

Date of Approval: 03/28/2025
Questionnaire Number: 1962

Basic Information/Executive Summary

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

LBI Data Capture System

Acronym:
LBI DCS

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.

The Large Business and International Data Capture System (LBI-DCS) is a critical infrastructure application within the Internal Revenue Service (IRS) that has maintained an authority to operate since 2005. It is primarily used by the research community across Large Business and International (LB&I), Research Applied Analytics and Statistics (RAAS), and Small Business Self-Employed (SBSE) divisions. The main functionalities within LBI-DCS include:

- Data Capture System (DCS)
- LB&I DataMart (DataMart)
- Taxpayer Information Gateway (TIG)
- LB&I Taxpayer Registry (LTR)
- Data Science Services (DSS)

DCS is the core infrastructure for digitizing LB&I paper-filed tax returns. It processes returns imaged through the Statistics of Income Distributed Processing System (SOI DPS), an LB&I Image Net (LIN) component. DCS also extracts data

from Modernized Electronically Filed (MeF) returns' Portable Document Format (PDF) attachments. Using commercial-off-the-shelf (COTS) software, DCS captures data from paper-filed tax returns and integrates it in database tables, enabling alignment with electronically filed returns for scoring and return selection purposes. By converting paper-filed returns into electronic format, this part of the system ensures that all LB&I taxpayers are treated consistently, regardless of how their returns were filed.

LB&I DataMart (DataMart)

DataMart provides access to business and trust income tax returns, Forms 1041, 1065, and 1120 families (e.g. F1120, F1120-S, F1120-PC, etc.). DataMart extracts, loads, and transforms (ETL) data from multiple sources to support research initiatives, workload selection, trend analysis and issue identification. It also provides restricted working spaces to approved users within the research community enabling them to optimize the data usages in their mission driven work. There are multiple transformations and supports that make DataMart data unique and allow for more effective and efficient use of data. Additionally, DataMart ETL products are used by various other LBI-DCS downstream processes, including Tier 3 and Tier 4 systems. DataMart data includes:

- Modernized electronically filed (MeF) returns
- Paper-filed returns
- Business Master File authoritative data
- Audit Information Management System (AIMS)
- Forms 1041, 1065, and 1120-S Schedules K1

Taxpayer Information Gateway (TIG) TIG provides case-building analytical reports to assist field Revenue Agents (RAs) in identifying potential examination issues and trends. TIG utilizes data from DataMart to generates pre-audit reports saving RAs hours of manual effort. LB&I Taxpayer Registry (LTR) LTR, developed and maintained by LB&I, is used for multiple strategic LB&I compliance initiatives including campaigns, Schedule UTP (Uncertain Tax Positions), Office of Tax Shelter Analysis (OTSA), and other projects such as Large Partnership Compliance, Partnership Model, Foreign Payments Practice, and Research Risk Review. LTR collates data from multiple existing systems to enable the filtering, classification, risk assessment, feedback, and reporting of taxpayer compliance for the LB&I population to various stakeholders e.g. management, analysts, technical specialists, and the field. Data is periodically extracted, transformed and loaded into LTR to enable the matching of published financial information with tax filing information, identification and analysis of compliance issues and reporting. Information is delivered to the field and feedback gathered from the field to improve upstream processes. LTR provides a platform for the development of the Automated Comparative Analytics Tool (ACAT), comparing five years of tax return data, identifying anomalies and Large, Unusual, or Questionable (LUQ) items in addition to delivering the LB&I Web Statistics. ACAT and the Web Statistics provides additional information for workload selection efforts and analysis.

Data Science Services (DSS)

DSS hosts tax data and provides advanced computational infrastructure to support the development and execution of data models that exceed the capacity of DataMart. DSS-hosted tax data include data replicated from DataMart and

project-specific data used by modeling teams. The DSS computational infrastructure consists of servers designed for data science research that house advanced analytical tools such as in-database Python, R, and specialized SQL Server data science and engineering tools. DSS currently supports the Large Corporate Compliance (LCC), Unified Corporate Compliance (UCC), Tax Trends Initiative (TTI), and Line Anomaly Recommender (LAR) models with more to follow.

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 LB&I Data Capture System tool provides for LB&I and other Business Operating Divisions to access data extracted from large and mid-size business tax returns that have been imaged by the Statistics of Income Distributed Processing System (SOI DPS) (LB&I Image Net (LIN) component) process. Additionally, LBI-DCS will provide a data store that will make Modernized E-file (MeF) and Master File authoritative data immediately available to researchers. Copies of source data are managed for up to 10 years and then purged from our systems.

