Department of the Treasury

2016 Annual Privacy Act and Data Mining Report

Message from the Deputy Assistant Secretary for Privacy, Transparency, and Records

On behalf of the Department of the Treasury Senior Agency Official for Privacy and Chief Privacy and Civil Liberties Officer, I am pleased to present Treasury's Annual Privacy and Data Mining Reports for Fiscal Year 2016, as required by Section 522 of the Consolidated Appropriations Act of 2005 and the Federal Agency Data Mining Reporting Act of 2007. For the fourth year in a row, Treasury is combining these two separate reporting requirements into a single report.

Inquiries about this report may be directed to privacy@treasury.gov. This report, as well as previous annual reports, can be found on the Department's Privacy Act website.

Ryan Law
Deputy Assistant Secretary
Privacy, Transparency, and Records
U.S. Department of the Treasury

2016 Annual Privacy and Data Mining Reports

Table of Contents

Message from the Deputy Assistant Secretary for Privacy, Transparency, and Records **2**

Statutory Requirements 4

The Reporting Period 4

The Annual Privacy Report 4

The Data Mining Report 4

SECTION ONE: DEPARTMENT OF THE TREASURY 2016 ANNUAL PRIVACY REPORT 6

Oversight and Compliance 6

System of Records Notices (SORN) 6

Privacy and Civil Liberties Impact Assessments 7

Federal Information Security Management Act of 2002 7

Section 803 of the Implementing Recommendations of the 9/11 Commission Act of 2007 7

Treasury's Compliance with Privacy-Related Requirements in OMB M-16-04 8

Elimination of the Unnecessary Use of Social Security Numbers 8

Privacy Awareness and Training 9

Leadership and Coordination within Treasury 11

Treasury Computer Matching Programs 11

SECTION TWO: DEPARTMENT OF THE TREASURY 2016 DATA MINING REPORT 13

The Role of the Treasury Chief Privacy and Civil Liberties Officer (CPCLO) 13

Treasury Data Mining Activities 14

Conclusion **35**

Statutory Requirements

In this report, Treasury consolidates the following two reporting requirements to reduce duplication and to provide Congress and the public with a more comprehensive overview of Treasury's privacy compliance and oversight activities:

(1) The Annual Privacy Report required by Section 522(a) of the Consolidated Appropriations Act of 2005; and,
(2) The Data Mining Reporting Act requirement contained in Section 803 of the Implementing Recommendations of the 9/11 Commission Act of 2007, 42 U.S.C. § 2000ee–3.

The Reporting Period

This report covers Treasury activities within the 2016 fiscal year (the reporting period).

The Annual Privacy Report

The Annual Privacy Report has been prepared in accordance with Section 522(a) of the Consolidated Appropriations Act of 2005, which includes the following requirement:

> Privacy Officer—
>
> Each agency shall have a Chief Privacy Officer to assume primary responsibility for privacy and data protection policy, including—
>
> * * *
>
> (6) preparing a report to Congress on an annual basis on activities of the Department that affect privacy, including complaints of privacy violations, implementation of section 552a of title 5, 11 United States Code, internal controls, and other relevant matters;
>
> * * *

The Data Mining Report

The Federal Agency Data Mining Reporting Act of 2007, 42 U.S.C. § 2000ee-3, includes the following requirement:

> (c) Reports on data mining activities by Federal agencies
>
> (1) Requirement for report - The head of each department or agency of the Federal Government that is engaged in any activity to use or develop data mining shall submit a report to Congress on all such activities of the department or agency under the jurisdiction of that official. The report shall be produced in coordination with the privacy officer of that department or agency, if applicable, and shall be made available to the public, except for an annex described in subparagraph (C).

(2) Content of report - Each report submitted under subparagraph (A) shall include, for each activity to use or develop data mining, the following information:

(A) A thorough description of the data mining activity, its goals, and, where appropriate, the target dates for the deployment of the data mining activity.

(B) A thorough description of the data mining technology that is being used or will be used, including the basis for determining whether a particular pattern or anomaly is indicative of terrorist or criminal activity.

(C) A thorough description of the data sources that are being or will be used.

(D) An assessment of the efficacy or likely efficacy of the data mining activity in providing accurate information consistent with and valuable to the stated goals and plans for the use or development of the data mining activity.

(E) An assessment of the impact or likely impact of the implementation of the data mining activity on the privacy and civil liberties of individuals, including a thorough description of the actions that are being taken or will be taken with regard to the property, privacy, or other rights or privileges of any individual or individuals as a result of the implementation of the data mining activity.

(F) A list and analysis of the laws and regulations that govern the information being or to be collected, reviewed, gathered, analyzed, or used in conjunction with the data mining activity, to the extent applicable in the context of the data mining activity.

(G) A thorough discussion of the policies, procedures, and guidelines that are in place or that are to be developed and applied in the use of such data mining activity in order to—

(i) protect the privacy and due process rights of individuals, such as redress procedures; and

(ii) ensure that only accurate and complete information is collected, reviewed, gathered, analyzed, or used, and guard against any harmful consequences of potential inaccuracies.

SECTION ONE: DEPARTMENT OF THE TREASURY 2016 ANNUAL PRIVACY REPORT

Oversight and Compliance

For Treasury to accomplish its mission, it must collect PII from its employees and the public, as well as from various organizations and other government agencies. The Department is responsible for managing and protecting the information it collects, maintains, and discloses. Federal law, regulations, and policies govern these activities and are designed to maintain the public's trust.

System of Records Notices (SORN)

A system of records is a grouping of paper or electronic records maintained by a federal agency from which information about an individual is retrieved by the name of the individual or another unique identifier assigned to the individual (e.g., Social Security number). Pursuant to 5 U.S.C. § 552a(e)(4), agencies are required to publish a SORN in the Federal Register for each system of records. Treasury has published regulations describing how it collects, maintains, and discloses records about individuals that are maintained in a system of records. These regulations provide procedures by which individuals may request access to their information maintained by Treasury.[1]

During FY 2016, the Department published three new SORNs in the Federal Register:

- Treasury/DO .016, Multiemployer Pension Reform Act of 2014 System of Records, March 16, 2016 (81 FR 14223);
- Treasury/IRS 10.008, Certified Professional Employer Organizations System of Records, July 11, 2016 (81 FR 44924); and
- Treasury/United States Mint .014, Denver Public Tour and Outreach Reservation System of Records, October 27, 2015 (80 FR 65870).

In FY 2016, Treasury also published 16 reissued and renewed SORNs in the Federal Register:

- Department of the Treasury .004 Freedom of Information Act/Privacy Act Request Records System of Records, September 16, 2016 (81 FR 63856).
- The Office of the Comptroller of the Currency (OCC) republished its systems of records inventory, including 15 SORNs, in the Federal Register on January 19, 2016 (81 FR 2945). These republished SORNs are available on Treasury's Privacy Act website.

Treasury maintains approximately 211 systems of records, nearly 42 percent of which are maintained by the Internal Revenue Service (IRS). The entire Treasury SORN collection was

[1] *See* 31 C.F.R. §§ 1.20-1.36.

updated in 2016 as part of the biennial review process required by the Privacy Act of 1974. A complete list of the Department's SORNs is available on Treasury's Privacy Act website.

Privacy and Civil Liberties Impact Assessments

A Privacy and Civil Liberties Impact Assessment (PCLIA) is an analysis of how information is handled in compliance with legal, regulatory, and policy privacy requirements. It assesses the risks and effects of collecting, maintaining, and disseminating information and discusses the mitigation strategies used to address those risks. Section 208 of the E-Government Act of 2002 (E-Gov Act) requires agencies to conduct PCLIAs for electronic information systems and collections that involve the collection, maintenance, or dissemination of information in identifiable form from or about members of the public.

In FY 2016, Treasury reviewed 234 PCLIAs. Treasury currently has 259 information technology systems that require a PIA. Pursuant to the E-Gov Act, agencies are required to make PCLIAs publicly available through the agency website, the Federal Register, or other means. The Department's PCLIAs are available on Treasury's Privacy website.

Federal Information Security Management Act of 2002

The Federal Information Security Management Act of 2002 (FISMA) requires each agency to develop, document, and implement an agency-wide program to provide security for the information and information systems that support its operations. In addition, FISMA requires Chief Information Officers, Inspectors General, and SAOPs to report to OMB on information security questions that address areas of risk. Federal agencies must report performance metrics related to the management of their privacy programs. This entails tracking and reporting the number of Treasury systems that contain PII, and the number of systems that require and/or have completed a PCLIA and/or SORN.

For FY 2016, the Department reported a total inventory of 281 FISMA systems that are used to create, collect, use, process, store, maintain, disseminate, disclose, or dispose of PII.

Section 803 of the Implementing Recommendations of the 9/11 Commission Act of 2007

Pursuant to Section 803 of the Implementing Recommendations of the 9/11 Commission Act of
2007, agencies must ensure that adequate processes exist to receive, investigate, respond to, and redress complaints from individuals who allege privacy or civil liberties violations. To meet the requirement, Treasury issued Treasury Directive (TD) 25-09, which directs heads of bureaus and relevant offices to establish internal procedures to ensure accurate and complete reporting to OPTR.

