



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

September 27, 2012

Reference Number: 2012-20-109

This report has cleared the Treasury Inspector General for Tax Administration disclosure review process and information determined to be restricted from public release has been redacted from this document

Phone Number | 202-622-6500

E-mail Address | TIGTACommunications@tigta.treas.gov

Website | <http://www.tigta.gov>



HIGHLIGHTS

THE CUSTOMER ACCOUNT DATA ENGINE 2 DATABASE WAS INITIALIZED; HOWEVER, DATABASE AND SECURITY RISKS REMAIN, AND INITIAL TIMEFRAMES TO PROVIDE DATA TO THREE DOWNSTREAM SYSTEMS MAY NOT BE MET

Highlights

Final Report issued on September 27, 2012

Highlights of Reference Number: 2012-20-109 to the Internal Revenue Service Chief Technology Officer.

IMPACT ON TAXPAYERS

The overall goals for the Customer Account Data Engine 2 (CADE 2) Program are to process individual taxpayer account data in a modernized environment and provide more timely and accurate data to front-line employees. A transactional database capable of supporting both tax processing and enterprise-wide data access is a cornerstone of that effort. In Transition State 1, the IRS will establish the database and processes will be developed to keep the database current with daily account information from the Individual Master File. The database will be able to provide daily updates to the IRS's key customer service database, the Integrated Data Retrieval System, and it will be able to populate the key compliance analytical database, the Integrated Production Model, with more timely data. Incomplete, inaccurate, and unsecured data on the CADE 2 database will prevent the IRS from providing quality customer service and could compromise taxpayer data.

WHY TIGTA DID THE AUDIT

The overall objective was to review the CADE 2 database implementation and ensure that the database was secure, accurate, and complete, and that prior weaknesses identified were corrected or mitigated. This review addresses the major management challenge of Modernization.

WHAT TIGTA FOUND

Our review determined that data integrity testing did not provide assurance that CADE 2 database data are consistently accurate and complete. Also, the CADE 2 database design has not fully met initialization, daily update, and downstream interface needs.

To address the issues identified during testing, the IRS developed version 2.2 of the CADE 2 database. The IRS spent up to \$22.3 million on database implementation including developing version 2.2 of the CADE 2 database from January through July 2012. The IRS does not track cost at the development activity level; therefore, TIGTA could not determine the actual cost for version 2.2 of the CADE 2 database.

Enhanced security is one of the goals of the CADE 2 Program. CADE 2 database security will be implemented via a role-based access model and the Resource Access Control Facility. However, vulnerabilities in the JAVA code could result in loss of sensitive taxpayer information, and remediation of identified security weaknesses is ineffective.

WHAT TIGTA RECOMMENDED

TIGTA recommended that the Chief Technology Officer: 1) ensure the CADE 2 Program does not exit Transition State 1 until the CADE 2 database can provide accurate and complete data to the three downstream systems; 2) ensure the database design process follows the Internal Revenue Manual and validate that the database design meets business requirements; 3) realign data validation and testing efforts with business functionality and processes; 4) ensure JAVA code weaknesses are remediated; 5) ensure privileged accounts are documented, administered, monitored, and reviewed in accordance with the Internal Revenue Manual or removed from the system; 6) ensure sample tables and default ports are disabled or removed; and 7) enhance the Online 5081 system.

The IRS agreed with three and partially agreed with one of the seven recommendations and corrective actions are planned. The IRS disagreed with three recommendations and TIGTA provided comments in the audit report.



TREASURY INSPECTOR GENERAL
FOR TAX ADMINISTRATION

DEPARTMENT OF THE TREASURY
WASHINGTON, D.C. 20220

September 27, 2012

MEMORANDUM FOR CHIEF TECHNOLOGY OFFICER

A handwritten signature in black ink, appearing to read "Michael E. McKenney".

FROM: Michael E. McKenney
Acting Deputy Inspector General for Audit

SUBJECT: Final Audit Report – The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met (Audit # 201220023)

This report presents the results of our review of the Customer Account Data Engine 2 Database Implementation to ensure that the database was secure, accurate, and complete, and that prior weaknesses identified were corrected or mitigated. This review addresses the major management challenge of Modernization.

Management's complete response to the draft report is included in Appendix IV.

Copies of this report are also being sent to the Internal Revenue Service managers affected by the report recommendations. Please contact me at (202) 622-6510 if you have questions or Alan R. Duncan, Assistant Inspector General for Audit (Security and Information Technology Services), at (202) 622-5894.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Table of Contents

Background	Page 1
Results of Review	Page 3
Design Issues Are Jeopardizing the Ability of the Customer Account Data Engine 2 Database to Serve As a Trusted Source of Data.....	Page 3
<u>Recommendation 1:</u>	Page 7
<u>Recommendation 2:</u>	Page 8
<u>Recommendation 3:</u>	Page 9
Security Weaknesses and Poor Coding Practices in the Customer Account Data Engine 2 Database Could Result in the Loss of Taxpayer Data.....	Page 10
<u>Recommendation 4:</u>	Page 13
<u>Recommendations 5 through 7:</u>	Page 14
Appendices	
Appendix I – Detailed Objective, Scope, and Methodology	Page 16
Appendix II – Major Contributors to This Report	Page 18
Appendix III – Report Distribution List	Page 19
Appendix IV – Management’s Response to the Draft Report	Page 20



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Abbreviations

CADE 2	Customer Account Data Engine 2
ETL	Extract, Transform, and Load
IDRS	Integrated Data Retrieval System
IMF	Individual Master File
IRM	Internal Revenue Manual
IRS	Internal Revenue Service
TIF	Taxpayer Information File
TS1	Transition State 1



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Background

The Customer Account Data Engine 2 (CADE 2) Program is the top information technology modernization project in the Internal Revenue Service (IRS). The CADE 2 mission is to provide state-of-the-art individual taxpayer account processing and data-centric technologies to improve service to taxpayers and enhance IRS tax administration. CADE 2 will replace the current Individual Master File (IMF) account settlement system with a modernized, relational database processing system and become a key component in the IRS’s enterprise-wide, data-centric information technology strategy. Figure 1 provides the CADE 2 system implementation phases.

