TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION

*Customer Account Data Engine 2 Database
Deployment Is Experiencing Delays and
Increased Costs*

September 23, 2013

Reference Number: 2013-20-125

CUSTOMER ACCOUNT DATA ENGINE 2 DATABASE DEPLOYMENT IS EXPERIENCING DELAYS AND INCREASED COSTS

Highlights

Final Report issued on September 23, 2013

Highlights of Reference Number: 2013-20-125 to the Internal Revenue Service Chief Technology Officer.

IMPACT ON TAXPAYERS

The Transition State 1 system deployment phase of the Customer Account Data Engine 2 (CADE 2) database, which included interfaces to downstream systems, was initially scheduled for implementation in September 2012. However, database deployment has been delayed, and deployment costs have risen an estimated 74 percent to $83 million. Taxpayer service improvements that were to be provided by the new transactional database have also been delayed. Deployment delays and cost overruns can decrease the public's confidence in the IRS's ability to develop, monitor, and use its resources effectively.

WHY TIGTA DID THE AUDIT

The overall objective was to determine whether the IRS has implemented adequate CADE 2 database downstream interface data validation to ensure that the data provided are accurate and complete. This audit is included in our Fiscal Year 2013 Annual Audit Plan and addresses the major management challenge of Modernization.

WHAT TIGTA FOUND

The CADE 2 database cross-functional triage team effectively managed and resolved more than 1,000 data defects. However, our review determined that the downstream system interfaces were not implemented due to data quality issues that exist with the CADE 2 database. The interfaces were also not implemented by the revised date of June 2013.

With a revised projected implementation date of January 2014, the overall total estimated cost of Transition State 1 system deployment rose from $47.7 million to $83 million.

The CADE 2 database's lack of accuracy, completeness, and availability prevents it from serving as the trusted source for the downstream systems. TIGTA also determined that the solution architecture of the CADE 2 database interfaces does not meet the IRS's business needs because it does not meet performance expectations and creates resource contention situations between servicing online transactions and query operations.

In addition, the lack of security systems integration prevents transaction-level tracking of employee access to the CADE 2 database.

WHAT TIGTA RECOMMENDED

TIGTA recommended that the Chief Technology Officer: 1) not exit Transition State 1 Milestone 5 until the interfaces with selected downstream systems are implemented into production; 2) ensure that the CADE 2 database is accurate, complete, timely, and available; 3) deploy the CADE 2 database as the transactional database architected for Transition State 2 and the Target State, as the authoritative data source for an enterprise data warehouse or a data mart, and not as a direct data source for downstream systems; and 4) ensure that user access to the CADE 2 database is traced at the transaction level by individual user's identification.

In its response to the report, the IRS agreed with one of the four recommendations and corrective action is planned. IRS management agreed to certify the database is accurate, complete, timely, and available to serve as the trusted source. IRS management disagreed with the three remaining recommendations. We believe risks remain and we provided Office of Audit Comments in the report.

DEPARTMENT OF THE TREASURY

WASHINGTON, D.C. 20220

TREASURY INSPECTOR GENERAL
FOR TAX ADMINISTRATION

September 23, 2013

MEMORANDUM FOR CHIEF TECHNOLOGY OFFICER

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

SUBJECT: Final Audit Report – Customer Account Data Engine 2 Database
                   Deployment Is Experiencing Delays and Increased Costs
                   (Audit #201320021)

This report presents the results of our review of the Customer Account Data Engine 2 database downstream system interfaces. The overall objective of this review was to determine whether the Internal Revenue Service has implemented adequate Customer Account Data Engine 2 database downstream interface data validation to ensure that the data provided are accurate and complete. This review, which was requested by the Chief Technology Officer, is included in the Treasury Inspector General for Tax Administration's Fiscal Year 2013 Annual Audit Plan and 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. If you have any questions, please contact me or Alan R. Duncan, Assistant Inspector General for Audit (Security and Information Technology Services).

Customer Account Data Engine 2 Database Deployment Is Experiencing Delays and Increased Costs

Table of Contents

Abbreviations

CADE 2	Customer Account Data Engine 2
CFOL	Corporate Files Online
DAS	Data Access Service
IDRS	Integrated Data Retrieval System
IMF	Individual Master File
IMFOL	Individual Master File Online
IRM	Internal Revenue Manual
IRS	Internal Revenue Service
MIPS	Million Instructions Per Second
MS	Milestone
TIGTA	Treasury Inspector General for Tax Administration
TS	Transition State
VSAM	Virtual Storage Access Method

Background

The Customer Account Data Engine 2 (CADE 2) Program is one of the top information technology modernization projects 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 tax administration. The CADE 2 will replace the current Individual Master File (IMF)[1] account settlement system with a relational database processing system and become a key component in the IRS's enterprisewide, 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 (TS) 1	The IRS will establish a relational database that will store all individual taxpayer accounts. Processing on the current IMF will be enhanced to include daily batch processing, allowing the key IRS customer service database, the Integrated Data Retrieval System (IDRS),[2] to have the benefit of more timely posted data. Enhanced data security will be in place. Interfaces between the CADE 2 database and selected downstream systems, *i.e.*, Corporate Files Online (CFOL)/IMF Online (IMFOL) and IDRS, will be developed.
TS2	A single processing system will be implemented. Applications will use the taxpayer account database. The solution will leverage elements of the current IMF and current CADE for some functions. The CADE 2 Program will make continued progress addressing the financial material weaknesses. In the TS2, a combination of current-state components and transitional components will be used to fill the functional needs of individual taxpayer account processing. The TS2 is scheduled to be completed in March 2015 with a total cost of $227.5 million. The total cost is subject to change.
Target State	Provides a complete data-centric solution, retires all transitional components, and addresses all financial and security material weaknesses identified at the inception of the Program. As of May 20, 2013, the IRS had not established a Target State implementation date.

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

The TS1 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

---

[1] The IRS system of record for individual taxpayer accounts.
[2] A mission-critical, steady state system consisting of databases and operating programs that support IRS employees working active tax cases within each business function across the entire IRS.

January 2012. Database Implementation, while not fully implemented, developed a relational database to store individual taxpayer account data migrated from IMF tape files on a daily basis. In March 2012, the IRS initialized version 2.1 of the CADE 2 database with 270 million individual taxpayer accounts and more than one billion tax modules. The IRS completed a second database initialization in October 2012 and kept the database current and in-sync with IMF data through December 2012.

In addition to building the CADE 2 database in TS1, certain downstream system interfaces are also to be developed:

- CFOL/IMFOL: The TS1 will provide the capability to view taxpayer account data stored in the CADE 2 database using CFOL/IMFOL commands.

