Date of Approval:
Questionnaire Number: 1828

Basic Information/Executive Summary

What is the name of your project (system, database, pilot, product, survey, social media site, etc.)?

Streaming Data Monitoring Tool

Acronym:

SDMT

Business Unit

IT - Cybersecurity

Preparer

For Official Use Only

Subject Matter Expert

For Official Use Only

Program Manager

For Official Use Only

Designated Executive Representative

For Official Use Only

Executive Sponsor

For Official Use Only

Executive Summary: Provide a clear and concise description of your project and how it will allow the IRS to achieve its mission.

Streaming Data Monitoring Tool-Cloud (SDMT-Cloud) is the project name for the implementation and integration of Splunk technology into the Treasury Cloud. The SDMT-Cloud project delivers end-to-end enterprise-wide auditable event identification, generation, collection, monitoring, and reporting capabilities by leveraging Splunk, key Splunk modules, and other capabilities available for multiple data sources. Splunk is a software product that enables SDMT-Cloud users to search, analyze, and visualize the data gathered from the components of the IRS IT infrastructure or business. This includes data from websites, applications, sensors, devices, etc. After the SDMT-Cloud project team defines the data source, Splunk indexes the data stream and parses it into a series of individual events that SDMT-Cloud users can view and search. SDMT-Cloud is the overall project and FISMA name for Splunk. The SDMT project has migrated

on-prem Splunk data to the Amazon Web Services (AWS) Treasury cloud. Therefore, this document may reference on-prem (SDMT), but this PCLIA is specific to the cloud implementation (SDMT-Cloud).

Personally Identifiable Information (PII)

Will this project use, collect, receive, display, store, maintain, or disseminate any type of Sensitive but Unclassified (SBU), Personally Identifiable Information (PII), or Federal Tax Information (FTI)?

Yes

Please explain in detail how this project uses sensitive data from inception to destruction (data lifecycle).

Personally Identifiable Information (PII), Taxpayer Information (TPI), Federal Tax Information (FTI), Standard Employee Identifier (SEIDs), Internet Protocol (IP) addresses, platform host names, and other similar data is collected by IRS technologies, and is used to validate and authenticate individuals trying to access IRS services. The information is required to ensure only valid and approved IRS taxpayers and Non-Filers may access IRS services. The following Internal Revenue Manuals (IRMs) provide requirements for external authentication of users to IRS systems, to include 10.5.1, 1.35.6, 10.8.1, and 10.8.2. It requires use of identity proofing elements such as taxpayer name, taxpayer address, taxpayer Social Security number and taxpayer date of birth and/or filing status. The other business use of the collected PII information is to conduct fraud analysis to identify and deter fraudulent usage of Electronic Authentication (eAuth) system by unauthorized users. The purpose of Splunk is to collect security audit information. Splunk assists Cybersecurity, Business Units, and Treasury Inspector General for Tax Administration (TIGTA) to detect intrusions and privileged access abuse. Splunk is used to alert business owners, and the stakeholders mentioned when unauthorized access to any PII occurs, or an actionable event may need to be escalated to an incident. The National Archives and Records Administration (NARA) approved the destruction of Splunk audit data when 7 years old (Job No. N1-58-10-22, approved 4/5/2011). Splunk retention requirements are published under IRS Document 12990, Records Control Schedule 19. Any new records generated by the system will be managed according to requirements under IRM 1.15.1 and 1.15.6, and as coordinated with the IRS Records and Information Management (RIM) Program and IRS Records.

Please select all types of Sensitive but Unclassified data (SBU)/Personally Identifiable Information (PII)/Federal Tax Information (FTI) that this project uses.

Address
Email Address
Employment Information
Federal Tax Information (FTI)

Individual Taxpayer Identification Number (ITIN)
Internet Protocol Address (IP Address)
Name
Standard Employee Identifier (SEID)
Telephone Numbers

Cite the authority for collecting SBU/PII/FTI (including SSN if relevant).
PII for personnel administration - 5 USC
SSN for personnel administration IRS employees - 5 USC and Executive Order 9397
SSN for tax returns and return information - IRC section 6109

Product Information (Questions)

1.1 Is this PCLIA a result of the Inflation Reduction Act (IRA)?
No

1.3 What type of project is this (system, project, application, database, pilot/proof of concept, power platform/visualization tool)?
System

1.35 Is there a data dictionary for this system?
Yes

1.36 Explain in detail how PII and SBU data flow into, through and out of this system.
Data from platforms and applications operating on the IRS network is collected via universal, intermediate, and heavy forwarders. These forwarders operate like agents and are installed on servers. In addition to collecting data, these forwarders also assist in the process of data normalization to ensure that data is compliant with the Common Information Model (CIM). Once collected and normalized, data is sent to SDMT where it is validated and indexed. At that point, data is available for use by developers to be visualized in dashboards or to generate reports within SDMT. Approver users can export data from SDMT to meet IRS and Treasury executive management needs.