Figure 1: CADE 2 System Implementation Phases

Phase	Description
Transition State 1 (TS1)	The IRS will establish a single database that will store all individual taxpayer accounts. Processing will be enhanced to include daily batch processing. The key IRS customer service operational database, the Integrated Data Retrieval System (IDRS), will have the benefit of more timely posted data. The solution will populate the Integrated Production Model analytical data store and provide business users with tools to more effectively use the data for compliance and customer service. Enhanced data security will be in place. Downstream systems that must be modified to support daily processing are included in the scope of the TS1.
Transition State 2	A single processing system will be implemented. Applications will directly access and update the taxpayer account database, and continued efforts will be made in addressing existing financial material weaknesses. The IRS planned to implement Transition State 2 in January 2014. This date is no longer viable, due to funding delays, but a new date has not been determined.
Target State	Implement a single system in which all transitional applications are eliminated. The complete solution is also planned to address all the financial material weaknesses. As of April 3, 2012, the IRS had not established a Target State implementation date.

Source: The CADE 2 Program Charter and meetings with the CADE 2 executives.

TS1 will move the IRS away from operating in two tax processing environments – the IMF and Current CADE – towards a single system for managing individual taxpayer accounts. It has two major implementation pieces: Daily Processing and Database Implementation. Daily Processing, which uses IMF files and not the CADE 2 database, went into production in



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

January 2012. IRS management stated that this has resulted in providing faster refunds for millions of taxpayers and that posted information was viewable on the IDRS within 48 hours of processing.

Database Implementation, which is the subject of this audit, is in the final testing stage for version 2.2, which is expected to be placed into production in late 2012. Version 2.1 of the database was initialized earlier in Calendar Year 2012. IRS management stated that this earlier version of the database successfully initialized 270 million individual taxpayer accounts and more than a billion tax modules while balancing to the penny.

Within TS1, the primary deliverable of the CADE 2 Database Implementation project is a relational database that will store individual taxpayer account data, currently being processed by the IMF. This database will serve as the trusted source of data for three critical downstream systems: Corporate File On-Line/Individual Master File On-Line, IDRS Taxpayer Information File (TIF), and the Integrated Production Model. In Transition State 2, the CADE 2 database will become the sole source of IMF data and become the system of record for individual tax account processing, as the IMF entity and tax module files will be retired.

This audit reviewed the steps taken by the IRS to prepare for the CADE 2 Database Implementation and examined the process for addressing weaknesses and issues within TS1. This review was performed at the IRS Information Technology headquarters office in New Carrollton, Maryland, during the period February through June 2012. We conducted this performance audit in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objective. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objective. Detailed information on our audit objective, scope, and methodology is presented in Appendix I. Major contributors to the report are listed in Appendix II.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Results of Review

Design Issues Are Jeopardizing the Ability of the Customer Account Data Engine 2 Database to Serve As a Trusted Source of Data

The CADE 2 database is the cornerstone for all CADE 2 system development. One of the primary goals of the CADE 2 database is for it to be a trusted source of data. To provide this, it needs a stable design built to support tax processing functions and the assurance of complete and accurate data. Without these, the CADE 2 Program will not be successful. In TS1, the CADE 2 database will be initialized with IMF entity and tax module data, updated on a daily basis to keep it synchronized with the IMF files, and serve as a trusted source of data for selected downstream systems. The database must, therefore, provide sufficient evidence that its data are accurate and complete, and that it will provide the design needed for continued data reliability. This will be critical when CADE 2 transitions into a transactional database processing system and the system of record for all individual taxpayer accounts.

Our review of the CADE 2 Database Implementation project determined that weaknesses were found in the data validation process and the database design was not fully validated against business needs.

Data integrity testing did not provide assurance that CADE 2 data are consistently accurate and complete

The IRS cannot ensure the data on the CADE 2 database are consistently accurate and complete despite current control procedures and data integrity testing efforts. The Internal Revenue Manual (IRM) defines data controls as activities or tasks employed to preserve the accuracy of data by either deleting, detecting, or preventing operator errors, and by providing assurances that data are not lost, added, or inadvertently changed.¹ The IRM also provides that testing be conducted to ensure system components are free of logic and design errors, and customer requirements are satisfied.² In addition, the IRM requires that business requirements and business functions be fully documented during the business analysis phase.³

The IRS has data integrity checks in place at several levels of the CADE 2 database: data field level, record level, account and file level, and Master File level. Figure 2 summarizes the data integrity checks performed at each of these levels.

¹ IRM 2.5.3, *Systems Development - Programing Techniques and Source Code Standards*.

² IRM 2.6.1, *Product Assurance - Test, Assurance and Documentation*.

³ IRM 2.5.13, *Systems Development - Database Design Techniques and Deliverables*.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Figure 2: Types of Data Validation Performed on the CADE 2 Database

Control Level	Approach	Description
Data Field Level	Data Integrity Validation Approach	Validates data values between the CADE 2 database and the IMF. Uses a combination of manual and systemic data compares and data transformation rule validation.
Record Level	Database Referential Integrity Checks	Ensures that every record inserted into a table has a valid relationship to an existing account on the database.
Account and File Level	Balance and Control Procedures	Checks counts and amounts between the IMF source files and the CADE 2 database. Record counts and module balance amounts are checked through the use of control records during the Extract, Transform, and Load (ETL) process.
Master File Level	Database Implementation Simplified Financial Report	Balances to the IMF Recap Report.

Source: Treasury Inspector General for Tax Administration analysis of IRS documents.

During the first initialization of the CADE 2 database using version 2.1.1 of the data model, documentation showed that data validation efforts were adequate at the record level, account and file level, and Master File level. Referential integrity was maintained within the database, file control records were balanced to the CADE 2 database counts and amounts, and total financial assessments, credits, and debits balanced to the IMF Recap Report. However, at the data field level, the Data Integrity Validation Approach did not provide assurance that all the data values loaded into the CADE 2 database were accurate and complete. This was due to the complexity of many of the data transformation rules and embedded business logic contained within IMF data fields.

Manual and systemic data comparisons, when combined, validated approximately 70 percent of the data columns on the CADE 2 database against their IMF source values. Systems Acceptability Testing tested the remaining 30 percent through data transformation rule tests. The IRS acknowledged that these tests could not ensure the accuracy of the remaining 30 percent because these tests were limited and did not cover all variations or conditions of transformation logic.