- IDRS: The TS1 will provide daily data extracts from the CADE 2 database to the IDRS.

The Treasury Inspector General for Tax Administration (TIGTA) conducted a prior review of the CADE 2 Database Implementation project[3] to ensure that the CADE 2 database is secure, accurate, and complete. Our review raised concerns that testing did not provide assurance that the CADE 2 database data are accurate and complete. In addition, we noted the CADE 2 database design had not fully met initialization, daily update, and downstream interface needs. As a result, the audit report contained the following recommendations:

1) Ensure that the CADE 2 Program does not exit the TS1 until the CADE 2 database can provide accurate and complete data to the three downstream systems.

2) Ensure that the database design process follows the Internal Revenue Manual (IRM) and validate that the database design meets business requirements.

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

This review was performed at the IRS Information Technology organization's offices in Lanham, Maryland, during the period December 2012 through April 2013. We conducted this 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.

---

[3] TIGTA, Ref. No. 2012-20-109, *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* p. 3 (Sept. 2012).

Results of Review

The Cross-Functional Triage Team Effectively Managed and Resolved Data Defects

One of the objectives of the TS1 is for the CADE 2 database to stay in-sync with the IMF system and become the data source for the CFOL/IMFOL and its downstream systems. To accomplish this objective, the IRS developed the Data Access Service (DAS) as an interface to the CADE 2 database for the CFOL/IMFOL and its 16 downstream systems. The CADE 2 database CFOL/IMFOL/DAS solution architecture includes Identify and Extract Account Changes modules to extract data from the IMF, modules to transform and load data into the CADE 2 database, and the DAS interface. Each system component was developed and supported by a different team.

Between July 2012 and December 2012, the IRS performed multiple data quality tests and recorded more than 1,000 defects, *i.e.*, data and code. The IRS planned to conduct a production deployment of the CADE 2 database and the CFOL/IMFOL/DAS interface on December 30, 2012, and the development teams were tasked to resolve the defects prior to this date. Defect management is a challenge for software development projects, and the final CFOL/IMFOL/DAS interface test in production could not be completed until the identified defects were resolved. To resolve the defects, the IRS established a cross-functional triage team to manage the defects. The triage team included members from all development teams and business users. It met daily to analyze defects, take ownership of the issues, and develop solutions. The collaborative working environment established by the IRS to resolve the defects increased productivity and expedited issue resolution. By the end of Calendar Year 2012, the IRS reduced the number of open defects to five.

The Customer Account Data Engine 2 Database Downstream System Interface Implementation Has Been Delayed and Incurred Additional Costs

The initial scope of the TS1 included the deployment of CADE 2 database downstream interfaces to the CFOL/IMFOL/DAS and the IDRS in September 2012. As development issues arose, it became clear the IRS could not deliver the intended scope by September 2012. On November 5, 2012, the CADE 2 Executive Steering Committee granted a conditional TS1 Milestone (MS) 5 (System Deployment Phase) exit with the following conditions:

1. Availability of CFOL/IMFOL/DAS/CADE 2 database interface in production by December 2012.

2. Deployment of the CADE 2 database/IDRS interface by the fourth quarter of Fiscal Year 2013.

The CADE 2 Database Implementation Design Specification Report requires that the CFOL/IMFOL/DAS interface provide 24-hour access seven days a week to the CADE 2 database for daily updated IMF data, support 400,000 transactions per peak hour, and serve up to 15,000 concurrent users. The IRS stated that it deployed the CFOL/IMFOL/DAS interface in the production environment in December 2012. During the deployment, the production environment was limited. For example, the IRS turned on the interface for six hours during the non-peak period between 1 a.m. and 7 a.m. Eastern Time. Only six employees participated as testers requesting 1,500 pre-selected taxpayer accounts from the CADE 2 database for data validation. In addition, none of the 16 downstream systems that use CFOL/IMFOL to request IMF taxpayer data participated in the December 2012 deployment. The interface has not been turned on again since December 2012. Based on the limited deployment of the interface, we believe the deployment was actually a test in the production environment. In addition, this test does not prove the CFOL/IMFOL/DAS interface fulfills the CADE 2 Database Implementation Design Specification Report requirements.

While the IRS is working to satisfy the TS1 MS 5 exit condition for deploying the CFOL/IMFOL/DAS interface, our review identified no evidence of progress towards deploying the CADE 2 database interface to the IDRS by the fourth quarter of Fiscal Year 2013. We reviewed the February 2012 CADE 2 TS1 Integrated Master Schedule which projected deployment of the IDRS interface to production in June 2012. On September 11, 2012, both the *Information Technology Capital Asset Summary for CADE 2* (Exhibit 300A) and the *Performance Measurement Report for CADE 2* (Exhibit 300B) showed delivery of the IDRS interface extract planned for May 2013. However, our review of the Integrated Master Schedule dated August 29, 2012, showed no entries for continued work on the IDRS interface and no planned completion date. IRS management stated that it reassessed the IDRS interface and made a risk-based decision to delay the development and deployment of the IDRS interface.

The IRS had not deployed the CFOL/IMFOL/DAS interface per Enterprise Life Cycle[4] MS 5 requirements stated in IRM 2.16.1 nor demonstrated progress towards deployment of a CADE 2 database/IDRS interface by the fourth quarter of Fiscal Year 2013. However, the CADE 2 Program Management office, with Governance Board concurrence, proposed in April 2013 that the CADE 2 Executive Steering Committee approve both of these interfaces for TS1 MS 5 conditional exit. The IRS concluded that the original intent was to "prove out" that the CADE 2 database can feed downstream systems, and that it met this requirement with the December 2012 database feed to the CFOL/IMFOL/DAS interface. TIGTA does not agree that "proving out" meets the IRM MS 5 exit requirement that the solution is put into use by all users

---

[4] The approach used by the IRS to manage and effect business change. The Enterprise Life Cycle provides the direction, processes, tools, and assets for accomplishing business change in a repeatable and reliable manner.

to conduct IRS business. In addition, TIGTA does not agree that CFOL/IMFOL/DAS interface testing in a production environment validates or "proves out" the design and implementation of the IDRS interface because both interfaces operate under different models. The IDRS interface will extract data daily from the CADE 2 database in batches and send the extract files to the IDRS. The IDRS itself will not connect to the CADE 2 database. The CFOL/IMFOL/DAS interface is a real-time, end user query access into the CADE 2 database to retrieve user requested taxpayer account data.