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

- Address
- Alien Registration Number
- Document Locator Number (DLN)
- Email Address
- Employer Identification Number
- Employment Information
- Federal Tax Information (FTI)
- Financial Account Number
- Internet Protocol Address (IP Address)
- Name
- Passport Number
- Photograph
- Preparer Taxpayer Identification Number (PTIN)
- Professional License Number
- Protected Information
- Standard Employee Identifier (SEID)
- Tax ID Number

Telephone Numbers

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 tax returns and return information - IRC section 6109

Product Information (Questions)

1 Is this PCLIA a result of the Inflation Reduction Act (IRA)?

No

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

Application

3 What Tier designation has been applied to your system?

1

4 Is this a new system?

No

4.1 Is there a previous Privacy and Civil Liberties Impact Assessment (PCLIA) for this project?

Yes

4.11 What is the previous PCLIA number?

7145

4.12 What is the previous PCLIA title (system name)?

Data Capture System DCS-2

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.)

Changes to PII and Form Types

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

Yes

5.1 Identify the parent system's approved PCLIA number.

7145

5.2 Identify the parent system's name as previously approved.

Data Capture System, DCS-2

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.

Execution

7 Is this a change resulting from the OneSDLC process?

No

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

Risk Inventory Control Board (RICB)

9 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.

210289

10 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?

No

10.1 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.

Disclosures of tax or employee or personnel information where the disclosure is made to IRS employees or with a need to know are excluded from the accounting requirement.

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

14 Describe the business process allowing an individual to access or correct their information. (Due Process)

The information is not collected directly from individuals. The information collected as part of the application is the information obtained from various IRS databases and files, which in turn are tax forms filed by tax entities. Notice, consent and due process are provided in the tax forms instructions, and pursuant to 5 USC.

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

No

16 Identify what role(s) the IRS and/or the contractor(s) performs; indicate what access level (to this system's PII data) each role is entitled to. (Include details about completion status and level of access of the contractor's background investigation was approved for.)

User: Read-Only

Manager: Read-Only

System Administrator: Administration

Access Level: Read Only, Read and Write, or an Administrator.

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.

Privacy Act Statement is not used; individuals are not given the opportunity to consent to the collection of their PII.

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?

The LBI DCS system utilizes the standard IRS on-line access application to document approvals for access. Data access is granted on a need-to-know basis. A potential user must submit a request for access to their local management for approval. Users are not permitted access without a signed form from an

authorized management official. Specific permissions (Read, Write, Modify, Delete, and/or Print) are defined on the form and set (activated) by the System Administrator prior to the user being allowed access. User privileges and user roles determine the types of data that each user has access to. Management monitors system access and removes permissions when individuals no longer require access. Users are assigned to specific modules of the application and specific roles within the modules and accounts follow the principle of least privilege which provide them the least amount of access to PII/SBU data that is required to perform their business function after receiving appropriate approval.

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.

None

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.

LBI-DCS is a COTS (Commercial Off the Shelf) software-based system and relies on GSS-30 for collecting Audit Information. SQL Audit logs are stored on the P:\SQLAudits folder on SQL Servers. Else, Windows Event Viewer for all servers.

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

Yes

27.1 Please upload the Approved Email and one of the following SBU Data Use Forms, Questionnaire (F14664) or Request(F14665) or the approved Recertification (F14659). Select Yes to indicate that you will upload the Approval email and one of the SBU Data Use forms.

Yes

Interfaces

Interface Type

State Agencies

Agency Name

New York City Department of Finance

Incoming/Outgoing

Incoming (Receiving)

Transfer Method
Electronic File Transfer Utility (EFTU)

Interface Type
IRS Systems, file, or database

Agency Name
Audit Information Management System (AIMS)

Incoming/Outgoing
Incoming (Receiving)

Transfer Method
Electronic File Transfer Utility (EFTU)

Interface Type
IRS Systems, file, or database

Agency Name
Compliance Data Warehouse (CDW)

Incoming/Outgoing
Both

Transfer Method
Electronic File Transfer Utility (EFTU)

Interface Type
IRS Systems, file, or database

Agency Name
Modernized Tax Return Data Base (MTRDB)

Incoming/Outgoing
Incoming (Receiving)

Transfer Method
Electronic File Transfer Utility (EFTU)

Systems of Records Notices (SORNs)

SORN Number & Name

IRS 42.008 - Audit Information Management System

Describe the IRS use and relevance of this SORN.

This information is necessary for the administration of the Tax Law pertaining to Business Tax Returns and Tax Returns containing International Business transactions or features.