Further, in a May 2012 meeting, IRS management acknowledged that not all variations or conditions in these complex data transformations had been identified. The IRS CADE 2 Full



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Data Coverage Mapping document mapped data elements to subject areas within the CADE 2 database, but it did not map data elements to core IMF business functions or processes. Thus, the IRS could not identify data elements that supported some business functions or determine where business logic may have been embedded in IMF data elements. As a result, the IRS is encountering unanticipated data values and hidden business logic during the database load and update process. Without a complete list of all data elements, values, and business processes, the IRS could not design an adequate strategy for data validation at the data field level.

The CADE 2 database will contain data elements from the IMF entity and tax module files. However, additional data elements necessary for the IDRS TIF will be loaded into the CADE 2 database during the daily update process. This TIF data will include notice data, which will be used to address existing financial material weaknesses. To validate this data, the IRS intends to use an IDRS TIF comparison tool to compare the data extracted from the CADE 2 database to the data currently being sent to the IDRS TIF by the IMF. Development of the tool was not completed as of May 2012; therefore, we could not verify the effectiveness of this planned data validation effort.

Although the IRS designed a fairly comprehensive strategy to check the data integrity of the CADE 2 database, it did not conduct a proper Business Analysis to align IMF data elements with business processes and business requirements before attempting to initialize and update the CADE 2 database. It is therefore impossible for the IRS to verify that the data transformation rules used to load and update the database are complete and that all embedded business logic and system conditions contained in IMF data fields are accounted for and tested. Without a documented inventory of business processes and their supporting data elements, the IRS cannot verify the accuracy and completeness of the CADE 2 database. The database should not be used as a trusted source of data until a method to validate the accuracy of the data is developed.

The CADE 2 database design has not fully met initialization, daily update, and downstream interface needs

A logical data model defines the structures of the data for a database. The logical data model is designed from data requirements that support a set of business processes derived from business requirements. The CADE 2 database was designed using the IMF's DB2 database, the Current CADE database, and the IMF's core record layouts. In TS1, the goal was to initialize and update the CADE 2 database on a daily basis with data from the IMF system. The CADE 2 data migration was performed through the ETL process, during which data were extracted from the source IMF system, transformed to fit into the destination CADE 2 database, and loaded into the CADE 2 database. The ETL process used rules and functions to transform the IMF source data into the data loaded into the CADE 2 database. Transformation rules and functions should be developed from an analysis of business functions, business processes, and business requirements.

During initialization of the CADE 2 database in January 2012, the IRS discovered that the Taxpayer Delinquent Account data field contained embedded business logic and was being used



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

for more than one purpose and for more than one business process. For years, the IRS policy has allowed changes to the IMF data structures only once a year; therefore, the IMF developers would use existing bits and/or bytes of the IMF data structures in order to support new business requirements. This practice of using embedded business logic was not always documented, and in this instance led to programming issues during the ETL process. The programming issue forced the IRS to redesign the database model. The IRS had recorded the Taxpayer Delinquent Account data field issue in October 2011. However, the database initialization phase proceeded with version 2.1.1 of the data model without any remedy for the issue. Proceeding with the database initialization was not in accordance with an independent contractor's recommendation that stated: "Database implementation teams do not compromise quality for the sake of hitting the schedule as this is likely to result in more painful re-work in the future."

The daily update testing, performed by the IRS in February 2012, revealed two other business requirements that were not accounted for in the database design. The first was processing where the IRS overlays data in an original transaction when that transaction is reversed. To accommodate the recording of the reversal, the IRS had to create a history table on the CADE 2 database. The second missed requirement dealt with taxpayer account merges. The database had to accommodate the situation where the Social Security Number of the taxpayer account changed, whether it was due to identity theft or other circumstances. The unique key⁴ used on the IMF sequential files could not be used on the CADE 2 relational database because a piece of the unique key had changed. This impacted database indexes and referential integrity checks. The IRS had to redesign the database with a new unique key that would not be impacted by account merges.

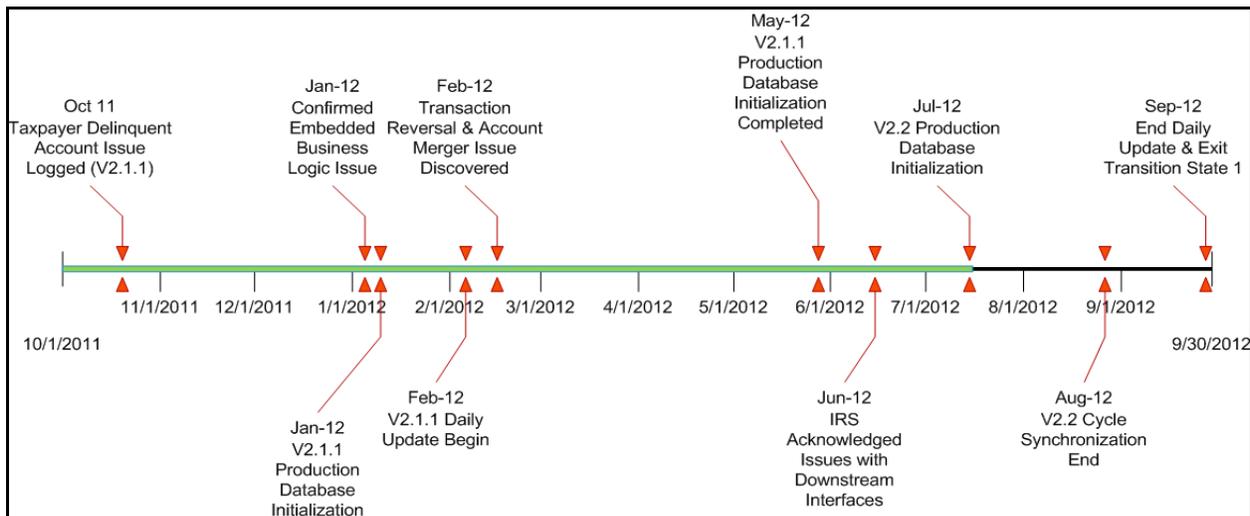
During a Program Management Office meeting in June 2012, the IRS acknowledged that it was having problems with its CADE 2 database interface to the IDRS TIF. The CADE 2 program's architecture solution planned to re-use the existing IMF to IDRS TIF interface for the CADE 2 database interface to IDRS TIF. However, the data types being extracted and sent from the CADE 2 database were not what the IDRS TIF system was expecting. Zeroes, blanks, and null values were not being transformed correctly; therefore, the IDRS TIF could not process the incoming CADE 2 database data successfully. As a result, the IRS is re-evaluating its data strategy for feeding downstream systems and is considering delaying the IDRS TIF interface. Figure 3 presents a partial CADE 2 database development timeline.