The CADE 2 database downstream system interfaces did not meet the TS1 MS 5 exit criteria due, in part, to the CADE 2 Program's ongoing challenges of assuring quality data on the CADE 2 database and meeting system performance requirements. These challenges are reflected in the CADE 2 Governance Board's recommendation in April 2013 that the CADE 2 Program shift its focus from further proving out database functionality to getting the data accurate and providing robust and sustainable system performance and operational readiness for the 2014 Filing Season. The CADE 2 Program acknowledged the importance of data quality assurance to the successful deployment of the TS1 and developed a new comprehensive data validation plan for 2013/2014. The Governance Board recommended that data assurance be a new MS 5 exit condition. While the CADE 2 database is not in production, and downstream systems are unable to retrieve and receive data from the CADE 2 database, CADE 2 TS1 implementation continues to be delayed and costs are increasing. Figure 2 illustrates the TS1 MS 5 deployment delays since September 2012 and the rise in deployment costs.

Figure 2: CADE 2 Timeline and Costs

CADE 2 TS 1 MS 5 Deployment Delays and Related Cost Estimates
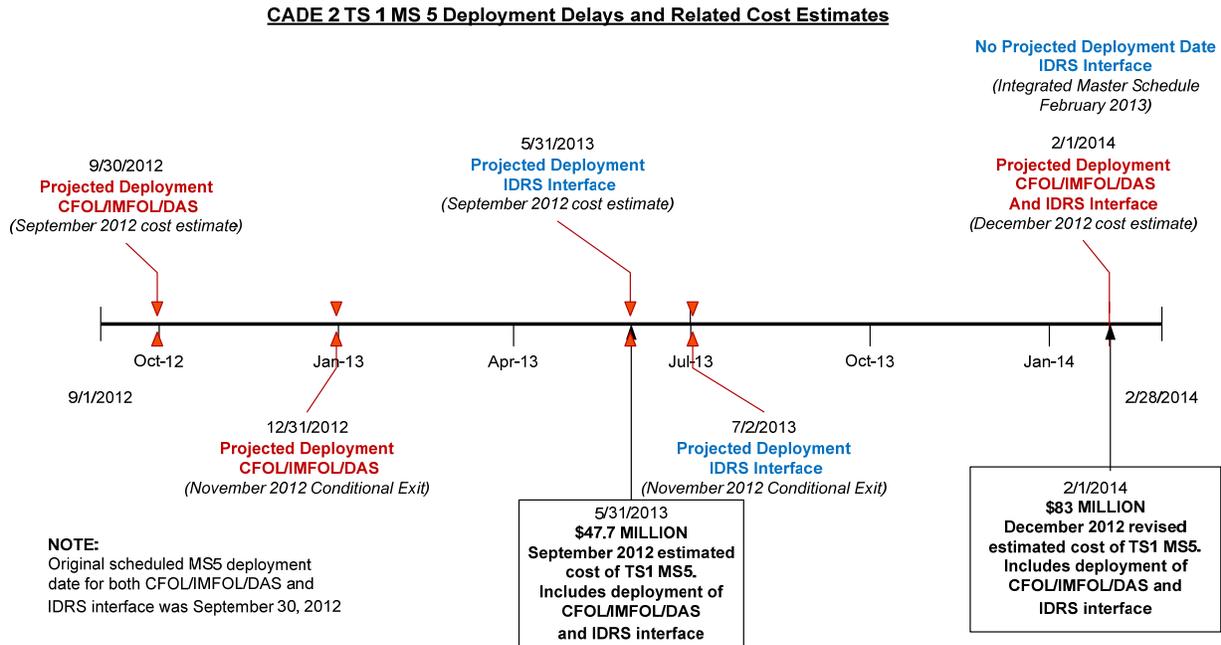
Source: Fiscal Year 2013 Performance Measurement Report; CADE 2 Transition State 1 Integrated Master Schedule Milestone 4b-5; Chief Technology Officer memo, CADE 2 MS 5 Exit Decision Document; Executive Steering Committee Briefing on December 20, 2012; and TIGTA's analysis of Enterprise Life Cycle MS 5 requirements.

CADE 2 TS1 is experiencing increased costs and implementation delays. TS1 MS 5 deployment, which included both the CFOL/IMFOL/DAS interface and the IDRS interface, was estimated in September 2012 to cost $47.7 million. This estimate projected deployment of the CFOL/IMFOL/DAS interface in September 2012 and deployment of the IDRS interface in May 2013. In November 2012, the CADE 2 Executive Steering Committee approved a conditional exit of TS1 MS 5, which delayed deployment of the CFOL/IMFOL/DAS interface to December 31, 2012, and deployment of the IDRS interface to the fourth quarter of Fiscal Year 2013. By December 2012, cost estimates needed to be revised to reflect the reevaluation of IDRS needs and the addition of CFOL, annual IMF conversion, and performance work. The new December 2012 cost estimate for TS1 MS 5 was $83 million, an increase of 74 percent over the September 2012 estimate.

Deployment of both the CFOL/IMFOL/DAS interface and the IDRS interface were included in the December 2012 cost estimate. However, IDRS interface deployment activities were dropped from the Integrated Master Schedule for the 2013 and 2014 Filing Seasons early in 2013. Cost estimates will need to be revised upward again if the IDRS is to be delivered as part of the TS1 as originally planned.

Recommendation

Recommendation 1: The Chief Technology Officer should not exit TS1 MS 5 until the CFOL/IMFOL/DAS and IDRS interfaces are implemented into production.

> **Management's Response:** The IRS disagreed with this recommendation because it has already exited TS1 MS 5 based on the CADE 2 Executive Steering Committee approval on November 5, 2012, with exit conditions. On April 4, 2013, the CADE 2 Executive Steering Committee approved the closure of both milestone exit conditions and approved two new conditions on data assurance, robust and sustainable system performance, and operational readiness. The Executive Steering Committee gave approval to proceed on a release plan and deployment approach for CADE 2 TS1 Database Implementation for the 2014 Filing Season that mitigated risks and ensured a clean and sustainable filing season production deployment. The model to make risk-based decisions exercised by the Executive Steering Committee on April 4, 2013, is part of the overall governance model that the IRS adopted at the beginning of the CADE 2 Program.

> **Office of Audit Comment:** Interfaces from the CADE 2 database to the CFOL/IMFOL/DAS and the IDRS are two of three TS1 interface deliverables documented in the CADE 2 Program Charter. These two interfaces were not delivered and were not in production prior to exiting the TS1. Therefore, the IRS did not meet the conditions for exiting MS 5. Also, during the 2014 Filing Season, the CADE 2 database will not be used to process tax returns.

The Customer Account Data Engine 2 Database Currently Cannot Be Used as a Trusted Source for Downstream Systems

According to industry standards, data quality assurance can be achieved only when the following criteria are met:

- Accuracy: Data must be correct and consistent.

- Completeness: All related data must be linked from all possible sources.

- Availability: Data must be available upon demand.

- Timeliness: Current data must be available.

