TREASURY INSPECTOR GENERAL FOR TAX ADMINISTRATION



The Customer Account Data Engine Release 2.2 Posted Tax Return Information Accurately

October 29, 2007

Reference Number: 2008-40-004

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/ 202-622-6500Email Address/ inquiries@tigta.treas.govWeb Site/ http://www.tigta.gov



FOR TAX ADMINISTRATION

DEPARTMENT OF THE TREASURY

WASHINGTON, D.C. 20220

October 29, 2007

MEMORANDUM FOR COMMISSIONER, WAGE AND INVESTMENT DIVISION

FROM:

Michael R. Phillips Deputy Inspector General for Audit

SUBJECT:Final Audit Report – The Customer Account Data Engine Release 2.2
Posted Tax Return Information Accurately (Audit # 200740039)

This report presents the results of our review to determine whether Customer Account Data Engine (CADE) Release 2.2 accurately posted tax return information, focusing on functionality not included in prior releases. This report is the fourth in a series of ongoing reviews of production releases of the CADE¹ meant to determine if the deployed production versions of the CADE are working as designed.

Impact on the Taxpayer

The CADE is the new computer system for taxpayer account information. It is being implemented in phases and will eventually replace the current aging Master File.² It is essential that the tax return information posted to the CADE is accurate because inaccurate information could affect subsequent activities or adjustments to taxpayer accounts. The Release 2.2 version of the CADE accurately posted tax return information to approximately 11 million taxpayer accounts. However, delays in implementing Release 2.2 prevented an additional estimated 22 million tax returns from posting to the CADE. As such, many taxpayers did not receive the benefits the CADE provides, such as faster refunds.

¹ Prior reports were *Individual Income Tax Return Information Was Accurately and Timely Posted to the Customer Account Data Engine* (Reference Number 2005-40-109, dated July 2005), *Individual Income Tax Return Information Was Successfully Posted to the Latest Release of the Customer Account Data Engine* (Reference Number 2006-40-134, dated August 29, 2006), and *The Customer Account Data Engine Release 2.1 Generally Posted Tax Return Information Accurately* (Reference Number 2007-40-131, dated August 10, 2007).

² The Internal Revenue Service database that stores various types of taxpayer account information. This database includes individual, business, and employee plans and exempt organizations data.



<u>Synopsis</u>

The CADE consists of current and planned databases and related applications that will eventually replace the Internal Revenue Service (IRS) Master File. It is a critical building block in the IRS' modernization program that will enable the development of subsequent modernized systems to improve customer service and compliance. The CADE is designed to post information to taxpayers' accounts daily rather than weekly, which will facilitate faster refunds to taxpayers and allow IRS employees to provide improved service to taxpayers because the employees will have up-to-date, accurate account information available.

The CADE is being implemented in a series of releases to be placed in production in the coming years. The first CADE release began posting the simplest individual tax returns³ in July 2004. The subject of this review, Release 2.2, began posting returns in March 2007 and contained significant additions over prior releases, including the ability to post returns with the most common filing statuses and the most common schedules such as those related to self-employment.⁴ The IRS successfully posted approximately 11 million tax returns to the CADE and generated approximately \$11.6 billion in refunds from March through August 2007.

The IRS originally estimated that it would post 33 million returns to the CADE for the 2007 Filing Season.⁵ However, implementation of Release 2.2 was postponed from January to March 2007 so the IRS could make required performance improvements and complete filing season upgrades. Accordingly, this delay resulted in the actual number of returns posted being approximately 22 million fewer than the original estimate.

Overall, tax return information was accepted and posted accurately to CADE accounts during the 2007 Filing Season. We reviewed samples of tax returns posted with the various attributes the CADE can now accept, focusing primarily on the attributes that were new to Release 2.2. However, we also verified some attributes from prior releases.⁶ In all cases, the tax return information was posted to taxpayers' accounts accurately.

