

A. SYSTEM DESCRIPTION

1. Enter the full name and acronym for the system, project, application and/or database. Big Data Analytics, BDA

2. Is this a new system? No

2a. If no, is there a PIA for this system? Yes

If yes, enter the full name, acronym, PIA ID Number and milestone of the most recent PIA.

Big Data Analytics, BDA#1017

Next, enter the **date** of the most recent PIA. 9/10/2014

Indicate which of the following changes occurred to require this update (check all that apply).

- | | |
|-----|--|
| Yes | Addition of PII |
| No | Conversions |
| No | Anonymous to Non-Anonymous |
| Yes | Significant System Management Changes |
| Yes | Significant Merging with Another System |
| No | New Access by IRS employees or Members of the Public |
| No | Addition of Commercial Data / Sources |
| No | New Interagency Use |
| No | Internal Flow or Collection |

Were there other system changes not listed above? Yes

If yes, explain what changes were made. The inclusion of the CADE2 Operational Data Store (ODS) onto the GreenPlum device will be leveraged by Integrated Production Model (IPM). Security change request have been submitted and reviewed for this inclusion (Security Change Management (SecChm)-2014-12-1417, SecChm-2014-12-1418, SecChm-2014-12-1420, SecChm-2014-12-1421.

3. Check the current ELC (Enterprise Life Cycle) Milestones (select all that apply)

- | | |
|-----|--|
| No | Vision & Strategy/Milestone 0 |
| No | Project Initiation/Milestone 1 |
| No | Domain Architecture/Milestone 2 |
| No | Preliminary Design/Milestone 3 |
| No | Detailed Design/Milestone 4A |
| No | System Development/Milestone 4B |
| No | System Deployment/Milestone 5 |
| Yes | Operations & Maintenance (i.e., system is currently operational) |

4. Is this a Federal Information Security Management Act (FISMA) reportable system? Yes

A.1 General Business Purpose

5. What is the general business purpose of this system? Provide a clear, concise description of the system, application or database, the reason for the system, and the benefits to the IRS to use the information, and how the information will be used. The BDA project is an infrastructure project that will provide the IRS with a massively parallel processing capability to support many projects that have a need for case identification, selection, prioritization, delivery, compliance and decision analytics. BDA will serve to perform advanced data analysis that can facilitate IRS audit selections, analyzing taxpayer filings, and more.

B. PII DETAIL

6. Does the system use, collect, receive, display, store, maintain, or disseminate IR Code 6103 taxpayer information; or any type of Sensitive but Unclassified (SBU) or Personally Identifiable Information (PII)? Yes

6a. If yes, does the system use, collect, receive, display, store, maintain, or disseminate Social Security Numbers (SSN s) or tax identification numbers (i.e. last 4 digits, etc.)? Yes

If yes, check who the SSN (or tax identification number) is collected on.

Yes On Primary Yes On Spouse Yes On Dependent

If yes, check all types SSNs (or tax identification numbers) that apply to this system:

Yes	Social Security Number (SSN)
Yes	Employer Identification Number (EIN)
Yes	Individual Taxpayer Identification Number (ITIN)
Yes	Taxpayer Identification Number for Pending U.S. Adoptions (ATIN)
Yes	Practitioner Tax Identification Number (PTIN)

Describe the planned mitigation strategy and forecasted implementation date to mitigate or eliminate the use of SSN s (or tax identification numbers). There is no alternative to the use of the SSN. The SSN is the significant part of the data being processed. This program is aware of and part of the IRS-wide SSN elimination and reduction program.

6b. Does this system contain other (non-SSN) PII that it uses, collects, receives, displays, stores, maintains, or disseminates? (i.e. Names, addresses, etc.) Yes

If **yes**, specify the information.

<u>Selected</u>	<u>PII Element</u>	<u>On</u>	<u>On</u>	<u>On</u>
		<u>Primary</u>	<u>Spouse</u>	<u>Dependent</u>
Yes	Name	Yes	Yes	Yes
Yes	Mailing address	No	No	No
Yes	Phone Numbers	No	No	No
No	E-mail Address	No	No	No
Yes	Date of Birth	Yes	Yes	Yes
No	Place of Birth	No	No	No
No	SEID	No	No	No
No	Mother's Maiden Name	No	No	No
Yes	Protection Personal Identification Numbers (IP PIN)	No	No	No
No	Internet Protocol Address (IP Address)	No	No	No
No	Criminal History	No	No	No
No	Medical Information	No	No	No
No	Certificate or License Numbers	No	No	No
No	Vehicle Identifiers	No	No	No
No	Passport Number	No	No	No
No	Alien (A-) Number	No	No	No
Yes	Financial Account Numbers	No	No	No
No	Photographic Identifiers	No	No	No
No	Biometric Identifiers	No	No	No
No	Employment (HR) Information	No	No	No
Yes	Tax Account Information	Yes	Yes	Yes

6c. Does this system contain SBU information that is not PII, it uses, collects, receives, displays, stores, maintains, or disseminates? Yes

If **yes**, select the types of SBU.

