

Compliance Data Warehouse (CDW) – Privacy Impact Assessment

PIA Approval Date – Feb. 23, 2011

System Overview:

The Compliance Data Warehouse (CDW) provides access to a wide variety of tax return, enforcement, compliance, and other data to support the query and analysis needs of the Research community. CDW provides a range of solutions to users, including data integration, computing services, data transfer, and educational services. CDW captures data from multiple production systems, migrating the data to the CDW environment, transforming, standardizing, and augmenting the data, and organizing the data in a way that is conducive to analysis.

Systems of Records Notice (SORN):

- IRS 22.034--Individual Returns Files, Adjustments and Miscellaneous Documents
- IRS 22.054--Subsidiary Accounting Files
- IRS 22.060--Automated Non-Master File (ANMF)
- IRS 22.062--Electronic Filing Records
- IRS 24.030--CADE Individual Master File (IMF)
- IRS 24.046--CADE Business Master File (BMF)
- IRS 26.020--Taxpayer Delinquency Investigation (TDI) Files
- IRS 34.037--IRS Audit Trail and Security Records System
- IRS 42.008--Audit Information Management System (AIMS)
- IRS 42.021--Compliance Programs and Projects Files

Data in the System

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

- A. Taxpayer data are extracted from filed tax returns, enforcement information, and narrative data that sequence taxpayer history.
- B. Employee data (as taxpayers) may be included as in A above, in addition to Standard Employee Identification Number (SEID); and, CDW account and logon information.
- C. Audit Trail Information is handled by RAS staff and contractor systems administrators (SAs) and contractor database administrators (DBAs) when making periodic activity log file checks to ensure normal system functioning and to look for warning messages that might indicate a problem. The activity log contains:
 - Logon/logoff by USER Identification (ID)
 - Password change action
 - Date and time of event
 - Success or failure of event
 - All other actions by SAs & DBAs

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

- A. IRS:

- Audit Information Management System Reference (AIMS–R) (formerly AIMS) – EXAM Division:
 - Location
 - Age
 - Status of returns in Examination
- Account Receivable Dollar Inventory (ARDI) – IMF and Business Master File (BMF) Masterfile (MF): Entities and tax modules that show a debit balance on the Master File
- Automated Underreporter (AUR) – IMF MF: The AUR program matches information returns against individual income tax returns to verify that income and deductions are correctly reported
- Business Master File (BMF) – BMF MF:
 - Collection Status Codes
 - Transaction Codes (TC) for business taxpayers
 - TC 590 File
- Business Returns Transaction File (BRTF) – BMF MF: Business entity and tax return data for corporations, partnerships, trusts, and other entities
- Collection Information (COLL) – BMF/IMF MF and Small Business/Self–Employed (SB/SE) Automated Collection System (ACS): Collection information to support analysis and delivery of more productive inventory
- Data Master–1 (DM–1) – National Account Profile Program – IMF and BMF Masterfile:
 - Date of Birth (DOB)
 - Date of Death (DOD)
 - Gender
 - Citizenship
 - Name Control Information for all issued Social Security Numbers (SSN)
 - Number of Name Controls for each SSN
- Earned Income Tax Credit (EITC) – IMF MF:
 - Extract used to report information about qualifying children for those claiming the Earned Income Tax Credit
 - In addition to line items from the Form 1040 and Schedule earned income credit (EIC), this table also includes Master File transaction codes and information from other 1040 schedules (e.g., Schedules A, C and F)
- Examination Operational Automation Database (EOAD) – EXAM Division – IMF and BMF Masterfile: The EOAD allows for data–driven operational analysis. It identifies issues and causality. It can be used to identify productive audit issues, improve training requirements, and improve resource usage
- Enforcement Revenue Information System (ERIS) – SB/SE COLLECTION DIVISION:
 - Enforcement activity on Appeals, Exam, and Collection cases since FY 1992
 - Information Reporting Program (IRP)/AUR closings since FY 1994
 - penalty amounts by Internal Revenue Code and Reference Number since FY 1996

