

Date of Approval: 06/03/2025
Questionnaire Number: 2235

Basic Information/Executive Summary

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

Digitalization Rapid Automation Framework - Historical Forms

Business Unit

Enterprise Digitalization

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.

This project involves Optical Character Recognition/Intelligent Character Recognition-based extraction of critical fields from Form 940, Form 941 and Form 1040 documents and transmission of the extracted data (from Forms 940, 941 and 1040) and images into an As-Received Data Storage. This solution reduces and eventually eliminates the backlog of Forms 940, 941 and 1040 documents (versions 2010 - 2024). In addition, it creates a scalable solution for extracting data from digitized paper forms and meets field level accuracy thresholds as defined by enterprise architecture service level agreement.

Benefits:

Reduce/eliminate backlog of Historical forms including F940, F941 & F1040.

Expedited Processing: Improves efficiency of processing paper forms 940, 941 and 1040.

Quality & Value Output: Increases data accuracy and eliminates duplication of efforts.

Maintenance is much easier as the data is digitized rather than storing bulkier paper forms.
Improves employee morale by streamlining back-office processes.

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).

The PII data from Forms 940, 941 and 1040 will be extracted and stored in an As-Received database. It will be viewable in DCR (Digital Content Retrieval). The PII data is stored only in the final As-Received database and not anywhere in between along the process flow. The PII data does not get destroyed but stored in the As-Received database.

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

Employer Identification Number
Federal Tax Information (FTI)
Individual Taxpayer Identification Number (ITIN)
Name
Other
Social Security Number (including masked or last four digits)

Please explain the other type(s) of PII that this project uses.

Name Control from IDRS (Integrated Data Retrieval System), DLN (Document Locator Number)

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

PII for federal tax administration - generally IRC Sections 6001 6011 or 6012
SSN for tax returns and return information - IRC section 6109

Product Information (Questions)

1 Is this PCLIA a result of a specific initiative or process improvement?

Yes

1.1 What is the name of the Business Unit (BU) or Agency initiative?

IRA 4.3

2 Describe in detail, the Robotic Process Automation (RPA) process; be sure to identify the project title and business unit owner; state what IRS Strategy or initiative it supports; identify the system or process it supports and if PII will be required for the RPA to run; identify activities and workflow controls with the type and capabilities that will be incorporated; lastly indicate how the service benefits from the use of this RPA. (Process, Library, Test Automation, Template.)

Project Title: Digitalization Rapid Automation Framework - Historical Forms

Business Unit: Taxpayer Services

PII Required: Yes

RPA Type: Unattended but requires human in the loop function to perform validation.

Process Description: Every form that gets scanned will be converted into a PDF image and a scanner manifest will be attached. This will be zipped into a single file and will be EFTU'd (Electronic File Transfer Utility) to the IEP (Integrated Enterprise Portal) AWS (Amazon Web Services) S3 bucket. The PDF will go through the UiPath service which digitizes the PDF, extracts the necessary data and presents it for validation in the UiPath service's Action Center; saves the validated extracted data in XML format in the S3 bucket. From the S3 bucket, this is transported downstream to the As-Received database through SFPS. The extracted data and the PDF image of the original form can be viewed through the DCR application.

Benefits:

Reduce/eliminate backlog of Historical forms including F940, F941 & F1040.

Expedited Processing: Improves efficiency of processing paper forms 940, 941 and 1040.

Quality & Value Output: Increases data accuracy and eliminates duplication of efforts.

Maintenance is much easier as the data is digitized rather than storing bulkier paper forms.

Improves employee morale by streamlining back-office processes.

3 Is this a new Robotic Process Automation (RPA) project?

No

3.1 Is there a PCLIA for this project?

Yes

3.11 What is the questionnaire number of the most recent approved PCLIA?

1511

3.12 What is the questionnaire title of the most recent approved PCLIA?

RPA Digitalization Enablement Platform for Historic Forms Data Extraction (2021-2024)

4 Identify the IRS IT systems, applications, projects, and/or databases this RPA is applied to; include the associated system name.

EFTU, AWS S3 bucket, UiPath (Action Center), SFPS, IDRS, DCR.

5 Identify why the use of SBU/PII/FTI is required; include any type of Sensitive But Unclassified (SBU), Personally Identifiable Information (PII), or Federal Tax Information (FTI) that this project will create, collect, receive, use, process, maintain, access, inspect, display, store, disclose, disseminate, or dispose of.