<u>Selected</u>	<u>SBU Name</u>	<u>SBU Description</u>
Yes	Agency Sensitive Information	Information which if improperly used or disclosed could adversely affect the ability of the agency to accomplish its mission
No	Procurement sensitive data	Contract proposals, bids, etc.
Yes	Official Use Only (OUO) or Limited Official Use (LOU)	Information designated as OUO or LOU is information that: is exempt under one of the statutory Freedom of Information Act exemptions; is prohibited by other laws or regulations; would significantly impede the agency in carrying out a responsibility or function; or would constitute an unwarranted invasion of privacy.
No	Proprietary data	Business information that does not belong to the IRS Information which if modified, destroyed or disclosed in an unauthorized manner could cause: loss of life, loss of property or funds by unlawful means, violation of personal privacy or civil rights, gaining of an unfair procurement advantage by contractors bidding on government contracts, or disclosure of proprietary information entrusted to the Government
Yes	Protected Information	Information which if modified, destroyed or disclosed in an unauthorized manner could cause: loss of life, loss of property or funds by unlawful means, violation of personal privacy or civil rights, gaining of an unfair procurement advantage by contractors bidding on government contracts, or disclosure of proprietary information entrusted to the Government
No	Physical Security Information	Security information containing details of serious weaknesses and vulnerabilities associated with specific systems and facilities
No	Criminal Investigation Information	Information concerning IRS criminal investigations or the agents conducting the investigations.

6d. Are there other types of SBU/PII used in the system? No

6e. Cite the authority for collecting SBU/PII (including SSN if relevant)

<u>Yes</u>	PII for federal tax administration is generally Internal Revenue Code Sections 6001, 6011, 6012e
<u>Yes</u>	SSN for tax returns and return information is Internal Revenue Code Section 6109
<u>No</u>	SSN for personnel administration (IRS Employees) is 5 USC & Executive Order 9397
<u>No</u>	PII for personnel administration is 5 USC
<u>No</u>	PII about individuals for Bank Secrecy Act compliance 31 USC
<u>No</u>	Information by CI for certain money laundering cases may be 18 USC

6f. Has the authority been verified with the system owner? Yes

B.1 BUSINESS NEEDS AND ACCURACY

7. Explain the detailed business needs and uses for the SBU/PII, and how the SBU/PII is limited only to that which is relevant and necessary to meet the mission requirements of the system. If SSNs (or tax identification numbers) are used, explicitly explain why use of SSNs meets the criteria. Be specific. The Business needs and uses the SBU/PII for case identification, selection, prioritization to conduct tax administration, delivery, compliance and decision analytics. BDA performs advanced data analysis that can facilitate IRS audit selections, analyzing taxpayer filings, and more. IPM is a common, read-only data store loaded onto the BDA-GSS infrastructure, containing core IRS data (e.g., tax accounts, tax returns, and information returns). The IPM database will be utilized by a wide range of pre-approved IPM client-project applications to support case identification, selection, prioritization, delivery, compliance analysis and decision analytics. IPM provides a centralized database, populated with current and historical return data from a variety of data sources. As a result of the analysis, case workloads are selected, notices may be generated, or transactions may be required to post back to the individual's tax account. IPM client-project applications performing analysis, and receiving data downloaded from the IPM BDA Greenplum database, use the taxpayer SSN as a major search and account matching key.
8. How is the SBU/PII verified for accuracy, timeliness, and completeness? Explain how steps are taken to ensure that all information maintained by the system that is used by IRS to make any adverse determination about an individual's rights, benefits, and/or privileges is maintained with such accuracy, relevance, timeliness, and completeness as is reasonably necessary to assure fairness to the individual in the determination. BDA is an appliance with Greenplum Database component. The client projects, such as IPM, have their own verification process for data accuracy, timeliness, completeness. BDA assumes that the data is accurate, timely, and complete when the client projects store the data to the Greenplum database. The data elements of each table in the IPM Greenplum database are loaded from the data source. The data is regarded as accurate when received. Timeliness and completeness are verified through a table on IPM, TABLE VIEW LOAD STATUS, which documents the date and time when data loads are completed. The IPM project and IPM client-project developers perform Unit Testing on the IPM Greenplum Development database. IT AD Enterprise System Testing (EST) owns the IPM Test & FIT databases on the BDA-GSS infrastructure, and are responsible for access control and audit logging. Information Technology Application Development (IT AD) EST performs both System Acceptability Test (SAT) and Final Integration Test (FIT) testing of the IPM database.

