Date of Approval: 12/20/2024
Questionnaire Number: 1673

Basic Information/Executive Summary

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

Enterprise Informatica Platform

Acronym:

EIP

Business Unit

Information Technology

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.

Overview: Enterprise Informatica Platform (EIP) is an enterprise infrastructure platform providing IRS projects access to Informatica Extract, Transform, and Load (ETL) software for use in their processing pipeline. EIP maintains a full set of servers in Development (DEV), TEST, Final Integration Test (FIT), Production (PROD), and Disaster Recovery (DR) environments for projects' use. EIP is not an application project in that no code is directly developed, however it provides a centralized, enterprise-wide environment for critical projects to perform, test, and deploy Informatica code as part of their overall application. IRS Strategic Objective: EIP supports the IRS Strategic Objective: Modernize the IRS through its People, Processes, and Technology. By providing a required server infrastructure component, the IRS reduces the deployment time of applications and therefore increases the overall business value of projects from each business organization. Ownership and Administration: -- Enterprise Services (ES), Technology Strategy Management (TSM), Managed Technology Shared Services (MTSS) owns and administrates the system. TSM does not have any contractor support for EIP. -- TSM partners with Enterprise Services, Solutions Engineering (SE), Data Engineering (DE), Data Engineering Platform Services (DEPS) to provide engineering support for the platform. DEPS has contractor support directly from Informatica to provide technical and engineering expertise as related to the Informatica products. -- TSM partners with Enterprise Operations (EOps), Infrastructure Services Division (ISD), Middleware Services Branch (MSB), Middleware Transformation Services Section (MTSS) for the O&M activities for the Informatica products and user-support. MSB:MTSS does have contractors as part of their team. -- TSM partners with Enterprise Operations (EOps), Enterprise Computing Center (ECC), Server Production and Application Support Branch 2, Unix/Linux Product and Application Support 2B for the administration of the Red Hat Linux (RHEL) operating system used by our platform. They do have contractors as part of their team. -- TSM partners with Enterprise Operations (EOps), Data Management Services and Support Division, Distributed Database Services Branch 1, Oracle Database Support Section 3 for the administration of our Oracle databases. They do not have contractors supporting the platform. System Location: All servers are virtual. The physical hosting hardware is in the XXXXXXXXXXX Computing Center for EIP's Sandbox, Development, Test, Final Integration Test, and Production environments. The physical hosting hardware is in the XXXXXXX Computing Center for EIP's backup Final Integration Test and Disaster Recovery environments. Purpose: EIP provides data integration and Extract, Transform, and Load (ETL) capabilities for application projects to utilize in their data processing pipeline to support their unique business requirements. EIP is currently utilized by several mission critical projects such as Customer Account Data Engine 2 (CADE2), Integrated Production Model (IPM), and Return Review Program (RRP). Method of Data Transmission: The PII data that EIP collects includes the username, email address, and phone number of the users within the Informatica product. This information comes into Informatica through Active Directory using TLS 1.2. IP addresses are captured when a connection is established to the Informatica product. Connections to the Informatica product are made using TLS 1.2. Access and Storage: The PII data is stored in the Informatica metadata database. Access to the database is restricted to the database administrator and the Informatica database service accounts. The username, phone number, and email address can be viewed for each user in the Informatica Administrator console in the security section. Access to the security section is restricted to Informatica support staff that have been approved for Informatica Admin access via BEARS. Acronyms: EIP - Enterprise Informatica Platform ETL - Extract, Transform, and Load TLS - Transport Layer Security DEV - Development FIT - Final Integration Test PROD - Production DR - Disaster Recovery BEARS - Business Entitlement Access Request System

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).

Informatica syncs with Active Directory to obtain the user information for the users in the Informatica application. The name, email address, and telephone number are part of the information that Active Directory sends back to Informatica. Informatica uses the username (SEID) for authentication and authorization of the user within the Informatica product. It also uses this name for event logging which is then sent to Splunk. Informatica captures the IP address when a connection is made to the Informatica product. This is used for event logging which is then sent to Splunk. Informatica uses the email address as part of a custom developed scripting solution to implement disabling inactive users. This scripted solution retrieves the email address for users that are 30 days away from being deactivated and for those that have been deactivated to send them notification emails. Informatica doesn't use the full name or phone number for anything, but since it is sent from Active Directory, it is included. The Informatica product is configured to request membership attributes for Active Directory and has no control over what information related to a user Active Directory sends back.