The IRS conducted data validation tests of the CADE 2 database between July 2012 and December 2012. Our review of the test results determined data in the CADE 2 database do not meet accuracy and completeness or availability criteria.

The IMF tape files are the current source of data for downstream systems. Figure 3 shows the current CFOL/IMFOL/Virtual Storage Access Method (VSAM) interface with the IMF tape files. This solution provides accurate daily updated IMF data to CFOL/IMFOL users.

Figure 3: The IMF VSAM CFOL/IMFOL Interface Data Flow
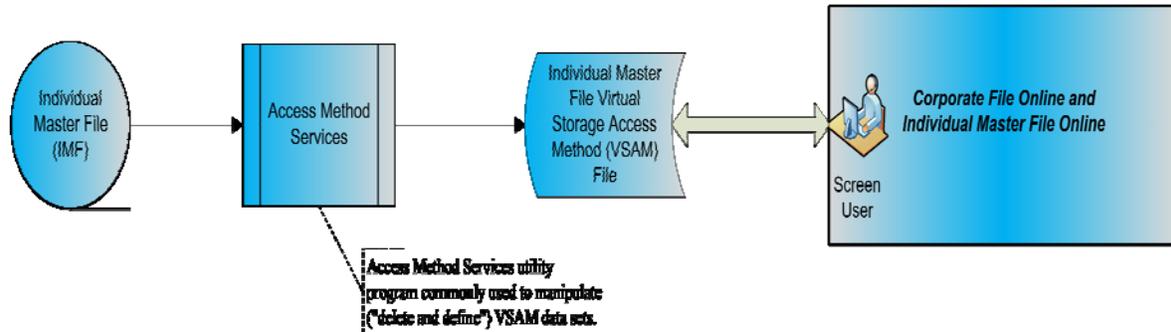
Source: Based on the IRS Computer Operator Handbook.

Figure 4 shows the data migration process and the flow of data through the CADE 2 database/CFOL/IMFOL/DAS interface.

Figure 4: The CADE 2 Database CFOL/IMFOL/DAS Interface Data Flow
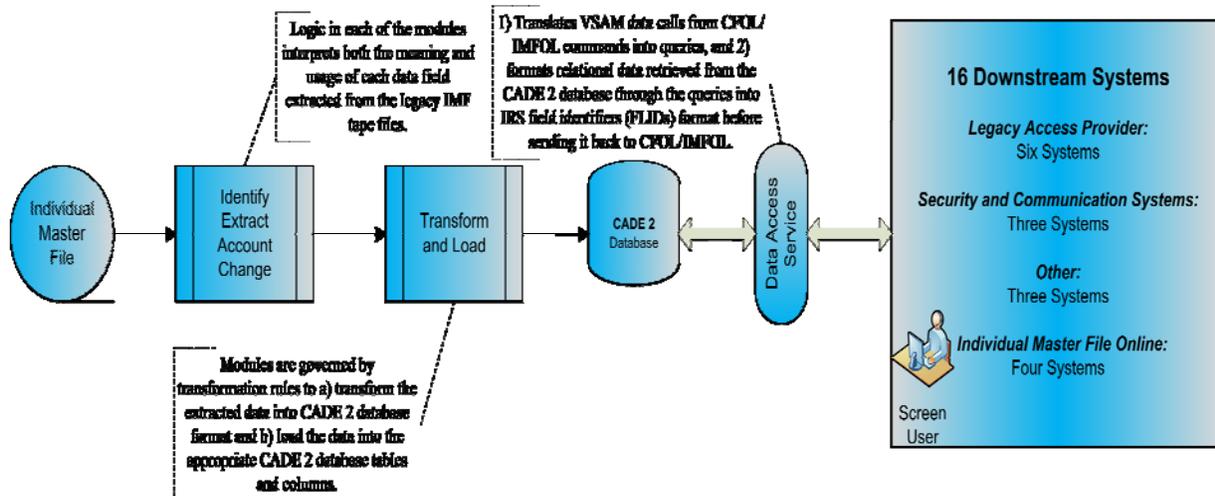
Source: TIGTA's review of the CADE 2 Database Implementation Design Specification Report.

Figures 3 and 4 show that the CFOL/IMFOL/VSAM interface and the CFOL/IMFOL/DAS interface data flows use the same IMF tape files as their original data source. While the IMF tape file data successfully migrates to the IMF VSAM files, the migration of data from the IMF tape files to the CADE 2 database has had data quality issues. Between July 2012 and

January 2013, data validation tests of the CADE 2 TS1 design resulted in 1,006 defect tickets. The data validation test results led to the IRS applying 2.4 million data corrections to the CADE 2 database.

Figure 5 provides a summary of the data correction counts. Of the 2.4 million data corrections, more than 2.2 million (almost 92 percent) resulted from the Extract, Transform, and Load process.

Figure 5: The CADE 2 Database Data Corrections

Component/Process	Data Correction Count
Identify and Extract Account Changes - Extract	1,383,265
Transform and Load	890,638
IMF Reel Replacements	11
Manual Refund Transaction, *i.e.*, Transaction Code 840	137,653
TOTAL	**2,411,567**

Source: CADE 2 Executive Status Update January 2, 2013, and CADE 2 development teams.

The Extract, Transform, and Load process is the core process of the CADE 2 database's continuous data migration process of Initialization, Cycle Synchronization, and Daily/Weekly Update. In order for the CADE 2 TS1 design to replicate the success of the IMF data migration process, the interpretation of the meaning and usage of data from the IMF files and fields must be consistent between the Identify and Extract Account Changes modules, the Transformation and Load modules, and the CADE 2 database tables and columns; otherwise, it will result in a data defect. Although the current CADE 2 data validation effort attempts to validate the data quality of the CADE 2 database, it compares only the IMF VSAM data against the end result of the data migration process: the contents of the CADE 2 database. The current CADE 2 data validation does not review the entire data migration process of Extract, Transform, and Load where data and code defects were identified. A validation that interprets the meaning and usage of IMF data and ensures that it is consistent throughout the entire data migration and retrieval process would prevent future data defects from occurring.

Another issue we identified from the data validation tests is that the IMFOL Screen Compare data validation test tool used by the IRS was capable of validating only 55 percent (533 of 964) of the columns in the CADE 2 database. As of April 2013, the IMFOL Screen Compare was the only automated high volume data validation tool available. Given that the IRS could not evaluate 431 columns for data accuracy, there are potentially more data defects yet to be discovered. The IRS is developing additional tools and implementing a new data validation testing methodology intended to achieve timeliness, accuracy, integrity, validity, reasonableness,

completeness, and uniqueness. The IRS requested that TIGTA review the new data validation testing methodology. The effectiveness of the IRS's new methodology will be evaluated in a separate audit.