SBU/PII/FTI data collected/used/processed: Name, SSN, EIN, DLN, Name Control, Tax Year, Tax Period, Quarter, Received Date. Purpose of PII collected/used/processed: Is to send these data downstream to the As-Received database.

6 Is your RPA Attended/Unattended?

Unattended automation with human in the loop function for validation purposes.

7 Is this RPA process converting from paper to electronic format or automating a process currently performed by a human?

Yes

7.1 Explain the process being replaced/automated.

Currently, tax technicians enter all the data from paper forms 940, 941 and 1040 into the As-Received database. This is a time-consuming, tedious and expensive process.

8 Indicate what level of complexity the RPA is classified as and if you were required to register with One Solution Delivery Lifecycle (OneSDLC) or not, or indicate if Information Technology's (ITs) Technical Insertion process was used for approval of this RPA.

Level of Complexity: High

Registered with OneSDLC: Yes

9 Will connections or interdependencies be established for this RPA?

Yes

9.1 Will the connection be encrypted?

Yes

9.2 Will authentication/credentials be required?

Yes

9.3 Please provide details for the connection/interdependency. Indicate if this occurs on the backend versus through the system/user interface.

Incoming Connections to the system: Various Campus Centers scanning via SCRIPS BanTec scanner for form versions 1990-2024. File transfer via EFTU from BanTec scanner storage location to AWS S3 bucket in IEP.

Outgoing Connections: UiPath service retrieves the individual PDF images of forms from the S3 bucket for OCR extraction, loads it to Action Center for validation using human in the loop, moves the data to the As-Received database.

10 Indicate who has or will have permission to access the data and how users are authenticated.

Tax Technicians with the Taxpayer Services Business Unit will have access to the data. They will be authenticated using BEARS Entitlements.

11 Indicate if Business Entitlement Access Request System (BEARS) entitlements are required for access and if Privileged User Management Access System (PUMAS) control management is applied for granting access to the system(s)? If BEARS/PUMAS are not applied, indicate what access controls are in place.

Yes, BEARS Entitlements are required for access.

12 Identify the maintenance tasks or updates performed; state whether or not the maintenance tasks are inherited from the host (UiPath Platform) or you are using customized maintenance activities.

No customized maintenance tasks or updates are performed. Maintenance tasks are inherited from the UiPath platform.

13 Indicate if this product or system shares data outside of the United States or its territories.

No

14 Indicate if this system or Robotic Process Automation (RPA) is trained through the use of algorithms; indicate if the algorithm used contains data with a sensitivity classification. (Sensitive but unclassified data might include algorithms, methods, system data, or PII/FTI that could be used to re-identify a person.)

No

15 Describe this system's (RPAs) audit trail process in detail; include location of supporting documents (SPLUNK). Note: Upload of this document is required.

UiPath provides the audit trails at the organization/tenant level and these logs are stored in Microsoft Sequential Query Language (MSSQL) database. UiPath also provides an integration to external log products like SPLUNK. And managed service uses SPLUNK as the log aggregator, and all the UiPath logs are fed into IEP SPLUNK and IEP SPLUNK is connected to IRS SPLUNK. Any logs going to IEP SPLUNK will be forwarded to IRS SPLUNK.

Interfaces

Interface Type

IRS Systems, file, or database

Agency Name

Electronic File Transfer Utility

Incoming/Outgoing

Both

Transfer Method

Electronic File Transfer Utility (EFTU)

Interface Type

IRS Systems, file, or database

Agency Name

DIG (Digitalization) 317

Incoming/Outgoing

Both

Transfer Method

Application to Application (A2A)

Interface Type

IRS Systems, file, or database

Agency Name

UiPath Action Center

Incoming/Outgoing

Both

Transfer Method

Application to Application (A2A)

Interface Type

IRS Systems, file, or database

Agency Name

Digital Content Retrieval (DCR)

Incoming/Outgoing

Incoming (Receiving)

Transfer Method

Application to Application (A2A)

Interface Type

IRS Systems, file, or database

Agency Name

As-Received Database (SYS 316)

Incoming/Outgoing

Incoming (Receiving)

Transfer Method

Other

Other Transfer Method

IRPS (Information Returns Processing System) is used to send the data to As Received.

