S. person associates that are identified in the TSCDB or in Accurint, FTTTF may look at them to see if they have derogatory information. STAR does not automatically search associates. This is a decision that the analyst would make during the course of his or her analysis.

Next, STAR takes the QTIP results and scores them by applying the weighted scoring rules developed by the analysts, assigning a risk score to each person of interest. The score is determined by the total weight of the rules that the person matches. STAR then returns the results to the analyst in the form of an Excel spreadsheet. It also displays the rules that the person matched (based on his/her data) and the underlying evidence that caused particular rules to apply. After scoring the names through STAR, the FTTTF conducts further in-depth analysis of those individuals according to the scoring priorities produced by STAR. This analysis is put in a written assessment of the threat and sent to the relevant FBI field office and/or JTTF for investigative follow-up.

> An example of a threat received by the FTTTF could be that an individual from Pakistan traveling through South Africa from January 1, 2008 through February 15, 2008 is suspected of planning to bomb the U.S. embassy. In this example certain factors are established: names, country of origin (Pakistan), location (South Africa), and time frame (January 1, 2008 through February 15, 2008). Based on this information, before STAR is applied, FTTTF analysts query available databases in order to develop the list of possible subjects meeting the above factors before assessing the threat represented by these individuals.

Once the original set of names is gathered, the analyst will then use STAR. After STAR has assessed, scored, and prioritized the subjects of interest, FTTTF analysts will perform in-depth analysis on prioritized individuals before sending the final results to effected field offices (in whose territory potential subjects meeting the above criteria may be located).

STAR results will all be kept within the FTTTF. The STAR system will not be used in the field and the STAR "scores" will not be sent to the field. STAR results will be stored in the FTTTF Data Mart. Data Mart has a completed PIA that was approved by the FBI's Senior Privacy Officer in October 2005 and was further submitted to the DOJ Chief Privacy Officer after that office was created in February 2006. A determination has not been made as to whether the names of all persons entered into the STAR program will be retained on a temporary basis as part of the record and eventually archived for research referral purposes only. That determination will be made as part of the PIA done on STAR specifically before it is deployed. The efficacy of STAR will be tested and refined as the investigative measures taken in response to its results provide new information about its reliability.

Access and Use of Information

This software is initially designed for use by FTTTF analysts only. In order to access the program, STAR requires that analysts enter the system with a separate login and password. All users of the STAR system must be authorized by FTTTF management for use. FTTTF analysts using STAR are trained in data protection policies of the FBI, and STAR results are classified at the SECRET level. The scoring results for individuals are not editable; however, if a subject receiving a high score is not verified as a threat, the information is not used.

(2) *Plan to use*: STAR is currently under development, and the target date for deployment will be determined once the PIA is approved by the FBI's Office of the General Counsel, but is anticipated sometime this year. A prototype version is ready to test, and the goal is to do so during the second quarter of 2007; however, this testing will not occur without full compliance with the PIA and any applicable retention policies. The prototype version will be tested with operational data. Currently, STAR is tested with old sample data.

(3) *Efficacy*: Once operational, the validity of STAR results will be tested by two means: 1) feedback from investigative follow-up; and 2) a continuous application of SPSS (Statistical Package for the Social Sciences) Clementine™, a COTS statistical data analysis tool, to evaluate the common characteristics in the existing FTTTF databases. Designed to analyze data, Clementine™ will support STAR by validating the analyst-written scoring rules by looking in the databases to determine if those rules are reflected in the data and in what strength. Analysts initially wrote 38 scoring rules. Clementine™ assisted in eliminating three of them, either because the rule did not work or there was insufficient data to support the rule. Clementine™ is a separate part of the risk assessment process and not technically integrated with STAR.

In addition to these two means, STAR analyzes data quickly and efficiently and is expected to lead to consistent analytical results. The STAR rules will be periodically re-examined and updated, based on investigative feedback from its use, by FTTTF staff personnel and reviewed by FTTTF management through a board composed of FTTTF analysts.

The severity of terrorist threats affects the timing and nature of analysis and threat response needed by the FBI. STAR will save valuable time in helping to narrow the field of individuals potentially meriting additional investigation with respect to severe and imminent threats. FTTTF did evaluate commercial solutions for priority analysis of threats but none was found suitable. Other government risk-analysis programs were looked at as well, but all were designed to assess the potential damage to critical infrastructure sites and property, not to assess the risk potential of suspected foreign terrorists. The STAR initiative is the best available method for intelligence and law enforcement purposes.

(4) *Privacy and civil liberties impact*: As noted in the introduction, while advanced analytical tools may present potential privacy implications, the FBI has lawfully obtained the information that is being analyzed and has the authority and responsibility to conduct law enforcement investigations and national security investigations, including threat assessments. The FBI may take no investigative action unless Attorney General Guideline requirements are met, and the program is compliant with privacy and IT security regulations. While the data analyzed by STAR with respect to a given threat may contain information about a large number of individuals, it is important to note that the potential of implicating a U.S. person's privacy rights is minimal as most of the analytical work conducted by FTTTF deals with incoming and outgoing foreign nationals.

It is also important to reiterate that STAR is not used to predict terrorist behavior or to label an individual a terrorist. The STAR initiative is a legitimate threat analysis method under Attorney General Guidelines. Individuals are identified as falling within the parameters of a threat, and depending on their risk assessment score, worthy of further investigation or not. The risk assessment scores are based on numerous indicators and require that a substantial amount of elements be met by an individual in order to be included in the data set. The broad range of factors and the scoring mechanism minimize the possibility of arbitrary standards and false identification associated with an individual.

Each factor is legitimately included in the analysis based on years of experience and analyses of thousands of cases. For example, country of origin, which refers to the country in which an individual resides or began his travels, has historically proven to be an important factor in certain situations. One can well imagine a threat originating from a specified country (as in the example set forth above), in which case determining the country of origin of individuals referred for analysis could be highly relevant. To ignore this indicator could result in significant gaps in the analysis, reducing the accuracy and value of this analytical tool. Moreover, the potential harms of including a more sensitive factor as an indicator of terrorism is mitigated by, *inter alia*, the breadth of the indicators analyzed.

STAR is used to respond to a foreign terrorist threat that is received by U.S. authorities which, based on all information available to the authorities at the time, is deemed sufficiently credible to justify further action. By prioritizing individuals identified within the data set, STAR provides a quick determination for follow-up analysis and possible investigation. Only those individuals who best fit the parameters of the list of indicators score high enough to warrant further scrutiny. STAR does no more than an analyst currently does now by searching through multiple databases and manually assessing the likelihood that a given person or group fits the parameters of a threat to an extent that justifies further investigation. STAR just does it faster, more thoroughly, and more consistently in order to protect the citizens of the United States. Names of persons who are not scored highly by STAR and, for that reason, are not subjected to further analysis or investigation will be retained in FTTTF as part of the record of that particular threat assessment. Those names will not be disseminated.

(5) *Law and regulations*: The proposed STAR initiative is based on the authority of the FBI to conduct lawful investigations, specifically threat assessments under the NSIG, and complies with relevant Privacy Act and FISMA requirements, and DOJ's guidelines, privacy policy and practices. There are, furthermore, no prohibitions on the FBI's ability to use advanced analytical tools and initiatives on lawfully obtained information, provided specified standards and other restrictions are met. Finally, this proposed data mining initiative does not implicate any privacy rights protected by the Fourth Amendment of the Constitution of the United States.