While data defects challenge CADE 2 data quality assurance, the completeness of the database is also in question. For example, when the daily update process cannot successfully update Taxpayer Identification Numbers, *i.e.*, taxpayer accounts, that have a known data problem, the daily update process continues and leaves those accounts incomplete on the CADE 2 database until they can be updated at a later time. While these accounts are awaiting update, the DAS has to retrieve its data from the IMF VSAM files, not the CADE 2 database. This is known as the Taxpayer Identification Number-bypass solution. The CADE 2 database is incomplete while the update issues are being resolved. The Taxpayer Identification Number-bypass solution is an acceptable process once the CADE 2 database is in production and is serving as a transactional database. This solution should not be used for the migration of processed data from the IMF to the CADE 2 database.

In addition, not all taxpayer account data requested by downstream systems are available on the CADE 2 database. Archived taxpayer data on the Recoverable Retention Register remain only on the IMF VSAM files; the data are not migrated to the CADE 2 database in the TS1. Therefore, the CADE 2 database is not a complete representation of IMF taxpayer account data. IMF VSAM files must be maintained and kept operational along with the CADE 2 database in order to ensure that CFOL/IMFOL users have access to complete taxpayer account data.

The CADE 2 database is not in daily operation in the production environment; therefore, it is not available for the downstream system CFOL/IMFOL interface, and TIGTA cannot evaluate its timeliness. The lack of data accuracy, completeness, and availability prevents the CADE 2 database from serving as the trusted data source for the downstream systems.

Recommendation

Recommendation 2: The Chief Technology Officer should certify the CADE 2 database is accurate, complete, timely, and available to serve as the trusted source.

> ***Management's Response:*** The IRS agreed with the recommendation, and controls are in place to ensure that this occurs. On April 4, 2013, the CADE 2 Executive Steering Committee approved the CADE 2 TS1 MS 5 exit conditions focusing on data assurance, robust and sustainable system performance, and operational readiness. These exit conditions will ensure that the database is accurate, complete, and can be updated timely before it is made available to serve as a trusted source for tax processing.

The Customer Account Data Engine 2 Database Interface Solution Architecture Does Not Meet the Business Needs

The DAS interface is designed to be an interface between the CADE 2 database and the existing CFOL/IMFOL user interface. In the TS1, 16 of 18 downstream systems that are currently using the CFOL/IMFOL with the IMF VSAM files as the data source will use the CFOL/IMFOL/DAS interface with the CADE 2 database as the data source. The CFOL/IMFOL/DAS interface assists these 16 systems to directly retrieve data from the CADE 2 database, and transform and deliver data in the same format that the 16 downstream systems' users are accustomed to receiving. The CFOL/IMFOL/DAS interface is expected to:

- Serve up to 15,000 concurrent users.

- Provide 24-hour access seven days a week to daily updated IMF data from the CADE 2 database.

- Deliver 400,000 transactions per peak hour and serve up to 3.8 million data calls a day.

To meet the business needs, the CADE 2 Program implemented a dual database solution architecture, *i.e.*, an active database and a replica database. The CADE 2 active database is designed to support all future IMF transaction processing and serve as the back-up data source for the CFOL/IMFOL interface. The CADE 2 replica database serves as the primary data source for the CFOL/IMFOL interface and its 16 downstream systems. The replica database is cloned from the CADE 2 active database. It requires 4.5 hours to complete cloning and consumes approximately 1,400 Million Instructions Per Second (MIPS) during cloning.

Figure 6 shows the CADE 2 database CFOL/IMFOL/DAS systems interface performance statistics during the December 2012 limited production deployments in comparison to its expected performance statistics, and with the IMF VSAM - CFOL/IMFOL interface performance statistics.

Figure 6: Comparison of IMF VSAM Interface Versus CADE 2 Database Interface Performances

	IMF VSAM - CFOL/IMFOL Interface	CADE 2 Database - CFOL/IMFOL/DAS Interface Expected	CADE 2 Database - CFOL/IMFOL/DAS Interface December 2012
MIPS	1.6	3.6-4.1	16.1
Response Time in Seconds	< 1	< 1	7.2

Source: The CADE 2 Database Implementation Performance Review Summary, the CADE 2 Database Implementation Design Specification Reports, the CADE 2 Database Implementation Design Effort Approach, and the development team's confirmation.

The CFOL/IMFOL/DAS interface performance statistics for response time and MIPS were several times that of CFOL/IMFOL/VSAM interface, and more than 700 percent greater than the expected performance statistics. However, due to the limited production deployment not accurately representing the production environment, we believe the actual performance numbers will be worse for the following reasons:

- The cloning process, which can consume 1,400 MIPS for 4.5 hours, was idle.

- The Daily/Weekly Update process including "insert," "delete," and "update" operations did not take place.

- The 16 downstream systems were not included in the tests, where the DAS could have retrieved and transformed 60 million taxpayer accounts in a single request.

- Only six users participated in the test versus the 15,000 concurrent users expected in production.

- Up to 100 report users who access the CADE 2 database via Business Object.

IRS management stated that system performance has improved since our analysis of the December 2012 test in production. The IRS made changes to the extract, transform, and load code and reduced the MIPS consumption per transaction to 8.8 MIPS during Final Integration Testing.

In addition, to resolve the CFOL/IMFOL/DAS interface performance issues experienced during the limited production deployments, the IRS implemented additional indexes to the CADE 2 database. However, indexes negatively affect the CFOL/IMFOL/DAS interface solution architecture and affect its ability to deliver business needs. Our review of best practices confirms that indexes applied to relational databases degrade the performance of "insert," "delete," and "update" operations. In TS1, the same CADE 2 database indexes implemented to improve the DAS's query performance would negatively affect the Daily/Weekly Update process of "insert," "delete," and "update" operations that update the CADE 2 active database with IMF data. It has been documented that on average, there would be 350,000 inserts and 5,142,857 updates to the CADE 2 database per day. If the average number of indexes for a CADE 2 database table is four, then there could be 1,750,000 insert operations per day (350,000 x 5, one for the data row plus four for index rows).

When the CADE 2 database is serving as the transactional database in production, its primary goal will be servicing the online transactions of "insert," "delete," and "update." By adding indexes to improve DAS query performance, the CADE 2 online transaction performance will be negatively affected as previously described. However, the performance impact will be directed toward the user community during office hours or peak hours, and magnified as these online transactions are not being executed in the background at night like the Daily Update. In addition, when the active database is serving as the data source, DAS queries will compete directly against the active database's transactional operations resulting in a resource contention and resource

constraint situation within the system. As a result, not only are online transactions affected, the downstream system performance will continue to get worse.