⁴ The IMF unique key is a combination of the Taxpayer Identification Number, the type of tax, and the tax year.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Figure 3: Partial CADE 2 Database Development Timeline



Source: The CADE 2 Integrated Master Schedule dated May 30, 2012, other IRS documentation, and interviews of IRS personnel.

To address the issues identified during testing, the IRS developed version 2.2 of the CADE 2 database. The IRS spent up to \$22.3 million on database implementation including developing version 2.2 of the CADE 2 database from January through July 2012. The IRS does not track cost at the development activity level. Therefore, we could not determine the actual cost for version 2.2 of the CADE 2 database. These costs could have been avoided by properly identifying the business requirements up front and including these requirements in the original design.

Recommendations

The Chief Technology Officer should:

Recommendation 1: Ensure that the CADE 2 Program does not exit TS1 until the CADE 2 database can provide accurate and complete data to the IDRS TIF, Corporate File On-Line/Individual Master File On-Line, and to the re-evaluated Integrated Production Model.

Management's Response: The IRS agreed with the recommendation. With appropriate approvals from the CADE 2 governance committee, the IRS plans to exit Milestone 5 according to schedule in September 2012, in order to deploy planned Corporate Files On-Line/Individual Master File On-Line and Integrated Production Model Reports functionality. The milestone exit will be conditional, however, until such time as the IRS has deployed planned IDRS TIF functionality, which will be done upon completion of the 2013 Filing Season peak.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Recommendation 2: Ensure that the database design process follows the IRM and validate that the database design meets business requirements.

Management's Response: The IRS disagreed with this recommendation. IRS management stated that the database design approach is extremely sound and meets the IRM standard. It leverages the legacy IMF, which has undergone years of refinement and embodies business requirements that are complete and accurate. The database fully supports the business requirements, as the CADE 2 data model was built using historical lessons learned from previous successes in the CADE and production data from the IMF. An added layer of confidence to the IRS approach was gained by running through transformations using real taxpayer data, which further proved out that the data model and database design was very strong. The CADE 2 data model design approach provided an in-depth understanding of the current nuances of the IRS's taxpayer data and allowed the IRS to easily introduce new fields to address the financial material weakness. In total, only seven material change requests to the data model have been approved since it was built three years ago. One of these included the change to upgrade to data model version 2.2, which was framed as a "redesign" in the Treasury Inspector General for Tax Administration's audit report. In fact, the IRS resolved the issue with a minor modification to the data model, which is another clear indicator that the CADE 2 data model is stable.

Office of Audit Comment: The IRM 2.5.13 standard requires that the database design process deliver a set of documents:

- a) Decision Analysis and Description Forms.
- b) Task Analysis and Description Forms.
- c) Task/Data Element Usage Matrix.
- d) Data Models.
- e) Entity-Attribute Lists.
- f) Data Definition Lists.
- g) Physical Database Specification Document.

This set of documents validates that the database design supports the business requirements. We did not receive the necessary documents to confirm that the database design supports the business requirements. Further, the Executive Status Update on September 12, 2012, stated there is a delay in clearing the backlog of defects identified during the data validation activities.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Recommendation 3: Realign data validation and testing efforts with business functionality and processes.

Management's Response: The IRS disagreed with this recommendation. The IRS stated that full data coverage mapping and its data validation and testing approach leveraged business requirements that are implicit in the IMF and have proven the test of time. Additional mapping exercises at the data element level as recommended would add little value to the process. As part of the validation approach, the IRS does mock testing in production simulation environments. It does testing using production data – a sampling of 2.5 million returns – to ensure integrity of data. A considerable portion of the integrity testing, for example, has been designed to ensure that outputs from the CADE 2 database – through Individual Master File On-Line or Taxpayer Identification File outputs (future) – either match the parallel output from the legacy Master File or fall into a small set of “acceptable” differences. With the business organization fully engaged throughout all phases of this data integrity testing and review, the comprehensive top-down and bottom-up approach for data verification has been extremely effective in discovering issues, which are certainly to be expected on projects the size and magnitude of the CADE 2 system. As correctly described in the report, the functional (Systems Acceptability Testing) tests are also verifying whether the transformation rules were implemented per specification. While it is not reasonable to think that any validation approach will cover every possible combination and permutation of those rules, it provides reasonable risk mitigation to complement the IRS’s high-volume data validation testing.

Office of Audit Comment: We agree that business requirements are implicit in IMF data. However, we found no evidence of these requirements being traced to either data elements on the CADE 2 database or the transformation rules used to load IMF data into the CADE 2 database. Without verifiable evidence, data elements and values may have been missed. The IRS’s data integrity plan noted that the more complex data transformation rules were tested by Systems Acceptability Testing and that, for many of these fields, Systems Acceptability Testing could not cover all the permutations or conditions of the transformation logic. Therefore, there is no way of knowing which data fields or values may have been missed or left untested, or what business requirements they were needed to support. Only an alignment of data elements to business functionality and a testing effort based on that alignment would ensure the degree of data validation necessary for the CADE 2 database to become the IMF’s authoritative file of record.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Security Weaknesses and Poor Coding Practices in the Customer Account Data Engine 2 Database Could Result in the Loss of Taxpayer Data

Enhanced security is one of the goals of the CADE 2 Program. CADE 2 database security will be implemented via a role-based access model and the Resource Access Control Facility.⁵ Security will remain a key concern until role-based access is developed and fully implemented across the IRS.

Vulnerabilities in the JAVA code could result in loss of sensitive taxpayer information

In designing systems, the IRS has several security requirements from multiple sources that need to be met. The National Institute of Standards and Technology publishes the Federal Information Processing Standards that provide the requirements for encryption to be used by governmental systems to prevent anyone without the necessary credentials from being able to ascertain the data stored on computer systems.

The IRS and the Department of the Treasury also have established standards for systems operating on their networks. For example, one IRS policy requires that passwords must be changed after a set number of days and that the password must exceed a specific number of characters and include certain types of characters. Another requirement is that test code and example database tables and components must be removed from an application.

In October 2011, two independent contractors conducted source code security reviews of the balance and control module. Figure 4 provides the JAVA code weaknesses identified by the source code security reviews.

