



*Better Use of Available Third-Party Data
Could Identify and Prevent More Than
One Billion Dollars in Potentially
Erroneous Refunds*

July 13, 2010

Reference Number: 2010-40-062

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.

Redaction Legend:

1 = Tax Return/Return Information

2(f) = Law Enforcement - Risk Circumvention of Agency Regulation or Statute

Phone Number | 202-622-6500

Email Address | inquiries@tigta.treas.gov

Web Site | <http://www.tigta.gov>



TREASURY INSPECTOR GENERAL
FOR TAX ADMINISTRATION

DEPARTMENT OF THE TREASURY

WASHINGTON, D.C. 20220

July 13, 2010

MEMORANDUM FOR DEPUTY COMMISSIONER FOR OPERATIONS SUPPORT
DEPUTY COMMISSIONER FOR SERVICES AND
ENFORCEMENT

Michael R. Phillips

FROM:

Michael R. Phillips
Deputy Inspector General for Audit

SUBJECT:

Final Audit Report – Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds (Audit #200840021)

This report presents the results of our review to determine how well the Internal Revenue Service (IRS) verifies and controls information provided by third parties and whether the information is effectively used where possible to verify information on individual income tax returns during and after processing. The audit was included in our Fiscal Year 2008 Annual Audit Plan and addresses the major management challenge of Erroneous and Improper Payments.

Impact on the Taxpayer

To verify the accuracy of the information that taxpayers report on their tax returns, the IRS uses data from various sources such as employers, government agencies, and financial institutions. These types of data are referred to as third-party data. Effective use of these data helps to ensure taxpayer compliance. We identified a number of areas in which the IRS could make more effective use of third-party data to identify and stop erroneous and improper claims for credits and refunds. Further, the IRS needs to better identify and control all the data that it receives from third parties.

Synopsis

The IRS' use of third-party data to verify information reported on U.S. Individual Income Tax Returns (Form 1040) is one of its most important tools to ensure compliance. However, the IRS is not fully using the third-party data it receives during returns processing. *****2(f)*****



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

*****2(f)*****

***** Further, additional action is needed to use third-party data to validate other EITC requirements or questionable claims.

In addition, *****2(f)*****

*****. These problems allow a substantial number of erroneous refunds and credits to be granted that are not allowable by law.

Finally, the IRS does not have a centralized data log or control point for all of the third-party data received, does not validate the data as they are received, and does not document the files to ensure the correct ones are used in processing.

Recommendations

We recommended that the Commissioner, Wage and Investment Division, use the authority provided in the law to 1) freeze refunds while contacting those taxpayers with potentially invalid EITC claims or questionable information on their tax returns, 2) require valid responses before allowing the EITC claims, and 3) adjust the returns if the taxpayers do not respond within a specific time period. The Deputy Commissioner for Operations Support should institute procedures to create and maintain a centralized database that contains all third-party data. Procedures should ensure all data are included in the database, data are received timely, the data elements and cost of the file are listed for each file, and the database is researchable. Procedures should also ensure that all data files received from outside sources are validated for format upon receipt and documented with at least the date and size of the data received on the file, and that only the contract administrators contact vendors when there are problems with the data.

Legislative Recommendations

*****2(f)*****

*****2(f)*****
*****2(f)*****



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

In addition, the Commissioner, Wage and Investment Division, should work with the Office of Tax Policy, Department of the Treasury, to obtain limited math error authority so the IRS can freeze refunds while contacting taxpayers *****2(f)*****

Response

IRS management agreed with two of our five recommendations. Management did not comment on the first legislative recommendation, but agreed to work with the Department of the Treasury on our second legislative recommendation. IRS management disagreed with the recommendation to freeze refunds and correspond with taxpayers who had a ***2(f)***** **2(f)*****. Management believes its current EITC examination strategy already focuses on selection of those cases most likely to be noncompliant and least likely to burden eligible taxpayers. Management also disagreed with the recommendation to centralize the third-party database, stating it has established procedures and processes to receive and maintain third-party data across the IRS. Management agreed with the recommendation to institute procedures to ensure all data files received from outside sources are validated for format and that only the contract administrators contact vendors when there are problems with data; however, management did not present any associated corrective action. Management’s complete response to the draft report is included as Appendix V.