First and foremost, the following authorities form the legal foundation for the FBI to conduct counter-terrorism investigations--and to collect and use information and employ lawful investigative and analytical techniques in furtherance thereof:

- 28 U.S.C. § 533 authorizes the FBI to investigate violations of federal law, which includes acts of terrorism (18 U.S.C. § 2332b, for which the FBI has primary investigative jurisdiction);
- 28 U.S.C. § 534 authorizes the FBI to collect and retain criminal information;
- 28 C.F.R. § 0.85 authorizes the FBI to conduct federal criminal investigations and to assume lead agency role in counter-terrorism investigations;
- The NSIG authorize the FBI to conduct investigations of threats to national security, including the preliminary assessments of these threats; to employ all lawful techniques in that pursuit; and to collect and retain information from lawful sources in compliance with the Constitution and federal law;
- Homeland Security Presidential Directives (HSPD) 2, 6, and 11—all of which direct the strengthening of screening and analysis program to detect, identify, and interdict individuals entering or within the United States who pose a terrorist threat to national security. HSPD 2 specifically directs the FTTTF to perform this function;
- National Security Presidential Directive 46 (War on Terror) also sets forth strengthening of terrorist screening tools as a major objective of national policy.

Second, the Department of Justice complies with current laws and regulations regarding privacy, such as the Privacy Act of 1974, FISMA and Federal Information Processing Standards (FIPS) published by the National Institute of Standards and Technology (NIST) documents. Each of these set forth requirements for securing agency information and IT systems. With respect to the Privacy Act, to the extent that the records concern U.S. citizens or lawful permanent residents, the information collected is part of the FBI's Central Records System, which has both published system of records notices and published exemptions from certain Privacy Act provisions requiring notice and individual access to records.[14] In addition, the Department's policy of

[14] The system of records notices appear at 63 Fed. Reg. 8671 (Feb. 20, 1998), amended 66 Fed. Reg. 8425 (Jan. 31, 2001), 66 Fed. Reg. 17,200 (Mar. 29, 2001), and 72 Fed. Reg. 3410 (Jan. 25, 2007), and the exemption regulations appear at 28 C.F.R. § 16.96 (2006).

conducting PIAs for IT systems ensures that the FBI is aware of and mitigates potential privacy risks.

Third, the Department of Justice has set forth guidance on implementing privacy policy and practices that apply to all data collection and use techniques in FBI investigations. These include, primarily, the Attorney General's Guidelines on General Crimes, Racketeering Enterprise and Domestic Security/Terrorism Investigations and NSIG. The NSIG authorize the use of techniques like STAR in the way STAR is set to be deployed and also describe how checking of lead information by non-intrusive means must be done before formal investigative activity is undertaken.

Fourth, data mining has been recognized by both the Executive and Legislative branches as a legitimate law enforcement analytical technique. *See, e.g.,* Homeland Security Act of 2002, Section 201 (e)(14); P.L. 107-296, Nov. 25. 2002; *Data Mining: An Overview,* Congressional Research Service, Dec.14, 2004.

The Department of Justice always must conduct itself within the bounds of the United States Constitution. The FBI's STAR initiative, which queries only information that already has been lawfully obtained by the Government, does not infringe upon any privacy right protected by the Fourth Amendment to the Constitution of the United States.

The Fourth Amendment guarantees "[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures." The Supreme Court repeatedly has held that this guarantee applies only where the individual invoking the protection of the Fourth Amendment has a "justifiable," "reasonable," or "legitimate" expectation of privacy that has been invaded by government action. *Minnesota v. Carter*, 525 U.S. 83, 88 (1998); *Smith v. Maryland*, 442 U.S. 735, 740 (1979); *United States v. Miller*, 425 U.S. 435, 442 (1976); *Katz v. United States*, 389 U.S. 347, 353 (1967). There is both a subjective and an objective component in evaluating whether such an expectation of privacy is protected by the Fourth Amendment. An individual must subjectively believe that he is entitled to privacy, and that individual's expectation also must objectively be "one that society is prepared to recognize as reasonable." *Smith*, 442 U.S. at 770 (quoting *Katz*, 389 U.S. at 361 (Harlan, J., concurring)). "Official conduct that does not 'compromise any legitimate interest in privacy is not a search'" within the meaning of the Fourth Amendment. *Illinois v. Caballes*, 543 U.S. 405, 408 (2005) (quoting *United States v. Jacobsen*, 466 U.S. 109, 123 (1984)).

The analysis of information by the FBI that has already been obtained lawfully (i.e., consistent with the Fourth Amendment) by the Government does not implicate any constitutionally recognized expectation of privacy. *See United States v. Joseph*, 829 F.2d 724, 729 (9th Cir. 1987) ("[E]xamination by another law enforcement agency is not a sufficiently distinct intrusion into the defendants' privacy to trigger the requirements of the Fourth Amendment.") (quoting *United States v. Romero*, 585 F.2d 391, 396 (9th Cir. 1978)); *Jabara v. Webster*, 691 F.2d 272, 278-79 (6th Cir. 1982) ("We do not believe that an expectation that information lawfully in the possession of a government agency will not be disseminated, without a warrant, to another government agency is an expectation that society is prepared to recognize as reasonable.") (holding that Fourth Amendment did not prohibit FBI from obtaining foreign intelligence information collected by NSA); *United States v. Hearst*, 563 F.2d 1331, 1347 (2d Cir. 1977) (declining "to mutate the prohibitions of the Fourth Amendment, which deal with government-instigated searches and seizures, into a code of regulations governing interagency transfer of evidence legitimately in government control"); *United States v. Gargotto*, 476 F.2d 1009, 1014 (6th Cir. 1973) ("Evidence legally obtained by one police agency may be made available to other such agencies without a warrant, even for a use different from that for which it was originally taken."); *see also Johnson v. Quander*, 440 F.3d 489, 499 (D.C. Cir. 2006) (government's "storage and use" of DNA information collected "in conformance with the Fourth Amendment" "does not give rise to an independent Fourth Amendment claim"). Likewise, courts have held that the Fourth Amendment does not prohibit the Government from analyzing information already lawfully collected by a third party. *See United States v. Miller*, 425 U.S. 435, 443 (1976) ("This Court has held repeatedly that the Fourth Amendment does not prohibit the obtaining of information revealed to a third party and conveyed by him to Government authorities."); *see also Jacobsen*, 466 U.S. at 118 (holding that the Government does not frustrate an individual's legitimate expectation of privacy by examining information already revealed to private parties: "The Fourth Amendment is implicated only if the authorities use information with respect to which the expectation of privacy has not already been frustrated.").

(6) *Privacy and accuracy protection policies*: First, NSIG, authorities cited above and others (for example, Executive Order 12333) prohibit the collection of information and related investigative activity based solely on a subject's exercise of rights guaranteed by the First Amendment. Homeland Security Presidential Directive 2, in particular, states that screening for terrorist threats shall be conducted in a manner that "safeguards legal rights, freedoms, civil liberties, and personal privacy," and, in a similar vein, the Attorney General's Guidance on the Use of Race in Law Enforcement prohibits investigative activity based solely on an individual's race and (by inference) ethnicity.

Second, FTTTF policies include privacy protections such as:

- limiting access to STAR to specified users who are trained in the proper uses of personal information and the predication requirement for conducting queries of the data sets, and who have executed non-disclosure agreements;
- training STAR users so that they fully understand what the STAR results mean and understand that the results must be analyzed in depth before taking any further action;
- ensuring that STAR results are not disseminated to individuals who do not understand and could misinterpret the results; and
- ensuring that STAR queries, results, and analyses are maintained only on authorized office computers.

Furthermore, analysts using STAR are working entirely within a secure office environment with limited physical access, password controlled computer access, an effective audit capability, and in-house dissemination only to those who need to know.

Third, FTTTF's charter dictates that it collect data and analyze them only in response to a credible, predicated foreign terrorist threat; meaning that FTTTF will not use STAR to score individuals for links to terrorism on its own initiative. Accuracy is ensured by the actions of the analyst in following up STAR prioritization with further analysis and corroboration from FBI case files, other routine database checks of the law enforcement and intelligence communities, and open source verification before any investigative action by the field office is taken.

Fourth, even if an entry "scores high" on the STAR spectrum, the NSIG prohibit further investigative activity beyond the initial threat assessment (which, itself, precludes intrusive investigative activity) unless the NSIG's investigative criteria for a preliminary inquiry or full investigation are satisfied.