The ASM, with the support of OPTR, continues to provide timely Section 803 metrics to Congress on behalf of the Department. For FY 2016, Treasury bureaus and its offices performed 348 reviews, provided advice 15 times, and responded to 27 privacy and civil liberties

complaints. The Department and its bureaus reviewed:

- 114 Privacy Threshold Assessments (to determine whether a Privacy and Civil Liberties Impact Assessment was required).
- 234 Privacy and Civil Liberties Impact Assessments.
- The elimination/redaction or masking of Social Security numbers from 83 forms or programs.
- Four Computer Matching Agreements for approval or extension.
- Five privacy complaints, all of which were resolved.
- 22 civil liberties complaints, of which 12 were resolved.

Treasury's Compliance with Privacy-Related Requirements in OMB M-16-04

In FY 2016, Treasury initiated processes to identify its High Value Assets (HVAs) to comply with OMB Memorandum 16-04, *Cybersecurity Strategy and Implementation Plan (CSIP) for the Federal Civilian Government* (OMB M-16-04). OMB M-16-04 defines HVAs as "assets, systems, facilities, data and datasets that are of particular interest to potential adversaries. These assets, systems, and datasets may contain sensitive controls, instructions or data used in critical Federal operations, or house unique collections of data (by size or content) making them of particular interest to criminal, politically-motivated, or state-sponsored actors for either direct exploitation of the data or to cause a loss of confidence in the U.S. Government."[2] Treasury identified as HVAs 176 Treasury systems containing PII. All HVA systems that require PCLIAs or updated PCLIAs either had them or were in the review process. Updated SORNs for all HVA systems were current and up-to-date.

Elimination of the Unnecessary Use of Social Security Numbers

Internal Revenue Service (IRS)

In FY 2016, the IRS continued its efforts to develop and implement written policy and is developing updated guidance and an Internal Revenue Manual section related to the collection or use of Social Security numbers (SSN). The IRS continues to address the use of SSNs and reduce the use of unnecessary SSNs through its SSN Elimination and Reduction program.

The IRS's SSN Elimination and Reduction Program made significant strides in eliminating or reducing the use of SSNs within systems, forms, notices, and letters where the collection or use of the SSN was not necessary. The IRS is systematically reviewing 100% of all existing and new notices, letters, and forms for unnecessary SSN use. As of FY 2016, the IRS eliminated or reduced the use of the SSNs on 138 payment and non-payment notices, with an estimated annual volume of 54 million taxpayer mailings. Wherever possible, the IRS is using a Truncated Taxpayer Identification Number (TTIN). A TTIN can be an alternative to using an SSN, IRS individual taxpayer identification number (ITIN), or IRS adoption taxpayer identification number

[2] OMB M-16-04, Cybersecurity Strategy and Implementation Plan (CSIP) for the Federal Civilian Government, October 30, 2015.

(ATIN). The filer of certain information returns may use a TTIN on the corresponding payee statements to identify the individual being furnished a statement. To further protect identities, the TTIN displays only the last four digits of an individual's identifying number and is shown in the format XXX-XX-1234 or ***-**-1234.

Financial Crimes Enforcement Network (FinCEN)

The Financial Crimes Enforcement Network (FinCEN) has statutory obligations under the Bank Secrecy Act (BSA) and the USA PATRIOT Act to deter and detect criminal activity and safeguard financial systems from abuse. This is achieved in large part by the collection, maintenance, and sharing of financial information with law enforcement and regulatory agencies. This information typically includes SSNs for the individuals who are the subjects of reporting mandated by FinCEN regulations. Given the unique role that SSNs currently play in identifying individuals in the United States, it is not currently practicable for FinCEN to eliminate the collection and use of SSNs in its data collection. If and when alternative mechanisms for identifying individuals become widely accepted, FinCEN will provide implementation plans with the associated timelines to eliminate the use of SSNs in the related processes. FinCEN will monitor its business processes such that a new or modified business process may necessitate a re-evaluation of the use of SSNs.

Privacy Awareness and Training

A Culture of Privacy Awareness

In M-07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information*, OMB required agencies to train employees on their privacy and security responsibilities before granting them access to agency information and information systems. Additionally, agencies must provide at least annual refresher training to ensure employees continue to understand their responsibilities. Ninety-six percent of all Treasury employees completed annual privacy awareness training during the reporting period.

Internal Revenue Service (IRS) Privacy Training

Annual Privacy and Unauthorized Access to Taxpayer Accounts (UNAX) training is mandatory for all IRS employees. Treasury continuous updates these training modules to address emerging issues, support new directives, and meet business needs. For example, the UNAX training includes three video vignettes depicting current, real-life workplace dilemmas and the resulting violations applicable when policies are not followed.

In FY 2016, the IRS developed a privacy training series to meet new OMB requirements and expectations. The courses are virtual, interactive and available on demand through the IRS Enterprise Learning Management (ELMS) system. The series includes:

1. IRS Privacy Foundation Training;
2. Privacy Training for Information Technology Specialists;
3. Privacy Training for Information Technology Designers;
4. Enterprise Architecture and Data Strategy Officers Privacy Training; and
5. Cybersecurity Privacy Training

The IRS Privacy Foundation Training is supplemented by four role-based, tactical privacy courses for technical employees including Project Managers, Privacy & Civil Liberties Impact Assessment Preparers, Business Subject Matter Experts, and Adaptive Privacy Impact Assessments Preparers.

Appropriate, on-demand privacy training is equally important for IRS contractors. The IRS overhauled its Privacy Act training to highlight emerging privacy policy issues and make the training more applicable to a wider audience including contractors. Additionally, 534 Contracting Officer Representatives were trained on how to ensure proper privacy, security, and disclosure clauses are included in contracts and enforced. Because these clauses are a critical step toward protecting privacy during contract execution and closeout, the IRS simultaneously conducted Contract Review training for key privacy staff members.

Fiscal Service Privacy Training

In FY 2016, Treasury's Bureau of the Fiscal Service focused on role-based privacy training. The Fiscal Service defined privacy roles and identified personnel who met the minimum requirements for role-based privacy training. The Fiscal Service's continued efforts include working with the Department to ensure the Treasury Learning Management System includes adequate privacy training modules.

IRS Advancements in Privacy Policy and Protection

The IRS takes a proactive approach to privacy policy development by monitoring emerging issues, identifying gaps, issuing policy, and establishing accountability. In 2016, the IRS updated and reinforced employee policy and guidance on three privacy issues: personal email, electronic messaging, and shared drives. Interim guidance on "Using IRS and Personal Email Accounts" reinforced existing policy by providing specific procedures employees must follow when sending emails to stakeholders, taxpayers, other IRS employees, and themselves. Guidance on "Electronic Message Usage and Preservation" reminded employees that the content of an electronic message may be a federal record and, if so, they must manage and protect the records accordingly. Interim Guidance on shared drives provided specific policy on methods to protect privacy when shared drives contain Sensitive But Unclassified (SBU) data, including tax information and PII.

Fiscal Service Significant Advancements in Privacy-Related Policy

The Bureau of the Fiscal Service drafted a general privacy policy. In fall 2016, the bureau's policy governance process was reviewing the draft policy, and the Fiscal Service anticipates publishing it prior to the end of the year.

In 2016, Fiscal Service began updating its "Privacy Program Handbook," which will include significant updates regarding the bureau's Privacy Program framework and additional guidance for the Fiscal Service's business area programs and bureau personnel who access Privacy Act-protected data.

Leadership and Coordination within Treasury

Executive Order (E.O.) 13636: Improving Critical Infrastructure Cybersecurity

In 2013, the President signed E.O. 13636, *Improving Critical Infrastructure Cybersecurity*, stating: "[i]t is the policy of the United States to enhance the security and resilience of the Nation's critical infrastructure and to maintain a cyber environment that encourages efficiency, innovation, and economic prosperity while promoting safety, security, business confidentiality, privacy, and civil liberties."

To ensure the inclusion of privacy and civil liberties protections in activities under the Order, section 5(a) of the E.O. required federal agencies to coordinate E.O. 13636-related cybersecurity activities with their SAOP. Section 5(b) further required the SAOP to conduct an assessment of their agency's activities under the Order. As required, OPTR conducted a privacy and civil liberties assessment of the Department's cybersecurity activities under the E.O. and submitted its assessment to the Department of Homeland Security for inclusion in a consolidated public report. The consolidated report is available on DHS' website.

Treasury Computer Matching Programs

Pursuant to the Computer Matching and Privacy Protection Act of 1988,[3] Treasury maintains a Data Integrity Board (DIB) to oversee its computer matching programs. Computer matching programs provide a direct benefit to the public by assisting in the elimination of errors and in monitoring waste, fraud, and abuse.

In FY 2016, the Treasury DIB reviewed and approved three extensions of computer matching programs and renewed seven of the Department's ongoing computer matching programs. Matching agreements expires in 18 months after execution unless extended for an additional 12-month period. After an extension expires, an agreement may be renewed for an additional 18 months.

Published notices for Treasury's ongoing computer matching programs are available on Treasury's Privacy Act website.

[3] Pub. L. No. 100-503.