1.4 Is this a new system?
No

1.5 Is there a Privacy and Civil Liberties Impact Assessment (PCLIA) for this system?
Yes

1.6 What is the PCLIA number?
6425

1.7 What are the changes and why?

When first deployed at the IRS, SDMT was configured to operate on premises (on prem), meaning that it was supported by servers housed in IRS facilities. Due to the large volume of data and processing power that SDMT requires to operate at maximum efficiency, the decision was made to migrate the system to the Treasury Cloud (T-Cloud) so that the IRS could take full advantage of the capacity to scale operations as well as the security that the cloud now offers. For technical reasons, some system components, such as data collectors and forwarders remain on prem. Eventually, these components are also expected to migrate to the cloud. As a system, SDMT collects and analyzes data to both identity and neutralize cyber threats before they happen. The system also provides critical forensics support when breaches or incursion do occur. Data within SDMT can help isolate when and where violations happen who is responsible.

1.8 If the system is on the As-Built-Architecture (ABA), what is the ABA ID number of the system? If this PCLIA covers multiple applications shown on the ABA, please indicate the ABA ID number(s) for each application covered separated by a comma. If the system is not in the ABA, then contact the ABA (https://ea.web.irs.gov/aba/index.html) for assistance.

211560

1.9 What OneSDLC State is the system in (Allocation, Readiness, Execution)?

Execution

1.95 If this system has a parent system, what is the four digit PCLIA Number of the parent system?

0

2.1 If this system discloses any PII to any third party outside the IRS, does the system have a process in place to account for such disclosures in compliance with IRC 6103(p)(3)(A) or Subsection c of the Privacy Act? Contact Disclosure to determine if an accounting is required. Enter "Yes" or "No". If Exempt, type "Exempt".

No

2.2 Please provide the full name of and acronym of the governance board or Executive Steering Committee (ESC) this system reports to.

SDMT Splunk Change Control Board (CCB)

3.1 Does your project/system involve any use of artificial intelligence (AI), including virtual assistant, chat bot, and robotic process automation, as defined in Executive Order 13960?

Yes

3.2 What is the methodology used and what database is training your AI?

SDMT Splunk uses a "Retrieval Augmented Generation" or RAG approach to training its AI. This approach leverages a large language model (LLM) to assist developers in fine tuning their Search Processing Language (SPL) queries in order to provide more contextually relevant responses.

3.3 Does this system use cloud computing?

Yes

3.31 Please identify the Cloud Service Provider (CSP), FedRAMP Package ID, and date of FedRAMP authorization.

AWS F1603047866 May 20, 2023

3.32 Who has access to the CSP audit data (IRS or 3rd party)?

IRS

3.32 Does the CSP allow auditing?

Yes

3.33 Please indicate the background check level required for the CSP (None, Low, Moderate or High).

High

3.4 Is there a breach/incident plan on file?

Yes

3.5 Does the data physically reside in systems located in the United States and its territories and is all access and support of this system performed from within the United States and its territories?

Yes

3.6 Does this system interact with the public through a web interface?

No

3.61 If the system requires the user to authenticate, was a Digital Identity Risk Assessment (DIRA) conducted?

No

3.7 Describe the business process allowing an individual to access or correct their information.

IRS systems that collect the information from users are responsible for managing access.

4.1 Who owns and operates the system (IRS Owned and Operated, IRS Owned and Contractor Operated, Contractor Owned and Operated)?

IRS Owned and Operated

4.2 If a contractor owns or operates the system, does the contractor use subcontractors?

No

4.5 Identify the roles and their access level to the PII data. For contractors, indicate whether their background investigation is complete or not.

The roles vary and include administration, operations, and maintenance. All contractor background investigations are complete.

4.51 How many records in the system are attributable to IRS Employees? Enter "Under 50,000", "50,000 to 100,000", "More than 100,000" or "Not Applicable".

More than 100,000

4.52 How many records in the system are attributable to contractors? Enter "Under 5,000", "5,000 to 10,000", "More than 10,000" or "Not Applicable".

More than 10,000

4.53 How many records in the system are attributable to members of the public? Enter "Under 5,000", "5,000 to 10,000", "More than 10,000" or "Not applicable".

More than 10,000

4.54 If records are attributable to a category not mentioned above in 4.51 through 4.53, please identify the category and the number of corresponding records to the nearest 10,000. If none, enter "Not Applicable".

Not applicable

4.6 How is access to SBU/PII determined and by whom?

Users are granted access by submitting entitlement requests through BEARS/PUMAS. Requests for roles with access to SBU/PII require approval from two separate approval groups (one from the user's line management and one from SDMT management).

5.1 Please describe any privacy risks, civil liberties and/or security risks identified for the system that need to be resolved and what is the mitigation plan?

There are no identified privacy risks to be resolved.

5.11 Is there a Risk Assessment Form and Tool (RAFT) associated with this system on file with your organization or the IRS Risk Office.