- EITC math error information was added to existing enforcement cases since calendar year 1999
- Electronic Tax Administration Marketing Database (ETA MDB) – ETA DIVISION: National database to profile individual and business return filers to support marketing and communications for e–submissions programs. Wage and Investment's (W&I's) Stakeholder Partnerships, Education and Communication (SPEC) organization and SBSE's Taxpayer Education & Communication (TEC) organization periodically provide spreadsheets. The Census Bureau provides the North American Industry Classification System codes
- Individual Master File (IMF) – IMF MF:
 - Collection Status Codes
 - IMF Transaction Code (TC) 150 information
 - IMF Transaction History
- Information Returns Master File (IRMF) – IMF MF: Third–party information documents. It includes Forms W–2, K–1, 1098, 1099, 103, 5498, 8300, and PASSPORT. These documents are used by AUR, nonfiler, and other programs
- Individual Returns Transaction File (IRTF) – IMF MF:
 - Transcribed tax returns for individuals, and include most Forms and Schedules.
 - Entity
 - Tax module
 - Other information associated with taxpayer accounts
- Notice Delivery System (NDS) – SUBMISSION PROCESSING (Service Centers):
 - Taxpayer notice data
 - Identifies a notice including physical stations for assembling and details about duplicate/culls during the printing process
 - Information specific to the inserts used to assemble a notice
- National Research Program (NRP):
 - Compliance study focused on Tax Year (TY) 2001 Form 1040 returns.
 - The second NRP study is for TY 2003 and TY 2004 Form 1120S returns
- United States Postal Service (USPS) ZIP Codes:
 - List of USPS ZIP codes
 - City
 - County
 - Post office names associated with those ZIP codes
 - Data Tables
 - Data Dictionary
 - Field–related ID's
 - OL5081 passwords
 - Data parameters
 - CDW account/login
 - ALL DataBases/Sources and their descriptions need to be returned to this section

B. Taxpayer

- Taxpayer Identification Number (TIN)
- SSN
- Unique Identifiers in lieu of TINs/SSNs
- Financial institutions
- Sources of income and expenses
- Business related information
- All other tax return data

C. Employee

- Data sources used
- SEID
- Field-related IDs
- CDW account/login

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

Yes. Data in CDW are used to support a wide range of ad hoc requests (for example requests from the Commissioner or Congress) unpredictable research queries, requests involving prediction, simulation, optimization, and sampling, by our customers & stakeholders.

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

The multiple databases on CDW contain one or more tables with data elements common across databases.

- Accuracy – Tables are frequently matched for inaccuracies that can occur during data extraction, transformation or loading.
- Timeliness – Data are updated according to the source life cycle and current year filings.
- Completeness – User-driven to ensure no response if fields contain inaccurate information.

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

No. There are no other sources.

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

Data is typically retrieved after users are given CDW access –based on established parameters – by query via third-party client tools, such as SQL, SAS, and Hyperion.

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

Yes, data is retrievable by personal identifiers, such as SSNs, TINs, Industry Codes and any other data element on CDW. However, access to personal identifiers minimally require, a bona fide business need, two-levels of managerial signatures, and approval by a RAS security official.

Access to the Data

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

IRS research analysts, managers, Dept of Treasury personnel, security-cleared contractors (developers & administrators), GAO, and Treasury Inspector General for Tax Administration (TIGTA) personnel in conjunction with: 1) need to know 2) proper approvals 3) security training 4) meeting all applicable policies, standards, investigations, procedures & safeguards in place.

Role: Treasury & GAO

Permission: use data to respond to Congress

Role: Contractors

Permission: achieve deliverables stated in their contracts

Role: IRS managers

Permission: lead projects and studies requiring data

Role: IRS users

Permission: query databases for solutions to the projects and studies

Role: System administrators

Permission: manage the hardware that house the data

Role: Developers

Permission: manage the software which communicates with the database

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

CDW is not accessible by the public; therefore requests for access are made only through internal (electronic or paper) channels with the applicable signatory requirements. All levels of data access are limited to what is specified on the approved request; and, by those parameters established through assigned rights and privileges.