Fifth, as re-enforced by the NSIG, departmental regulations, and relevant statutes, a National Security Letter, surveillance pursuant to the Foreign Intelligence Surveillance Act, or the issuance of a grand jury subpoena against a STAR- identified individual may not be used unless and until the legal predicates and the procedures for those techniques are met and followed.

Sixth, the STAR initiative will be subjected to a thorough privacy impact assessment under the privacy policies of the Department of Justice and the FBI before it is implemented in its final form—an assessment which will address such privacy issues as lawful and appropriate collection, retention, use and dissemination of personal information as well as the use of a control program to verify the accuracy of the score results and the integrity of the process. FBI's Office of the General Counsel attorneys will remain engaged in the implementation and evaluation of STAR and, in particular, the validity of the scoring rules.

B. Identity Theft Intelligence Initiative

(1) *Description*: This initiative uses Microsoft Excel, Microsoft Access, and Analyst Notebook I2 to extract consumer complaints from the Federal Trade Commission's Identity Theft Clearinghouse[15] into an FBI database to develop clusters of common identities, phone numbers and e-mail addresses of subjects of complaints in a given geographical area.

Once imported into the FBI's database, this information is then compared by FBI analysts—using basic analytical tools such as Microsoft Access and I2 Analyst Notebook—to internal FBI case complaints of identity theft and reports of suspicious financial transactions filed with the Financial Crime Enforcement Center (FinCEN) for further verification and corroboration. Subjects identified by the FTC and FinCEN data are also run against private data aggregators such as LexisNexis, Accurint, and Autotrack to further verify the accuracy of the information. Privacy protection and data accuracy are staples of the business practices of these nationally known companies, each of which has a comprehensive privacy policy published on its web site. The result of these comparisons is a knowledge base from which the FBI can evaluate identity theft typologies, identify theft rings through subject relationships, and send leads to the affected field offices. The knowledge base includes a written analysis supporting the

[15] Given the Federal Trade Commission's enforcement authority, this information is arguably collected for law enforcement purposes, in which case this initiative would fall outside the scope of section 126(b)(1)(A); in the interests of full disclosure, however, the Department has included information on this initiative.

identification of the subjects, a spreadsheet of the complaints, and charts showing activity relationships among them.

(2) *Plans for use*: This initiative is an ongoing project that has been used to generate leads for field offices to pursue since it was first introduced in late 2003.

(3) *Efficacy*: This initiative is designed to identify individuals and their associates who are the subjects of multiple, similar consumer complaints in a given geographic area. Because this form of crime can only succeed with a victim for a short period of time and requires multiple victims, this technique is an efficient and effective way to identify the more serious and prolific offenders of identity theft. As with the other financial crimes initiatives, the effectiveness of the technique is evaluated primarily by investigative means. To this point, more than 13 targeting packages have been developed as a result of this initiative. Going forward, effectiveness will continue to be evaluated by determining whether the identified names lead to viable investigations and prosecutions. Given the type of technology that is being used (widely available COTS), the validity or efficacy of the computer software or technology have already proven reliable and efficient.

(4) *Privacy and civil liberties impact*: As noted above, the potential privacy and civil liberties impacts of the initiatives described in this report vary. For example, as compared to STAR, the information involved in this initiative is far more limited and focused, and is provided by the crime victims themselves. Although the complaints extracted from the FTC database contain victim identities, this personal information is provided voluntarily by the victim with the express warning that it may be forwarded to, and acted upon, by law enforcement. The victim identities are only incorporated into the FBI database for the purpose of determining relationships among subject identities, phone numbers, and e-mail addresses.

In the event additional investigation is merited, victim information and the complaints themselves are forwarded to the relevant FBI field office as part of a lead and treated with the same privacy protection as any victim information pertaining to any offense investigated by the FBI. *See, e.g.*, 18 U.S.C. § 3771; Attorney General Guidelines for Victim and Witness Assistance. Consumer information that does not indicate criminal activity is retained as part of the record but is not forwarded to the field office for investigation. With respect to the privacy and civil liberties impact on the subjects, the analytical technique is designed to identify identity thieves with a high degree of probability and to eliminate persons named in complaints who do not appear to be associated with crime.

As to the potential implication of the complainant's privacy rights, the name of the individual who filed a complaint would be retained in FBI files only if the analysis reveals the complaint is worthy of investigative follow-up by a field office. However, thereafter, the name of the individual who files the complaint is only in the case file as a victim (like the victim of any federal crime reported to the FBI)—not a subject—and a search of main files for investigative subjects would not come up with the victim's name.

Furthermore, before a potential identity thief becomes the subject of a field office investigation, logical lead follow-up is pursued to verify and corroborate the analytical conclusions. An investigation is opened, and various investigative techniques are subsequently pursued, only if the criteria in the Attorney General Guidelines and relevant statutes are met.

(5) *Law and regulations*: The legal foundation for the FBI to conduct such investigations is derived from:

- 28 U.S.C. § 533 authorizes the FBI to investigate violations of federal law; which include credit card and mail fraud (18 U.S.C. §§ 1030; 1341);
- 28 U.S.C. § 534 authorizes the FBI to collect and retain criminal information;
- 28 C.F.R. § 0.85 authorizes the FBI to conduct federal criminal investigations; and
- The Attorney General Guidelines for General Crimes authorizes the FBI to conduct investigations of violations of federal crimes; to employ all lawful techniques in that pursuit; and to collect and retain information from lawful sources in compliance with the Constitution and federal law.

A complete description of the applicable law and regulations is set forth in Part II.A.5 above.

(6) *Privacy and accuracy protection policies*: First, the FBI complies with laws and regulations regarding privacy, such as the Privacy Act of 1974, FISMA, Federal Information Processing Standards (FIPS) published by the National Institute of Standards and Technology (NIST), and privacy policies established by the Department of Justice in 28 C.F.R. §§ 16.40 and 16.54 and DOJ Order 2640.1 (Privacy Act Security Regulations for Systems of Records). Each of these set forth requirements for securing agency information and IT systems. The information in this initiative becomes part of the FBI's Central Record System, which has both published system of records notices and published exemptions from the notice and record access requirements of the Privacy Act.[16] A PIA will be completed for this initiative.

Second no investigative activity is initiated by an FBI field office against any transaction participant identified by this initiative unless the criteria established in the relevant Attorney General Guidelines are met—which includes the logical evaluation of lead information through other non-intrusive, lawful means. If fraud victims are contacted and interviewed as part of an ensuing investigation, their privacy is protected in accordance with 18 U.S.C. § 3771 and the Attorney General Guidelines for Victim and Witness Assistance.

Third, the use of other techniques is regulated by law and procedure (such as the Federal Wiretap Act for wiretaps and Rule 41 of the Federal Rules of Criminal Procedure for search warrants) designed to ensure that such techniques are lawfully and appropriately employed. These additional investigative techniques would be undertaken separate and apart from any pattern-based data mining initiatives.

Fourth, the personally identifiable information collected by the analyst from FTC records is maintained by the analyst in a password-controlled computer within a restricted access space

[16] The system of records notices appear at 63 Fed. Reg. 8671 (Feb. 20, 1998), amended 66 Fed. Reg. 8425 (Jan. 31, 2001), 66 Fed. Reg. 17,200 (Mar. 29, 2001), and 72 Fed. Reg. 3410 (Jan. 25, 2007), and the exemption regulations appear at 28 C.F.R. § 16.96 (2006).

in FBI Headquarters. Data are only shared with other FBI personnel within the particular analysis unit who have a need to know. Investigative leads produced by the analysis (containing personally identifying information about the subjects and the persons who were victimized) are entered into the FBI Automated Case System (ACS) where they can be accessed by authorized field office personnel. ACS is also password controlled and is internal to the FBI. All Privacy Act requirements that apply to ACS are followed and ACS has a vigorous audit capability. There are distinct access and security restrictions on the use of ACS in the FBI's Manual of Investigative and Operational Guidelines and the FBI's Security Policy Manual. In addition to entry into ACS, the data are transmitted to field offices via secure internal FBI e-mail where they can only be accessed by password within restricted FBI field office spaces.

