

Date of Approval: 03/19/2026
Questionnaire Number: 2798

Basic Information/Executive Summary

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

OCI DAS Extraction Tool

Acronym:

ODExT

Business Unit

Large Business and International

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.

Processing documents from foreign banks, foreign trusts, and other international entities is challenging due to the unstructured information embedded in PDFs. Manual extraction is slow, resource-intensive, and prone to inconsistencies that hinder workflow efficiency. Offshore Compliance Initiative Data Analytics Services Extraction Tool (ODExT) addresses this challenge by automating the extraction of unstructured data from complex PDFs. The tool uses 22 specialized business logic modules tailored to institutions such as Belize Bank, British Caribbean Bank, Nevis Trust and comparable institutions. These modules enable accurate parsing of documents including personal and corporate bank applications, trust application forms, trust deeds, and Integrated Data Retrieval System (IDRS) transcripts. ODExT converts unstructured PDF content into structured datasets through automated parsing and standardized output formats.

Extracted data is saved as text files that can be easily converted to Excel or loaded into databases such as Oracle, SQL, or Microsoft Access. This supports smooth integration with existing analytical workflows. ODEXT also generates structured search results that help users identify relevant information that may not be immediately visible in the source documents, improving analytical efficiency and accuracy. Implementing ODEXT significantly accelerates document processing. Tasks that once required months or years can now be completed in minutes or seconds, greatly improving record-production timelines. The tool enhances accuracy and consistency by applying standardized extraction logic and reducing manual effort. This allows staff to focus on higher-value analytical work and supports faster, more informed decision-making across the organization.

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).

This is a tool that enhances research skills by automating the search function on massive data files obtained via summons when conducting examinations or investigations. The tool has 16 business logics that a user can choose from depending on the type of documents being searched, e.g., personal bank application form, corporate bank application form, trust application form, or trust deed. The extracted data results are saved as a text file format which can be converted to an Excel file. The results of this tool often point the way to additional useful information for the user. The tool converts from PDF to text files format and gathers the data then uploaded to the Offshore Compliance Initiative (OCI) database. All data in Offshore Compliance Initiative is organized and presented in a similar fashion. All data is organized first by a data source, and second by a table. Each table is presented in a grid format. Consequently, each grid can be described as belonging to a single data source-table combination. Data sources are financial institutions which have provided the IRS with data relating to offshore accounts, the types of data sources include banks, credit card processors, and merchants.

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

Address
Citizenship or Migration Status
Credit Card Number
Email Address

Employer Identification Number
Employment Information
Family Members
Financial Account Number
Individual Taxpayer Identification Number (ITIN)
Name
Other
Passport Number
Social Security Number (including masked or last four digits)
Tax ID Number
Telephone Numbers

Please explain the other type(s) of PII that this project uses.
Date of Birth, Trust Information, Formation of Business

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 a process improvement?
Yes

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

2 What type of project is this (system, project, application, database, pilot/proof of concept/prototype, power platform/visualization tool)?
Offshore Compliance Initiative system tool

3 What Tier designation has been applied to your system? (Number)
3

4 Is this a new system?
Yes

4.1 Is there a previous Privacy and Civil Liberties Impact Assessment (PCLIA) for this project?
No

4.12 What is the previous PCLIA title (system name)?
No

4.2 You have indicated this is not a new system; explain what has or will change and why. (Expiring PCLIA, changes to the PII or use of the PII, etc.)

The tool is an enhancement to an existing OCI's record production database under the PIA 6135 and renewal PCLIA 2154.

5 Is this system considered a child system/application to another (parent) system?

No

6 Indicate what OneSDLC State is the system in (Allocation, Readiness, Execution) or indicate if you go through Information Technology's (IT) Technical Insertion Process and what stage you have progressed to.

Readiness

7 Is this a change resulting from the OneSDLC process?

Yes

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

LB&I Governance Board

9 Is this System listed on As-Built-Architecture (ABA)? If the system is not in the ABA, then contact the ABA (<https://ea.web.irs.gov/aba/index.html>) for assistance.

Yes

9.1 What is the ABA ID?

211977

10 Does this system disclose any PII to any third party outside the IRS?

No

10.1 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?

Yes

10.2 You have indicated that you do not have an "accounting of disclosures" process in place; please indicate a projected completion date or explain the steps taken to develop your accounting of disclosures process. Note: The Office of Disclosure should be contacted to develop this system's accounting of disclosures process.

Project completed 10/25/2023.

11 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 and 14110?