Interface Type

IRS Systems, file, or database

Agency Name

Amazon Web Services S3 Bucket

Incoming/Outgoing

Both

Transfer Method

Amazon Web Services Platform (AWS)

Interface Type

IRS Systems, file, or database

Agency Name

API (Application Programming Interface) Gateway

Incoming/Outgoing

Both

Transfer Method

Application to Application (A2A)

Interface Type

IRS Systems, file, or database

Agency Name

Integrated Data Retrieval System (IDRS)

Incoming/Outgoing

Both

Transfer Method

Application to Application (A2A)

Interface Type

Forms

Agency Name

940

Incoming/Outgoing

Both

Transfer Method

Electronic File Transfer Utility (EFTU)

Interface Type

Forms

Agency Name
941
Incoming/Outgoing
Both
Transfer Method
Electronic File Transfer Utility (EFTU)

Interface Type
IRS Systems, file, or database
Agency Name
Scanned Form Processing Service (SFPS)
Incoming/Outgoing
Both
Transfer Method
Application to Application (A2A)

Interface Type
Forms
Agency Name
1040
Incoming/Outgoing
Both
Transfer Method
Electronic File Transfer Utility (EFTU)

Systems of Records Notices (SORNs)

SORN Number & Name
IRS 24.046 - Customer Account Data Engine Business Master File
Describe the IRS use and relevance of this SORN.
Information related to businesses is looked up on IDRS and the value of Name Control is retrieved from IDRS.

SORN Number & Name
IRS 34.037 - Audit Trail and Security Records
Describe the IRS use and relevance of this SORN.
To identify and track any unauthorized access to sensitive/classified information and potential breaches.

SORN Number & Name
IRS 24.030 - Customer Account Data Engine Individual Master File

Describe the IRS use and relevance of this SORN.

Information related to individual taxpayers is looked up on IDRS and the value of Name Control is retrieved from IDRS.

Records Retention

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

Infrastructure Project Records

What is the GRS/RCS Item Number?

GRS 3.1 Item 010

What type of Records is this for?

Both (Paper and Electronic)

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

GRS 3.1 item 010- Infrastructure Project Records. Infrastructure project records. Information Technology (IT) infrastructure, systems, and services project records document the basic systems and services used to supply the agency and its staff with access to computers and data telecommunications. Includes requirements for and implementation of functions such as: • maintaining network servers, desktop computers, and other hardware, • installing and upgrading network operating systems and shared applications, and • providing data telecommunications; and infrastructure development and maintenance such as acceptance/authorization of infrastructure components, analysis of component options, feasibility, costs and benefits, and work associated with implementation, modification, and troubleshooting.

What is the disposition schedule?

Destroy 5 years after project is terminated, but longer retention is authorized if required for business use.

Data Locations

What type of site is this?

System

What is the name of the System?

As-Received Database

What is the sensitivity of the System?

Federal Tax Information (FTI)

Please provide a brief description of the System.

Storage and staging of data extracted from forms 940,941 and 1040.

What are the incoming connections to this System?

UiPath automation performs OCR extraction on 709 forms retrieved from S3 bucket. The output of the extraction will go through schema validation, after that API call occurs to grab validated Form 709 extracted data to As-Received database

What are the outgoing connections from this System?

As Received database is the final location where the 709 extracted data will go to. Data can be retrieved via the Digital Content Retrieval (DCR) system and a summary report will be generated indicating downstream routing transactions.

What type of site is this?

System

What is the name of the System?

AWS (Amazon Web Services) S3 bucket in IEP

What is the sensitivity of the System?

Federal Tax Information (FTI)

Please provide a brief description of the System.

This is a data storage for extracted data from forms 940, 941 and 1040.

What are the incoming connections to this System?

Data is coming in from the scanned files.

What are the outgoing connections from this System?

Extracted data gets sent to the bot machines to perform validations and formatting.

What type of site is this?

System

What is the name of the System?

SPLUNK

What is the URL of the item, if applicable?

https://staging2.splunk.iam.int.for.irs.gov/en-US/app/irs_dig_app/dig_dcr_audit

Please provide a brief description of the System.

SPLUNK collects and reports all the logs from the UiPath Orchestrator which can be used to create metrics about the automation.

What are the incoming connections to this System?

From UiPath through Nlog.