C. Health Care Fraud Initiative:

(1) *Description*: This initiative enables FBI analysts to research and investigate health care providers who may be continually over-billing Medicare for patient care.

Specifically, this initiative uses Microsoft Excel and Microsoft Access to examine Medicare summary billing records extracted from the Centers for Medicare and Medicaid Services (CMS), supported by the CMS Fraud Investigative Database, Searchpoint (DEA's pharmaceutical claims database, discussed below at Part III.A.1), and the National Health Care Anti-Fraud Association Special Investigative Resource and Intelligence System (private insurance data).

In summary, Medicare claims information is extracted into a spreadsheet by FBI analysts from the CMS database. That information is provided to CMS with the patient's consent to enable processing of the claims.[17] No personally identifying patient information is extracted from CMS as part of this process; instead, the information consists of summaries of the basic provider identification, billing amounts, and summaries taken from claims forms submitted to CMS. In the event analysis indicates that certain providers merit additional investigation, a provider number may be matched, through publicly available information, to the provider name. The data are then subjected to statistical analysis and exported to statistical analysis programs in order to rank the data based on national aberrancy information and to conduct frequency analysis by an analyst at FBI Headquarters. The statistical analysis is designed to give the percent and the dollar amount above the national average and the frequency of a provider's submissions that are above the national averages. CMS provides the national aberrancy information. The data are then exported to a database to accommodate queries for provider names and addresses. Records vary in number depending on the type of provider and whether it is part of a national chain, group practice, or an individual practitioner operating in a single locality.

Finally, the provider information is placed back into an Excel spreadsheet and sent to each FBI field office for investigation of the providers who billed the largest cumulative above-average amounts. There is no new or additional information sent in the returned Excel file. When made available by CMS, peer billing comparisons (comparing a provider's billing to that

[17] HIPAA and its implementing regulations permit access by the FBI by virtue of its health care fraud oversight and law enforcement role. *See, e.g.*, 45 CFR 164.512(d) (permitting FBI access to information for "health oversight purposes").

of his peers) were also conducted by specialty code, in which the billing rates of a medical specialist (for example, a cardiologist) are compared to the rates of a physician with the same specialty for the same diagnosis.

Quality assurance in the execution of this technique consists of manual calculations by the analysts of the billing rates and the number of records followed by further review by field office personnel who receive the information to ensure it is consistent with their own information and FBI case files.

(2) *Plans for use*: This technology was introduced in its present form in 2003 and has been used for the last three years for 54 of the 56 FBI field offices.

(3) *Efficacy*: In order to investigate, prosecute, and deter health care fraud, reliable standards for identifying billing irregularities that exceed national averages so substantially that fraud is a likely explanation have been developed in recent years. These standards allow analysts to assess the vast amount of health care billing information contained in both government and private insurance claims databases. This initiative has resulted in the initiation of more than 50 FBI investigations and nearly 200 referrals to state and other federal agencies. Of these, several providers identified by the initiative were already under FBI investigation, confirming the accuracy of the initiative's results. Many of these investigations and referrals have led to criminal convictions and civil settlements for violations of health care fraud statutes. As with the identity fraud initiative, efficacy is measured not with respect to the widely available and tested technology used by the analysts, but through subsequent investigations.

(4) *Privacy and civil liberties impact*: This initiative does import personally identifying information of health care providers whose volumes of claims, services, and/or billing patterns are so unusual as to prompt further examination for possible referral to FBI field offices for investigation; however, any impact on health care providers is minimal for three reasons. First, the providers have voluntarily initiated the claims upon which the initiative is based, using forms that specifically indicate that false and misleading entries on the form may subject the claimant to criminal and civil penalties.

Second, as noted below, the data that support the analytical indications of fraud are so extensive that the chances of those indicators proving correct are high. The result is that persons reasonably suspected of fraud are identified and those whose data do not indicate fraud are eliminated as subjects of investigations.

Third, any investigation that results from an initiative lead is pursued under the Attorney General Guidelines, which impose threshold criteria that must be met at the outset (no case is opened against a provider unless those criteria are met) and privacy and civil liberties protections that must be adhered to as the investigation progresses. Additional protections inuring from the Attorney General Guidelines and statutory restrictions are built into the investigative progress—such as, for example, opening a case on adequate predication of criminality and applying for a search warrant or Title III surveillance order.

Throughout the process, sources are protected along with case sensitive information, as they would be in cases that did not involve the use of pattern-based data mining techniques. Health care providers who do not meet the excess billing criteria are not further identified to field investigators, although their identities are retained for approximately one year in a database as part of the record and then archived to a disk solely for research referral.

With regard to the impact on patient privacy, as noted above, the records that are extracted from CMS to the FBI do not contain personally identifying information on patients. That information can be obtained through CMS if it becomes necessary in an investigative follow-up; it is not necessary to include the patient's identity as part of the analysis however.

(5) *Law and regulations*: The legal foundation for the FBI to conduct such investigations is derived from the following:

- 28 U.S.C. § 533 authorizes the FBI to investigate violations of federal law; which includes health care fraud (18 U.S.C. § 1035);
- 28 U.S.C. § 534 authorizes the FBI to collect and retain criminal information;
- 28 C.F.R. § 0.85 authorizes the FBI to conduct federal criminal investigations; and
- The Attorney General Guidelines for General Crimes authorizes the FBI to conduct investigations of violations of federal crimes; to employ all lawful techniques in that pursuit; and to collect and retain information from lawful sources in compliance with the Constitution and federal law.

A complete description of the applicable law and regulations is set forth in Part II.A.5 above.

(6) *Privacy and accuracy protection policies*: First, patient privacy is protected because patient identities are not included in statistical analyses in the first instance. In the event that an investigation is opened by the field office, agents must adhere to Department and FBI policy on patient confidentiality. This policy requires that identities be isolated and kept confidential, and, if public exposure is planned at trial, that the patient's consent be obtained.

Second, the FBI complies with current laws and regulations regarding privacy, such as the Privacy Act of 1974, FISMA, Federal Information Processing Standards (FIPS) published by the National Institute of Standards and Technology (NIST), and privacy policies established by the Department of Justice in 28 C.F.R. §§ 16.40 and 16.54 and DOJ Order 2640.1 (Privacy Act Security Regulations for Systems of Records). Each of these set forth requirements for securing agency information and information technology systems. With respect to the Privacy Act, the information collected falls within the FBI's Central Records System, which has both published system of records notices and published exemptions from the notice and personal right of access provisions of the Privacy Act.[18] A PIA will be conducted for this initiative.

[18] The system of records notices appear at 63 Fed. Reg. 8671 (Feb. 20, 1998), amended 66 Fed. Reg. 8425 (Jan. 31, 2001), 66 Fed. Reg. 17,200 (Mar. 29, 2001), and 72 Fed. Reg. 3410 (Jan. 25, 2007), and the exemption regulations appear at 28 C.F.R. § 16.96 (2006).

Third, no investigative activity is initiated by an FBI field office against any health care provider identified by this initiative unless the criteria established in the Attorney General Guidelines are met—which includes the logical evaluation of the lead information through other non-intrusive, lawful means (including FBI record checks; private insurance industry record checks; and the use of developed sources). If patients need to be contacted and interviewed as part of an ensuing investigation, their privacy is protected in accordance with 18 U.S.C. § 3771, the Attorney General Guidelines for Victim and Witness Assistance, and other departmental guidelines respecting patient privacy.

Fourth, the use of more intrusive techniques (such as grand jury subpoenas or electronic surveillance) is regulated by law and procedure designed to ensure that the technique is lawfully and appropriately employed.