⁵ An IBM security system that provides access control and auditing functionality for the z/OS and z/VM operating systems.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Figure 4: JAVA Code Weaknesses

High Risk	Moderate Risk	Low Risk
<ul style="list-style-type: none"> SQL Injection 	<ul style="list-style-type: none"> Bug: Incorrect Logical Operator Insecure Algorithm Insufficient Input Validation Insufficient Password Management Use of Inner Classes 	<ul style="list-style-type: none"> Dead Code Detailed Error Messages Improper Logging Information Exposure Test Code Unreleased Resources

Source: Contractor Source Code Review, IRS Wage and Investment Business Unit's CADE 2 Database Implementation TS-1 JAVA Code, Balance and Control Module Core Module.

Both contractors recommended that the weaknesses identified in Figure 4 be corrected. The CADE 2 Governance Board deemed the overall risk to the database application as low stating the code is hard to exploit and that it will be removed after the second database initialization. The Governance Board accepted the security weaknesses contained in the JAVA initialization code and will not take any remediation actions. IRS management also advised that the JAVA code was developed for one-time use.

However, the IRS has used this JAVA code multiple times in testing. The JAVA code was also used to initialize the production database in March 2012 with data model version 2.1.1 and it will be used to initialize the database with data model version 2.2 in the summer of 2012. Based on the utilization of this JAVA code, it does not appear to have been developed for one-time use.

Remediation of the weaknesses will enhance the JAVA database initialization balancing and control code and enhance the security of the database. Ineffective password, incorrect logical operator statement, dead code, and test code could result in the loss of Personally Identifiable Information data, loss of reputation, and loss of taxpayers' trust. Dead code also could impact performance of the database initialization, and test code could be executed during initialization. This could result in data not being accurate or complete in the CADE 2 database.

Remediation of identified security weaknesses is ineffective

The IRS performed mainframe database security testing on its IBM mainframe systems using the IBM Guardium scanner in December 2011 and March 2012. The Guardium scanner reviewed all sub-systems on the database management system, including CADE 2. As the scan encompassed more than just CADE 2, the weaknesses related specifically to the CADE 2 subsystem could not be easily identified. The March 2012 Guardium scan identified 67 weaknesses, of which 49 were deemed critical and 18 were deemed major. We compared the critical weaknesses identified in the December 2011 and March 2012 scans and concluded the weaknesses were mostly repeat findings. Figure 5 summarizes the comparison of the scan results.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Figure 5: December 2011 Versus March 2012 Guardium Scan Identified Weaknesses

Date of Scan	Scan Type	Number of Critical Weaknesses	Number of Major Weaknesses	Total Identified Weaknesses
December 2011	Privileged Users	47	2	49
	Configuration	2	16	18
March 2012	Privileged Users	47	2	49
	Configuration	2	16	18

Source: Treasury Inspector General for Tax Administration analysis of the Guardium scan results.

Weaknesses identified among privileged user accounts included users with unauthorized access to tables, packages, and files. Configuration weaknesses are related to default ports and an enabled demo table. The IRM states that default sample databases, along with any associated objects and user accounts are to be removed.⁶ These default databases and tables utilize default accounts, passwords, and ports. In addition, default ports with known vulnerabilities should not be utilized.⁷ Figure 6 provides examples of repeat weaknesses that were identified by the Guardium scanner.

Figure 6: Examples of Repeat Weaknesses Identified By Guardium

Rule Description	Number of Exceptions	
	December 2011	March 2012
LOAD privilege has been granted to unauthorized users.	104	111
SYSADM privilege has been granted to unauthorized users.	3	3
CREATEDBA privilege has been granted to unauthorized users.	13	13
One or more sample databases have been found.	6	6

Source: Treasury Inspector General for Tax Administration analysis of the Guardium scan results.

⁶ IRM 10.8.21, Information Technology (IT) Security - Database Security Policy.

⁷ IRM 10.8.21, Information Technology (IT) Security - Database Security Policy.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Privileged user accounts are those accounts with elevated privileges which are used to maintain and administer systems or to perform tasks. Privileged user accounts include service, database administrator, and system administrator accounts. The Online 5081 system is used to record all access requests and document the semi-annual review of privileged user accounts as required by the IRM.⁸

We reviewed the administration of privileged user accounts by selecting a judgmental sample of five service accounts and five database administrator accounts.⁹ Supporting evidence for the five service accounts were not documented in the Online 5081 system. Therefore, we could not identify the purpose of these accounts. The Online 5081 system retains only the last review date so we were unable to verify that the semi-annual reviews were performed on all 10 privileged user accounts. Further, we were unable to determine if a privileged user access authority is appropriate and commensurate with job role and responsibilities as this information was not available in the Online 5081 system. This could result in unauthorized access and loss of Personally Identifiable Information, unauthorized changes to the database, and loss of data integrity.

The CADE 2 database was developed and implemented in a short time period and accounts were migrated from existing legacy systems. As a result, default tables and ports were overlooked and were not removed. In addition, when the IRS migrated to the Online 5081 system, validation for accuracy and completeness was not conducted and historical records were lost. In addition, using default ports and enabling a demo table increases the IRS's vulnerability.

Recommendations

The Chief Technology Officer should ensure:

Recommendation 4: JAVA code weaknesses are remediated to enhance security and efficiency of the JAVA code.

Management's Response: The IRS agreed with this recommendation. Although the CADE 2 governance has assessed the actual code weakness as low and a risk-based decision was made to accept it, there are processes now in place where the developers provide the code to the Cybersecurity organization for review prior to code promotion. The Cybersecurity organization provides weakness feedback and the cycle is repeated until all code weaknesses have been addressed. Additionally, the developers now use an automated code review tool as part of their own development process. Finally, a decision (*i.e.*, fix the code, remove it permanently, or accept the risk) will be made as to the final disposition of this code prior to the CADE 2 Milestone 5 exit.

⁸ IRM 10.8.1, *Information Technology (IT) Security - Policy and Guidance*.

⁹ A judgmental sample is a nonstatistical sample, the results of which cannot be used to project to the population.



The Customer Account Data Engine 2 Database Was Initialized; However, Database and Security Risks Remain, and Initial Timeframes to Provide Data to Three Downstream Systems May Not Be Met

Recommendation 5: Privileged user accounts are properly documented, administered, monitored, and reviewed in accordance with the IRM or removed from the system.

Management's Response: The IRS agreed with this recommendation. The IRS stated that it should be noted that the Treasury Inspector General for Tax Administration did not take into account, as part of this audit, any risk-based decisions around privileged user accounts or the fact that the IBM Guardium scanner's predefined "critical" weaknesses levels are not necessarily correct for the CADE 2 environment. Notwithstanding, enterprise-level remediation plans are being developed to address validated scan findings.

