

# Customer Account Data Engine (CADE) – 2 Database Implementation (DI) – Privacy Impact Assessment

PIA Approval Date – Oct. 4, 2010

## System Overview

Individual taxpayer accounts are currently contained in two systems: the Individual Master File (IMF) and the Customer Account Data Engine. As a result of this split, the Current CADE application must save data to both the Current CADE database, and return data to IMF in a process known as Return to Current Processing Environment (R2CPE). Additionally, both IMF and Current CADE must maintain connections with downstream systems, and both systems must be maintained every year with tax law updates.

CADE–2 DI Transition State 1 (TS1) will establish a single database that will house all individual taxpayer accounts, which will provide the ability for IRS employees to view the updated account information on–line, will allow for enhanced daily batch processing, and will enhance data security. Additionally, the key IRS customer service operational database, Integrated Data Retrieval System (IDRS), will have the benefit of more timely posted data. Further, CADE–2 DI will populate the Integrated Production Model (IPM) analytical data store to provide business users with tools to more effectively use the data for compliance and customer service. The data model and database design of Current CADE provided a foundation for the CADE–2 DI target database, which can be used to manage all individual taxpayer accounts. Current CADE infrastructure and models from the software architecture will also be used in later transition states.

The overall scope of the DI Project is to:

- Develop the end–state database data model that meets the business needs such as transaction posting, financial settlement, and post analysis
- Create and implement the database that supports both the database system functions (e.g. record, read, write, update, performance) and business requirements (e.g. data security, account viewing and reporting)
- Initialize the database with the Individual Master File (IMF) and Current CADE data and update active records from Daily Processing
- Provide data to both the Integrated Data Retrieval System (IDRS) and the Integrated Production Model (IPM)
- Provide a viewing capability for the taxpayer account data residing in the CADE 2 database via Corporate Files On Line (CFOL)/Individual Master File On–Line (IMFOL) commands

## Systems of Records Notice (SORN):

- CADE–2 DI is currently in transition state one and does not require a System of Records Notice (SORN).

## Data in the System

**1. Describe the information (data elements and fields) available in the system in the following categories:**

- A. Taxpayer – The data in CADE–2 DI includes all of the data from the IRS 1040, 1040A, 1040EZ, and 1040X forms and schedules. The data includes the following individually identifiable data elements:
  - Taxpayer Identification Number including:
    - Social Security Number (SSN)
    - Employer Identification Number (EIN)

- Taxpayer Identification Number (TIN)
- Taxpayer name
- Taxpayer address
- Taxpayer phone number
- Taxpayer account/routing number
- Taxpayer income
- Taxpayer liability

B. Employee – There is no employee data, outside of their individual taxpayer data, stored on CADE–2 DI.

C. Audit Trail Information – CADE–2 DI is currently in development and no audit plan is available. Therefore, audit trail information will be determined.

**2. Describe/identify which data elements are obtained from files, databases, individuals, or any other sources.**

A. IRS – Individual taxpayer accounts are currently contained in two systems: the IMF and CADE. CADE–2 DI TS1 will establish a single database that will house all individual taxpayer accounts. The CADE–2 DI TS1 database will be initialized with IMF and current CADE data and update active records from Daily Processing.

B. Taxpayer – The tax filing data is reported by the taxpayer on IRS forms 1040, 1040A, 1040X and schedules in the 1040 Individual Income Tax family of returns.

C. Employee – There is no data obtained from employees. The only data stored on CADE–2 DI are individual taxpayer data.

- Form 1040 U.S. Individual Income Tax Return:

- Full name
- SSN
- Spouse's full name
- Spouse's SSN
- Address
- Phone number
- Marital status
- Dependents' names
- Dependents' SSN
- Occupation
- Spouse's occupation
- Taxable income
- Routing number
- Account number
- Designee's name
- Designee's phone number
- Designee's PIN
- Paid preparer's name
- Paid preparer's SSN or PTIN
- Preparing firm's EIN
- Preparing firm's phone number

Form 1040A U.S. Individual Income Tax Return:

- Full name
  - SSN
  - Spouse's full name
  - Spouse's SSN
  - Address
  - Phone number
  - Dependents' names
  - Dependents' SSN
  - Marital status
  - Taxable income
  - Routing number
  - Account number
  - Designee's name
  - Designee's phone number
  - Designee's PIN
  - Occupation
  - Spouse's occupation
  - Paid preparer's name
  - Paid preparer's PTIN or SSN
  - Preparing firm's EIN
  - Preparing firm's address
  - Preparing firm's phone number
- Form 1040EZ U.S. Individual Income Tax Return:
    - Full name
    - SSN
    - Spouse's full name
    - Spouse's SSN
    - Address
    - Phone number
    - Marital status
    - Taxable income
    - Routing number
    - Account number
    - Designee's name
    - Designee's phone number
    - Designee's PIN
    - Occupation
    - Spouse's occupation
    - Paid preparer's name
    - Paid preparer's PTIN or SSN
    - Preparing firm's EIN
    - Preparing firm's address
    - Preparing firm's phone number

**3. Is each data item required for the business purpose of the system? Explain.**

Yes. The name, TIN, and address are required to identify the taxpayer's account. The income, deductions credits, other data reported on the tax return, as well as any payments received, are required to settle the taxpayer's account, and maintain a record of taxes assessed, abated and collected.

**4. How will each data item be verified for accuracy, timeliness, and completeness?**

The front–end input systems, listed in question #10 below, validate and correct data entries. The CADE–2 DI process reformats this information for processing and storage. The front–end systems adhere to the established (and published) input and posting cycles to ensure that returns received from the taxpayers are processed, validated and posted to the Individual Master File within required time frames.