Finally, the data that are extracted from CMS to the FBI are transferred via private intranet transit to FBI controlled space. The data are then housed on FBI password-controlled computers, located in restricted space and placed in a drive that is only shared within the analysis unit. As noted above, personal identifiers of patients are not included in these data. Once an investigation is opened by a field office, the information is entered into in FBI ACS, which has its own restricted and password-controlled access. ACS does contain personally identifying information and is in full compliance with the Privacy Act. There are distinct access and security restrictions on the use of ACS in the FBI's Manual of Investigative and Operational Guidelines and the FBI's Security Policy Manual. This information is also sent to field offices, as appropriate, via the FBI's secure, password-controlled e-mail system over FBINet.

D. Internet Pharmacy Fraud Initiative

(1) *Description*: This initiative uses commercial off-the-shelf software such as Microsoft Access and Analyst Notebook I2 to search consumer complaints involving alleged fraud by Internet pharmacies to develop common threads indicative of fraud by such pharmacies. The consumer complaints are originally made to the Food and Drug Administration (FDA) and to the Internet Fraud Complaint Center. Additional information regarding Internet pharmacies that may be involved in the distribution of illegal or counterfeit medications is obtained from open source aggregators and the resulting analyses are compared to FBI case files.

The complaint population data are large enough to develop reliable common threads based on information derived from the analysis of the data. The foci of the analysis are time frame, web sites, sponsor identities, drugs prescribed, amounts charged, financial methods used, and methods of operation. Search and analysis tools, such as Analyst Notebook, are used to identify these commonalities among the complaint population and to create useful link charts and diagrams. The results are also reviewed by FBI intelligence analysts and coordinated with the DEA and FDA to ensure thorough corroboration, deconfliction, and compliance with information sharing directives. The FBI has a high degree of confidence that fraud or other criminality has been identified prior to the lead being sent to an FBI field office for further investigation.

(2) *Plans for use*: The Internet Pharmacy Fraud Initiative was created in December 2005 to identify and prosecute licensed and unlicensed Internet pharmacies involved in the illegal distribution of diverted, counterfeit, or unapproved pharmaceuticals to consumers in the U.S.

(3) *Efficacy*: Because consumer complaints vary considerably in their reliability and their viability as a basis to investigate federal crime, it is necessary and effective to group them together and analyze what they have in common in terms of the subjects of the complaints, their locations, the drugs involved, and the means of payment—and, in addition, to verify the particulars of the complained-of Internet sites through public source data. Only through this means is the quality of the information as a basis for investigation elevated to an acceptable level and the inaccurate, misleading, and possibly bogus complaints eliminated as investigative leads. This initiative performs what analysts and agents used to perform manually. Data and analytical accuracy are primarily assured by analysts' manual calculations and corroboration with public source and other reliable data. Finally, feedback from investigations provides the most reliable assurance of the efficacy of the technique. Thus far, more than 40 leads have been generated as a result of this initiative. As with the financial crimes initiatives, the technology is widely available COTS that has demonstrated and tested reliability.

(4) *Privacy and civil liberties impact*: Names of persons who allege fraud and the names and commercial identities of the subjects of the complaints are included in the data set and the analysis in order to establish links and multiple indicators of fraud (or in the alternative, to eliminate those who appear to be the targets of specious claims). The impact on on-line pharmacy owners and associates should be minimal because they hold themselves out as commercial vendors in a regulated industry and the data indicative of fraud are sufficiently broad and corroborated so as to reduce the risk of including innocent vendors. Furthermore, before investigative action is taken by the field offices, logical non-intrusive lead measures are pursued and the applicable Attorney General Guideline criteria to open an investigation and pursue intrusive investigative techniques must be satisfied.

As to consumer or medical patient data, the complaints are taken by the FDA and the Internet Fraud Complaint Center with the understanding that they will or may be referred to law enforcement for investigation—which is one of the primary purposes of lodging the complaint. Consumer victim information is only used in the analyses to develop relationships showing the same on-line pharmacies as the common thread and thereafter retained in the FBI database at Headquarters solely as part of the record of the analysis, which is the disk or shared drive on which the results of analysis are maintained. Although this information is retained as a record of the analysis, access is restricted to members of the unit involved in the analysis. As a result, there is virtually no impact on the privacy of a consumer who lodged a complaint beyond the voluntary act of the consumer filing the complaint in the first instance.

If a lead is sent to a field office, the consumer information is sent as part of the case file for appropriate follow-up by the case agent. These complainants are either victims or witnesses or both and, as such, their identities are essential to these investigations and eventual prosecutions. Such complainants are also entitled to the same protections that victim or witness identities would be in cases that do not involve data mining. *See, e.g.*, 18 U.S.C. § 3771(a)(8) (requiring federal officials to respect and protect victim privacy and dignity) and the Attorney General Guidelines for Victim and Witness Assistance.

(5) *Law and regulations*: The legal foundation for the FBI to conduct such investigations is derived from the following:

- 28 U.S.C. § 533 authorizes the FBI to investigate violations of federal law; which includes health care fraud (18 U.S.C. § 1035);
- 28 U.S.C. § 534 authorizes the FBI to collect and retain criminal information;
- 28 C.F.R. § 0.85 authorizes the FBI to conduct federal criminal investigations; and
- The Attorney General Guidelines for General Crimes authorizes the FBI to conduct investigations of violations of federal crimes; to employ all lawful techniques in that pursuit; and to collect and retain information from lawful sources in compliance with the Constitution and federal law.

A description of the applicable law and regulations is set forth in Part II.A.5 above.

(6) *Privacy and accuracy protection policies*: First, consumer and patient privacy is protected because patient information (other than the medication ordered which, again, has been voluntarily submitted) is not included in the statistical analysis in the first instance, and, when an investigation is opened by the field office, the FBI adheres to Department and FBI policy on patient confidentiality. This policy requires identities to be isolated, kept confidential, and if public exposure is planned at trial, to obtain the patient's consent.

Second, the FBI complies with current laws and regulations regarding privacy, such as the Privacy Act of 1974, FISMA, Federal Information Processing Standards (FIPS) published by the National Institute of Standards and Technology (NIST), and privacy policies established by the Department of Justice in 28 C.F.R. §§ 16.40 and 16.54 and DOJ Order 2640.1 (Privacy Act Security Regulations for Systems of Records). Each of these set forth requirements for securing agency information and information technology systems. With respect to the Privacy Act, the information collected falls within the FBI's Central Records System, which has both published system of records notices and published exemptions from the notice and personal right of access provisions of the Privacy Act.[19] A PIA will be conducted for this initiative.

Third, pursuant to Attorney General Guidelines, no investigative activity is initiated by an FBI field office against any on-line pharmacy identified by this initiative unless the criteria established in the Attorney General Guidelines are met—which includes the logical evaluation of the lead information through other non-intrusive, lawful means (FBI record checks, private insurance industry record checks, and checking with developed sources) and, in addition, use of more intrusive techniques (such as grand jury subpoenas or electronic surveillance) is regulated by law and procedure designed to ensure that the technique is lawfully and appropriately employed.

[19] The system of records notices appear at 63 Fed. Reg. 8671 (Feb. 20, 1998), amended 66 Fed. Reg. 8425 (Jan. 31, 2001), 66 Fed. Reg. 17,200 (Mar. 29, 2001), and 72 Fed. Reg. 3410 (Jan. 25, 2007), and the exemption regulations appear at 28 C.F.R. § 16.96 (2006).

Finally, all personally identifying information is retained on a password controlled desktop computer and server in FBI restricted spaces. Access is limited to those in the unit with the need to know. Information of investigative value to the field is entered into the FBI ACS, a password-controlled databases with a robust audit capability. There are distinct access and security restrictions on the use of ACS in the FBI's Manual of Investigative and Operational Guidelines and the FBI's Security Policy Manual.

E. Housing Fraud Initiative

(1) *Description*: This initiative uses public source data containing buyer, seller, lender, and broker identities and property addresses purchased from ChoicePoint, Inc. in order to uncover fraudulent housing purchases.

A set of data purchased for this initiative contained real estate transactions in which properties were purchased and sold within a short-time period with a differential price above a set amount. These data were exported to a Microsoft Access database and include buyer, seller, lender, address and values. Analysts then reviewed the data and identified suspicious behavior. Other sources such as ChoicePoint On-Line were accessed by FBI Headquarters analysts to determine related transactions. These connections were researched and developed solely by an analyst, not by a program.