The ability to accept and post returns with self-employment information was added to the CADE on May 30, 2007. We reviewed judgmental samples of returns posted after this date that had self-employment schedules. In all cases, the posted information was accurate, and we did not identify any discrepancies. Also, the self-employment information being provided in a weekly report to the United States Census Bureau was accurate. Finally, the process used to determine the audit potential of returns posting to the CADE was adequately tested.

³ Income Tax Return for Single and Joint Filers With No Dependents (Form 1040EZ).

⁴ Profit or Loss From Business (Schedule C), Profit or Loss From Farming (Schedule F), and Self-Employment Tax (Schedule SE).

⁵ The period from January through mid-April when most individual income tax returns are filed.

⁶ See Appendix V for a list of all attributes reviewed.



Although the information posted to the CADE was accurate, we did identify one condition that could have resulted in \$400,000 of overpaid taxes being incorrectly refunded to taxpayers. Accounts with special processing codes that were intended to stop refunds so any overpayments could be applied to amounts owed on related business accounts were incorrectly added to the CADE database. However, because the CADE was not yet programmed to accept these codes, the overpayments on these accounts would have been incorrectly refunded.

We identified this issue soon after the CADE began posting returns in March 2007 and referred it to the CADE Project Office. The prompt corrective action taken at that time prevented these accounts from having the overpayments refunded. However, had this condition not been identified and corrected early, the CADE would have allowed these overpayments to refund rather than be applied to the related balance-due accounts.

Recommendations

We made no recommendations in this report. However, key IRS management officials reviewed the report prior to issuance and agreed with the facts and conclusions presented.

Copies of this report are also being sent to the IRS managers affected by the audit results. Please contact me at (202) 622-6510 if you have questions or Michael E. McKenney, Assistant Inspector General for Audit (Wage and Investment Income Programs), at (202) 622-5916.



Table of Contents

Background	1
Results of Review	4
Tax Return Information Was Posted AccuratelyPage	4
Taxpayer Accounts With Special Conditions Were Incorrectly Added to the Customer Account Data EnginePage	5

Appendices

Appendix I – Detailed Objective, Scope, and Methodology	.Page	7
Appendix II – Major Contributors to This Report	.Page	10
Appendix III – Report Distribution List	.Page	11
Appendix IV – Outcome Measure	.Page	12
Appendix V – Tax Return Attributes Reviewed	.Page	13



Abbreviations

CADE	Customer Account Data Engine
IRS	Internal Revenue Service



Background

The Customer Account Data Engine (CADE) is the foundation for managing taxpayer accounts in the Internal Revenue Service's (IRS) modernization plan. It consists of current and planned databases and related applications that will eventually replace the IRS Master File.¹ The CADE is a critical building block in the IRS' modernization program that will enable the development of subsequent modernized systems to improve customer service and compliance. When fully operational, the CADE will house tax information for more than 200 million individual and business taxpayers.

Tax returns, both paper and electronic, are received by the IRS at various Submission Processing sites across the country. At these sites, the tax return information is input to the IRS return processing computer system,² which validates certain taxpayer identifying information and checks the tax returns for mathematical errors. After the tax return information has been validated and errors have been corrected, the sites send the information to the IRS Computing Centers³ to be posted to taxpayers' accounts. In the past, this account information was posted to the Master File, the official repository for all tax accounts.

The age and complexity of the Master File causes inaccuracies and delays in providing service to taxpayers. Updates to taxpayers' account information on the Master File (such as posting payments) occur weekly, and some updates require multiple weeks to complete. Because current data are not always available to IRS employees, taxpayers requesting help with their accounts may be given outdated information. In contrast, the CADE is designed to post information to taxpayers' accounts daily rather than weekly. Taxpayers should receive refunds faster, and IRS employees should be able to provide improved service to taxpayers because the employees will have up-to-date, accurate account information available.

The first phase of the CADE, which is for individual taxpayer accounts, is being implemented in a series of releases to be placed in production in the coming years. The simplest taxpayer accounts were moved to the CADE first, and each successive release adds a more complex segment of taxpayer accounts. The first CADE release began posting the simplest tax returns⁴ in July 2004; subsequent releases have added new functionality and tax form types. The subject of this review, Release 2.2, began posting returns in March 2007 and contained significant additions

¹ The IRS database that stores various types of taxpayer account information. This database includes individual, business, and employee plans and exempt organizations data.