The DIB's actions included:

Agreement Title	Agencies Involved	Action	Date of Action
Data Loss Prevention Program	Internal Revenue Service	Renewal	January 5, 2016
Taxpayer Address Request Program	Internal Revenue Service – Department of Justice	Extension	June 6, 2016
Disclosure of Information to Federal, State, and Local Agencies	Internal Revenue Service – Social Security Administration	Renewal	October 1, 2015
Disclosure of Information to Federal, State, and Local Agencies	Internal Revenue Service – Department of Veteran Affairs	Extension	October 1, 2015
Disclosure of Information to Federal, State, and Local Agencies	IRS – (50) States *This includes 50 separate agreements.	Renewal	October 1, 2015
CMA 1038-Supplemental Security Income	Bureau of the Fiscal Service – Social Security Administration	Renewal	May 2, 2016
CMA 1304-Medicare Part D Prescription Benefit Program	Bureau of the Fiscal Service – Social Security Administration	Renewal	February 26, 2016
Verification of Household Income and Family Size for Insurances Affordability Programs and Exemptions	Internal Revenue Service – Department of Health and Human Services Exemptions	Renewal	September 20, 2016
Prescription Drug Subsidy Program IRS Project 692 SSA Match #1305	Internal Revenue Service – Social Security Administration	Extension	September 27, 2016
Title XVI Benefits	Internal Revenue Service – Social Security Administration	Renewal	November 20, 2015

SECTION TWO: DEPARTMENT OF THE TREASURY 2016 DATA MINING REPORT

The Role of the Treasury Chief Privacy and Civil Liberties Officer (CPCLO)

The Department of the Treasury (Treasury or the Department) is providing this report to Congress pursuant to Section 803 of the *Implementing Recommendations of the 9/11 Commission Act of 2007* (9/11 Commission Act), entitled the *Federal Agency Data Mining Reporting Act of 2007* (Data Mining Reporting Act or the Act). This report discusses activities currently deployed or under development in the Department that meet the Data Mining Reporting Act's definition of data mining. The report also provides the information the Act requires with respect to each data mining activity.

Definitions

(1) DATA MINING. The term "data mining" means a program involving pattern-based queries, searches, or other analyses of one or more electronic databases, where:

a. a department or agency of the Federal Government, or a non-Federal entity acting on behalf of the Federal Government, is conducting the queries, searches, or other analyses to discover or locate a predictive pattern or anomaly indicative of terrorist or criminal activity on the part of any individual or individuals;
b. the queries, searches, or other analyses are not subject-based and do not use personal identifiers of a specific individual, or inputs associated with a specific individual or group of individuals, to retrieve information from the database or databases; and
c. the purpose of the queries, searches, or other analyses is not solely—
    i. the detection of fraud, waste, or abuse in a Government agency or program; or
    ii. the security of a Government computer system.

(2) DATABASE. The term "database" does not include telephone directories, news reporting, information publicly available to any member of the public without payment of a fee, or databases of judicial and administrative opinions or other legal research sources.[4]

Three Treasury bureaus maintain systems using applications that meet the definition of data mining: the Financial Crimes Enforcement Network (FinCEN), the Internal Revenue Service (IRS), and the Alcohol and Tobacco Tax and Trade Bureau (TTB). These IRS, FinCEN, and TTB systems were discussed in previous Treasury data mining reports.

[4] 42 U.S.C § 2000ee-3(b)(1). "[T]elephone directories, news reporting, information publicly available to any member of the public without payment of a fee, or databases of judicial and administrative opinions or other legal research sources" are not "databases" under the Act. § 2000ee-3(b)(2).

Treasury Data Mining Activities

FinCEN Data Mining Activities

(A) A thorough description of the data mining activity, its goals, and, where appropriate, the target dates for the deployment of the data mining activity.

FinCEN's mission is to safeguard the financial system from illicit use, combat money laundering, and promote national security through the collection, analysis, and dissemination of financial intelligence and strategic use of financial authorities. To accomplish its mission, FinCEN provides financial intelligence, data stewardship, and support for law enforcement. FinCEN also engages in the detection of trends and typologies of money laundering and terror finance. FinCEN strives to respect privacy and civil rights and the exercise of civil liberties while overseeing the data it maintains and uses in fulfillment of its mission as set forth under the USA PATRIOT Act, Public Law 107-56, October 26, 2001.

In furtherance of this goal, FinCEN is required to maintain a government wide data access service with a range of financial transaction information; to conduct analysis and dissemination of information in support of law enforcement at the federal, state, local, and international levels; to identify emerging trends and methods in money laundering and other financial crimes; to serve as the Financial Intelligence Unit of the United States; and to carry out other delegated regulatory responsibilities. FinCEN's legal authorities are codified at 31 U.S.C. § 310. FinCEN works to achieve its mission while avoiding the collection and indexing of information on persons exercising their constitutional rights and civil liberties.

FinCEN's analysts use various data analysis techniques for generating leads on subjects or institutions whose activities warrant outreach, investigation, or other statutorily mandated activities.

FinCEN has successfully developed algorithms to identify activity associated with specific types of financial crimes, such as check cashing activity associated with health insurance fraud. FinCEN also uses algorithms to examine filing patterns across financial sectors. This analysis supports a broad range of objectives from the identification of trends and patterns of illicit financial activity to the detection of institutions that may require additional regulatory oversight.

FinCEN continues to develop and expand the use of automated business rules to rapidly surface high value reports of illicit financial activity on a daily basis. The term "business rule" refers to automated queries or algorithms designed to screen incoming Bank Secrecy Act (BSA) findings against known criteria to identify high priority filings likely to require further review or analysis. Rule findings are reviewed internally by FinCEN and distributed to external stakeholders, such as law enforcement and our foreign financial intelligence unit (FIU) partners. FinCEN's business rules play a vital role in the identification and dissemination of timely financial intelligence to combat threats such as terrorist financing, money laundering, cyber threats, and other illicit financial activity.

(B) A thorough description of the data mining technology that is being used or will be used, including the basis for determining whether a particular pattern or anomaly is indicative of terrorist or criminal activity.

FinCEN leverages two principal methods for deriving information relevant to illicit financial activity from the BSA data. The first is content driven, that is, searching for specific entity names, or term combinations used in reporting that are associated with various types of illicit financial activity. The second method is pattern driven and can take various forms. Patterns may be derived from searches for a particular type of subject in the data. FinCEN then identifies subjects that fit that same pattern and have certain filing profiles. Matching filing patterns across different types of BSA reports highlights anomalous behavior that leads to the identification of subjects for potential investigation.

For content driven data analysis, FinCEN staff leverage a web-based application called FinCEN Query. The application provides analysts with the capability to search for specific entity names and term combinations across all of FinCEN's records. For pattern driven analysis, staff leverage FinCEN's "Advanced Analytics" system. This system is comprised of commercial off-the-shelf (COTS) and custom developed tools with capabilities including statistical, social network, and geospatial analysis, data modelling and visualization, and text analytics that aid in the analysis of BSA data.

(C) A thorough description of the data sources that are being or will be used.

BSA reports administered by FinCEN, e.g., a report by a financial institution of a suspicious transaction relevant to a possible violation of law or regulation,[5] form the underlying data for FinCEN's manual and automated search methods and trend analysis activities.

To accomplish its mission and give context to the data FinCEN extracts from its BSA database, FinCEN must consider other information available to it through a variety of sources, including open source material, law enforcement information, other government information, and subscription services. This information is used to support or amplify conclusions or hypotheses derived from the analysis of BSA data. For example, commercially available databases are used to support or further identify information and to aid in the identification of potential illicit activity based on suspicious trends, patterns, or methods. FinCEN's trend analysis uses any records available to it in fulfilling its mission, including subpoenaed financial records, public source information, commercial database information, and third party data sources, such as Census Bureau, Social Security Administration,[6] and Office of Foreign Assets Control data.

---

[5] 31 U.S.C. § 5318(g).

[6] The Death Master File is Social Security Administration (SSA) information used by medical researchers, hospitals, medical programs, and law enforcement agencies and other government agencies to verify a person's death and to prevent fraud. Although it is SSA information, the National Technical Information Service in the Department of Commerce maintains the database. For more information, please visit the NTIS website.

(D) An assessment of the efficacy or likely efficacy of the data mining activity in providing accurate information consistent with and valuable to the stated goals and plans for the use or development of the data mining activity.

FinCEN provides strategic and tactical products for several audiences: law enforcement, foreign Financial Intelligence Unit (FIU) partners, financial regulators, the financial industry, and the general public. Each of these sets of consumers has different restrictions or guidelines under which FinCEN can provide BSA data or BSA data derived analysis.

In FY 2016, FinCEN produced a total of 1,110 intelligence products for law enforcement partners and responded to 996 requests for BSA information from foreign FIU partners. For domestic and foreign law enforcement partners, FinCEN provides high value data analytics. FinCEN annually receives the results of surveys of its foreign Egmont member counterparts and domestic law enforcement agencies regarding the utility of its analytical products. These survey results consistently reflect positive feedback from our foreign and domestic stakeholders. FinCEN also receives feedback on individual reports from law enforcement and regulatory agencies on our efforts to combat terrorism financing, healthcare, mortgage, and government programs fraud, southwest border narcotics, and bulk cash smuggling. Examples of several analytical projects that received significant positive feedback are outlined below:

- To combat threats related to ISIL and Foreign Terrorist Fighters (FTFs), FinCEN has conducted extensive network analysis designed to identify potential FTFs and facilitators, developed targeted business rules to identify individuals who may seek to travel abroad to fight with ISIL, and partnered with domestic law enforcement agencies such as Customs and Border Protection to help interdict individuals returning from conflict areas that may pose a threat to the United States. These efforts have received significant positive feedback from both domestic law enforcement agencies and the intelligence community.