The statistical analyses of suspected fraud were forwarded to the field office as leads for investigative follow-up. Originally, database information itself was saved to a disk and sent to the affected field office. Beginning in late 2005, the database was made available to FBI field offices to access directly through the FBI Intranet to use as appropriate in future investigations.

(2) *Plans for use*: This initiative was first completed in 1999. However, it continues to be updated by ChoicePoint as new real estate transactions that meet the criteria take place.

(3) *Efficacy*: Microsoft Access is a widely available and highly reliable tool. As an operational matter, this initiative is very effective at identifying those real estate transactions most likely to be fraudulent in a given area—especially in situations where the same lenders and brokers are consistently associated with a similar fraudulent process (commonly known as "property flipping"). Leads sent from the initiatives to FBI field offices have led to several investigations and, in some cases, convictions. Currently, the field offices themselves can access the database via the FBI Intranet to either develop viable cases for investigation through link analysis of the suspected "flips" in their geographical area or to corroborate investigative information on mortgage fraud received from another source.

(4) *Privacy and civil liberties impact*: The impact on the privacy and civil liberties of those individuals fitting the "flip" criteria is virtually nonexistent for several reasons. First, this initiative relies on public record information to obtain buyer, seller, lender, and broker identities and property addresses and does not further expose that information at the analytical stage. In fact, personal information in the data is afforded greater protection than was provided in the public records from which it was obtained because it is protected to the same degree of privacy

protection applicable to all FBI record information. In the interests of accuracy, the personal and transactional information is also cross-checked against the general, publicly-accessible ChoicePoint database for accuracy before it is transferred to the field offices.

Second, even when a transaction is identified as likely fraudulent, no action is taken without further authorization. The personal information associated with that transaction is maintained solely within the "property flipping" database unless and until an FBI investigation was or is actually opened and pursued, in which case, the information would also be entered into ACS. An investigation would identify inaccuracies in amounts, dates and names, if they existed, before criminal action was taken against any individuals.

Personal information that does not lead to an investigation is retained in the database for future access, but the information is not subject to wider access or action unless and until the participants became subjects of an FBI investigation.

(5) *Law and regulations*: The legal foundation for the FBI to conduct such investigations is derived from the following:

- 28 U.S.C. § 533 authorizes the FBI to investigate violations of federal law; which includes mortgage fraud (18 U.S.C. §§ 1001, 1011, 1341, & 1342);
- 28 U.S.C. § 534 authorizes the FBI to collect and retain criminal information;
- 28 C.F.R. § 0.85 authorizes the FBI to conduct federal criminal investigations; and
- The Attorney General Guidelines for General Crimes authorizes the FBI to conduct investigations of violations of federal crimes; to employ all lawful techniques in that pursuit; and to collect and retain information from lawful sources in compliance with the Constitution and federal law.

A full description of the applicable law and regulations is set forth in Part II.A.5 above.

(6) *Privacy and accuracy protection policies*: First, the FBI complies with current laws and regulations regarding privacy, such as the Privacy Act of 1974, FISMA, Federal Information Processing Standards (FIPS) published by the National Institute of Standards and Technology (NIST), and privacy policies established by the Department of Justice in 28 C.F.R. §§ 16.40 and 16.54 and DOJ Order 2640.1 (Privacy Act Security Regulations for Systems of Records). Each of these set forth requirements for securing agency information and IT systems. With respect to the Privacy Act, the information collected on individuals for whom fraud is indicated are entered into the ACS which is part of the FBI Central Records System for which there is both published system of records notices and published exemptions from the notice and personal right of access provisions of the Privacy Act.[20] A PIA was conducted for this initiative and has been published on the FBI Website.

[20] The system of records notices appear at 63 Fed. Reg. 8671 (Feb. 20, 1998), amended 66 Fed. Reg. 8425 (Jan. 31, 2001), 66 Fed. Reg. 17,200 (Mar. 29, 2001), and 72 Fed. Reg. 3410 (Jan. 25, 2007), and the exemption regulations appear at 28 C.F.R. § 16.96 (2006).

Second, pursuant to Attorney General Guidelines, no investigative activity is initiated by an FBI field office against any transaction participant identified by this initiative unless the criteria established in the Attorney General Guidelines are met—which includes the logical evaluation of lead information through other non-intrusive, lawful means (FBI record checks, private lender record checks, developed sources). In addition, the use of more intrusive techniques (such as grand jury subpoenas or electronic surveillance) is regulated by law and procedure designed to ensure that the techniques are lawfully and appropriately employed.

Third, all personally identifying information is retained in a database accessible only through the FBI password-controlled Intranet. The FBI Intranet is accessed only from FBI computers located within restricted spaces in FBI field offices and FBI Headquarters. Access is limited by FBI policy to those in the unit with the need to know. Information of investigative value to the field is entered into the FBI ACS, a password-controlled databases with a robust audit capability. There are distinct access and security restrictions on the use of ACS in the FBI's Manual of Investigative and Operational Guidelines and the FBI's Security Policy Manual.

F. Automobile Accident Insurance Fraud Initiative

(1) *Description*: This initiative was designed to identify and analyze information regarding possible staged automobile accident cases as well as other automobile insurance fraud schemes. The analysis is expected to reveal the national scope of staged accident frauds, identify the major perpetrators and organized groups, and to identify multi-city clusters where the staged accidents are occurring.

Analysts from the private insurance industry's National Insurance Crime Bureau (NICB) extracted potential fraudulent claims for insurance reimbursement from its industry-wide database and submitted them to the FBI in ready-to-use format (meaning, with the claimants identified). NICB's analysts are well trained and all come to NICB with insurance industry experience. NICB will update its information to the FBI on an "as needed" basis, although the exact process to be used for such updates has not been finalized.

Using commercial-off-the-shelf software, such as Microsoft Access and Analyst Notebook I2, FBI analysts compare the subject identities to other data sources, including FBI case reporting and commercial data aggregators. The information the FBI analysts are analyzing at this stage does not include claimant information. This information is also compared to health care insurance claims information from HHS and the chiropractic industry, as there are usually fraudulent medical and chiropractic claims associated with staged accidents. Following this verification and corroborative analysis, which identify the claimants most likely to be engaged in fraud, the results are sent to the affected FBI field office as an investigative lead to pursue.

Analysts in the appropriate divisions will be afforded access to the program. The analysts make the connections based on the information within the program. Once the connections are made, the analysts will conduct link analysis of common information that may connect different offenders—for example, phone numbers, aliases, etc.

Based on the initial analysis by the NICB analysts, the portion of the database that is imported to the FBI will identify the most likely fraudulent claims.

(2) *Plans for use*: This initiative is new and currently used only with one field office, where it was sent in January 2007. The goal is to use this initiative to target staged automobile accidents in major metropolitan areas throughout the United States, although the target date for national deployment has not been determined.

(3) *Efficacy*: By comparing these data to existing FBI case data, as well as medical insurance fraud data, a high degree of probability is reached that the resulting data set will identify probable offenders. The technique uses basic software, the efficacy and accuracy of which have been tested through wide public and private sector use. As with other financial crimes initiatives, the efficacy will be tested primarily through subsequent investigative activity. It will also be checked against FBI investigations and public source data such as state DMV records before any formal investigation is initiated. Whether the initiative and the use of advanced analysis is effective in this area will be assessed continually over time and implementation.

(4) *Privacy and civil liberties impact*: The only personal information involved in this initiative is that provided by the claimant to his or her insurance company. This information is voluntarily provided to the insurance industry as part of a claim. As described in more detail below, insurance industry policies, including policies concerning the privacy of information submitted to insurance companies, is regulated by state law, which provides for consent forms addressing the use of the information submitted. The information is transmitted to the FBI from NICB via a password-controlled secure on-line network account and is received and entered into FBI computers by one FBI analyst or an associate in the same unit. In addition, the analyst only analyzes a limited number of claims from one geographic area at a time. Accordingly, the risk of misuse or loss is minimal.