No

12 Does this system use cloud computing?

No

13 Does this system/application interact with the public?

No

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

No

15 Is this system owned and/or operated by a contractor?

No

17 The Privacy Act of 1974 (5 USC § 552a(e)(3)) requires each agency that maintains a system of records, to inform each individual requested to supply information about himself or herself. Please provide the Privacy Act Statement presented by your system or indicate a Privacy Act Statement is not used and individuals are not given the opportunity to consent to the collection of their PII.

SSN for tax returns and return information is Internal Revenue Code Section 6109

SSN for personnel administration (IRS Employees) is 5 USC & Executive Order

9397 PII for personnel administration is 5 USC

18 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

19 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

20 How many records in the system are attributable to members of the public? Enter "Under 100,000", "100,000 to 1,000,000", "More than 1,000,000" or "Not applicable".

More than 1,000,000

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

Access is determined by manager of the analyst (OCI group).

23 Is there a data dictionary on file for this system? Note: Selecting "Yes" indicates an upload to the Attachment Section is required.

No

24 Explain any privacy and civil liberties risks related to privacy controls.

There are no residual system privacy, civil liberties, and/or security risks identified.

25 Please upload all privacy risk finding documents identified for the system (Audit trail, RAFT, POA&M, Breach Plan, etc.); click "yes" to confirm upload(s) are complete.

No

26 Describe this system's audit trail in detail. Provide supporting documents.

No automatic data extraction It can run on any Windows machine. Data can be processed on external, local and network drives. File processing can be done locally. Output files can be easily converted to Excel or Access. No programming skills required to run data extraction.

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

No

Interfaces

Interface Type

Other Organization

Agency Name

Federal Courts

Incoming/Outgoing

Incoming (Receiving)

Transfer Method

Compact Disc (CD)

Systems of Records Notices (SORNs)

SORN Number & Name

IRS 00.001 - Correspondence Files and Correspondence Control Files

Describe the IRS use and relevance of this SORN.

Official use only. The tool is used to convert pdf file to text file.

SORN Number & Name

DO .149 - Foreign Assets Control Legal Files

Describe the IRS use and relevance of this SORN.

These records are maintained to assist in providing legal advice to the OFAC and the Department of the Treasury regarding issues of compliance, enforcement, investigation, and implementation of matters related to OFAC and the statutes and regulations administered by the agency. These records are also maintained to assist in litigation related to OFAC and the statutes and regulations

administered by the OFAC. The tool is used to convert pdf file to text file.

SORN Number & Name

IRS 34.037 - Audit Trail and Security Records

Describe the IRS use and relevance of this SORN.

Official use only. The tool is used to convert pdf file to text file.

Records Retention

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

5.2 Transitory and Intermediary Records

What is the GRS/RCS Item Number?

020

What type of Records is this for?

Electronic

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

Intermediary records. Records that meet the following conditions:

- They exist for the sole purpose of creating a subsequent record and
- They are not required to meet legal or fiscal obligations, or to initiate, sustain, evaluate, or provide evidence of decision-making.

What is the disposition schedule?

Destroy upon creation or update of the final record, or when no longer needed for business use, whichever is later.

Data Locations

What type of site is this?

System

What is the name of the System?

OCI Application

What is the sensitivity of the System?

Personally Identifiable Information (PII) including Linkable Data

What is the URL of the item, if applicable?

<https://vp0smemappociw1.ds.irsnet.gov/OCI/index.aspx>

Please provide a brief description of the System.

All data in Offshore Compliance Initiative (OCI) is organized and presented in a similar fashion. All data is organized first by a data source, and second by a table. Each table is presented in a grid

format. Consequently, each grid can be described as belonging to a single data source-table combination. Data sources are financial institutions which have provided the IRS with data relating to offshore accounts, the types of data sources include banks, credit card processors, and merchants. Examples of data sources are Credomatic, UBS-AG, and Stanford. Tables contain groupings of related attributes. Examples of tables would include Names, Transactions, and Monthly Summaries. Related attributes for the Names table might include First Name, Last Name, Full Name, DOB (Date of Birth). Each individual table, therefore, is expressed as a combination of Data Source and Table. Examples of the Data Source-Table combination would be Credomatic Names, UBS AG Monthly Summary, and Stanford Transactions. Tables and Table attributes are Data Source specific, and therefore, each data source's tables and table attributes can and will be different. To utilize ODEXT, the user would perform the following tasks: Perform a search on the Search screen, review the results on the Search Results screen, and then examine the detailed data in the tables.

What are the incoming connections to this System?

None

What are the outgoing connections from this System?

None