- To combat threats related to transnational organized crime (TOC), FinCEN, working in conjunction with a number of federal law enforcement agencies, developed an algorithm designed to identify high volume funnel account activity. FinCEN's law enforcement partners have indicated that TOC organizations often use funnel accounts to move their illicit proceeds. To combat this threat, FinCEN designed an algorithm that law enforcement can use to target the highest volume funnel accounts. Initial feedback on the algorithm from law enforcement agencies, such as IRS-CI, has been very positive.

- To combat terrorist financing threats, FinCEN has developed more than 31 business rules designed to identify and disrupt these organizations' revenue streams and target their financial support networks. These rules generate more than 1,250 leads per month that FinCEN disseminates to the law enforcement, intelligence, and FIU communities via expedited "Flash Reports." Flash Reports are designed to provide critical intelligence to FinCEN's stakeholders on a timely basis. Since the inception of the program in late 2014, FinCEN has disseminated more than 1,500 counter terrorist

financing Flash Reports. Feedback on these reports has been extremely positive, with stakeholders noting that the reports helped corroborate information related to investigations and provided new leads, assisted investigators in identifying targets, cultivated sources, resulted in the identification of more than 30 previously unknown foreign terrorist fighters, and assisted in the real-time interdiction of an outbound shipment of concern.

FinCEN narrowly tailors its business rules to achieve its mission, and each rule is developed, tested, implemented, and re-tested for efficacy throughout its deployment. The Office of Chief Counsel and the Technology Division are engaged during the development of all business rules. FinCEN continues to receive strong positive feedback both from our domestic and international partners on the value of the intelligence derived from our business rules program.

Finally, FinCEN provides annual aggregated statistics on SAR data by sector to the public in a publication titled "SAR Stats" and provides an interactive SAR Stats module for SAR Statistical data searches. The most recent version of SAR Stats was published on FinCEN's website in October 2015. Data indicate that in FY15, readers accessed the publication 271,276 times, and that the interactive SAR Stats module was accessed 295,810 times. This is a total of 567,086 accesses for SAR statistical information, an indication of the data's high utility.

(E) An assessment of the impact or likely impact of the implementation of the data mining activity on the privacy and civil liberties of individuals, including a thorough description of the actions that are being taken or will be taken with regard to the property, privacy, or other rights or privileges of any individual or individuals as a result of the implementation of the data mining activity.

The impact of FinCEN's congressionally mandated mission on the privacy and civil liberties of individuals has been and will continue to be minimal. As a threshold matter, the Supreme Court has ruled that the financial information that banks and other financial institutions hold, and that FinCEN collects and analyzes pursuant to its authority in 31 U.S.C. § 310 and the BSA (discussed in more detail in item (F) below), carries no constitutionally protected "expectation of privacy." Moreover, the Right to Financial Privacy Act of 1978 expressly provides that it gives no protection for financial records or information required to be reported in accordance with any federal statute or regulation, which includes information contained in BSA reports.

Significantly, FinCEN takes no adverse actions against individuals based on the existence of, or information contained in, BSA data. Since a BSA report itself is not necessarily indicative of criminal activity, it is only useful when viewed in conjunction with other evidence. Therefore, FinCEN provides the data, or analytical products analyzing the data, to outside agencies where the information may be relevant to current or potential investigations or proceedings under the jurisdiction of those agencies.

During the development of all business rules, analytical models, and algorithms, FinCEN

documents a description of why the analytics will not adversely affect an individual or entity's privacy or civil liberty rights. This documentation is shared with both FinCEN's Technology Division and its Office of Chief Counsel.

The BSA provides standards for proper use of the financial data collected by FinCEN. The collected information is also generally subject to the Privacy Act of 1974, discussed in more detail under item (F) below. FinCEN has developed extensive policies and procedures to ensure, to the extent reasonably possible, that: (1) the analyzed information is used for purposes authorized by applicable law; and (2) the security of the information is adequately maintained. Analytical products produced by FinCEN are subject to clearly specified restrictions regarding use and further dissemination of the products to ensure that the products will only be used by appropriate agencies for statutorily authorized purposes. To the extent such products reference information collected pursuant to the BSA, FinCEN has issued guidelines requiring user agencies to attach warning language to such products and to follow specific procedures for further dissemination of the BSA information. These procedures aim to ensure that: (1) only appropriate agencies will have access to the information; (2) the information will be used for statutorily authorized purposes; (3) agencies with access are aware of the sensitivity of the material; and (4) FinCEN will be able to track which agencies have such materials in their possession.

FinCEN posts Privacy Impact Assessments (PIA) on their public website, which informs the public of FinCEN activities and practices related to the collection, processing, retention, and distribution of personally identifiable information (PII).[7] The PII these systems handle is necessary to assist regulators and law enforcement in identifying and monitoring the financial activities of individuals who are potentially committing financial crimes.

(F) A list and analysis of the laws and regulations that govern the information being or to be collected, reviewed, gathered, analyzed, or used in conjunction with the data mining activity, to the extent applicable in the context of the data mining activity.

1) The Bank Secrecy Act, 31 U.S.C. § 5311, et seq. (BSA) and Implementing Regulations, 31 C.F.R. Chapter X, et seq:

31 U.S.C. § 5311— Declaration of Purpose

This section specifies that the purpose of the recordkeeping and reporting requirements in the BSA is to, "require certain reports where they have a high degree of usefulness in criminal, tax, or regulatory investigations or proceedings, or in the conduct of intelligence or counterintelligence activities, including analysis, to protect against international terrorism." FinCEN strives to ensure that all uses of information are consistent with this purpose.

31 C.F.R. § 1010.301 — Determination by the Secretary

[7] For more information about FinCEN PIAs, please visit FinCEN's website.

This regulation provides the determination that the reports collected pursuant to the BSA have a, "high degree of usefulness," in the areas covered by 31 U.S.C. § 5311.

31 U.S.C. § 5319 — Availability of Reports

This section makes it clear that, upon request, the Secretary (as delegated to FinCEN) is required to provide BSA information for the purposes described in this section, to agencies including state financial institutions supervisory agencies, United States intelligence agencies, or self-regulatory organizations registered with the Securities and Exchange Commission or the Commodity Futures Trading Commission. This list of types of agencies is not exhaustive, but those listed are clearly covered. This section also provides that reports collected pursuant to the BSA are exempt from disclosure under the Freedom of Information Act, 5 U.S.C. § 552.

31 C.F.R. § 1010.950 — Availability of Information

This section authorizes the Secretary to make BSA information available to appropriate agencies for purposes specified in the BSA, and specifies that the requesting agency is to receive the information, "in confidence."

31 U.S.C. § 5313 — Reports on domestic coins and currency transactions

This section provides for the reporting by financial institutions of reports of certain currency transactions involving more than an amount specified by the Secretary (as delegated to FinCEN).

31 C.F.R. §§ 1010.311; 1021.311 — Reports of transactions in currency

These regulations implement the reporting requirement of 31 U.S.C. § 5313 and specify the amount of reportable transactions in currency at more than $10,000.

31 U.S.C. § 5316 — Reports on exporting and importing monetary instruments

This section requires reports by those that transport currency or other monetary instruments of more than $10,000 at one time from outside the United States into the United States, or from the United States outside the United States.

31 C.F.R. § 1010.340 — Reports of transportation of currency or monetary instruments

This regulation implements the reporting requirement of 31 U.S.C. § 5316 with respect to currency or other monetary instruments of more than $10,000 imported into the United States or exported outside the United States.

31 U.S.C. § 5314 — Records and reports on foreign financial agency transactions

This section authorizes the Secretary (as delegated to FinCEN) to prescribe regulations requiring the reporting of certain types of foreign transactions and relationships with foreign institutions.

31 C.F.R. § 1010.350 — Reports of foreign financial accounts

This regulation, implementing 31 U.S.C. § 5314, requires that U.S. persons file reports of foreign bank accounts.

31 U.S.C. § 5318(g) — Reporting of suspicious transactions

This section authorizes the Secretary (as delegated to FinCEN), to require the reporting of suspicious transactions relevant to a possible violation of law. The section also provides for the confidentiality of such reports, barring financial institutions from notifying anyone involved in the transaction that the transaction has been reported. Government employees are subject to the same confidentiality restrictions, except as "necessary to fulfill the official duties" of such employees. The policies and procedures detailed above in response to item (E) are aimed, in large part, at maintaining the confidentiality of these reports.

31 C.F.R. §§1010.320; 1020.320; 1021.320; 1022.320; 1023.320; 1024.320; 1025.320; 1026.320 — Reports of Suspicious Transactions

These regulations implement 31 U.S.C. § 5318(g), requiring covered financial institutions to file suspicious activity reports and requiring the maintaining of strict confidentiality of the reports.

31 U.S.C. § 5331 — Reports relating to coins and currency received in nonfinancial trade or business

This section provides for the reporting of currency transactions of more than $10,000 by businesses other than financial institutions.