Office of Audit Comment

*****2(f)*****

*****. Further, the lack of adequate corrective action is not in accordance with the intent of the Executive Order³ to help reduce improper payments and eliminate waste in Federal programs. We will advise the Assistant Secretary for Management and Chief Financial Officer, Department of the Treasury, of our concerns related to this issue.

Without a centralized control log of all third-party data, IRS functions may not be aware of information already being received by the IRS. As has happened previously, this may result in different IRS functions requesting and paying for duplicate information. It may also result in the IRS not making the best use of the data that it has in its compliance efforts. For management’s response to our last recommendation, appropriate corrective actions and completion dates are needed.

³ Executive Order 13520 – Reducing Improper Payments, signed November 20, 2009.



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Copies of this report are also being sent to the IRS managers affected by the report recommendations. Please contact me at (202) 622-6510 if you have questions or Michael E. McKenney, Assistant Inspector General for Audit (Returns Processing and Account Services), at (202) 622-5916.



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Table of Contents

BackgroundPage 1

Results of ReviewPage 3

 Use of Third-Party Data Provides Some Authentication
 During and After Initial ProcessingPage 3

 Available Third-Party Data Could Be Used More
 Effectively to Reduce Erroneous RefundsPage 4

Recommendations 1 and 2:Page 7

Recommendation 3:Page 11

 The Internal Revenue Service Does Not Have a Centralized
 Data Log for the Third-Party Data It Receives and
 Does Not Validate All Received DataPage 12

Recommendation 4:Page 14

Recommendation 5:Page 15

Appendices

 Appendix I – Detailed Objective, Scope, and MethodologyPage 16

 Appendix II – Major Contributors to This ReportPage 20

 Appendix III – Report Distribution ListPage 21

 Appendix IV – Outcome MeasuresPage 22

 Appendix V – Management’s Response to the Draft ReportPage 24



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Abbreviations

EITC	Earned Income Tax Credit
IRS	Internal Revenue Service
SSN	Social Security Number
TY	Tax Year



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Background

To verify the accuracy of the information reported on U.S. Individual Income Tax Returns (Form 1040), the Internal Revenue Service (IRS) uses data from various sources. For purposes of this report, data from outside of the IRS are referred to as third-party data. The availability of this type of data has a significant impact on taxpayer compliance. For example, only about 1 percent of all wage, salary, and tip income, which is subject to significant third-party reporting, is misreported. In contrast, nonfarm sole proprietor income, which is reported on a Profit or Loss From Business (Schedule C) and is subject to little third-party reporting or withholding, has a net misreporting percentage of 57 percent, contributing about \$68 billion to the tax gap.

The tax gap measures the extent to which taxpayers do not voluntarily file their tax returns and voluntarily pay the correct tax on time. In Tax Year (TY) 2001, the IRS estimated the net tax gap was \$290 billion. The tax gap can be divided into three components:

- **Nonfiling** – taxpayers who are required to file a return but do not do so on time.
- **Underreporting** – taxpayers either understate their income or overstate their deductions, exemptions, and credits on timely filed returns.
- **Underpayment** – taxpayers file their return, but fail to remit the full amount due by the payment due date.

Of these 3 components, underreporting taxes (income taxes, employment taxes, and other taxes) represents about 80 percent of the tax gap. The largest subcomponents of underreporting involve individuals understating their incomes, taking improper deductions, overstating business expenses, or erroneously claiming credits. Underreporting on individual income tax returns accounts for about half of the total tax gap.

This review was performed at the Fresno, California, Submission Processing Site, and included data provided by management and program analysts in Lanham, Maryland, and Cincinnati, Ohio, and at the offices of the Commissioner, Wage and Investment Division, in Atlanta, Georgia, and the Commissioner, Small Business/Self-Employed Division, the Chief Technology Officer, and the Chief Financial Officer, in Washington, D.C., during the period June 2008 through September 2009. There was a significant impairment to the scope of our audit work. The IRS was unable to provide a complete accounting of all the data received from third-party sources. As such, we used a sample of four data files to evaluate the accuracy and completeness of the various data files received by the IRS. This matter is discussed in further detail in the Results of Review section. With the exception of this impairment, 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



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

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.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

Results of Review

Use of Third-Party Data Provides Some Authentication During and After Initial Processing

For TY 2006, the IRS received almost 1.8 billion information documents (third-party data) from various sources. According to a Government Accountability Office report,¹ the IRS is able to process and use 88 percent of the third-party data it receives.