The IRS identified MIPS consumption as a risk to the CADE 2 Program. The query performance, database cloning, Daily/Weekly update, IMF VSAM interface operation, CFOL/IMFOL/DAS interface operation, and potential large amounts of data retrieved by downstream systems will compete for MIPS/computing resources and CADE 2 database services on a regular basis. This will create resource contention that could affect the system performance. The dual database solution has a fundamental design issue because the transactional CADE 2 database also serves as the reporting database for the downstream systems. In reviewing the identified issues, we conclude that the CADE 2 database CFOL/IMFOL/DAS interface solution architecture does not meet the IRS's business needs. These issues will affect IRS tax administration processes and the quality of the service provided to taxpayers.

Recommendation

Recommendation 3: The Chief Technology Officer should deploy the CADE 2 database as the transactional database. Architect the CADE 2 database for the TS2 and the Target State, as the authoritative data source for an enterprise data warehouse or a data mart, and not as a direct data source for downstream systems.

> **Management's Response:** The IRS disagreed with this recommendation. The CADE 2 database in the TS1 is being deployed as the transactional database foundation and is built to support the TS2 and the Target State. In the Target State, the CADE 2 database will serve as the consolidated authoritative source of taxpayer account data for an integrated data warehouse and for common services that provide for Online Transaction Processing. The IRS acknowledges that during the transition phases, the CADE 2 database may serve indirectly as a data source for downstream systems through the CFOL/IMFOL/DAS interface. The IRS is making architecture decisions to use transition components to bridge to existing systems to achieve incremental modernization.

> **Office of Audit Comment:** Our review of the solution architecture indicated the CADE 2 database will serve as a transactional database and a direct source for the downstream systems in the production environment. In addition, the IRS identified MIPS consumption as a risk to the CADE 2 Program. The query performance, database cloning, Daily/Weekly update, IMF VSAM interface operation, CFOL/IMFOL/DAS interface operation, and potential large amounts of data retrieved by downstream systems will compete for MIPS/computing resources and CADE 2 database services on a regular basis.

The Lack of Security Systems Integration Prevents Transaction-Level Tracking of Employee Access to the Customer Account Data Engine 2 Database

The CADE 2 – CFOL/IMFOL systems interface uses two security systems to provide user authentication and access control and auditing functionality. The two security systems are:

- Security and Communication System: Authentication system for end users in the CFOL/IMFOL systems.

- Resource Access Control Facility: Authentication for the system account used to access the CADE 2 database.

We reviewed logs generated by both security systems to verify the CADE 2 data calls (requests) and user authentication processes. The Security and Communication System logs trace individual user authentication in the CFOL/IMFOL systems prior to accessing the DAS system account. However, the database audit logs track the Resource Access Control Facility system account that accesses the CADE 2 database. For example, the database audit logs we reviewed indicate that for all recorded transactions, the Source User Name and Destination User Name are the same system account with no individual users identified.

Our tests determined access to taxpayer data in the CADE 2 database does not comply with IRM 10.8.32.2.1, *Mainframe System Security Requirements*, which ensures that the users are who they say they are and identifies the resources, datasets, and transactions that they are allowed to access. TIGTA attempted to trace a transaction from the CFOL/IMFOL systems data call to the CADE 2 database and was unable to follow the transaction once it passed through the DAS system account. Figure 7 illustrates a data call and the authentication process.

Figure 7: CFOL/IMFOL Command Call to CADE 2 Database
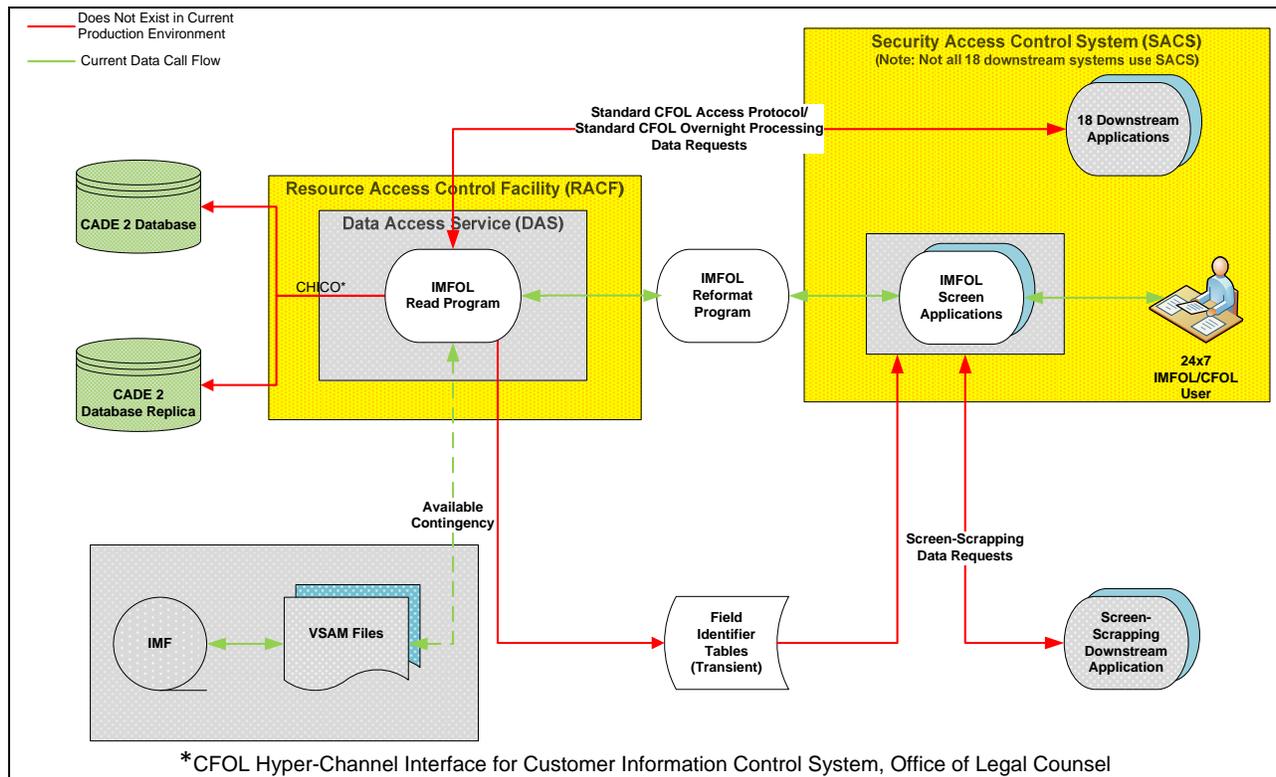
Source: CADE 2 Simulation/Monitoring Room Tabletop Exercise: CFOL/IMFOL DAS – Initial Deployment Process Steps, November 4, 2012.