C. PRIVACY ACT AND SYSTEM OF RECORDS

9. Are 10 or more records containing SBU/PII maintained, stored, and/or transmitted by or through this system? Yes

9a. If yes, are records in the system retrieved by any personal identifier (e.g., name, SSN, Photograph, IP Address) for an individual? Yes

If yes, is there a System of Records Notice(s) or SORNs that addresses the PII records in this system? Yes

If **yes**, enter the SORN number(s) and the complete the name of the SORN.

<u>SORNS Number</u>	<u>SORNS Name</u>
Treasury/IRS 24.030	Customer Account Data Engine Individual Master File
Treasury/IRS 42.021	Compliance Programs and Project Files
Treasury/IRS 24.046	Customer Account Data Engine Business Master File
Treasury/IRS 34.037	Audit Trail and Security Records System
Treasury/IRS 22.061	Information Return Master File
Treasury/IRS 22.026	Form 1042S Index by Name of Recipient

If **yes**, does the System of Records Notice(s) (SORN) published in the Federal Register adequately describe the records as required by the Privacy Act? Yes

D. RESPONSIBLE PARTIES

10. Identify the individuals for the following system roles. ## Official Use Only

E. INCOMING PII INTERFACES

11. Does the system receive SBU/PII from other system or agencies? Yes

11a. If **yes**, does the system receive SBU/PII from IRS files and databases? Yes

If **yes**, enter the files and databases.

<u>System Name</u>	<u>Current PIA?</u>	<u>PIA Approval Date</u>	<u>SA & A?</u>	<u>Authorization Date</u>
ACA Information Returns (AIR)	Yes	10/17/2016	No	
Affordable Care Act Coverage Data Repository (ACA-CDR)	No		Yes	02/27/2014
CP2100	No		No	02/27/2014
SCHEDULE R	No		No	02/27/2014
Duplicate TIN On-Line (DUP TIN)	No		No	02/27/2014
Information Returns Database (IRDB)	Yes	01/13/2016	No	02/27/2014
Individual Master file (IMF)	Yes	02/28/2017	No	02/27/2014
Information Returns Master File Processing (IRMF)	No		No	02/27/2014
Individual Return Transaction File (IRTF)	No		No	02/27/2014
CENSUS American Industry Classification System (NAICS)	No		No	02/27/2014
Employee Plans Master File (EPMF)	Yes	12/08/2015	No	02/27/2014
REIC	No		No	02/27/2014
ZIP-Place Of Duty (POD)	No		No	02/27/2014
SS8 Integrated Case Processing (SS8ICP)	Yes	11/07/2014	No	02/27/2014
WITHHOLDING COMPLIANCE	Yes	04/24/2015	No	02/27/2014
Employer Tax Employee Project (ETEP)	No		No	02/27/2014
Business Master File-DB2	Yes	11/27/2012	No	02/27/2014
Compliance Data Environment (CDE)	Yes	03/23/2016	Yes	04/17/2013
Payer Master File Processing (PMF)	Yes	03/09/2017	Yes	02/07/2013
Combined Annual Wage Reporting (CAWR)	Yes	02/19/2016	No	02/07/2013
Audit Information Management System Reference (AIMS-R)	Yes	12/15/2015	Yes	02/08/2012
ALLOWABLE LIVING EXPENSE	No		No	02/08/2012
Chapter Three Withholding System (CTW)	Yes	11/04/2014	No	02/08/2012
Business Return Transaction File (BRTF)	No		No	02/08/2012
Data Master One (DM-1)	No		No	02/08/2012

11b. Does the system receive SBU/PII from other federal agency or agencies? Yes

If **yes**, for each federal interface, identify the organization that sends the SBU/PII, how the SBU/PII is transmitted and if there is an Inter-Agency Agreement (ISA)/Memorandum of Understanding (MOU).