Yes

5.2 Does this system use or plan to use SBU data in a non-production environment?

No

Interfaces

Interface Type
IRS Systems, file, or database
Agency Name
Operating System
Incoming/Outgoing
Incoming (Receiving)
Transfer Method
Electronic File Transfer Utility (EFTU)

Interface Type
IRS Systems, file, or database
Agency Name
Web Based (Intranet)
Incoming/Outgoing
Both
Transfer Method
Secured channel via HTTPS

Interface Type
IRS Systems, file, or database
Agency Name
Internet - IRS.gov, through Enterprise Remote Access Project (ERAP)
Incoming/Outgoing
Both
Transfer Method
Secured channel via HTTPS
Other Transfer Method

Interface Type
IRS Systems, file, or database
Agency Name
Client / Server
Incoming/Outgoing
Both
Transfer Method
Electronic File Transfer Utility (EFTU)

Systems of Records Notices (SORNs)

SORN Number & Name

IRS 34.037 - Audit Trail and Security Records

Describe the IRS use and relevance of this SORN.

To both help prevent and investigate system intrusions.

SORN Number & Name

IRS 36.003 - General Personnel and Payroll Records

Describe the IRS use and relevance of this SORN.

The IRS using HCO information to track and monitor user behavior.

SORN Number & Name

IRS 24.046 - Customer Account Data Engine Business Master File

Describe the IRS use and relevance of this SORN.

To maintain accurate records of business tax returns, return transactions, and authorized taxpayer representatives

SORN Number & Name

IRS 24.030 - Customer Account Data Engine Individual Master File

Describe the IRS use and relevance of this SORN.

Interface, match, and verify data with the Social Services Administration and the Census Bureau

Records Retention

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

IRS Document 12990, Records Control Schedule 19

What is the GRS/RCS Item Number?

Item 88

What type of Records is this for?

Electronic

Please provide a brief description of the chosen GRS or RCS item.

The Security Audit and Analysis 88 System (SAAS) enables the IRS and TIGTA to detect potential unauthorized accesses to IRS systems and enables users to analyze and report on audit log data for both Modernized and Current Processing Environment (CPE) applications.

What is the disposition schedule?

Destruction of SDMT data at 7 years old

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

IRS Document 12829, General Record Schedule 3.2

What is the GRS/RCS Item Number?

Item 30

What type of Records is this for?

Electronic

Please provide a brief description of the chosen GRS or RCS item.

System access records. These records are created as part of the user identification and authorization process to gain access to systems. Records are used to monitor inappropriate systems access by users.
Includes records such as:
Includes records such as:

- user profiles
- log-in files
- password files
- audit trail files and extracts
- system usage files
- cost-back files used to assess charges for system use

Exclusion 1. Excludes records relating to electronic signatures.
Exclusion 2. Does not include monitoring for agency mission activities such as law enforcement.
Systems not requiring special accountability for access.
These are user identification records generated according to preset requirements, typically system generated. A system may, for example, prompt users for new passwords every 90 days for all users.

What is the disposition schedule?

Temporary. Destroy when business use ceases.

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

IRS Document 12829, General Records Schedule 3.2

What is the GRS/RCS Item Number?

Item 31

What type of Records is this for?

Electronic

Please provide a brief description of the chosen GRS or RCS item.

System access records. These records are created as part of the user identification and authorization process to gain access to systems. Records are used to monitor inappropriate systems access by users. Includes records such as:

- user profiles
- log-in files
- password files
- audit trail files and extracts
- system usage files
- cost-back files used to assess charges for system use

Exclusion 1. Excludes records relating to electronic signatures.
Exclusion 2. Does not include monitoring for agency mission activities such as law enforcement.

What is the disposition schedule?

Temporary. Destroy 6 years after password is altered or user account is terminated, but longer retention is authorized if required for business use.

Data Locations

What type of site is this?

Data Gateway

What is the name of the Data Gateway?

TCloud

What is the sensitivity of the Data Gateway?

Federal Tax Information (FTI)

What is the URL of the item, if applicable?

https://core.splunk.iam.int.for.irs.gov

Please provide a brief description of the Data Gateway.

SDMT Splunk Analytics Platform is repository for data from across the IRS Enterprise. Though primarily in the cloud, there are some on premise components.

What are the incoming connections to this Data Gateway?

Heavy or Intermediate Forwarders TCLOUD:
hpcsispkhfw001.treasury.local hpcsispkhfw002.treasury.local
hpcsispkhfw003.treasury.local hpcsispkhfw004.treasury.local On Prem Memphis: vp2smemschfwd00.lxcs.csirc.irs.gov
vp2smemschfwd01.lxcs.csirc.irs.gov
vp2smemschfwd02.lxcs.csirc.irs.gov
vp2smemschfwd03.lxcs.csirc.irs.gov
vp2smemschfwd04.lxcs.csirc.irs.gov On Prem Martinsburg:
vp2smtbschfwd00.lxcs.csirc.irs.gov
vp2smtbschfwd01.lxcs.csirc.irs.gov

vp2smtbschfwd02.lxcs.csirc.irs.gov
vp2smtbschfwd03.lxcs.csirc.irs.gov
vp2smtbschfwd04.lxcs.csirc.irs.gov