The current system design does not integrate the two security systems. Therefore, the original user identification from the Security and Communication System cannot be passed to the Resource Access Control Facility system to ensure traceability from the CFOL/IMFOL systems to the CADE 2 database. As a result, all DAS system account requests from the CFOL/IMFOL systems to the CADE 2 database have the system account as the valid user, *i.e.*, all CFOL/IMFOL systems users access the CADE 2 database by going through a single DAS system account. This can lead to unauthorized access to taxpayer data being undetected by audit logs.

Recommendation

Recommendation 4: The Chief Technology Officer should ensure that user access to the CADE 2 database is traced at the transaction level by individual user's identification.

> **Management's Response:** The IRS disagreed with this recommendation. Users of the downstream applications such as CFOL/IMFOL are not authorized to directly access the CADE 2 database. Therefore, there are no permissions or access provisioned in the

Resource Access Control Facility system. Auditing for CFOL/IMFOL users is performed at the application level, and access is granularly controlled with checks validating the user is authorized to execute the command code in question from the terminal the user is on. The CFOL service account is granted access and appropriate permissions in the Resource Access Control Facility system to execute queries that access the CADE 2 database; these queries are audited by the Guardium appliance. This architecture provides auditing of user activity at the application level, auditing of application tier activity at the database level, and enforcement of the principles of least privilege and separation of duties because end users are not provisioned permissions to the database that are not required.

Office of Audit Comment: We believe user access risk to sensitive data is not properly mitigated. Users of downstream systems access the CADE 2 database through a system account. Users' credentials are not passed to the system account along with the command codes. Therefore, the system account does not allow traceability of individual users at the transaction level.

Detailed Objective, Scope, and Methodology

The overall objective of this review was to determine whether the IRS has implemented adequate CADE 2 database downstream interface data validation to ensure that the data provided are accurate and complete. To accomplish our objective, we:

I.   Determined if the CADE 2 – CFOL/IMFOL interface solution architecture delivers the business users' needs.

   A. Interviewed the Subject Matter Expert to understand the CADE 2 – CFOL/IMFOL interface solution architecture.

   B. Obtained and reviewed the CADE 2 – CFOL/IMFOL interface design and testing documents to ensure that the interface supports downstream systems.

   C. Determined if the CFOL/IMFOL performance statistics meet the required system performance objectives.

II. Determined if the operation of the CADE 2 – CFOL/IMFOL interface will ensure the availability, accuracy, and timeliness of the data.

   A. Obtained and reviewed the Change Control Management procedure of CFOL/IMFOL software and data components.

   B. Determined if support procedures and organizations have been implemented to ensure the existing CFOL/IMFOL systems availability.

   C. Determined the impact of the CFOL/IMFOL on downstream systems.

III. Determined if Personally Identifiable Information is properly secured and CFOL/IMFOL systems access is properly managed to prevent its exposure, mitigate system downtime, and identify the potential for fraud, waste, and abuse.

   A. Obtained and reviewed downstream system documents and procedures.

   B. Obtained and reviewed documents of access control and change control management regarding the security of CFOL/IMFOL systems software and data components.

   C. Obtained and reviewed the security integration of the IDRS, Security and Communication System, CFOL/IMFOL, DAS, CADE 2 database, and the Resource Access Control Facility system.

IV. Determined the progress of the CADE 2 database/IDRS interface implementation to resolve the TS1 MS 5 conditional exit.

A. Determined if the CADE 2 database/IDRS interface development progress is consistently communicated and on schedule to resolve the MS 5 conditional exit.

B. Interviewed the Subject Matter Expert to understand the CADE 2 database/IDRS interface solution architecture.

C. Obtained and reviewed the interface design and testing documents to ensure that the interface supports downstream systems.

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 and interfaces. We evaluated these controls by conducting interviews and meeting with IRS management and staff; attending CADE 2 meetings; and reviewing the CADE 2 Program Charter, CADE 2 Solution Architecture, CADE 2 Database Implementation Design Specification Report, CADE 2 Database Implementation Design and Performance Overview, CADE 2 Database Implementation Data Validation Plan, Data Validation Strategy – Smart Sampling, DAS Performance Summary, and other documents that provided evidence of whether the IRM database development and systems testing processes were followed and if those processes were adequate and operating as designed.

Major Contributors to This Report

Alan R. Duncan, Assistant Inspector General for Audit (Security and Information Technology Services)
Danny Verneuille, Director
K. Kevin Liu, Audit Manager
Hung Q. Dam, Lead Auditor
Myron Gulley, Senior Auditor
Arlene Feskanich, Information Technology Specialist
Nicholas Reyes, Information Technology Specialist

Report Distribution List

Acting Commissioner
Office of the Commissioner – Attn: Chief of Staff C
Deputy Commissioner for Operations Support OS
Deputy Chief Information Officer for Strategy and Modernization OS:CTO
Associate Chief Information Officer, Application Development OS:CTO:AD
Associate Chief Information Officer, Cybersecurity OS:CTO:C
Associate Chief Information Officer, Enterprise Information Technology Program Management Office OS:CTO:MP
Associate Chief Information Officer, Enterprise Operations OS:CTO:EO
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

Management's Response to the Draft Report

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

CHIEF TECHNOLOGY OFFICER

September 9, 2013

MEMORANDUM FOR DEPUTY INSPECTOR GENERAL FOR AUDIT

FROM: Terence V. Milholland
Chief Technology Officer

SUBJECT: Customer Account Data Engine 2 Database
Deployment Is Experiencing Delays and Increased
Costs
(Audit 201320021) (e-trak # 2013-46420)

Thank you for the opportunity to review your draft audit report and discuss earlier draft report observations with the audit team. I was pleased to read your observation acknowledging the use of cross-functional triage teams by the Customer Account Data Engine 2 (CADE 2) program to effectively manage and resolve data and code defects on the database.

Attached is our corrective action plan. While we agree with your recommendation around certifying readiness of the database to serve as the trusted source, we disagree with the other three recommendations that question aspects of our program governance process, guiding principles and solution architectural design. We believe these recommendations question foundational decisions and management judgments rather than audit the processes we used to come to the decisions and judgments.

We would like to acknowledge that with TIGTA's active involvement with the program every step of the way, the IRS fully leveraged its governance process and used responsible decision-making and guiding principles to deploy a sound database solution. And with the use of risk-based decisions along the way, the IRS was able to put timely and necessary emphasis on data quality, organizational readiness, and performance, which were instrumental to reaching Transition State 2 and our target state, versus substantially increasing risks by following an approach driven by a hard and fast schedule to deploy.