**5. Is there another source for the data? Explain how that source is or is not used.**

No. There is no other source for the data.

**6. Generally, how will data be retrieved by the user?**

CADE–2 DI data is not directly accessible by users. Data is extracted on a daily basis and loaded into other IRS systems, which provide read–only access to the data. These other systems provide the authorization, control and monitoring of access and IDRS through command codes.

**7. Is the data retrievable by a personal identifier such as name, SSN, or other unique identifier?**

Yes. Data is retrieved by name, SSN, and TIN.

**Access to the Data**

**8. Who will have access to the data in the system (Users, Managers, System Administrators, Developers, Others)?**

The only access to the data is to execute batch files and receive data extracts; there is no user interface and no direct users.

**Role:** System Administrators

**Permission:** Only authorized IRS system administrators run the batch files.

**Role:** Developers

**Permission:** May be granted temporary/emergency access on an as–needed basis.

Authorization is required via the Online (OL) 5081 process and accounts are automatically terminated after six hours. Temporary accounts are reviewed and authorized via the OL5081 process. Temporary access could include write, read, and execute on the CADE–2 DI application if required and authorized by the OL5081 process.

*.Note: Contractors do not have access to the application.*

**9. How is access to the data by a user determined and by whom?**

The only access is to execute batch files and receive data extracts; there is no user interface and no direct users. Data access is granted on a need to know basis. Potential System Administrators and Database Administrators must submit requests for access through the OL5081 application which is submitted to their local management for approval consideration. Administrators are not permitted access without a signed OL5081 application from an authorized management official.

**10. Do other IRS systems provide, receive, or share data in the system? If YES, list the system(s) and describe which data is shared.**

Yes. Individual taxpayer accounts are currently contained in two systems: the Individual Master File and Custer Account Data Engine. CADE–2 DI TS1 will establish a single database that will house all individual taxpayer accounts. The CADE–2 DI TS1 database will be initialized with IMF and current CADE data and update active records from Daily Processing. CADE–2 DI TS1 will provide taxpayer account data extracts to Integrated Data Retrieval System and Integrated Production Model.

Additionally, CADE–2 DI TS1 will provide a capability to view the taxpayer account data stored in the CADE–2 database via IMFOL commands.

- Customer Account Data Engine (CADE) – Taxpayer and tax return information
- Integrated Data Retrieval System (IDRS) – Taxpayer and tax return information
- Individual Master File (IMF) – Taxpayer and tax return information
- Individual Master File On–Line (IMFOL) – Taxpayer and tax return information
- Integrated Production Model (IPM) – Taxpayer and tax return information

## **11. Have the IRS systems described in Item 10 received an approved Security Certification and Privacy Impact Assessment?**

Customer Account Data Engine (CADE)

- Authorization to Operate (ATO) – April 2, 2009
- Privacy Impact Assessment (PIA) – April 30, 2008

Integrated Data Retrieval System (IDRS)

- Authorization to Operate (ATO) – March 10, 2009
- Privacy Impact Assessment (PIA) – November 6, 2008

Individual Master File (IMF)

- Authorization to Operate (ATO) – June 21, 2007
- Privacy Impact Assessment (PIA) – June 7, 2007

Individual Master File On–Line (IMFOL)

- Authorization to Operate (ATO) – June 21, 2007
- Privacy Impact Assessment (PIA) – June 7, 2007

Integrated Production Model (IPM)

- Authorization to Operate (ATO) – August 4, 2008
- Privacy Impact Assessment (PIA) – September 12, 2008

## **12. Will other agencies provide, receive, or share data in any form with this system?**

No other agencies will provide, receive, or share data in any form with CADE–2 DI TS1.

### **Administrative Controls of Data**

## **13. What are the procedures for eliminating the data at the end of the retention period?**

CADE–2 is unscheduled. A request for records disposition authority for CADE–2 and associated records is currently being drafted with the assistance of the IRS Records and Information Management (RIM) Program Office. When approved by the National Archives and Records Administration (NARA), disposition instructions for CADE–2 inputs, system data, outputs, and system documentation will be published under IRM 1.15.19 Records Control Schedule for the Enterprise Computing Center – Martinsburg, item number to be determined.

## **14. Will this system use technology in a new way?**

Yes. This is the first time a relational management database will be used to manage the flat files that comprise the legacy IMF environment.

**15. Will this system be used to identify or locate individuals or groups? If so, describe the business purpose for this capability.**

Yes. The purpose of the system is to identify individuals based on specific criteria for special processing such as collections.

**16. Will this system provide the capability to monitor individuals or groups? If yes, describe the business purpose for this capability and the controls established to prevent unauthorized monitoring.**

No. This system does not monitor individuals or groups. Other systems extract data from this repository. Extracts are performed only based on an approved request for information services.

**17. Can use of the system allow IRS to treat taxpayers, employees, or others, differently?**

No. The CADE-2 DI data allows IRS to ensure that taxpayers are treated fairly and equitably. The automated method of generating notices based on specific criteria eliminates the possibility of one taxpayer receiving preferential treatment over another. The CADE-2 DI also enables all taxpayers to receive the notifications that are applicable to them when there is a change in tax law, such as the recent changes affecting the child tax credit.

**18. Does the system ensure "due process" by allowing affected parties to respond to any negative determination, prior to final action?**

No. CADE-2 transition state 1 is a record repository and does not have due process function.

**19. If the system is web-based, does it use persistent cookies or other tracking devices to identify web visitors?**

No. The system is not web-based.

[View other PIAs on IRS.gov](#)