<u>Organization Name</u>	<u>Transmission method</u>	<u>ISA/MOU</u>
Social Security Administration	SSA Data Master-1	No
United States Postal Service	ZIP POD	No

11c. Does the system receive SBU/PII from State or local agency(s)? No

11d. Does the system receive SBU/PII from other sources? No

11e. Does the system receive SBU/PII from **Taxpayer** forms? Yes

If **yes**, identify the forms

<u>Form Number</u>	<u>Form Name</u>
1040	U.S. Individual Income Tax Return
940	Employer's Annual Federal Unemployment (FUTA) Tax Return
941	Employer's Quarterly Federal Tax Return
1120	U.S. Corporate Tax Return
1120S	U.S. Income Tax Return for an S Corporation
1065	U.S. Return of Partnership Income
706	U.S. Estate Tax Return
2141	Child and Dependent Care Expenses
5329	Additional Taxes on Qualified Plans (including IRAs) and other Tax Favored Accounts

11f. Does the system receive SBU/PII from **Employee** forms (such as the I-9)? No

F. PII SENT TO EXTERNAL ORGANIZATIONS

12. Does this system disseminate SBU/PII? Yes

12a. Does this system disseminate SBU/PII to other IRS Systems? Yes

If **yes**, identify the full name and acronym of the IRS system(s) that receive SBU/PII from this system.

<u>System Name</u>	<u>Current PIA?</u>	<u>PIA Approval Date</u>	<u>SA & A?</u>	<u>Authorization Date</u>
Enterprise Informatica Platform (EIP)	Yes	05/30/2013	No	
Business Objects Enterprise (BOE)	Yes	06/28/2013	No	

Identify the authority and for what purpose? IPM is a common, read-only data store (database) containing core IRS data (e.g., tax accounts, tax returns, and information returns) needed by a wide range of Tier B modernization projects to support case identification, selection, prioritization, delivery, compliance analysis and decision analytics. The purpose of the IPM is to provide one central relational database, populated with current and historical data from a wide variety of data sources. Client connections to IPM submit as a change request (CR). These client access requests are first pre-screened by Data Strategy Implementation (DSI), which may include meetings with the potential clients and informal discussions with the IPM team.

12b. Does this system disseminate SBU/PII to other Federal agencies? No

12c. Does this system disseminate SBU/PII to State and local agencies? No

12d. Does this system disseminate SBU/PII to IRS or Treasury contractors? No

12e. Does this system disseminate SBU/PII to other Sources? No

G. PRIVACY SENSITIVE TECHNOLOGY

13. Does this system use social media channels? No

14. Does this system use privacy-sensitive technologies such as mobile, global position system (GPS), biometrics, RFID, etc.? No
15. Does the system use cloud computing? No
16. Does this system/application interact with the public? No

H. INDIVIDUAL NOTICE AND CONSENT

17. Was/is notice provided to the individual prior to collection of information? Yes
- 17a. If **yes**, how is notice provided? Was the individual notified about the authority to collect the information, whether such is mandatory or voluntary, the purpose for which the information will be used, with whom the information may be shared, and the effects, if any, if they decide not to provide any of the requested information? This system does not collect any information directly from taxpayers. All information that is stored in IPM comes from the submission of tax returns submitted directly to the IRS through other internal IRS systems. Information from the tax returns is subsequently shared from the other internal applications to this system. The Tax return form provides taxpayers information regarding the opportunity to decline or consent to providing the information.
18. Do individuals have the opportunity to decline from providing information and/or from consenting to particular uses of the information? Yes
- 18a. If **yes**, describe the mechanism by which individuals indicate their consent choice(s): As previously indicated other internal IRS systems provide all data to IPM. Specifically, information from tax returns existing in other internal systems is subsequently shared by those systems and then ingested by IPM. The Tax return form provides taxpayers information regarding the opportunity to decline or consent to providing the information.
19. How does the system or business process ensure due process regarding information access, correction and redress? The system will allow affected parties the opportunity to clarify or dispute negative information that could be used against them. Due process is provided pursuant to 5 USC.

I. INFORMATION PROTECTION

20. Identify the owner and operator of the system (could be IRS owned and Operated; IRS owned, contractor operated; contractor owned and operated). IRS Owned and Operated

21. The following people have access to the system with the specified rights:

<u>IRS Employees?</u>	<u>Yes</u>	<u>Yes/No</u>	<u>Access Level (Read Only/Read Write/Administrator)</u>
Users	No		
Managers	Yes		Read-Only
Sys. Administrators	Yes		Read and Write
Developers	Yes		Read and Write

Contractor Employees? Yes

<u>Contractor Employees?</u>	Yes/No	Access Level	Background Invest. Level
Contractor Users	No		
Contractor Managers	No		
Contractor Sys. Admin.	Yes	Read and Write	High
Contractor Developers	Yes	Read and Write	High

21a. How is access to SBU/PII determined and by whom? Access can be requested through Online 5081. The perspective client applications must present a current SD-ATO/ATO and a Project Engagement Form (PEF) that details the project needs and the connection information. Anyone with a managerial approved Business need, with a current approved Live Data Waiver and managerial approved OL5081 can be granted access by BDA to the PII data on the IPM Development database. IPM Production access to PII data is limited to the IPM Database Administrators (read/write) and approved IPM client-applications (read only). It is the responsibility of the IPM client-application Project Manager and Database Administrator to keep the "Application System Account Credentials" secure.