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.

- Audit Information Management System Reference (AIMS–R) (formerly AIMS)
- Automated Underreporter (AUR)
- Business Master File (BMF)
- Enforcement Revenue Information System (ERIS)
- Individual Master File (IMF)
- Electronic Tax Administration Marketing Database (ETA–MDB)
- Notice Delivery System (NDS)
- National Research Program (NRP)

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

Audit Information Management System Reference (AIMS–R) (formerly AIMS)

- Authority to Operate (ATO) – May 1, 2009
- Privacy Impact Assessment (PIA) – February 11, 2009

Automated Underreporter (AUR)

- Authority to Operate (ATO) – May 6, 2009
- Privacy Impact Assessment (PIA) – February 27, 2009

Business Master File (BMF)

- Authority to Operate (ATO) – June 14, 2010
- Privacy Impact Assessment (PIA) – Marcy 16, 2010

Enforcement Revenue Information System (ERIS)

- Authority to Operate (ATO) – May 7, 2009
- Privacy Impact Assessment (PIA) – March 5, 2009

Individual Master File (IMF)

- Authority to Operate (ATO) – March 8, 2010
- Privacy Impact Assessment (PIA) – November 10, 2009

Electronic Tax Administration Marketing Database (ETA–MDB)

- Authority to Operate (ATO) – May 22, 2009
- Privacy Impact Assessment (PIA) – December 23, 2008

Notice Delivery System (NDS)

- Authority to Operate (ATO) – May 3, 2010
- Privacy Impact Assessment (PIA) – March 29, 2010

National Research Program (NRP)

- Authority to Operate (ATO) – May 21, 2008
- Privacy Impact Assessment (PIA) – April 7, 2008

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

No other agencies provide, receive, or share data in any form with CDW.

Administrative Controls of Data

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

CDW data is approved for destruction 10 years after end of the Processing Year or when no longer needed for operational purposes, whichever is later (Job No. N1–58–10–7). Retention requirements for CDW inputs, outputs and system documentation are also stipulated under that NARA–approved authority. When next updated, CDW disposition instructions will be published as item 54 under IRM 1.15.27 Records Control Schedule for Compliance Research. All records housed in the system will be erased or purged from the system at the conclusion of their retention period(s) as required under IRM 1.15.6.

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

No. CDW will not use technology in a new way.

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

Yes. CDW is used to perform research studies that may identify or predict taxpayer's non-compliance; to evaluate the impact of program or policy changes; or, to develop workload models that optimize the use of resources. Other identified groupings may include locating and identifying taxpayers; for example, those affected by Hurricane Katrina, Tsunami's, et al – by geographical location.

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.

Yes, CDW has monitoring capability only as far as using zip codes, SSNs, TINs, or the unique identifier used in lieu of the actual TIN/SSN, to track data elements. Other business purposes may

include: impact analyses on program changes; taxpayer filing activities; predictive modelling; longitudinal surveys; and, other research activities. Controls established to prevent unauthorized access and/or monitoring are based on permissions assigned and properly approved via the Form 5081 application, as well as annual recertification for security, privacy, disclosure, Sensitive But Unclassified (SBU), Official Use Only (OUO), and other personally identifiable information.

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

No, disparate treatment of taxpayers, employees, or others is not probable since CDW provides “read only” data. No taxpayer case selections are conducted. Compliance treatments, for purposes of estimation, prediction, simulation, optimization, and other statistical activities, may be used to result in new methods or approaches of tax administration.

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

No, CDW is read-only. Users can not make account changes or decisions on any taxpayer-type information. CDW only provides data needed to effect program/policy or treatment-based decisions. Should the taxpayer file subsequent corrected returns, or make any other adjustments, the resulting return data is eventually appended to the CDW system, but never replaced.

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

No, CDW is a web-based hybrid with no public access; therefore, it does not have visitors or have need for persistent cookies or other tracking devices.

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