Office of Audit Comment: Our analysis was based on the aggregation of the weaknesses identified in each rule by the Guardium scan dated December 2011 and March 2012 and not the default rating by the IBM Guardium scan.

Recommendation 6: Sample tables and default ports are disabled or removed prior to the CADE 2 Program exiting TS1.

Management's Response: The IRS disagreed with this recommendation. The IRS stated that IRM 10.8.21.5.4.2 does not explicitly list the use of default ports as forbidden. Changing the default DB2 port is a massive technology undertaking and does not add significantly to the level of security; therefore, doing so should not be taken lightly. Additionally, the default DB2 port impacts all risk-based applications on the Master File platform, not just CADE 2, and changes could jeopardize access to vital tax administration applications. Nonetheless, as stated in Corrective Action 5, changing the default port will be taken into consideration as part of an enterprise risk mitigation remediation plan.

Office of Audit Comment: Ports with known vulnerabilities should not be used when possible. If these ports are to be used in production, the port setting should be set to "disable broadcast." In addition, default tables were identified in the December 2011 Guardium scan. The same default tables were identified in the March 2012 Guardium scan. The IRM states default tables should be disabled or deleted.

Recommendation 7: The Online 5081 system should be enhanced to retain and display the last two review dates.

Management's Response: The IRS partially agreed with this recommendation. The IRS is reviewing the possibility of loss of historical records during the migration to the Online 5081 system, as reflected in the audit report. If, upon completion of that review, the IRS finds significant risks to the CADE 2 database, the Chief Technology Officer will work with the Director, Agency-wide Shared Services, and the owner of the Online 5081 system to consider ways to mitigate the vulnerabilities. If the mitigation strategy suggests enhancements to the Online 5081 system, the IRS will make that decision



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

weighing the risks to the CADE 2 database against the costs in time and resources to do the system enhancements.

Office of Audit Comment: The Online 5081 system is missing historical information such as account creation date, purpose of the account, and the last two review dates. The IRM requires this information to be documented and maintained.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Appendix I

Detailed Objective, Scope, and Methodology

The overall objective was to review the CADE 2 database implementation and ensure that the database was secure, accurate, and complete, and that prior weaknesses identified were corrected or mitigated. To accomplish our objective, we:

- I. Reviewed the architectural configuration for the database environment to identify control points and ensure weaknesses are identified and mitigated.
 - A. Reviewed the architectural diagram for the application environment and used it to identify potential control weaknesses.
 - B. Determined the impact to the database, impacted systems, and taxpayers of any control weaknesses not mitigated.
- II. Determined if the database is properly secured.
 - A. Reviewed the results of two independent assessments performed by contractors.
 - B. Determined if weaknesses identified by the December 2011 Guardium scan were corrected.
 - C. Determined if the database is secured and privileged user accounts are limited, monitored, and reviewed. There were 383 privileged user accounts and we judgmentally selected¹ five service accounts and five database administrator accounts for review.
 - D. Determined if default (demo) tables were properly secured, removed, or disabled.
- III. Determined if data integrity controls are developed and operating as designed to ensure data are accurate and complete.
 - A. Interviewed a balancing and control subject matter expert and the ETL subject matter expert for the database and obtained documents detailing the balance and control policy, procedures, and processing.
 - B. Determined if weaknesses identified in the ETL process were corrected or mitigating controls were developed and implemented.
 - C. Reviewed any other code reviews performed on the CADE 2 database, including cycle synchronization and daily updates.

¹ A judgmental sample is a nonstatistical sample, the results of which cannot be used to project to the population.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

- D. Ensured that data transfers between input data sources and the audited database are complete and accurate.
 - E. Determined if the processes for ensuring database consistency during cycle synchronization and daily update address the accuracy and completeness of data.
- IV. Reviewed downstream system/application interfaces and impact(s).
- A. Interviewed the subject matter expert for system interfaces to gain an understanding of how system interfaces and impact are determined.
 - B. Determined if interfaces are secured.

Internal controls methodology

Internal controls relate to management's plans, methods, and procedures used to meet their mission, goals, and objectives. Internal controls include the processes and procedures for planning, organizing, directing, and controlling program operations. They include the systems for measuring, reporting, and monitoring program performance. We determined the following internal controls were relevant to our audit objective: the IRM, related CADE 2 documents, and guidelines and processes in the development of the CADE 2 database. We evaluated these controls by conducting interviews and meetings with management and staff, attending CADE 2 Database Implementation meetings, and reviewing CADE 2 Program documentation and CADE 2 Database Implementation documents such as the CADE 2 Program Charter, CADE 2 Solution Architecture, CADE 2 Database Implementation Test Plan, CADE 2 Program Management and Integration Plan, CADE 2 Program Road Map, and CADE 2 Interface Control Document, and other documents that provided evidence of whether IRM systems testing processes were followed and if those processes were adequate and operating as designed.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Appendix II

Major Contributors to This Report

Alan R. Duncan, Assistant Inspector General for Audit (Security and Information Technology Services)

Danny R. Verneuille, Director

Larry W. Reimer, Audit Manager

Mark K. Carder, Senior Auditor

K. Kevin Liu, Lead Information Technology Specialist

Hung Q. Dam, Information Technology Specialist

Arlene Feskanich, Information Technology Specialist



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Appendix III

Report Distribution List

Commissioner C
Office of the Commissioner – Attn: Chief of Staff C
Deputy Commissioner for Operations Support OS
Associate Chief Information Officer, Applications Development OS:CTO:AD
Associate Chief Information Officer, Enterprise Operations OS:CTO:EO
Associate Chief Information Officer, Cybersecurity OS:CTO:C
Associate Chief Information Officer, Modernization Program Management Office OS:CTO:MP
Director, Security Risk Management OS:CTO:C:SRM
Chief Counsel CC
National Taxpayer Advocate TA
Director, Office of Legislative Affairs CL:LA
Director, Office of Program Evaluation and Risk Analysis RAS:O
Office of Internal Control OS:CFO:CPIC:IC
Audit Liaisons:
 Commissioner, Wage and Investment Division SE:W:S:PRA:PEI
 Director, Risk Management Division OS:CTO:SP:RM



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Appendix IV

Management's Response to the Draft Report¹



CHIEF TECHNOLOGY OFFICER

DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
WASHINGTON, D.C. 20224

SEP 11 2012

MEMORANDUM FOR DEPUTY INSPECTOR GENERAL FOR AUDIT

FROM: Terence V. Milholland *Terence V. Milholland*
Chief Technology Officer

SUBJECT: Draft Audit Report – *Database Implementation of
the Customer Account Data Engine 2 Program Is At
Risk of Not Meeting All of Its Transition State 1
Goals* (Audit #201220023) (e-trak #2012- 34860)

Thank you for the opportunity to review your draft audit report and discuss earlier draft observations with the audit team.