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

Email Address
Internet Protocol Address (IP Address)
Other
Standard Employee Identifier (SEID)

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

Username and Phone Number

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

PII for personnel administration - 5 USC

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)?

EIP is a Data Integration and Transformation Commercial Off-The-Shelf (COTS) Application. The COTS application is a pass-through and not responsible for storing any client data.

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.

The PII data that Enterprise Informatica Platform (EIP) collects includes the username, email address, and phone number of the users within the Informatica product. This information comes into Informatica through Active Directory using Transport Layer Security (TLS) 1.2. Internet Protocol (IP) addresses are captured when a connection is established to the Informatica product. Connections to the Informatica product are made using TLS 1.2. The data moves out of the system into Splunk for security tracking and auditing. This is accomplished by Linux shell scripts that are executed via Control-M every 10 minutes. The scripts extract information from Informatica's internal logs based on auditable events as determined by Cybersecurity. This information is then appended to the Linux operating system's syslog file which is ingested directly by Splunk.

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?

389

1.7 What are the changes and why?

The previous PCLIA for this system, #389, was created in 2013. Since then, the platform owner for Enterprise Informatica Platform (EIP) has changed and PCLIA #389 has become outdated through multiple version upgrades.

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.

EIP ABA ID: 211244 Link: https://ea.web.irs.gov/ABA/SA/ea-panel1_main-application_441851.htm#

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

Execution (The system has been in an O&M state for several years.)

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".

Exempt, this system does not disclose any PII.

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

Enterprise Services Governance Board (ESGB)

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?

No

3.3 Does this system use cloud computing?

No

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

No

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

Individuals are not allowed to access and correct their IP Address and SEID information, as that is part of the security audit trail for EIP.

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 Personally Identifying Information (PII) data is stored in the Informatica metadata database. Access to the database is restricted to the database administrator and the Informatica database service accounts. The username, phone number, and email address can be viewed for each user in the Informatica Administrator console in the security section. Access to the security section is restricted to Informatica support staff that have been approved for Informatica

Admin access via Business Entitlement Access Request System (BEARS). Any contractor support staff are cleared via the standard IRS Minimum Background Investigation (MBI) process.

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".

Under 50,000

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".

Under 5,000

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".

Not applicable

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

The Enterprise Informatica Platform (EIP) uses Business Entitlement Access Request System (BEARS) entitlements for access to the system. Access to Personally Identifying Information (PII) data via the Informatica Administrator console are controlled by the Informatica Admin role BEARS entitlements. There are two types of Informatica Admin BEARS entitlement: 1) For Platform ownership and engineering support (including contractor support from Informatica) and 2) for Enterprise Operations (EOPS) Middleware operations and maintenance support staff (including contractors). Approval of entitlement #1 is controlled jointly by the platform owner lead and the engineering support lead. Approval of entitlement #2 is controlled by the Middleware operations support lead. Red Hat Enterprise Linux (RHEL) access to the PII data captured by the Informatica product include the same groups as listed above and the Linux System Administration team that manages the RHEL operating system on the servers. The Linux System Administration team has their own BEARS entitlement controlled by their team.

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?

No risks are identified for the system, the PII data is required as a part of the Enterprise Security Audit Trail to map auditable events with individual users.

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

Interfaces

Interface Type

IRS Systems, file, or database

Agency Name

Active Directory

Incoming/Outgoing

Incoming (Receiving)

Transfer Method

Other

Other Transfer Method

Lightweight Directory Access Protocol (LDAP) Sync via Transport Layer Security (TLS) 1.2

Interface Type

IRS Systems, file, or database

Agency Name

Enterprise Security Audit Trail

Incoming/Outgoing

Outgoing (Sending)

Transfer Method

Other

Other Transfer Method

Splunk Universal Forwarder

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.

The system captures the SEID and IP address of users that access the COTS and server to map auditable events to specific individuals as part of the Enterprise Security Audit Trail.

Records Retention

What is the Record Schedule System?

Non-Record