² Also referred to as the Generalized Mainline Framework.

³ IRS Computing Centers support tax processing and information management through a data processing and telecommunications infrastructure.

⁴ Income Tax Return for Single and Joint Filers With No Dependents (Form 1040EZ).



over the prior releases. This Release accepted, for the first time, the most common filing statuses, the most common schedules, and limited dependency exemptions. Figure 1 outlines the basic additions to Release 2.2.

Return Characteristics	Prior Releases	Release 2.1	Release 2.2	
Tax Returns	Forms 1040EZ, 1040A, and 1040 ⁵	Same as prior releases	Form 1040EZ-T ⁶	
Filing Status	Single only	Single and Head of Household without dependents	Single, Married Filing Jointly, Married Filing Separately, and Head of Household	
Dependents	No dependents	No dependents	Limited dependents	
Schedules	No schedules	Schedules A, B, and R^7	Schedules C, D, E, F, and SE ⁸	
Tax Credits	Withholding credit	Withholding credit	Withholding credit Credit for Federal Telephone Excise Tax paid	
Name Changes	No name changes	Limited name changes	Same as Release 2.1	
Address Changes	Address changes on return	Address changes on return, United States Postal Service address change updates	Same as Release 2.1	

Figure 1: Functionality Added to CADE Release 2.2

Source: CADE Release Content Master Plan dated October 18, 2005.

⁵ U.S. Individual Income Tax Return (Form 1040 and Form 1040A).

⁶ Request for Refund of Federal Telephone Excise Tax (Form 1040EZ-T).

⁷ Schedule A-Itemized Deductions (Schedules A&B), Schedule B-Interest and Ordinary Dividends

⁽Schedules A&B), and Credit for the Elderly or the Disabled (Schedule R), along with their Form 1040A equivalents.

⁸ Profit or Loss From Business (Schedule C), Capital Gains and Losses (Schedule D), Supplemental Income and Loss (Schedule E), Profit or Loss From Farming (Schedule F), and Self-Employment Tax (Schedule SE).



This review was performed at the CADE Project Office in New Carrollton, Maryland, and at the Submission Processing Site in Fresno, California, during the period March through August 2007. The audit was conducted in accordance with *Government Auditing Standards*. Detailed information on our audit objective, scope, and methodology is presented in Appendix I. Major contributors to the report are listed in Appendix II.



Results of Review

Tax Return Information Was Posted Accurately

The IRS successfully posted approximately 11 million tax returns to the CADE and generated approximately \$11.6 billion in refunds from March through August 2007. The original estimate for the 2007 Filing Season⁹ was for the IRS to post

33 million returns to the CADE. However, implementation of Release 2.2 was postponed from January to March 2007 so the IRS could make required performance improvements and complete filing season upgrades. Accordingly, this delay resulted in the actual number of returns posted being approximately 22 million fewer than the original estimate.

Approximately 11 million tax returns posted to the CADE, generating \$11.6 billion in refunds from March through August 2007.

Overall, tax return information was accepted and posted accurately to CADE accounts during the 2007 Filing Season. We reviewed samples of tax returns posted with the various attributes the CADE can now accept, focusing primarily on the attributes that were new to Release 2.2. However, we also verified some from prior releases.¹⁰

CADE Release 2.2 accurately posted returns with all of the attributes we reviewed, and the posted information related to the specific attributes matched the information from the sample tax returns. In addition, we verified the following general items for every sample return: name, address, adjusted gross income, taxable income, tax, withholding, Earned Income Tax Credit, and refund amount. All the tax return information was posted to taxpayers' accounts accurately.

Self-employment schedule information posted accurately

The ability to accept and post returns with self-employment information was added to the CADE on May 30, 2007. Accordingly, we reviewed judgmental samples of returns posted after this date that had self-employment schedules (Schedules C, F, and SE). We verified the general information listed above on these returns, as well as the following specific information: primary and secondary taxpayer's Social Security income, Medicare income, and self-employment tax amounts. In all cases, the posted information was accurate, and we did not identify any discrepancies. Also, the self-employment information being provided in a weekly report to the United States Census Bureau was accurate.