While we appreciate TIGTA's concerns about the CADE 2 database as outlined in your draft report, we believe it is important to recognize that the IRS has met virtually all of its goals to date in relation to the scope elements of the Transition State 1 as shown on page 1 of your report, providing business functionality and benefits as follows:

- "The IRS established a single database that stores all individual taxpayer accounts." This capability has been delivered – fully initialized (270 million accounts and over a billion tax modules) and balanced to the penny in both testing and production environments for version 2.1 and an expanded version 2.2;
- "Processing has been enhanced to include daily batch processing." This capability was delivered in time for Filing Season 2012 and performed flawlessly in production, providing faster refunds for millions of taxpayers;
- "Key IRS customer service operational database, IDRS, has the benefit of more timely posted data." This capability was delivered on time with 100 percent of posted transactions viewable in Integrated Data Retrieval System (IDRS) within 48 hours of processing during filing season (versus 10 days in previous years under legacy Individual Master File (IMF));
- "Populate the Integrated Production Model (IPM) analytical data store and provide business users with tools to more effectively use the data for compliance and customer service." This functionality is being satisfied, in the absence of a longer-term strategy for our IPM solution, with eleven CADE 2 reports (five are already completed, six are in progress), pulling data from the CADE 2 database, using standard Business Objects reporting tools.

¹ The final audit report title was revised based on discussions with the IRS after the issuance of the draft report.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

2

- "Enhanced data security will be in place." Enhanced security was addressed by adding 18 NIST 800-53 security controls on top of the standard security control baseline for a database such as CADE 2.
- "Downstream systems modified to support daily processing," which is currently being satisfied through:
 - daily "IMF" feeds to IDRS, which is completed;
 - daily updates to the CADE 2 database with the launch of cycle synchronization, which began on August 29, 2012;
 - feeds from the CADE 2 database to downstream systems allow for online viewing of the taxpayer account data stored in the new CADE 2 database, which is on track for delivery in September 2012; and
 - daily CADE 2 database feeds to the IDRS Taxpayer Information File (TIF) for online updates to taxpayer account data by customer service representatives to be deployed after the 2013 Filing Season, in order to mitigate risks to the Filing Season.

In addition, we believe the design of the CADE 2 database is extremely sound, as the CADE 2 data model was built using historical lessons learned from CADE 1, and also production data from IMF. Only seven material change requests to the data model have been approved since it was built three years ago, including the change to upgrade to Version 2.2. The CADE 2 data model has demonstrated its stability and viability.

Moreover, we believe our testing approach is sound and appropriate for the CADE 2 Program. With our business organization fully engaged throughout all phases of testing and review, our approach is comprehensive, both top-down and bottom-up, ensuring quality of our data verification.

And finally, all CADE 2 database design, development, testing and implementation efforts have been accomplished under an unprecedented collaborative model with our IT suppliers. Our collaboration with subject matter experts from IBM, CSC, Informatica, Computech, Deloitte, McKinsey, and MITRE, helped us ensure our design and development decisions were well-vetted, built on sound technology and risk based.

In closing, we have been delivering the benefits of CADE 2 Transition State 1 and are on track to complete the downstream production use of the CADE 2 database.

Our responses to TIGTA's specific recommendations in the report are attached.

We value your continued support and the assistance and guidance your team provides. If you have any questions, please contact me at (202) 622-6800, or a member of your staff may contact Karen Mayr at (240) 613-1431.

Attachment



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Draft Audit Report – Database Implementation of the Customer Account Data Engine 2 Program Is At Risk of Not Meeting All of Its Transition State 1 Goals (*Audit #201220023*) (*e-trak #2012-34860*)

RECOMMENDATION #1: The Chief Technology Officer should ensure that the CADE 2 Program does not exit TS 1 until the CADE 2 database can provide accurate and complete data to the IDRS TIF, Corporate Files On-Line/Individual Master File On-Line, and to the re-evaluated Integrated Production Model.

CORRECTIVE ACTION #1: The IRS agrees with the recommendation. With appropriate approvals from CADE 2 governance committee, the IRS plans to exit Milestone 5 according to schedule in September 2012, in order to deploy planned Corporate Files On-Line/Individual Master File On-Line and IPM Reports functionality. The milestone exit will be conditional, however, until such time as the IRS has deployed planned Integrated Data Retrieval System Taxpayer Identification File (TIF) functionality, which will be done upon completion of Filing Season peak 2013.

IMPLEMENTATION DATE: September 30, 2013

RESPONSIBLE OFFICIAL: Associate Chief Information Officer, CADE 2 PMO/Modernization

CORRECTIVE ACTION MONITORING PLAN: We enter accepted Corrective Actions into the Joint Audit Management Enterprise System (JAMES). These Corrective Actions are monitored on a monthly basis until completion.

RECOMMENDATION #2: The Chief Technology Officer should ensure that the database design process follows the IRM and validate that the database meets business requirements.

CORRECTIVE ACTION #2: The IRS does not agree that this recommendation is necessary as a result of this audit. Our database design approach is extremely sound and it meets the Internal Revenue Manual (IRM) standard. It leverages the legacy Individual Master File (IMF), which has undergone years of refinement and embodies business requirements that are complete and accurate. Our database fully supports the business requirements, as the CADE 2 data model was built using historical lessons learned from previous successes in CADE, which ran in production for eight years, and production data from IMF, which has run in production for five decades. An added layer of confidence to our approach was gained by running through transformations using real taxpayer data, which further proved out that our data model and database design were very strong. This CADE 2 data model design approach provided an in-depth understanding of the current nuances of the IRS's taxpayer data, and allowed us to introduce new fields easily to address our financial material weakness. In total, only seven material change requests to the data model have been approved since it was built three years ago. One of these included the change to upgrade to data model V. 2.2, which TIGTA framed as a "redesign" in their audit report. In fact, the IRS resolved the issue with a minor modification to the data model, which is another clear indicator that the CADE 2 data model is stable.