SORN Number & Name

IRS 34.037 - Audit Trail and Security Records

Describe the IRS use and relevance of this SORN.

Monitor for security risk and compliance violations ensuring only authorized users access system.

SORN Number & Name

IRS 42.027 - Data on Taxpayers' Filing on Foreign Holdings

Describe the IRS use and relevance of this SORN.

This information is necessary for the administration of the Tax Law pertaining to Business Tax Returns and Tax Returns containing International Business transactions or features.

SORN Number & Name

IRS 22.054 - Subsidiary Accounting Files

Describe the IRS use and relevance of this SORN.

This information is necessary for the administration of the Tax Law pertaining to Business Tax Returns and Tax Returns containing International Business transactions or features.

SORN Number & Name

IRS 22.027 - Foreign Information System (FIS)

Describe the IRS use and relevance of this SORN.

Foreign Payments and International Information Reporting Forms

SORN Number & Name

IRS 42.001 - Examination Administrative Files

Describe the IRS use and relevance of this SORN.

This information is necessary for the administration of the Tax Law pertaining to Business Tax Returns and Tax Returns containing International Business transactions or features.

SORN Number & Name

IRS 42.017 - International Enforcement Program Information Files

Describe the IRS use and relevance of this SORN.

This information is necessary for the administration of the Tax Law pertaining to Business Tax Returns and Tax Returns containing International Business transactions or features.

SORN Number & Name

IRS 42.021 - Compliance Programs and Projects Files

Describe the IRS use and relevance of this SORN.

This information is necessary for the administration of the Tax Law pertaining to Business Tax Returns and Tax Returns containing International Business transactions or features.

Records Retention

What is the Record Schedule System?

Record Control Schedule (RCS)

What is the retention series title?

Compliance Research

What is the GRS/RCS Item Number?

27 / Item 54

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 maintained in the conduct of IRS compliance research.

What is the disposition schedule?

As required under IRM 1.15.6

What is the Record Schedule System?

Record Control Schedule (RCS)

What is the retention series title?

ENTERPRISE COMPUTING CENTER - XXXXXXXXXXXX
(ECC-XXX)

What is the GRS/RCS Item Number?

19 / Item 81

What type of Records is this for?

Electronic

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

Modernized e-File (MeF) System

What is the disposition schedule?

Delete/Destroy electronic data after successful entry and verification into the MTRDB repository.

What is the Record Schedule System?

Record Control Schedule (RCS)

What is the retention series title?

TAX ADMINISTRATION - WAGE AND INVESTMENT (W&I)
RECORDS

What is the GRS/RCS Item Number?

29 / Item 55

What type of Records is this for?

Electronic

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

Electronically Filed Individual, Partnership and Fiduciary Income Tax Returns

What is the disposition schedule?

Destroy on or after January 16, six years after the end of the processing year unless needed for Collection Statute Expiration Date (CSED) Extract due to a balance due.

What is the Record Schedule System?

Record Control Schedule (RCS)

What is the retention series title?

TAX ADMINISTRATION - WAGE AND INVESTMENT (W&I)
RECORDS

What is the GRS/RCS Item Number?

29 / Item 344

What type of Records is this for?

Both (Paper and Electronic)

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

Information Return with Respect to a Foreign Corporation (Form 5471).

What is the disposition schedule?

Retire to Records Center 3 years after the end of the processing year. Destroy 5 years after the end of the processing year.

Data Locations

What type of site is this?

System

What is the name of the System?

LBI-DCS

What is the sensitivity of the System?

Federal Tax Information (FTI)

Please provide a brief description of the System.

The LB&I Data Capture System tool provides for LB&I and other Business Operating Divisions to access data extracted from large and mid-size business tax returns that have been imaged by the Statistics of Income Distributed Processing System (SOI DPS) (LB&I Image Net (LIN) component) process. Additionally, DCS will provide a data store that will make Modernized E-file (MeF) and Master File authoritative data immediately available to researchers.

What are the incoming connections to this System?

Incoming Interface/s include AIMS (Audit Information Management System), MTRDB (Modernized Tax Return Data Base), CDW (Compliance Data Warehouse). Additionally, the New York City Department of Finance. All represent copies of 'authoritative' data.

What are the outgoing connections from this System?

Outgoing Interface/s include SWC (Selection and Workload Classification) - A rules-based engine that performs batch processing of LB&I tax data, to identify potentially high-risk returns.

Splunk - Splunk is a Security Information and Event Management (SIEM) software solution tool composed of various dashboards that more and more IRS employees are using to aggregate and/or analyze security data for systems/applications.