21b. If computer matching occurs, can the business owner certify that it meets requirements of IRM 11.3.39 Disclosure of Official Information, Computer Matching & Privacy Protection Act?

Not Applicable

I.1 RECORDS RETENTION SCHEDULE

22. Are these records covered under the General Records Schedule (GRS), or have a National Archives and Records Administration (NARA) archivist approved a Record Control Schedule (RCS) for the retention and destruction of official agency records stored in this system? Yes

22a. If yes, how long are the records required to be held under the corresponding RCS and how are they disposed of? In your response, please include the complete IRM number 1.15.XX and specific item number and title. The IPM Greenplum database, stored on the BDA-GSS infrastructure, provides a single point of access to core taxpayer data (such as taxpayer accounts and tax returns) and other data derived from authoritative systems used by a wide range of IRS internal business applications to support case identification, selection, prioritization and research, compliance analysis, and the ability to identify trends and implement measures to address non-compliance. It is not the authoritative data source for data and documents and does not require National Archives and Records Administration (NARA) to effect data disposition. Data loaded into IPM Greenplum database from the various data sources listed in this questionnaire will be updated with current information from weekly, monthly or annual loads. Official records used for case selection to address non-compliance are disposed of in accordance with Records Control Schedule 23 for Examination. Records associated with tax processing functions at the Campuses are disposed of in accordance with Records Control Schedule 29 for Tax Administration - Wage and Investment Records.

I.2 SA&A OR ECM-R

23. Has the system been through SA&A (Security Assessment and Authorization) or ECM-R (Enterprise Continuous Monitoring Reauthorization)? Yes

23a. If yes, what date was it completed? 6/6/2017

23.1 Describe in detail the system's audit trail. Enterprise Security Audit Trails (ESAT) completed and signed the Greenplum Platform Audit Plan April 2014. Audit records will capture all of the required elements contained with the IRS IRMs such as account logon, activity of admin users, as well as failed logon attempts.

J. PRIVACY TESTING

24. Does the system require a System Test Plan? Yes

24b. If yes, Is the test plan in process or completed: Completed

24.3 If **completed/ or in process**, describe what testing and validation activities have been conducted or are in progress to verify and validate that the applicable Privacy Requirements (listed in header) have been met? Continuous Monitoring (eCM) (now called Annual Security Control Assessment (ASCA)) occurs annually to ensure that controls remain in place to properly safeguard PII.

24b.1. If **completed**, where are the test results stored (or documentation that validation has occurred confirming that requirements have been met)? DocIT

24b.2. If **completed**, were all the Privacy Requirements successfully tested? Yes

24.2 If **completed**, are there any residual system privacy, civil liberties, and/or security risks identified that need to be resolved? No

K. SBU Data Use

25. Does this system use, or plan to use SBU Data in Testing? Yes

25a. If yes, was permission granted per the requirements of Form 14664, SBU Data Use Questionnaire or Form 14665, SBU Data Use Request? Yes

If yes, provide the date the permission was granted. 10/19/2017

25b. If yes, was testing performed in conformance with IRM 10.5.8, Sensitive But Unclassified (SBU) Data Policy: Protecting SBU in Non-Production Environments? Yes

L. NUMBER AND CATEGORY OF PII RECORDS

26. Identify the number of individual records in the system for each category:

26a. IRS Employees: 50,000 to 100,000

26b. Contractors: 5,000 to 10,000

26c. Members of the Public: More than 1,000,000

26d. Other: No

M. CIVIL LIBERTIES

27. Does the system maintain any information describing how any individual exercises their rights guaranteed by the First Amendment? No

28. Is the system information used to conduct data-mining as defined in the *Implementing the 9/11 Commission Recommendations Act of 2007, Public Law 110-53, Section 804?* No

29. Will this system have the capability to identify, locate, and monitor individuals or groups of people? No

N. ACCOUNTING OF DISCLOSURES

30. Does the system include or require disclosure of tax or employee information to anyone other than IRS employees in the performance of their duties, or to the person to whom the information pertains or to a 3rd party pursuant to a Power of Attorney, tax or Privacy Act consent? No

End of Report
