

Date of Approval: 05/02/2024
Questionnaire Number: 1104

Basic Information/Executive Summary

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

SCRIPS Modernization

Business Unit

Taxpayer Services

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.

The Service Center Recognition/Image Processing System Modernization (SCRIPS MOD) Project initiative, under IRA Project 1.2.4 - Expand Tax Return Digitalization, will leverage a vendor-driven solution to accelerate the digitalization of scanned paper tax returns. IT will implement modernized technology to digitalize the intake of paper tax returns. The solution will process paper returns by encompassing data ingestion, classification, extraction, validation/perfection, storage, and output of a complete submission via multiple data formats as needed for downstream processing for broad business usage. There are no external facing interfaces with taxpayers.

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).

SCRIPS MOD is targeted towards processing US taxpayer physical paper-based submissions. This means that live taxpayer data is ingested at the service centers by high-speed scanners, converted to PDF files with original images. These files are packaged into zip files for Secure Data Transfer (SDT) via the Enterprise File Transfer Utility (EFTU) from the service center(s) to the IEP AWS cloud environment. Original files are retained for 3 days. Zip files are unzipped and ingested by the solution software for form classification and data extraction. Extracted data is stored in DynamoDB as key/value pairs. DynamoDB is not a system of record and data will be purged per IRS business requirement (Retention period is TBD). The original PDF file, NARA metadata and changelog files are uploaded to IRS "As-Received" datastore (via the Information Returns Processing System (IRPS)) for permanent retention. Extracted data is formatted in compliance with downstream systems and sent to on-premises using EFTU for transport to the IRS datacenters for pick-up/ingestion.

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

Address

Document Locator Number (DLN)

Employer Identification Number

Employment Information

Family Members

Individual Taxpayer Identification Number (ITIN)

Name

Preparer Taxpayer Identification Number (PTIN)

Social Security Number (including masked or last four digits)

Telephone Numbers

Vehicle Identification Number (VIN)

Cite the authority for collecting SBU/PII/FTI (including SSN if relevant).

PII for federal tax administration - generally IRC Sections 6001 6011 or 6012

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)?

Yes

1.2 What is the IRA Initiative Number?

1.2

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?

No

1.36 Explain in detail how PII and SBU data flow into, through and out of this system.

From the On-Premise (On-Prem) IRS Network the Scanner Operators at the SCRIPS sites will scan in the data via highspeed scanners (Bantec XD54 or IBML). From there the data will transit to Virtual Storage Area Network (vSAN) Storage via Server Message Block (SMB) file share. Working its way from On-Prem to IRS IEP AWS Cloud the data will be sent via Secure Data Transport (SDT) via the Enterprise File Transfer Utility (EFTU) to the S3 storage Bucket. From the S3 Storage Bucket, the solution software platform will access the data via HTTPs. The solution software platform will then output to Dynamo DB storage and output formatter (translation layer) to Data files Store (S3 storage). From there the Data Files Store S3 will send the data via EFTU to On-Premise IRS Network Data file stores and Downstream IRS systems (again via EFTU). NOTE: AWS GovCloud processing will also output to IRS "As Received Datastore" (Information Returns Processing System (IRPS), leveraging the Enterprise API GW via HTTPS.

1.4 Is this a new system?

Yes

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

No, this is a new system that will eventually replace the legacy SCRIPS/ISRP system (which does have a PCLIA)

1.6 What is the PCLIA number?

Not Applicable

1.7 What are the changes and why?

Not Applicable - this is a PCLIA for a new system.

1.8 If the system is on the As-Built-Architecture, what is the ABA ID of the system? If this PCLIA covers multiple applications shown on the ABA, please indicate the ABA ID for each application covered separated by a comma.

SCRIPS Mod not ABA currently.

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 PCLIA Number of the parent system?

No parent system

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.

WebApps Governance Board and Strategic Development Executive Steering Committee

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?

Not Applicable - COTS software solution

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.

IRS IEP is the CSP for SCRIPS MOD, Date of FedRamp Auth is 6/21/2016 per their PCLIA

3.32 Does the CSP allow auditing?

Yes - Per the IEP PCLIA (#6800 - dated Sep 16, 2022) it states they do audit.

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

IRS audits CSP data - the owner of CSP data is IRS and CSP does allow auditing.

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

Moderate - Per the IEP PCLIA (dated Sep 16, 2022)

3.4 Is there a breach/incident plan on file?

For the CSP (IEP) yes - please see attachments.

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?