31 C.F.R. § 1010.330 — Reports related to currency in excess of $10,000 received in a trade or business

This regulation implements 31 U.S.C. § 5331.

2) The Privacy Act of 1974 (Privacy Act), 5 U.S.C. § 552a

Generally, the Privacy Act protects reports that FinCEN collects pursuant to the BSA as these reports are "records" contained in a "system of records."[8] The Privacy Act provides that covered records may be disclosed without the permission of the individual to whom the record

[8] 5 U.S.C. § 552a(a)(3) (defining a "record" to mean any item, collection, or grouping of information about an individual that is maintained by an agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph and a "system of records" to mean a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual);

pertains if they are disclosed pursuant to a "routine use."[9] FinCEN includes sets of routine uses in its published Systems of Records Notices (SORNs) as the Privacy Act requires. These routine uses identify the individuals and organizations external to Treasury with which FinCEN routinely shares BSA information. Sharing with these specified recipients is consistent with the purposes for which the information is collected, as specified in the BSA.

FinCEN has three SORNs that cover the information it collects under the BSA:

(1) Treasury/FinCEN .001, *FinCEN Investigations and Examinations System*[10];
(2) Treasury/FinCEN .002, *Suspicious Activity Report (SAR) System*[11]; and,
(3) Treasury/FinCEN .003, *Bank Secrecy Act (BSA) Reports System.*[12]

FinCEN followed Privacy Act procedures (including appropriate public notice and comment periods) to exempt certain records maintained in the SARs and BSA systems of records from specific provisions of the Privacy Act, including those allowing for subject's access to the reports, notification to the subject when reports are shared, requests for correction of the contents of such reports by the subject, and the civil remedies covering these areas. These exemptions prevent individuals who are planning crimes from avoiding detection or apprehension or structuring their operations to avoid detection or apprehension.

3) Other Relevant Provisions

31 U.S.C. § 310— Financial Crimes Enforcement Network

This section establishes FinCEN as a bureau in the Department of the Treasury, sets out the duties and powers of the Director, and empowers the Director to administer the BSA to the extent delegated by the Secretary of the Treasury.[13] This section also requires FinCEN to maintain a "government-wide data access service" for the information collected under the BSA, as well as records and data maintained by other government agencies and other publicly and privately available information.[14] FinCEN is required to "analyze and disseminate" the data for a broad range of purposes consistent with the law.[15] These purposes include identifying possible criminal activity; supporting domestic and international criminal investigations (and related civil proceedings); determining emerging trends and methods in money laundering and other financial crimes; supporting the conduct of intelligence and counterintelligence activities, including analysis, to protect against international terrorism; and supporting government initiatives against money laundering.

The section further requires that FinCEN furnish research, analytical, and informational

[9] 5 U.S.C. § 552a(b)(3).
[10] 79 Fed. Reg. 20969 (April 14, 2014).
[11] *Id.* at 20972.
[12] *Id.* at 20974.
[13] Treasury Order 180-01, *Financial Crimes Enforcement Network* (July 1, 2014) (delegating to the Director of FinCEN various duties and responsibilities, including the authority to administer, implement, and enforce the BSA).
[14] 31 U.S.C.§ 310(b)(2)(B).
[15] *Id*. at § 310(b)(2)(C)(i)-(vii).

services to financial institutions and domestic and foreign law enforcement agencies for the, "detection, prevention, and prosecution of terrorism, organized crime, money laundering and other financial crimes," and provide, "computer and data support and data analysis to the Secretary of the Treasury for tracking and controlling foreign assets."[16] The section also provides for the establishment of standards for making the information available through efficient means, and to screen appropriate users and appropriate uses.[17] The activities and procedures described in this report adhere to the requirements of this statute.

(G) A thorough discussion of the policies, procedures, and guidelines that are in place or that are to be developed and applied in the use of such data mining activity in order to:

(i) protect the privacy and due process rights of individuals, such as redress procedures; and

A description of the policies, procedures, and guidance in place to ensure the privacy and due process rights of individuals that are the subject of FinCEN data mining activities is provided in subsection (E) above.

(ii) ensure that only accurate and complete information is collected, reviewed, gathered, analyzed, or used, and guard against any harmful consequences of potential inaccuracies.

FinCEN, through its data perfection procedures, ensures that information contained in the database of BSA reports is accurate and complete. In addition, as discussed in item (E) above, FinCEN does not take adverse actions against individuals (outside the context of enforcing the requirements of the BSA itself) based on the information contained in BSA reports. In addition, because user agencies only use BSA information in conjunction with other evidence, a BSA report in itself is not used as the sole basis for adverse actions by user agencies. Accordingly, there is an inherent system of "checks and balances" in the use of BSA information that greatly reduces the risk of harmful consequences from inaccuracies that may be contained in BSA reports.

As noted earlier in this report, FinCEN's BSA data contains no constitutionally protected, "expectation of privacy" and FinCEN takes no adverse actions against individuals based on the BSA data collected. Therefore, FinCEN's BSA analyst training does not focus on civil liberties. However, FinCEN has mandatory training for its data users that includes the privacy component of secure handling and safeguarding of the information. FinCEN provides on-line training for all external users as a requirement for system access. Biennially, at a minimum, users must complete training as a requirement of continued system access. In addition to this online training, FinCEN hosts webinars as requested. All FinCEN staff are required to complete Privacy Awareness training annually that includes explanation of the staff's civil liberties responsibilities. Accountability for the security and confidentiality of the BSA data and its handling are prominently articulated in all course materials.

[16] *Id.* at § 310(b)(2)(E), (G).
[17] *Id*. at § 310(c)(1) and (c)(2).

IRS Data Mining Activities

(A) A thorough description of the data mining activity, its goals, and, where appropriate, the target dates for the deployment of the data mining activity.

Three divisions of the IRS are engaged in data mining activities covered by the Act: IRS Criminal Investigation organization (IRS-CI); the IRS Small Business/Self-Employed Division (SB/SE); and the IRS Wage and Investment Division (W&I). In FY 2016, each of these IRS divisions used one or more of six data mining applications to search for specific characteristics that are indicators of potential criminal activity:

- Reveal - retired in March 2016;
- Investigative Data Examination Application (IDEA) - formerly known as Investigative Data Analytics;
- Lead and Case Analytics (LCA);
- Electronic Fraud Detection System (EFDS);
- Return Review Program (RRP); and
- FinCEN Query

IRS-CI is tasked with protecting IRS revenue streams by detecting fraudulent activity and preventing recurrences. In FY 2016, IRS-CI used the Reveal, IDEA, LCA, EFDS, and RRP systems to support this work. Data uncovered using these systems may be reflected in indictments and criminal prosecutions.

Retired in March 2016, Reveal was a data query and visualization tool allowing CI analysts and agents to query and analyze large and potentially disparate sets of data through a single access point. CI streamlined the data analytics program to shut down Reveal and ensure data sets and features were consolidated with IDEA and LCA.

IDEA is a data query tool currently in use at the CI Lead Development Centers (LDC), Scheme Development Centers (SDC), and field offices, and it provides CI analysts and special agents with the ability to quickly search electronic data through a single access point. By using the IDEA application, special agents and investigative analysts can proactively identify patterns indicative of illegal activities. This tool enhances investigation selection and supports investigative priorities in tax law enforcement, counterterrorism, and other high-priority criminal investigations. The IDEA application uses data for both reactive and proactive queries. Reactive queries are a result of specific, targeted investigations; proactive queries are the result of pattern matching to generate leads. Data available in the IDEA application enable users to detect suspicious financial transactions indicative of money laundering, terrorism, and other financial crimes. IDEA query results are used exclusively for the purpose of generating leads. Any investigative process that results from these leads uses the corresponding data from the originating systems.

LCA is a data query and visualization application that allows CI investigative analysts and agents to query and analyze large and disparate sets of data through a single access point. This enhances the analyst's ability to develop a comprehensive picture of suspicious or criminal activity. The LCA application uses data for both reactive and proactive queries. Reactive

queries are a result of specific, targeted investigations; proactive queries are the result of pattern matching to generate leads. Data available in the LCA application enable users to detect suspicious financial transactions indicative of money laundering, terrorism, and other financial crimes. The application presents information to the user visually, exposing associations between entities in the data that might otherwise remain undiscovered. The software used to create LCA allows users from the LDCs, SDCs, and field offices to create visualization diagrams, graphs, spreadsheets, reports, timelines, and maps to enhance investigation selection and supports investigative priorities to proactively identify and develop leads for refund fraud, identity theft, counterterrorism, money laundering, offshore abusive trust schemes, and other financial crime, as well as Bank Secrecy Act (BSA) Suspicious Activity Report (SAR) reviews and Financial Crimes Task Force activity.

IRS-CI and W&I use RRP and EFDS to maximize detection of tax return fraud, tax noncompliance, and identity theft. EFDS compiles, cross-references, and verifies information indicative of potentially fraudulent tax returns. As EFDS receives returns, it loads and assigns a score to each tax return. Scores range from 0.0 to 1.0, with a higher score indicating a greater potential for fraud. RRP expands on EFDS' capabilities by providing multiple model scores, rule breaks, and linking characteristics. In both RRP and EFDS, IRS-CI does not directly examine the scores, but does use returns that W&I determines to be potentially fraudulent as a basis for its criminal investigations.