Furthermore, the FBI does not have access to the personal information of persons who are not suspected by NICB as engaged in fraud because NICB does not grant the FBI access to the overall claims database. In addition, effective initial screening by NICB analysts will reduce the risk that innocent claimants will be provided to the FBI. Possible risks of misuse or inaccuracy of information is further minimized by the subsequent comparative analysis with medical fraud data and FBI case files.

Finally, the identification of potential investigative subjects is not acted upon by a field office until the lead is logically pursued with non-intrusive preliminary measures and an investigation is opened according to criteria set forth in the Attorney General Guidelines. Names of persons who are provided by the NICB as fraud suspects but who do not meet the FBI's suspected fraud level in the FBI's independent analysis are retained as part of the record of the analysis but are not forwarded to the field office for investigation.

(5) *Law and regulations*: The legal foundation for the FBI to conduct such investigations is derived from the following:

- 28 U.S.C. § 533 authorizes the FBI to investigate violations of federal law; which includes fraud arising from staged accidents (18 U.S.C. §§ 1035 & 1341);
- 28 U.S.C. § 534 authorizes the FBI to collect and retain criminal information;
- 28 C.F.R. § 0.85 authorizes the FBI to conduct federal criminal investigations; and
- The Attorney General Guidelines for General Crimes authorize the FBI to conduct investigations of violations of federal crimes; to employ all lawful techniques in that pursuit; and to collect and retain information from lawful sources in compliance with the Constitution and federal law.

A complete description of the applicable law and regulations is set forth in Part II.A.5 above.

In addition to the Department's compliance with federal law, the insurance industry itself is regulated by state law. These privacy protections and practices therefore are based upon individual state legislation and regulations imposed by each state's insurance commissioner. Additionally, since 2001, insurers are required to provide customers with privacy notices required by the Gramm-Leach-Bliley Act (GLBA) for financial protections. An outline of each state's laws can be found on the Coalition Against Insurance Frauds website. These state laws ensure the legality of the information and practices used by the insurance industry.

(6) *Privacy and accuracy protection policies*: First, the FBI complies with current laws and regulations regarding privacy, such as the Privacy Act of 1974, FISMA, Federal Information Processing Standards (FIPS) published by the National Institute of Standards and Technology (NIST), and privacy policies established by the Department of Justice in 28 C.F.R. §§ 16.40 and 16.54 and DOJ Order 2640.1 (Privacy Act Security Regulations for Systems of Records). Each of these set forth requirements for securing agency information and IT systems. With respect to the Privacy Act, the information collected falls within the FBI's Central Records System, which has published system of records notices and published exemptions from the notice and personal right of access provision of the Privacy Act.[21] A PIA will be conducted for this initiative.

Second, pursuant to Attorney General Guidelines, no investigative activity is initiated by an FBI field office against any accident claimant or associate identified by this initiative unless the criteria established in the Attorney General Guidelines are met—which includes the logical evaluation of the lead information through other non-intrusive, lawful means. In addition, the use of more intrusive techniques (grand jury subpoenas, administrative subpoenas, tasking of sources, undercover operations, and electronic surveillance) is regulated by law and procedure designed to ensure that the technique is lawfully and appropriately employed.

Finally, all personally identifying information is retained on a password controlled desktop computer and server in FBI restricted spaces. Access is limited to those in the unit with the need to know. Information of investigative value to the field is entered into the FBI ACS, a password-controlled databases with a robust audit capability. There are distinct access and

[21] The system of records notices appear at 63 Fed. Reg. 8671 (Feb. 20, 1998), amended 66 Fed. Reg. 8425 (Jan. 31, 2001), 66 Fed. Reg. 17,200 (Mar. 29, 2001), and 72 Fed. Reg. 3410 (Jan. 25, 2007), and the exemption regulations appear at 28 C.F.R. § 16.96 (2006).

security restrictions on the use of ACS in the FBI's Manual of Investigative and Operational Guidelines and the FBI's Security Policy Manual.

III. Advanced Analytical Tools that Do Not Meet the Definition in Section 126

The Department of Justice has developed additional initiatives that do not meet the definition set forth in section 126, but may be perceived as involving "data mining" based on some understandings of that term. Information as to some of these initiatives is provided below, in the interests of providing full and useful information. In addition, information on certain systems that have the capacity to allow advanced analysis is also included below. Where applicable, components have completed or are in the process of completing PIAs for these programs, and Privacy Act compliance issues have been addressed.

A. Drug Enforcement Administration (DEA) Initiatives

1. *SearchPoint:* SearchPoint is a DEA project which utilizes information obtained commercially from ChoicePoint, a private data aggregate corporation. ChoicePoint procures prescription data (both insurance and cash transactions). The information provided to DEA consists of only filled prescriptions for controlled substances, which includes the prescribing official (practitioner) the dispensing agent (pharmacy, clinic, hospital, etc.) and the name and quantity of the controlled substance (drug information). No patient information is made available to DEA.

DEA utilizes this SearchPoint database to conduct queries on practitioners, pharmacies and controlled substances. The database enables DEA to identify the volume and type of controlled substances a practitioner is prescribing or the volume and type of controlled substances a pharmacy is dispensing. For example, through the use of the SearchPoint database, DEA can quickly corroborate a complaint raised about a practitioner (i.e., the practitioner prescribes only pain medications). Similarly, DEA can use SearchPoint to determine whether a pharmacy is operating as an Internet pharmacy looking at indicators such as use of all cash transactions, only one or two drugs being dispensed, and the prescribing official is in a different state than the pharmacy. DEA, utilizing the SearchPoint database, can also identify in which regions of the country sales of a particular type of controlled substance(s) is increasing in volume (e.g., OxyContin).

With the SearchPoint database, DEA has been able to identify current trends, prescribing and dispensing practices and other patterns of activity, thus enabling DEA to identify probable anomalies outside of the normal prescribing practices, either locally or nationally. Using this tool, DEA is capable of quickly identifying potential violations of the Controlled Substances Act and can more effectively deploy its resources and manpower to those situations demanding the greatest and most urgent attention.

ChoicePoint data is available to and used by many facets of the public sector who purchase it. Accordingly, the information DEA obtains from ChoicePoint is "available . . . to any member of the public" and therefore does not qualify as a "database" within the meaning of Public Law 109-177, Section 126(b)(2).

2. *Automation of Reports of Consolidated Orders System (ARCOS):* Under applicable DEA regulations, manufacturers and distributors of schedule I, II, or III narcotic controlled substances must report the sale, purchase, loss, or inventory adjustment of these controlled substances to DEA. This data, which is collected in the ARCOS database, enables DEA to monitor the flow of these controlled substances from their point of manufacture through commercial distribution channels to point of sale or distribution at the dispensing/retail level (hospitals, retail pharmacies, practitioners, and teaching institutions). DEA reviews this data to ensure that purchase, sale and other transaction reports match. It also reviews the data for suspicious activity, such as massive or recurrent losses of controlled substances. Such suspicious activity could lead DEA to investigate a target previously unknown to DEA. Because the database that is being queried consists of information reported to DEA by registrants pursuant to DEA regulation, the ARCOS system does not meet the congressional requirement that "at least one of the databases was obtained from or remains under the control of a non-Federal entity," as required by section 126(b)(1)(A). Similarly, the information in ARCOS was not acquired by DEA "for purposes other than intelligence or law enforcement." *Id.*

3. *Drug Theft Loss (DTL) Database:* Similar to ARCOS reporting, DEA registrants at all levels (including practitioners and pharmacies) must report all losses of controlled substances to the DEA. This information is maintained in the DTL database. As with ARCOS, this database is reviewed for suspicious activities, which may lead DEA to investigate a previously unknown target. Also as with ARCOS, this information is not part of a database obtained from or under the control of a non-Federal entity, *see* section 126(b)(1)(A), and it was acquired by DEA for law enforcement purposes. *Id.*

4. *Online Investigative Project (OIP):* OIP is a tool used to identify Internet pharmacies. This program enables DEA to scan the Internet using search terms that might indicate the operation of an illegal Internet pharmacy (such as "Vicodin," "no prescription necessary"). Leads developed through the OIP can be further examined by investigative personnel to determine whether the website, indeed, is operating as an illegal Internet pharmacy. The OIP is an effort to identify targets through a search of databases using terms that can be, but not necessarily are, indicative of criminal activity. All OIP searches are conducted exclusively of "information publicly available via the Internet," which Congress specifically excluded from the definition of "database" in section 126(b)(2).