*****2(f)*****

During its processing of tax returns, the IRS Submission Processing function uses names, Social Security Numbers (SSN), and in some cases age data provided by the Social Security Administration to verify certain entries on tax returns. Data provided by the Department of Defense are used to identify military personnel in combat zones so the IRS can ensure the proper special procedures are used to process their tax returns.

The IRS uses other third-party data (real estate transactions, pensions, miscellaneous income, etc.) after initial processing of tax returns to identify underpaid taxes. The IRS Automated Underreporter function is the first place most third-party data are used after processing to validate income and deductions. In Fiscal Year 2007, this function identified 15 million tax returns with potential underreporting. About 4.5 million returns were actually selected to be reviewed.

The IRS is considering other uses of third-party data as well. For example, the State Reverse File Match Initiative compares data on tax returns filed with the IRS to data on tax returns filed with various State taxing agencies. Any Taxpayer Identification Numbers that appear in the States' data but not the IRS' data are identified. The IRS is currently developing the logic to evaluate these data and then forward the identification numbers to the appropriate Compliance functions for action. According to the IRS, the preliminary results are promising, and it has been able to obtain new information that will be available to compliance employees to assist in identifying where additional assessments could be made.

¹ Tax Administration: Costs and Uses of Third-Party Information Returns (GAO-08-266, dated November 2007).
*****2(f)*****



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

While the IRS does use much of the third-party data it receives to help ensure taxpayers pay the correct amount of tax, more use can be made of the third-party data currently available. We believe that expanding the use of third-party data during processing would allow the IRS to decrease erroneous refunds, increase revenues, and promote voluntary compliance.

In addition, while the IRS does have authority to disallow certain credits claimed under specific conditions on tax returns during processing, it does not have the authority to disallow all types of questionable credits or items claimed on tax returns.

Available Third-Party Data Could Be Used More Effectively to Reduce Erroneous Refunds

Only a portion of the information from the Social Security Administration is used during the processing of a tax return. Most of the IRS' effort is focused on validity of the taxpayers' SSNs and the ages of their dependents. *****2(f)*****
***** In some circumstances, the IRS has authority to immediately adjust tax liabilities and refunds and continue to process the return. This authority is referred to as "math error authority."

Recognizing that some errors on tax returns did not justify the expense of an audit, Congress authorized the IRS to correct certain math errors that appeared on the face of a tax return. Initially, taxpayers had no right to appeal these corrections. In the Tax Reform Act of 1976,³ Congress provided that a taxpayer who received a math error notice had 60 days to file a request for abatement of the assessment. The Act also codified the first five definitions of "mathematical or clerical errors" to include:

- An error in addition, subtraction, multiplication, and division shown on a tax return.
- An incorrect use of an IRS table related to the return.
- Inconsistent entries on a return.
- Omitted information required to substantiate an entry on a return.
- An entry that claims a deduction or credit in excess of the statutory limit.

Congress has expanded the IRS' math error authority over the years to include denial of benefits for exemptions, filing status, and certain tax credits when taxpayers failed to provide valid Taxpayer Identification Numbers.

If the IRS effectively used the authority it has, and if its math error authority was expanded to address the issues discussed in this report, we believe it would help prevent the payment of additional improper claims.

³ Pub. L. No. 94-455, 90 Stat. 1520.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

*****2(f)*****

*****2(f)*****

*****2(f)*****

*****2(f)*****

4 The EITC is a refundable credit used to offset the impact of Social Security taxes on low-income families and to encourage them to seek employment rather than welfare. The amount of the Credit individuals receive depends on earned income, the number of qualifying children, and filing status. This Credit is also referred to as the Earned Income Credit.