IRS-CI and SB/SE users access the FinCEN Query system (see FinCEN report) as the system of record for BSA data.

(B) A thorough description of the data mining technology that is being used or will be used, including the basis for determining whether a particular pattern or anomaly is indicative of terrorist or criminal activity.

IDEA and LCA do not provide IRS with the ability to determine indicators of terrorist or criminal activity. Special agents and investigative analysts can query based on experience. Agents and analysts determine indicators of fraudulent activity based on previous successful investigations of money laundering, counterterrorism, and BSA violations.

W&I employees use RRP and EFDS to identity potentially fraudulent, noncompliant, and identity theft activity. IRS-CI uses the fraudulent tax returns identified by W&I as a basis for its criminal investigations. Paper refund returns come to EFDS from the Generalized Mainline Framework (GMF)[18] and Questionable Refund Program[19]. Paper returns come into RRP via multiple feeds from GMF. This allows W&I and SDC employees to review those returns for

[18] The Generalized Mainline Framework is a service center pipeline processing system that validates and perfects data from a variety of input sources. Tax returns, remittances, information returns, and adjustment and update transactions in the system are controlled, validated, corrected, and passed on for master file posting.

[19] The Questionable Refund Program (QRP) is a subsystem of EFDS. QRP is a nationwide multifunctional program designed to identify fraudulent returns, to stop the payment of fraudulent refunds, and to refer identified fraudulent refund schemes to CI field offices.

suspicious activities.

EFDS employs a data mining technology called IBM SPSS Modeler. Using this tool, EFDS creates rule sets using a standard built-in algorithm called C5.0. Using examples of current and prior year verified fraud and non-fraud data, the machine-learning model discerns patterns or rules indicative of fraud. The output of the model is a score where a higher score (in the range of 0.0 to 1.0) represents a higher risk or a higher likelihood of a return being fraudulent.

If a return meets designated score tolerances and other criteria, W&I and IRS-CI personnel examine the return for fraudulent activity. Once a return is verified to be false via screening, Taxpayer Protection Program authentication and/or the wage verification process, the fraudulent returns are added via EFDS systemically or by W&I and CI-IRS users to the Scheme Tracking and Retrieval System (STARS) component. IRS-CI investigative analysts review the returns in Discoverer and STARS to find possible schemes, or fraudulent patterns, which may result in a referral to a CI field office for investigation.

RRP employs multiple technologies for data mining activities. Each of these technologies uses current and prior year examples of identity theft (IDT), non-IDT tax fraud, and non-fraud to develop supervised models, unsupervised models, rules, and network analytics:

- SAS – RRP uses SAS as the workbench for developing and evaluating supervised and unsupervised models as well as for data exploration activities. RRP uses multiple SAS machine learning algorithms (e.g., decision trees, neural networks, logistic regression) to uncover patterns in the data associated with fraud. RRP also includes components of SAS' High Performing Analytics (e.g., SAS Grid, SAS in-database analytics) to develop and deploy models with greater complexity than what could be built on a traditional infrastructure. Greater complexity allows RRP models to display greater accuracy and robustness. Supervised models produce a score from 0.000 to 1.000 where a higher score represents a higher likelihood of a return being fraud.
- Greenplum Data Computing Appliance (DCA) – All RRP models are deployed and run directly in the database. Deploying models directly to the database removes the network latency required to move data to a separate application tier server containing the models. Moreover, the Greenplum DCA provides massively parallel processing capabilities across multiple segment servers. In addition to models developed using SAS, RRP also develops models in the form of custom user-defined functions in the Greenplum DCA.
- RRP's network analytics tool – Linked Return Analysis (LRA) – uses multiple custom built Greenplum functions to link returns that display common, suspicious characteristics.
- RRP builds "identity theft filters" using Greenplum functions. These functions combine the outputs of RRP models, rules and LRA to flag suspicious cases of identity theft treatment.
- FICO Blaze Advisor (FICO BA) – RRP builds and maintains business rules using FICO Blaze Advisor. FICO BA provides transparency into the logic that drives business decisions. FICO BA houses the logic that drives RRP's Systemic Verification process – the rule logic that matches taxpayer submitted Income Documents (IDOCs) to the document submitted by withholding party(ies) (e.g., employer submitted W-2s containing income and withholding information).

(C) A thorough description of the data sources that are being or will be used.

The IRS-CI applications IDEA and LCA leverage the following data sources:

- **Taxpayer:** The source is the electronically filed return, as transmitted through the Modernized E-File Program (MeF), or a paper filed tax return.
- **Employers/Payers:** Information from employers/payers captured on various forms as stored in the Information Returns Master File (IRMF).
- **Other Treasury sources:** BSA data provided by FinCEN, Specially Designated Nationals' data provided by the Office of Foreign Assets Control.
- **Other IRS sources:** Tax Exempt Organizations data, Voluntary Disclosures, Criminal Investigations data.

The EFDS and RRP application leverages the following data sources.

- **Taxpayer**: The source is the electronically filed return (as transmitted through the MeF) or a paper filed tax return. EFDS and RRP also load taxpayer data contained on the IRS Master File.
- **Employers/Payers**: Information from employers/payers captured on Form W-2 and/or Form 1099 as stored in the IRMF.
- **Other federal agencies**: Federal Bureau of Prisons for prisoner information; Social Security Administration for National Accounts Profile data for dates of births and deaths.
- **State and local agencies**: Prison systems in all states and the District of Columbia deliver prisoner-listing information annually to IRS-W&I in electronic format.

(D) An assessment of the efficacy or likely efficacy of the data mining activity in providing accurate information consistent with and valuable to the stated goals and plans for the use or development of the data mining activity.

The data uncovered during the query searches are only leads and require additional investigative steps for quality verification. There is no empirical data on the efficacy of searches by the IDEA and LCA applications.

The efficacy of the data mining on EFDS can be measured in terms of fraud detection. A key overall measure of efficacy is "hit: scan," which represents the number of returns selected for verification that, upon inspection by IRS employees, are found to be fraudulent. The overall "hit: scan" for the EFDS system is 1:1.75 for FY 2016. This means that the data mining program accurately predicts fraudulent returns in 10 of 17 cases.

The efficacy of RRP can be measured in terms of identity theft detection. Two key metrics are used to assess RRP's efficacy: lead generation and True Positive Rate. In 2016, RRP generated over 693,000 identity theft leads at a true positive rate of 62 percent. This means that over 6 out of every 10 returns flagged as IDT by RRP never receive a legitimate taxpayer identity authentication via the IRS's web, phone, or in-person authentication processes. In addition to identity theft detections, RRP deployed a new set of models in 2016 to detect fraud

that does not involve identity theft. During the year, RRP generated over 103,000 non-identity theft fraud leads (above all identity theft leads) at a true positive rate of 49 percent.

The efficacy of the FinCEN Query system is discussed in Section (D) of that report.

(E) An assessment of the impact or likely impact of the implementation of the data mining activity on the privacy and civil liberties of individuals, including a thorough description of the actions that are being taken or will be taken with regard to the property, privacy, or other rights or privileges of any individual or individuals as a result of the implementation of the data mining activity.

Once evidence of fraud is discovered, laws and administrative procedures, policies, and controls govern the ensuing actions. IDEA and LCA applications use PII for pattern matching, but the results of a query are used for further investigation. IRS-CI follows the IRS security and privacy IRM standards and regulations for the use and protection of PII.

The impact or likely impact of the EFDS and RRP data mining activities on privacy and civil liberties of individuals is governed by 26 U.S.C. § 6103, which provides general rules of maintaining confidentiality and permissible disclosures. Under this statute, all taxpayer data are private and confidential and protected from disclosure except under specific conditions. Additional laws provide for civil and criminal penalties for any unauthorized disclosure of taxpayer data. The penalties include (1) felony for the willful unauthorized disclosure of tax information, (2) misdemeanor for the unauthorized inspection of tax information, and (3) civil cause of action for the taxpayer whose information has been inspected or disclosed in a manner not authorized by Section 6103. The CI special agents receive periodic training on maximum sentencing and penalties for each criminal violation. Access to the system requires a background check. IRS has a system, Online 5081, governing program access authorization.

Further, EFDS and RRP data mining activities, including its machine learning and scoring process, do not directly use any PII in determining whether a return is likely to be fraudulent. Scoring occurs on the characteristics of the return in question, not on the PII. When performing investigative techniques, PII associated with the return is pulled to assist in validating the return was filed using the taxpayer account in question and to determine venue of the investigation.

The tax returns that IRS-CI reviews are the subjects of criminal investigations and actions based on tax laws, policies, and criminal procedures. Other tax returns are subjected to IRS civil treatments and examination procedures that provide for due process and redress procedures through taxpayer notification, appeals, and tax court options.

(F) A list and analysis of the laws and regulations that govern the information being or to be collected, reviewed, gathered, analyzed, or used in conjunction with the data mining activity, to the extent applicable in the context of the data mining activity.