We believe this audit does not take into account the comprehensive CADE 2 Program approach as it relates to implementation of the database and the underlying risk mitigations that account for the delays and increased cost. Nonetheless, we are committed to continuously improving our information technology systems and processes. 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 Karen Mayr at (240) 613-1431.

Attachment

Customer Account Data Engine 2 Database Deployment Is Experiencing Delays and Increased Costs (Audit 201320021

RECOMMENDATION #1: The Chief Technology Officer should not exit Transition State 1 Milestone 5 until the CFOL/IMFOL/DAS and IDRS interfaces are implemented into production.

CORRECTIVE ACTION #1: The IRS disagrees with this recommendation since the IRS has already exited Transition State 1, Milestone 5. On November 5, 2012, the CADE 2 Executive Steering Committee (ESC) approved the exit of Transition State 1, Milestone 5 with the following conditions: 1) availability of Corporate Files Online/Individual Master File Online/Data Access Service (CFOL/IMFOL/DAS) to CADE 2 Databases in Production (December 2012), and 2) deployment of CADE 2/Integrated Data Retrieval System (IDRS) Interface (Q4 FY 2013).

On April 4, 2013, the CADE 2 Executive Steering Committee approved closure of both milestone exit conditions, agreeing with the CADE 2 Governance Board that the original scope intent of the conditions had been satisfied and the functionality demonstrated with the production database feed to CFOL/IMFOL/DAS in December 2012.

They also approved the adoption of two new conditions on data assurance and robust and sustainable system performance and operational readiness. Lastly, they gave approval to proceed on a release plan and deployment approach for CADE 2 Transition State 1 Database Implementation for the 2014 Filing Season that mitigated risks and ensured a clean and sustainable filing season production deployment. The approved plan reflected a shift in priority for Database Implementation, from further proving out functionality, to a keen focus on "statistics accuracy" and "robust and sustainable system performance and operational readiness".

The model to make risk-based decisions exercised by the ESC on April 4, 2013, is part of the overall governance model that the IRS adopted at the beginning of the CADE 2 program and has used throughout its lifecycle. This model allows flexibility for the governance bodies to make critical decisions on projects based on new discoveries or unexpected circumstances to ensure the program stays on course to deliver intended business value and achieve target state within reasonable costs and time span.

IMPLEMENTATION DATE: N/A

RESPONSIBLE OFFICIAL: N/A

CORRECTIVE ACTION MONITORING PLAN: N/A

RECOMMENDATION #2: The Chief Technology Officer should certify the database is accurate, complete, timely, and available to serve as the trusted source.

CORRECTIVE ACTION #2: The IRS agrees with the recommendation, and controls are in place to ensure this occurs. On April 4, 2013, the CADE 2 ESC approved the current CADE 2

1

Customer Account Data Engine 2 Database Deployment Is Experiencing Delays and Increased Costs (Audit 201320021

Transition State 1 Milestone 5 exit conditions focusing on data assurance, robust and sustainable system performance, and operational readiness. These exit conditions will ensure that the database is accurate, complete and can be updated timely before it is made available to serve as a trusted source for tax processing.

IMPLEMENTATION DATE: September 25, 2014

RESPONSIBLE OFFICIAL: ACIO Enterprise IT Program Management Office

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 #3: The Chief Technology Officer should deploy the CADE 2 database as the transactional database. Architect the CADE 2 database for the TS 2 and Target State, as the authoritative data source for an enterprise data warehouse or a data mart; and not as a direct data source for downstream systems.

CORRECTIVE ACTION #3: The IRS disagrees with the recommendation. The design of the CADE 2 database is consistent with the IRS Enterprise Data Management Roadmap version 4 (May 2009), which calls for a series of transitional phases in accordance with the overall IRS data strategy. The CADE 2 database in Transition State (TS) 1 is being deployed as the transactional database foundation and is built to support TS 2 and the Target State. In the Target State, CADE 2 will serve as the consolidated authoritative source of taxpayer account data for an integrated data warehouse and for common services that provide for Online Transaction Processing, including an account view via data access services that provide a data abstraction layer.

The IRS acknowledges that during the transition phases, CADE 2 may serve indirectly as a data source for downstream systems through the CFOL/IMFOL/DAS interface. The overall CADE 2 transition strategy and resulting architecture decisions for TS 1 were created with the "do no harm" (e.g., protect the filing season) guiding principle in mind. During the transition from the CADE 2 Beginning State to Target State, the IRS is making architecture decisions (in some cases) to use transitional components to bridge to existing systems to achieve incremental modernization. In the process, the IRS gains more confidence and understanding on how the database can be used as a source of critical business data in a modernized architecture.

In addition, it is important to understand that the IRS balances the need to minimize risk and manage scope with potential performance tradeoffs. The IRS does not believe there is sufficient evidence to serve as a reasonable basis for the assertion that CADE 2 "indexes negatively impact the CFOL/IMFOL/DAS interface solution". Indexes are equally necessary for CADE 2 "update" operations and CFOL/IMFOL/DAS queries. There are few exceptions where an index was solely created to support the CFOL/IMFOL/DAS interface and the example cited of four indexes per table is inconsistent with the Design Specification Report version 1.5.1 which specifies an average of less than 2 indexes per table.

2

Customer Account Data Engine 2 Database Deployment Is Experiencing Delays and Increased Costs (Audit 201320021

IMPLEMENTATION DATE: N/A

RESPONSIBLE OFFICIAL: N/A

CORRECTIVE ACTION MONITORING PLAN: N/A

RECOMMENDATION #4: The Chief Technology Officer should ensure that user access to the CADE 2 database is traced at the transaction level by individual user's identification.

CORRECTIVE ACTION #4: The IRS disagrees with this recommendation. Users of downstream applications such as CFOL/IMFOL are not authorized to directly access the CADE 2 database. Therefore, there are no permissions or access provisioned in Resource Access Control Facility (RACF) and auditing for CFOL/IMFOL users is performed at the application level and access is granularly controlled with checks validating the user is authorized to execute the command code in question from the terminal the user is on. Command codes executed by users are audited and security violations are flagged and referred for review. CFOL/IMFOL users do not access the service account to directly execute queries to retrieve data from the CADE 2 database; the CFOL application uses the service account to execute queries. The CFOL service account is granted access and appropriate permissions in RACF to access the CADE 2 database and execute queries and its queries are audited by the Guardium appliance. This architecture provides auditing of user activity at the application level, auditing of application tier activity at the database level and provides enforcement of the principles of least privilege and separation of duties because end users are not provisioned permissions to the database that are not required.

IMPLEMENTATION DATE: N/A

RESPONSIBLE OFFICIAL: N/A

CORRECTIVE ACTION MONITORING PLAN: N/A

3