*****2(f)*****

*****2(f)*****

*****2(f)*****

*****2(f)*****



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

*****2(f)*****

*****2(f)*****

Because of a change in the Social Security Administration's policy, *****2(f)*****

*****2(f)*****

We used the available IRS data to identify *****2(f)***** status of the primary and secondary taxpayers on all individual income tax returns filed over a 5-year period ending in TY 2007. *****2(f)*****

***** Figure 1 contains return information for this 5-year period.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

Figure 1: *****2(f)*****

***	*****	*****	*****

Legislative Recommendation

Recommendation 1: *****2(f)*****

Management’s Response: The IRS did not comment on this recommendation. Matters of tax policy are within the jurisdiction of the Office of Tax Policy, Department of the Treasury.

Recommendation

Recommendation 2: Until clarifying legislation is passed, the Commissioner, Wage and Investment Division, should use the authority already provided in the law to 1) freeze refunds while contacting those taxpayers with potentially invalid EITC claims or questionable information on their tax returns, 2) require valid responses from the taxpayers before allowing the EITC, and 3) adjust the returns if the taxpayers do not respond within a specific time period.

Management’s Response: IRS management disagreed with this recommendation. The existing authority of the IRS to address this issue can only be exercised through an examination. The EITC examination strategy already focuses on selection of those cases most likely to be noncompliant and least likely to burden eligible taxpayers. Management does not believe that refocusing other resources on this issue will meet the IRS goal of a balanced compliance program that strategically addresses noncompliance with the resources available.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

Office of Audit Comment: *****2(f)*****

Additional validation is needed on claims for the EITC to prevent improper claims

Each year, a substantial number of taxpayers claim the EITC. In a population of 154 million TY 2007 individual income tax returns, there were 24.5 million returns claiming the EITC for \$48.5 billion. EITC fraud has been a problem for the IRS for years, and the IRS estimates that there are more than \$10 billion in erroneous EITC claims each year. Numerous changes have been made to the EITC qualifications in order to reduce the amount of fraud associated with the claims.

In a prior audit, 10 we reported that taxpayers were using the SSNs of individuals who were at least 20 years older than the primary taxpayer for purposes of claiming the EITC. In September 2006, the IRS performed a study to verify this problem, but it did not correct the problem. We identified 50,535 TY 2007 returns filed by single 11 taxpayers, claiming 1 or more dependents (at least 20 years older than the primary taxpayer) on the Earned Income Credit (Schedule EIC) for \$165.6 million in credits.

Subsequent to legislation enacted in October 2008 12 that further defined a child as younger than the taxpayer (unless disabled), the IRS planned to submit requests for changes to its computer programming to enforce this requirement. 13 We reviewed the requested change documentation and identified a problem with how the new requirement would be programmed. The issue was

*****2(f)*****

10 Better Use of the National Account Profile During Returns Processing Can Eliminate Millions of Dollars in Erroneous Payments (Reference Number 2004-40-098, dated May 12, 2004).

11 Single meaning no spouse was claimed, so filing status could be Single, Head of Household, or Widow(er).

12 Fostering Connections to Success and Increasing Adoptions Act of 2008, Pub. L. No. 110-351, 122 Stat. 3949.

13 Based on the requirements of the law, a dependent older than the taxpayer does not qualify for the EITC unless he or she is permanently and totally disabled, and then only if he or she is a son, daughter, stepchild, foster child, brother, sister, half brother, half sister, stepbrother, stepsister, or a descendant of any of them.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

discussed with the IRS, and the IRS agreed to change the programming request. This programming was scheduled to be implemented for the 2010 Filing Season.

We analyzed the ages of the primary and secondary taxpayers on the 24.5 million FY 2007 returns based on IRS and Social Security Administration data and found the *****1*****

***** In addition, approximately 3,300 of the primary taxpayers did not have a date of birth shown on their accounts on the IRS Master File,¹⁴ and *****2(f)*****

***** per Social Security Administration data should be scrutinized to determine if the returns are incorrect or fraudulent.

Taxpayers' ages indicate that information reported on some tax returns is questionable

*****2(f)*****

***** In a population of 154 million¹⁶ TY 2007 individual income tax returns, the IRS processed 4,454 tax returns for primary taxpayers ***2(f)*
*****2(f)*****who were filing as 1) Married Filing Jointly, 2) Head of Household, 3) Married Filing Separately, or 4) Qualifying widow(er). These returns claimed more than \$8.1 million in credits. The taxpayer's ***2(f)*****should raise concerns about the validity of the information claimed on a tax return. Figure 2 provides accepted return information for taxpayers ***2(f)***
*****.