⁹ The period from January through mid-April when most individual income tax returns are filed.

¹⁰ See Appendix V for a list of all attributes reviewed.



The audit scoring process used by the CADE worked as intended

CADE Release 2.2 also included other functionality not directly related to the information on a taxpayer's account. One of these items was an expanded scoring process used to determine the audit potential of the returns being posted. This process has been performed by previous releases of the CADE; however, it is updated with each release because information from the additional tax form types and schedules needs to be added to the scoring process.

Because the IRS tested this revised process extensively, we validated its test methodology and results rather than performing detailed testing of the process. Our validation showed the testing methodology to be comprehensive. The test results appear to be accurate and support the IRS' conclusion that the expanded audit scoring process worked as intended.

Taxpayer Accounts With Special Conditions Were Incorrectly Added to the Customer Account Data Engine

Due to the inclusion of some accounts with special conditions on the CADE, approximately \$400,000 in overpaid taxes would have been incorrectly refunded to taxpayers. These overpayments should have been applied to other taxes owed by the taxpayers.

Taxpayer accounts are marked with freeze codes if they require special processing. These freeze codes are used for a variety of reasons, including to stop a refund from being issued if there is a pending additional assessment on the taxpayer's account or if there is a balance due on a related account. Approximately \$400,000 in refunds would have been issued incorrectly due to a programming error.

Some accounts with a freeze code that indicated the taxpayer had a tax liability on a related business account were incorrectly added to the CADE. Normally this freeze code holds any overpayment credit(s) from refunding so the credit(s) can be applied to the business tax liability. However, in the case of the CADE, the overpayments were not frozen because the CADE had not yet been programmed to allow freeze codes and the overpayments would have been refunded.

We identified this issue soon after the CADE began posting returns and referred it to the CADE Project Office. The prompt corrective action taken at that time prevented these accounts from having refunds issued incorrectly. However, the CADE would have released the refunds from these accounts if the situation had not been identified and corrected.

At the time we identified this issue, many of the accounts with freeze codes did not have Tax Year 2006 returns posted yet. Accordingly, we delayed performing an analysis of the amount of overpayments that would have been incorrectly refunded until the end of the audit. We identified 664 accounts that had the specific freeze condition with a posted Tax Year 2006 return



as of the week ended July 20, 2007. Of these returns, 338 had overpayments totaling approximately \$400,000 that would have been incorrectly refunded to the taxpayers.¹¹

Only accounts meeting specific criteria should be added to the CADE. However, in this case, a programming error resulted in accounts with these freeze conditions being incorrectly included in the CADE.

<u>Management Action</u>: We reported this issue to the CADE Project Office on April 12, 2007, and it took immediate corrective action. The accounts with freeze codes were transferred from the CADE back to the Master File, which enabled the freeze codes to work as intended. In addition, the CADE Business Assurance Executive stated that the programming would be reviewed to identify and correct what caused these accounts to be included in the CADE.

¹¹ See Appendix IV for details.



Appendix I

Detailed Objective, Scope, and Methodology

The overall objective of this review was to determine whether CADE Release 2.2 accurately posted tax return information, focusing on functionality not included in prior releases. Specifically, we:

- I. Determined whether tax return information was accurately posted to the CADE.
 - A. Obtained computer extracts of all 833,502 returns processed by the Submission Processing Sites in Fresno, California, and Austin, Texas, that were posted to the CADE during the weeks of March 26 and April 9, 2007. We validated the data from our computer extract by examining a random sample of 40 records. The validation test results demonstrated the data were reliable and could be used to meet the objectives of this audit.
 - B. Selected judgmental samples from the population identified in Step I.A. We used judgmental samples because we were trying to identify systemic processing problems and did not plan to project our outcomes. We selected a total of 20 individual judgmental samples of 40 returns each (for a total of 800 returns) with the following characteristics:
 - 1. Income Tax Return for Single and Joint Filers With No Dependents (Form 1040EZ) with a filing status of Single or Married Filing Jointly.
 - 2. Request for Refund of Federal Telephone Excise Tax (Form 1040EZ-T) with the Telephone Excise Tax Refund.
 - 3. U.S. Individual Income Tax Return (Form 1040A) with a filing status of Single, Married Filing Jointly, Married Filing Separately, or Head of Household.
 - 4. Form 1040A with Schedule 1, Schedule 3,¹ or dependency exemptions.
 - 5. U.S. Individual Income Tax Return (Form 1040) with a filing status of Single, Married Filing Jointly, Married Filing Separately, or Head of Household.

¹ Interest and Ordinary Dividends for Form 1040A Filers (Schedule 1) and Credit for the Elderly or Disabled for Form 1040A Filers (Schedule 3).



- 6. Form 1040 with Schedule A, Schedule B, Schedule D, Schedule E, Schedule R,² or dependency exemptions.
- C. Obtained the paper returns for the samples in Step I.B. We obtained the original data for electronically filed returns by accessing the Tax Return Database.³
- D. Researched the sample returns using the Integrated Data Retrieval System⁴ and compared the posted account information with the original tax return information to determine whether the information was posted accurately. The general return information verified was name, address, adjusted gross income, taxable income, tax, withholding, Earned Income Tax Credit, and refund amount. The specific information verified was the related filing status, exemptions, and schedule information.
- E. Obtained computer extracts of all returns processed nationally that were posted to the System between May 30 and June 29, 2007. The ability to post returns with self-employment schedules was implemented in May 2007. A total of 4,329 returns posted during this time period with a Schedule C and 176 returns posted with a Schedule F.⁵
 - 1. Validated the data from our computer extract by examining a random sample of 20 records. The validation test results demonstrated the data were reliable and could be used to meet the objectives of this audit.
 - 2. Selected individual judgmental samples of 40 returns each (for a total of 80 returns) that were filed with a Schedule C or Schedule F. We used judgmental samples because we were trying to identify systemic processing problems and did not plan to project our outcomes.
 - 3. Obtained the original return information for the samples in Step I.E.2. by accessing the Tax Return Database.
 - 4. Researched the returns using the Integrated Data Retrieval System and compared the posted account information with the original tax return information to determine whether they were posted accurately. We verified the same general return information as in Step I.D. The specific information verified was the

² Schedule A-Itemized Deductions (Schedules A&B), Schedule B-Interest and Ordinary Dividends (Schedules A& B), Capital Gains and Losses (Schedule D), Supplemental Income and Loss (Schedule E), and Credit for the Elderly or the Disabled (Schedule R).

³ This database contains the original (uncorrected) tax return information for electronically filed returns.

⁴ The IRS computer system capable of retrieving or updating stored information; it works in conjunction with a taxpayer's account records.

⁵ Profit or Loss From Business (Schedule C) and Profit or Loss From Farming (Schedule F).



self-employment schedule information, including Social Security and Medicare incomes and taxes.

- 5. Verified the posted Schedule SE⁶ information for those returns that had one attached. There were 18 returns from the samples selected in Step I.E.2. that had a Schedule SE attached.
- 6. Verified whether the information included in the weekly report issued to the United States Census Bureau was accurate by comparing it to the actual numbers of returns posted to the CADE with the associated self-employment schedules. This was accomplished by using the raw data from Step I.E.
- II. Determined whether the expanded audit scoring process used by the CADE was comparable to that used currently by the Master File. We obtained and evaluated information on the testing of the CADE scoring process to determine the audit potential of tax returns.
- III. Determined the cause and effect of accounts with freeze codes being included on the CADE.
 - A. Computer identified all 1,366 accounts on the System as of the last week of February 2007 (the week prior to the start of CADE processing) that had a freeze code.
 - B. Analyzed the types of freeze codes included on the CADE and discussed this issue with the CADE Project Office.
 - C. Obtained a computer extract of all accounts identified in Step III.A. and determined whether Tax Year 2006 returns had been filed as of July 20, 2007. A total of 664 returns had been filed as of this date. We analyzed the accounts and determined whether the overpayments were potentially refunded erroneously.