DIRA Not required for this project.

3.62 Please upload the approved DIRA report using the Attachments button.

DIRA Not Required; Email concurrence attached.

3.63 If individuals do not have the opportunity to give consent to collect their information for a particular use, why not?

The SCRIPS Modernization System collects data via tax forms collected by the IRS and is not taxpayer facing.

3.64 If the individual was not notified of the following items prior to the collection of information, why not? 1) Authority to collect the information 2) If the collection is mandatory or voluntary 3) The purpose for which their information will be used 4) Who the information will be shared with 5) The effects, if any, if they don't provide the requested information.

Information is collected via tax forms and is digitized and pushed downstream to the SCRIPS Modernization system which is not taxpayer facing.

3.65 If information is collected from third-party sources instead of the individual, please explain your decision.

Information is not collected from 3rd party sources.

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

Not Applicable - SCRIPS will not have any impact on taxpayers as the system is just accessible to IRS employees for processing of paper tax returns.

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

IRS Owned, Contractor developed, IRS operated.

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

4.3 What PII/SBU data does the subcontractor have access to?

Yes - The vendor development team will have read access to the Production Support Environment (PSU). The PSU environment functions as the continuous AI Model Training environment and is isolated from the Live Production Environment activity. Like other steps taken by other production deployed AI projects/products, the SCRIPS MOD project team is pursuing a Live Data Waiver.

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

Roles: Development and O&M support. Only fully cleared MBI contractors have access to any IRS PII/SBU. All development is taking place within the IRS IEP API Gateway (managed service).

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".

Not Applicable

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".

Not Applicable

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?

Access to the data is determined by the appropriate IRS manager(s) based on a user's roles, responsibilities, and need-to-know. The manager will request a user to be added. They must submit the request via the Business Entitlement Access Request System (BEARS) to request access to the System.

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?

System is still under development, please refer to additional information section for path forward.

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.

No

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

No

5.3 Please upload the approved SBU Data Use Questionnaire or Request. If the request has been recertified, please upload the approved recertification form. Select Yes to indicate that you will upload the SBU Data Use form.

Yes – uploaded.

Interfaces

Interface Type

Forms

Agency Name

LONG LIST WILL BE PROVIDED IN EXCEL SHEET

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.

This SORN is applicable due to IRS Enterprise Security Audit Trails (ESAT) requirements and mandates and IRM 10.8.1 AU Controls applicable to the system

SORN Number & Name

IRS 24.030 - Customer Account Data Engine Individual Master File

Describe the IRS use and relevance of this SORN.

Return data will be stored in IMF; all associated information would need to be maintained and managed by IRS standards by authorized representatives

SORN Number & Name

IRS 22.054 - Subsidiary Accounting Files

Describe the IRS use and relevance of this SORN.

In SCRIPS Modernization, data from each line of the return (t-line data) is transmitted to GMF for distribution to the appropriate database. Return data is then stored in Master Files (IMF/BMF) depending on if it's an individual or business return. Process return data is available for IRS Business users to retrieve via Integrated Data Retrieval System (IDRS).

SORN Number & Name

IRS 24.046 - Customer Account Data Engine Business Master File

Describe the IRS use and relevance of this SORN.

Return data will be stored in BMF; all associated information would need to be maintained and managed by IRS standards by authorized representatives

SORN Number & Name

IRS 22.061 - Information Return Master File

Describe the IRS use and relevance of this SORN.

Return data will be stored IMF

SORN Number & Name

IRS 22.060 - Automated Non-Master File

Describe the IRS use and relevance of this SORN.

In SCRIPS Modernization, data from each line of the return (t-line data) is transmitted to GMF for distribution to the appropriate database. Return data is then stored in Master Files (IMF/BMF) depending on if it's an individual or business return.

Records Retention

What is the Record Schedule System?

Record Control Schedule (RCS)

What is the retention series title?

RCS 29

What is the GRS/RCS Item Number?

56

What type of Records is this for?

Electronic

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

The records covered by this schedule are created and/or accumulated in the territory offices of W&I Business Unit and/or in the processing campus operations, formerly service centers

What is the disposition schedule?

7

What is the Record Schedule System?

Record Control Schedule (RCS)

What is the retention series title?

RCS 29

What is the GRS/RCS Item Number?

65

What type of Records is this for?

Electronic

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

The records covered by this schedule are created and/or accumulated in the territory offices of W&I Business Unit and/or in the processing campus operations, formerly service centers

What is the disposition schedule?

7 years