B. Bureau of Alcohol, Tobacco, Firearms and Explosives

1. *Bomb Arson Tracking System:* BATS is an Internet-accessible system that permits state, local and other federal law enforcement agencies to share information related to bomb and arson investigations and incidents. ATF owns the BATS database, but each participating agency manages and controls its own information. The type of information queried via BATS is similarities of components, targets, or methods, and can be used, for example, to make connections between multiple incidents with the same suspect. The BATS database consists of

information collected by ATF, or another law enforcement entity, as part of an investigation, and falls outside the scope of section 126(b)(1)(A).

2. *GangNet:* GangNet is an Internet-accessible commercial-off-the-shelf system owned by ATF. GangNet tracks gang members, gangs, and gang incidents in a granular fashion and allows for sharing of this information across departments, agencies, states, and regions. This system provides gang, gang members, and gang incident tracking, and also provides for gang intelligence analysis to discern trends, relationships, patterns and demographics with respect to gangs. GangNet consists of information collected by ATF, or another law enforcement entity, as part of an investigation, and falls outside the scope of section 126(b)(1)(A).

C. Federal Bureau of Investigation

1. *Durable Medical Equipment (DME) Initiative*: This initiative is designed to help set investigative priorities for the FBI based on preliminary analysis of suspicious claims (submitted by DME providers) by contractors for the Centers for Medicare and Medicaid Services (CMS). These analyses, which identify DME providers engaged in the most egregious fraud and providers who have abnormal results from CMS billing audits, are provided to the FBI where they are analytically compared (using COTS) by FBI analysts to FBI databases as well as other complaints submitted to CMS and the HHS Inspector General's office. The results (analyses, provider lists, and billing information) are forwarded to the affected FBI field office for further investigation as enclosures to an electronic communication that becomes part of FBI case files. In each case, the search is conducted in a manner that falls outside the scope of section 126(b)(1)(B) because the queries are subject-based, rather than pattern-based.

D. Additional Department of Justice Systems

The Department also has systems or data warehouses that could be capable of supporting advanced analytical tools, but do not themselves fall within the requirements set forth in section 126. Distinct technical and operational differences exist when comparing, on the one hand, a data warehouse that utilizes search tools and, on the other hand, a warehouse that is part of an initiative within the meaning of 126. DOJ law enforcement components employ numerous search tools and databases to help accomplish a variety of missions. Various groups collect data, others analyze data, and still others report data to DOJ law enforcement entities, as well as trusted federal, state, local, and tribal law enforcement partners. These systems are used to save time and enable law enforcement properly and accurately to connect the dots, as prescribed by the 9-11 Commission, the Markle Foundation and others. The systems listed below are data systems with search and analytical tools used to conduct investigations, but they do not perform data mining, as defined in Section 126. In addition, several of the systems mentioned below have either completed or are in the process of completing PIAs. Because these systems are national security systems, a PIA is not required under the eGov Act; however, the Department still requires certain projects to complete PIAs as a risk mitigation step (although the PIA is not publicly available) and this policy is enforced by the Chief Information Officers in each component. Of course, the Department also ensures that it complies with the requirements of the Privacy Act where applicable to these systems.

1. The Organized Crime and Drug Enforcement Task Force (OCDETF) Fusion Center maintains a data warehouse named "Compass" that contains relevant drug and related financial intelligence information from numerous law enforcement organizations. DOJ centrally manages the group and its contributors in the OCDETF Fusion Center. These contributors include:

Current Contributors	**Future Contributors**
DEA	DHS-ICE
FBI	IRS
USMS	
BOP	
Treasury-FinCEN	
DHS-US Coast Guard	
DOJ-Joint Automated Booking System	
ATF	

The goal of the data warehouse is to use cross-case analysis tools to transform multi-agency information into actionable intelligence in order to support major investigations across the globe. The *Compass* system does not fall within section 126 given the sources of the information assessed, although it does use a powerful analytical tool that allows analysts to search multiple combined data sets. These data sets are law enforcement controlled, and no commercial or private sector data is merged with the law enforcement data.

2. The Investigative Data Warehouse (IDW) is an FBI-managed program that enables investigators to search many FBI data sources across organizations within the FBI. The IDW is a robust analytical tool used by both analysts and agents within the FBI across data from more than 45 sources including the FBI, DOS, and FinCEN. IDW users search data contained in intelligence reports, suspicious activity reports, watch lists, and FBI investigative files.

The IDW provides capability for distributed search and presentation of integrated results to the agents and analysts that use its capabilities. Prior to the deployment of IDW, each of the sources of information would have to have been searched independently, which was inefficient. By contrast, an IDW user today signs on to a single system and enters a search across the sources specified by the user with integrated search results provided to the user. The integration of such search results allows IDW users to efficiently examine the relationships between items of interest including persons, places, communication devices, organizations, financial transactions, and case-related information across significantly larger amounts of data.

IDW is not pattern based data mining within the meaning of section 126 because it is not automated to conduct pattern based searches and all of the databases within IDW contain information that is collected for law enforcement purposes. Although there is access to databases such as ChoicePoint or Accurint through IDW, in order to query those particular commercial databases, the analyst must use specific subject based identifiers such as a name.

3. Internet Crime Complaint Center (IC3) is a partnership between the FBI and the National White Collar Crime Center (NW3C).

The mission of IC3 is to serve as a vehicle to receive, develop, and refer criminal complaints regarding the rapidly expanding arena of cybercrime. The IC3 gives the victims of cybercrime a convenient and easy-to-use reporting mechanism to provide authorities with tips on suspected criminal or civil violations. For law enforcement and regulatory agencies at the federal, state, local and international level, IC3 provides a central referral mechanism for complaints involving Internet related crimes.

The IC3 database contains complaints, submitted by the public, crossing the spectrum of cybercrime matters, to include online fraud in its many forms including intellectual property rights matters, computer intrusions (hacking), economic espionage (theft of trade secrets), online extortion, international money laundering, identity theft, and a growing list of internet facilitated crimes.

The FBI maintains the database with all of the cybercrime complaints, and if a complaint turns into a case, that information is loaded into the FBI's central case management system, ACS. IC3 is merely a referral system, and data is not mined under the definition in Section 126. Moreover, the information is acquired by the FBI for law enforcement purposes. Cyber investigators perform searches on the IC3 database to look for commonalities.

4. Computer Analysis and Response Team (CART) Family of Systems (FOS) include the tools needed to support computer forensics work across the country. CART maintains its own Storage Area Network to handle the large amount of data that it processes. The data obtained and stored is data covered under a valid search warrant, as a result of a criminal investigation. CART takes all data from the hard-drive of a computer and makes an evidence-ready copy of the data. Advanced analytical tools are used to search the data on each system and to look for similarities across properly confiscated hard-drives. Because the data that is analyzed is all obtained through a search warrant as part of a criminal investigation, the CART tools and capabilities do not meet the definition of data mining under Section 126.

IV. Conclusion

As set forth above, the Department of Justice takes very seriously its obligation to prevent terrorism and investigate criminal conduct using all available and lawful tools, while also respecting the privacy and civil liberties of Americans. The use of advanced analytic tools is extremely valuable and is and should be undertaken with due regard for the privacy concerns of each individual. We believe that the Department's use of advanced analytical tools meet these standards.