The use of all tax data is governed by 26 U.S.C. § 6103. Subsection (a) sets out the general rule of confidentiality. Subsection (b) sets forth definitions of terms commonly used throughout Section 6103. Subsections (c) through (o) of Section 6103 contain exceptions to the general rule

of confidentiality. These subsections permit disclosures as described generally below:

- Section 6103(c) – Disclosures to taxpayer's designees (consent);
- Section 6103(d) – Disclosures to state tax officials and certain state and local law enforcement agencies;
- Section 6103(e) – Disclosures to the taxpayer and persons having a material interest;
- Section 6103(f) – Disclosures to certain committees of Congress;
- Section 6103(g) – Disclosures to the President and certain other persons;
- Section 6103(h) – Disclosures to Federal employees and the courts for tax administration purposes;
- Section 6103(i) – Disclosures to Federal employees for non-tax criminal law enforcement purposes and to combat terrorism, as well as the Government Accountability Office;
- Section 6103(j) – Disclosures for statistical purposes;
- Section 6103(k) – Disclosures for certain miscellaneous tax administration purposes;
- Section 6103(l) – Disclosures for purposes other than tax administration;
- Section 6103(m) – Disclosures of taxpayer identity information (generally for Federal debt collection purposes);
- Section 6103(n) – Disclosures to contractors for tax administration purposes; and
- Section 6103(o) – Disclosures with respect to certain taxes.

In addition to disclosures permitted under the provisions of Section 6103, other provisions of the Internal Revenue Code also authorize disclosure of tax information. For example, Section 6104 authorizes disclosure of certain tax information regarding tax-exempt organizations, trusts claiming charitable deductions, and qualified pension plans. Section 6110 authorizes disclosure of certain written determinations and their background files.

(G) A thorough discussion of the policies, procedures, and guidelines that are in place or that are to be developed and applied in the use of such data mining activity in order to:

(i) protect the privacy and due process rights of individuals, such as redress procedures; and

All tax information is protected as required in 26 U.S.C. § 6103 (see E and F above). All employees who interact with tax return and other protected information are required to undergo yearly refresher training that details their responsibilities with respect to information protection and disclosure. In addition to covering 26 U.S.C. § 6103 disclosure provisions, this training module also includes information on the Privacy Act, E-Government Act, Freedom of Information Act, and policies related to protecting PII and other sensitive information. The use of BSA information is strictly controlled under the statute that directs its collection.

The data uncovered during query in IDEA and LCA applications are used as a lead and requires additional investigative steps to verify the quality of the information, as discussed above. IRS maintains an audit trail on all users' access to case data. In addition, a full system log is maintained for any system level activities, including new data loads to the IDEA and LCA application.

Neither EFDS nor RRP determines whether a return is fraudulent or whether a person is going to be subject to criminal prosecution. Once fraud is suspected, laws and administrative

procedures, policies, and controls govern criminal investigations or any other ensuing actions. Due process is provided during any ensuing criminal investigation or civil action.

(ii) ensure that only accurate and complete information is collected, reviewed, gathered, analyzed, or used, and guard against any harmful consequences of potential inaccuracies.

An individual/entity self-reports tax data when submitting the information to the government. FinCEN's data are gathered from information compiled by the reporter based on information provided by their customer or based on the reporter's personal experience. Investigators scrutinize the Suspicious Activity Reports filed by the subject companies and request grand jury subpoenas for the underlying documentation. The supporting records are examined and individuals of interest are identified.

The IDEA and LCA applications are not the authoritative owners of data. However, the data are used for investigative analysis purposes under the IRS Internal Revenue Manual (IRM) standards and guidelines. The data uncovered during query searches are only used as a lead and require additional investigative steps to verify the quality of the information. Therefore, IRS-CI uses these data for generating leads and the special agents verify it through further investigative work.

The tax return information and other information stored in EFDS and RRP used for data mining are based on outside data sources. The only data generated directly in EFDS are the processing steps and the results of examinations of possibly fraudulent returns. The only data generated in RRP are for system monitoring and diagnostics. Through a series of test case procedures executed through Application Qualification Testing (AQT), Systems Acceptability Testing (SAT), and Final Integration Test (FIT), the IRS verifies that the data loaded into EFDS and RRP match the data from the input source and that the system accurately displays the data in the EFDS and RRP end user applications. AQT, SAT, and FIT perform verification with each release of the system. IRS applications are required to have internal auditing capabilities. The internal audits track user access and queries performed with checks against misuse.

TTB Data Mining Activities

(A) A thorough description of the data mining activity, its goals, and, where appropriate, the target dates for the deployment of the data mining activity.

TTB's analytics program performs three types of activities that, together, qualify as data mining as defined by the Federal Agency Data Mining Reporting Act of 2007:

- Queries of commercial transactions recorded by tax and trade databases maintained by
- TTB and other federal agencies;
- Searches of public records and law enforcement databases for indications of illicit dealings; and,
- Link analysis of connections between businesses and individuals.

TTB conducts these activities primarily for the purpose of discovering or locating patterns or anomalies indicative of activity by individuals or businesses that violate federal statutes and regulations administered by TTB. Many of the statutory provisions have criminal

sanctions for their violation. The data used in these activities are, for the most part, gathered with queries of registered individuals or businesses. However, subsequent analysis of the data is primarily pattern-based, seeking anomalies in compiled records. The data mining activities also include some queries and searches that are solely pattern-based, e.g., queries of all tobacco product imports over a given time period.

The goals of TTB's data mining activities are to automate certain routine oversight processes and improve detection of common violations. These activities support predictive models and business intelligence that identify compliance risks and potential fraudulent or criminal activity that may be subject to further field review and action. TTB has predictive models in place that score the risk of tax diversion in the tobacco industry and evaluate businesses seeking a TTB permit. TTB compiles business intelligence that highlights patterns in tax and trade data. These models and business intelligence are in regular use today with improvements and expansions planned for fiscal year 2017.

(B) A thorough description of the data mining technology that is being used or will be used, including the basis for determining whether a particular pattern or anomaly is indicative of terrorist or criminal activity.

TTB uses commercially available data mining technologies to access and analyze information. The experience of intelligence analysts and investigators provides the basis for determining whether a particular pattern or anomaly is indicative of violations. The ability to identify patterns and anomalies is supplemented with statistical analysis and machine learning techniques.

Most data mining is conducted with a combination of SAS statistical analysis software and Oracle relational database systems. Data are retrieved with SAS data step programming and/or Structured Query Language (SQL) queries. Data fields are transformed with procedures that aggregate, correlate, cluster, and otherwise simplify available variables.

Once data are collected and transformed, predictive models use the data to estimate the expected violation risk of a particular individual, business, or incident. The estimates today are based primarily on business rules and templates defined by experienced analysts. These estimates are then implemented in the SAS programming language. Patterns identified through these methods are vetted with experienced analysts and evaluated against randomized test cases.

(C) A thorough description of the data sources that are being or will be used.

TTB uses data from its own databases, the databases of other federal agencies, and commercial data providers. The data sources include:

Internal Data:

- Integrated Revenue Information System (IRIS) – tax data submitted by TTB industry members;
- Permits Online (PONL) –application data from businesses requesting a TTB permit;

- AutoAudit – data from TTB's audits and investigations;
- Formulas Online (FONL) – data from businesses submitting formula approval requests;
- COLAS Online – data from business submitting label approval requests;

External Data:

- Automated Commercial Environment (ACE) / Automated Commercial System (ACS) / Automated Export System (AES) – data regarding imports and exports of products regulated by TTB;
- Census Export Data – data regarding exports of products regulated by TTB;
- Financial Crimes Enforcement Network Query (FinCEN Query) – data submitted in compliance with the Bank Secrecy Act transcripts such as Suspicious Activity Reports (SARs), Currency Transaction Reports (CTRs), etc.; and,
- LexisNexis Accurint – public records data of court proceedings (including some criminal cases), property holdings, licenses, and registrations. This is a for fee service.

These databases are either in use today or being evaluated for inclusion in predictive models. Further integration of the sources is ongoing, as is identification of potential new data sources.

(D) An assessment of the efficacy or likely efficacy of the data mining activity in providing accurate information consistent with and valuable to the stated goals and plans for the use or development of the data mining activity.

TTB's data mining activity is valuable for automating certain routine oversight processes, and improving detection of compliance violations. Initial evaluations indicate that data mining enables more regular oversight and produces indicators for further field review, including investigation and audit. This evaluation is continuing and generating new improvements as the data mining activity matures.

The data mining activity, models, and business intelligence supported by the activity have been effective at helping to automate oversight processes. Predictive models automatically screen approximately 4,000 original permit applications, 2,000 active tobacco businesses, and 2,100 active distilled spirits manufacturers. The models verify information and monitor patterns in operations, tax payments, and international trade activity. The models also automatically monitor financial and trade databases for indications of activity by unregistered businesses. Automating basic screens enables TTB to provide oversight to a wider section of its regulated industries.

The ability of predictive models to detect compliance violations depends greatly on the accuracy of the source data available for the models. Data quality and mining techniques continue to improve with increased use and scrutiny over data. Predictive models that rely on data mining activity are showing promise in detecting violations, though the evaluation is still ongoing. The Tobacco Importer Risk Model and Small Wine Producers Model has demonstrated accuracy of approximately 0.90 (i.e., 9 out of 10 cases recommended by these models result in the detection of previously undisclosed tax liability) and provided a positive

return on investment. The accuracy rate (based on precision and recall) for the Tobacco Risk Model is approximately 0.87; the majority of leads have found compliance issues, and possible undisclosed tax liability. Evaluation of these and other models will continue as part of TTB's ongoing effort to improve model accuracy.