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Draft Audit Report – Database Implementation of the Customer Account Data Engine 2 Program Is At Risk of Not Meeting All of Its Transition State 1 Goals (*Audit #201220023*) (*e-trak #2012-34860*)

IMPLEMENTATION DATE: Not Applicable

RESPONSIBLE OFFICIAL: Not Applicable

CORRECTIVE ACTION MONITORING PLAN: Not Applicable

RECOMMENDATION #3: The Chief Technology Officer should realign data validation and testing efforts with business functionality and processes.

CORRECTIVE ACTION #3: The IRS does not agree with this recommendation. Our full data coverage mapping and our data validation and testing approach leveraged business requirements that are implicit in the Individual Master File and have proven the test of time. Additional mapping exercises at the data element level recommended by TIGTA would have added little value to the process. As part of our validation approach, we do mock testing in production simulation environments. We do our testing using production data – a sampling of 2.5 million returns – to ensure integrity of data. A considerable portion of our data integrity testing, for example, has been designed to ensure that outputs from the CADE 2 database – through Individual Master File On-Line (IMFOL) or Taxpayer Identification File (TIF) outputs (future) – either match the parallel output from the legacy Master File or fall into a small set of “acceptable” differences. With the business organization fully engaged throughout all phases of this data integrity testing and review, our comprehensive top-down and bottom-up approach for data verification has been extremely effective in discovering issues, which are certainly to be expected on projects the size and magnitude of CADE 2. As correctly described in the TIGTA report, the functional (SAT) tests are also verifying whether the transformation rules were implemented per specification. While it is not reasonable to think that any validation approach will cover every possible combination and permutation of those rules, it provides reasonable risk mitigation to complement our high-volume data validation testing.

IMPLEMENTATION DATE: Not Applicable

RESPONSIBLE OFFICIAL: Not Applicable

CORRECTIVE ACTION MONITORING PLAN: Not Applicable

RECOMMENDATION #4: The Chief Technology Officer should ensure JAVA code weaknesses are remediated to enhance security and efficiency of the JAVA code.

CORRECTIVE ACTION #4: The IRS agrees with this recommendation. Although the CADE 2 governance has assessed the actual code weakness as low and a risk-based decision was made to accept it, there are processes now in place where the developers provide the code to Cybersecurity for review prior to code promotion. Cybersecurity provides weakness feedback and the cycle is repeated until all code weaknesses have been addressed. Additionally, the developers now use an automated code review tool as part of their own development process. And finally, a decision will be made as to the final disposition of this code prior to the CADE 2



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Draft Audit Report – Database Implementation of the Customer Account Data Engine 2 Program Is At Risk of Not Meeting All of Its Transition State 1 Goals (Audit #201220023) (e-trak #2012-34860)

MS 5 exit. At that time the decision will be to fix the code, remove it permanently or accept the risk.

IMPLEMENTATION DATE: September 30, 2012

RESPONSIBLE OFFICIAL: Associate Chief Information Officer, Applications Development

CORRECTIVE ACTION MONITORING PLAN: We enter accepted Corrective Actions into JAMES. These Corrective Actions are monitored on a monthly basis until completion.

RECOMMENDATION #5: The Chief Technology Officer should ensure privileged user accounts are properly documented, administered, monitored, and reviewed in accordance with the IRM or removed from the system

CORRECTIVE ACTION #5: The IRS agrees with this recommendation. It should be noted, however, that TIGTA did not take into account as part of this audit any risk-based decisions around privileged user accounts or the fact that IBM Guardium scanner predefined “critical” weaknesses levels are not necessarily correct for the CADE 2 environment. Notwithstanding, enterprise-level remediation plans are being developed to address validated scan findings.

IMPLEMENTATION DATE: December 31, 2013

RESPONSIBLE OFFICIAL: Associate Chief Information Officer, Enterprise Operations

CORRECTIVE ACTION MONITORING PLAN: We enter accepted Corrective Actions into JAMES. These Corrective Actions are monitored on a monthly basis until completion.

RECOMMENDATION #6: The Chief Technology Officer should ensure sample tables and default ports are disabled or removed prior to the Program exiting TS 1.

CORRECTIVE ACTION #6: The IRS disagrees with this recommendation. Internal Revenue Manual 10.8.21.5.4.2 does not explicitly list the use of a default ports as forbidden. Changing the default DB2 port is a massive technology undertaking, and it does not add significantly to the level of security, doing so should not be taken lightly. Additionally, the default DB2 port impacts all risk-based applications on the Master File platform, not just CADE 2, and changes could jeopardize access to vital tax administration applications. Nonetheless, as stated above for Corrective Action #5, changing the default port will be taken into consideration as part of an enterprise risk mitigation remediation plan.

IMPLEMENTATION DATE: Not Applicable

RESPONSIBLE OFFICIAL: Not Applicable



*The Customer Account Data Engine 2 Database Was Initialized;
However, Database and Security Risks Remain, and Initial
Timeframes to Provide Data to Three Downstream Systems
May Not Be Met*

Draft Audit Report – Database Implementation of the Customer Account Data Engine 2 Program
Is At Risk of Not Meeting All of Its Transition State 1 Goals (Audit #201220023) (e-trak #2012-
34860)

CORRECTIVE ACTION MONITORING PLAN: Not Applicable

RECOMMENDATION #7: The Chief Technology Officer should ensure the Online 5081
system should be enhanced to retain and display two review dates.

CORRECTIVE ACTION #7: The IRS partially agrees with this recommendation. We are
reviewing the possibility of loss of historical records during the migration to the Online 5081
system, as reflected in the audit report. If, upon completion of that review, we find significant
risks to the CADE 2 database, the Chief Technology Officer will work with the Director,
Agency-wide Shared Services, and the owner of the Online 5081 system, to consider ways to
mitigate the vulnerabilities. If the mitigation strategy suggests enhancements to the Online 5081
system, the IRS will make that decision weighing the risks to the CADE 2 database against the
costs in time and resources to do the system enhancements.

IMPLEMENTATION DATE: January 31, 2013

RESPONSIBLE OFFICIAL: Associate Chief Information Officer, Cybersecurity

CORRECTIVE ACTION MONITORING PLAN: We enter accepted Corrective Actions into
JAMES. These Corrective Actions are monitored on a monthly basis until completion.