⁶ Self-Employment Tax (Schedule SE).



Appendix II

Major Contributors to This Report

Michael E. McKenney, Assistant Inspector General for Audit (Wage and Investment Income Programs) Scott A. Macfarlane, Director Richard J. Calderon, Audit Manager Steven D. Stephens, Lead Auditor Karen Fulte, Senior Auditor Mary Jankowski, Senior Auditor Sandra Hinton, Auditor Arlene Feskanich, Information Technology Specialist



Appendix III

Report Distribution List

Acting Commissioner C Office of the Commissioner – Attn: Acting Chief of Staff C Deputy Commissioner for Operations Support OS Deputy Commissioner for Services and Enforcement SE Chief Information Officer OS:CIO Deputy Commissioner, Wage and Investment Division SE:W Associate Chief Information Officer, Applications Development OS:CIO:AD Director, Customer Account Services, Wage and Investment Division SE:W:CAS Chief, Performance Improvement, Wage and Investment Division SE:W:S:PI Director, Strategy and Finance, Wage and Investment Division SE:W:S:PI Director, Submission Processing, Wage and Investment Division SE:W:CAS:SP 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: Senior Operations Advisor, Wage and Investment Division SE:W:S Chief, GAO/TIGTA/Legislative Implementation Branch SE:S:CLD:PSP:GTL



Appendix IV

Outcome Measure

This appendix presents detailed information on the measurable impact that our findings will have on tax administration. This benefit will be incorporated into our Semiannual Report to Congress.

Type and Value of Outcome Measure:

• Revenue Protection – Actual; approximately \$400,000 would have been incorrectly refunded to 338 taxpayers (see page 5).

Methodology Used to Measure the Reported Benefit:

We identified 1,366 accounts with freeze codes on the CADE as of the last week of February 2007. We obtained a computer extract as of July 20, 2007, to identify whether those 1,366 taxpayers had filed Tax Year 2006 returns as of that date and determined 664 taxpayers had filed returns.

Of the 664 taxpayers that had filed Tax Year 2006 returns, 326 had balances due or were even-balance returns. The remaining 338 accounts had potential incorrect refunds totaling approximately \$400,000. To calculate the amount, we totaled the amounts of the overpayments from these returns.



Appendix V

Tax Return Attributes Reviewed

The following is a list of CADE Release 2.2 attributes reviewed during our audit.

Form Type ¹	Attribute	Form Type	Attribute
1040EZ	Single Filing Status	1040	Single Filing Status
1040EZ	Married Filing Jointly Filing Status	1040	Married Filing Jointly Filing Status
1040EZ-T	Telephone Excise Tax Refund	1040	Married Filing Separately Filing Status
1040A	Single Filing Status	 1040	Head of Household Filing Status
1040A	Married Filing Jointly Filing Status	1040	Schedule A-Itemized Deductions (Schedules A&B)
1040A	Married Filing Separately Filing Status	1040	Schedule B-Interest and Ordinary Dividends (Schedules A&B)
1040A	Head of Household Filing Status	1040	Capital Gains and Losses (Schedule D)
1040A	Interest and Ordinary Dividends for Form 1040A Filers (Schedule 1)	1040	Supplemental Income and Loss (Schedule E)
1040A	Credit for the Elderly or the Disabled for Form 1040A Filers (Schedule 3)	1040	Credit for the Elderly or the Disabled (Schedule R)
1040A	Dependency Exemptions	1040	Dependency Exemptions

Source: These attributes were identified through Treasury Inspector General for Tax Administration analysis of the CADE Release 2.2 Candidate Requirement List.

¹ Income Tax Return for Single and Joint Filers With No Dependents (Form 1040EZ), Request for Refund of Federal Telephone Excise Tax (Form 1040EZ-T), U.S. Individual Income Tax Return (Form 1040A), and U.S. Individual Income Tax Return (Form 1040).