(E) An assessment of the impact or likely impact of the implementation of the data mining activity on the privacy and civil liberties of individuals, including a thorough description of the actions that are being taken or will be taken with regard to the property, privacy, or other rights or privileges of any individual or individuals as a result of the implementation of the data mining activity.

TTB's data mining activity has little impact on the privacy and civil liberties of individuals. Insights gained from the activity primarily result in actions against property, or the privilege to operate in regulated industries, after thorough review by experienced specialists with oversight authorities mandated by federal laws and regulations. The data sources mined are also limited to include only tax records, regulatory records, commercial records, and law enforcement records authorized for use in oversight and enforcement.

Any data concerning individuals or businesses are vigorously protected against unauthorized use and disclosure. Policies and procedures prohibit the search of any database for reasons other than providing authorized oversight or enforcement. In cases when patterns in data are thought to be indicative of compliance issues, the data and circumstances are carefully reviewed by experienced staff before any adverse action is taken. TTB also continues to protect data against any unauthorized disclosure through all investigation and enforcement actions.

Data gathered in data mining activities is considered private and confidential and 26 U.S.C. § 6103 protects it from disclosure. TTB handles these data consistent with that statute. Privacy protections are further assured by additional laws that provide for civil and criminal penalties for any unauthorized disclosure of taxpayer data. There are criminal penalties including: (1) felony for the willful unauthorized disclosure of tax information; (2) misdemeanor for the unauthorized inspection of tax information; and (3) civil cause of action for the taxpayer whose information has been inspected or disclosed in a manner not authorized by Section 6103.

(F) A list and analysis of the laws and regulations that govern the information being or to be collected, reviewed, gathered, analyzed, or used in conjunction with the data mining activity, to the extent applicable in the context of the data mining activity.

TTB administers the provisions of the Internal Revenue Code (IRC) relating to distilled spirits, wine, and beer (26 U.S.C. Chapter 51), tobacco (26 U.S.C. Chapter 52), firearms and ammunition excise taxes (26 U.S.C. sections 4181, 4182, and related portions of chapter 32), and the general rules of tax procedure with respect to these commodities (including related criminal provisions at 26 U.S.C. Chapters 68 and 75). In addition, TTB administers the Federal Alcohol Administration Act (27 U.S.C. chapter 8, subchapter I), which covers basic permits, unfair trade practices, and labeling and advertising of alcohol beverages; the Alcoholic Beverage Labeling Act of 1988 (27 U.S.C. chapter 8, subchapter II), which requires a specific "Government Warning" statement on alcohol beverage labels; and the Webb-Kenyon Act (27 U.S.C. sections 122-122b), which prohibits the shipment of liquor into a state in violation of state

law.

The IRC establishes qualification criteria to engage in the businesses relating to manufacturing and importing or exporting tobacco products, and manufacturing or importing processed tobacco, and require that persons obtain permits to engage in these activities. *See* 26 U.S.C. § 5713. A permit qualification requirement also applies to the production of distilled spirits and wine, as well as to the wholesaling and importation of all beverage alcohol products. *See* 26 U.S.C. §§ 5171(c) and (d), 5271; *see also* 27 U.S.C. §§ 201 *et seq*.

Through an agreement with FinCEN, dated May 3, 2005, TTB is granted direct electronic access to data collected pursuant to provisions of the Bank Secrecy Act, 31 U.S.C. § 5311 *et seq*. The direct access is granted for tax or regulatory purposes relevant to the mission of TTB.

The authority to collect excise taxes on imported alcohol and tobacco products was originally retained by the Secretary of the Treasury through the Homeland Security Act of 2002. *See* 6 U.S.C. §§ 212 and 215. Through Treasury Order 100–16, the Secretary of the Treasury delegates authority over "Customs revenue functions" to the Secretary of the Department of Homeland Security. The Homeland Security Act of 2002 defines these functions as "assessing and collecting customs duties (including antidumping and countervailing duties and duties imposed under safeguard provisions), excise taxes, fees, and penalties due on imported merchandise, including classifying and valuing merchandise for purposes of such assessment." 6 U.S.C. § 215(a)(1).

TTB is authorized pursuant to the Homeland Security Act of 2002, Pub. L. 107-296; Executive Order 13439, July 18, 2007; the Internal Revenue Code of 1986 (IRC); and the Federal Alcohol Administration Act, 27 U.S.C. chapter 8 (FAA Act) to access data within Customs and Border Protection (CBP) data systems necessary to fulfill its statutory mission. TTB is working in conjunction with CBP to fulfill its statutory mission as it relates to imported products subject to various taxes and to ensure taxpayers understand their tax responsibilities related to these products. Cooperative efforts across federal agency lines will accommodate the collection of data as it relates to imported commodities subject to federal taxes, including but not limited to retail, excise, manufacturers, and environmental taxes.

When the data analyzed by the models consists of taxpayer information, 26 U.S.C. § 6103 governs the use of all tax related data. Subsection (a) sets out the general rule of confidentiality. Subsection (b) sets forth definitions of terms commonly used throughout Section 6103. Subsections (c) through (o) of Section 6103 contain exceptions to the general rule of confidentiality. The use of confidential commercial, financial, or trade secrets information is governed by the Trade Secrets Act, 18 U.S.C. § 1905, which prohibits the unlawful disclosure of this information by any federal official, employee, or contractor.

(G) A thorough discussion of the policies, procedures, and guidelines that are in place or that are to be developed and applied in the use of such data mining activity in order to:

(i) protect the privacy and due process rights of individuals, such as redress procedures; and

All of TTB's information collections are subject to the OMB review process and any forms that request personal information include a Privacy Act Statement. In addition, TTB's privacy policy is posted on TTB's website and is referenced on TTB's Online Applications. TTB's systems of record notice can be found in the Federal Register at 80 F.R. 4637 (January 28, 2015).

TTB data mining activities do not determine whether a person or entity will be subject to administrative enforcement action or criminal prosecution. Any audit or investigation that is initiated based, in part, upon data from the activities are governed by the laws, administrative procedures, policies, and controls that govern criminal investigations or any other ensuing actions.

Information generated and accessed by the data mining activities is protected by internal controls that limit access to persons whose official duties require inspection of such information for tax administration purposes. The information is further protected by 26 U.S.C. § 6103, governing the confidentiality of returns and return information, and the Trade Secrets Act, 18 U.S.C. § 1905, which protects confidential commercial, financial, or trade secrets information collected by the federal government.

TTB notifies system operators of the requirements and legal consequences of accessing predictive models in production. The message states:

> 26 U.S.C. 6103 Data Warning. Information contained in this report is tax return information protected from disclosure by 26 U.S.C. 6103. By accessing this report, you hereby certify that your official duties require you to inspect such information for tax administration purposes.

Users of predictive models in production receive training in the proper handling of information. Users receive system demonstrations of the model and have access to a user guide. The same process will be followed for future models when successful testing and evaluation has been completed. Field Operations staff receive 26 U.S.C. § 6103 and disclosure training. In addition, all TTB employees complete the annual Privacy Awareness and Cyber Security Awareness training. Finally, system sponsors and IT staff supporting development, maintenance, and operations of IT systems are required to take additional specialized security training each year.

(ii) ensure that only accurate and complete information is collected, reviewed, gathered, analyzed, or used, and guard against any harmful consequences of potential inaccuracies.

The data mining activities rely on information collected through systems that have their own accuracy related checks and balances. TTB does not rely solely on information gathered

through predictive models to take any adverse action against any individual or entity. Rather, the models are the first step in gathering data and this information is verified through subsequent research and audits of companies and importers before any adverse action is taken.

TTB documents and manages all data sets associated with its systems using the TTB Systems Development Life Cycle (SDLC). Checks and balances are inherent to the data correction process ensuring different teams handle different steps of the effort and include oversight by the Office of the Chief Information Officer Quality Assurance (OCIO QA) Team. When the system owner identifies inconsistencies with data, TTB's OCIO QA Team may initiate a data correction. All changes are documented via the Request for Change process managed by the Configuration Management Team and work orders track the correction through its lifecycle (from request to development and through implementation), which includes confirmation of successful completion by the system owner. The process includes specific identification of the data to be corrected along with rationale for the change. SDLC artifacts (e.g., database scripts) supporting data corrections conform to Data Management (DM) standards. The Software Maintenance Team verifies analysis, development, and testing through a quality review process conducted by the DM Team to ensure the data correction is thoroughly documented. Once the DM Team has approved the data correction, the Operations Team executes the correction and the system owner verifies the correction.

The Memorandum of Understanding with CBP contains language that both parties will notify one another if either agency discovers data issues. Also, the ACS and ACE data import processes in support of the Tobacco Importer Risk Model were documented and tested using TTB's SDLC. For all available governmental data sources, users must sign a non-disclosure agreement before receiving access.

Conclusion

The Department of the Treasury is pleased to provide to Congress its Annual Privacy and Data Mining Reports for Fiscal Year 2016. OPTR has reviewed the activities and programs described in this combined report and will continue to work closely with all Treasury bureaus and offices to protect individual privacy and civil liberties in all Treasury activities.