¹⁴ The IRS database that stores various types of taxpayer account information. This database includes individual, business, and employee plans and exempt organizations data.

¹⁵ While the IRS did have date of birth information for these 3,300 taxpayers from the Social Security Administration, it did not use this information during the processing of the tax returns.

¹⁶ The population of TY 2007 tax returns is larger than normal due to economic stimulus payment returns filed.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

Figure 2: *****2(f)*****

***** *****	***** *****	***** *****	***** *****
*****	*****	*****	*****
*****	***	*****	*****
*****	*****	*****	*****
*****	*****	*****	*****

Source: IRS Return Transaction File (individual tax return data).

In addition, we identified more than 43,000 tax returns with primary or secondary taxpayers ranging in age from 100 years old to 345 years old based on IRS and Social Security Administration data. Based on the same data, the primary or secondary taxpayers on 1,952 of these returns were 120 years old or older. The 120-year-old and older taxpayers reported almost \$59 million in wages and \$11 million in withholding, and claimed more than \$1.3 million in credits. The Office of the Inspector General, Social Security Administration, recently released a report¹⁷ discussing fraud associated with the use of SSNs of persons 100 years or older. Tax returns filed by taxpayers whose SSN information indicates they are 120 years old or older should be scrutinized to determine if the returns are incorrect or fraudulent.

The Government Accountability Office also recently issued a report¹⁸ on the IRS' limited ability to identify fraud related to identity theft. According to the report, the IRS does not know the amount of refund or employment fraud that goes undetected and does not know how well its current strategy is working.

In order to reduce identity theft and mitigate the possibility of fraud and erroneous refunds, we believe that any tax return claiming a filing status other than Single with the primary or secondary taxpayer listed as *****2(f)***** should be considered questionable, as well as any return filed by a taxpayer who is 120 years old or older.¹⁹

By using Social Security Administration data currently available, the IRS could reduce the number of erroneous refunds and credits issued and identify cases of possible identity theft. The credits claimed by taxpayers *****2(f)***** on TY 2007 returns were almost \$8.1 million (\$5.5 million refundable credits and \$2.6 million other credits); credits for those 120 years old and older totaled more than \$1.3 million (\$.5 million refundable credits and \$.8 million other credits) in the same time period.

¹⁷ *Potential Social Security Number Misuse in Certain Unique Populations* (A-08-08-28060, dated May 2009).

¹⁸ *Tax Administration: IRS Has Implemented Initiatives to Prevent, Detect, and Resolve Identity Theft-Related Problems, but Needs to Assess Their Effectiveness* (GAO-09-882, dated September 2009).

¹⁹ As of January 2007, the oldest person in the world is 115 years old as reported on seniorjournal.com.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

*****2(f)*****

In a prior audit report,²⁰ we recommended that the IRS use the date of death information contained on the National Account Profile database to identify and correct returns using the SSN of a deceased individual. The IRS agreed to update manuals and procedures to require that the SSN used to support a claim for a tax benefit be checked against the date of death field for appropriate verification. A review of the manuals listed by the IRS shows they have been changed indicating employees should take specific steps if an SSN has a date of death listed in the National Account Profile database. *****2(f)*****

*****2(f)*****

*****2(f)*****

Legislative Recommendation

Recommendation 3: The Commissioner, Wage and Investment Division, should work with the Office of Tax Policy, Department of the Treasury, to obtain limited math error authority so that the IRS can freeze refunds while contacting taxpayers *****2(f)***** This

20 Better Use of the National Account Profile During Returns Processing Can Eliminate Millions of Dollars in Erroneous Payments (Reference Number 2004-40-098, dated May 12, 2004).

21 There are instances when it is proper to have a deceased taxpayer's SSN appear on a tax return (e.g., when the person dies during the tax year).



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

authority should allow the IRS to adjust the returns if the taxpayers do not respond within a specific time period.

Management's Response: The Wage and Investment Division, with the cooperation of the IRS Office of Chief Counsel, will discuss with the Department of the Treasury the merits of an administrative proposal to the Internal Revenue Code to obtain limited math error authority to freeze refunds while contacting taxpayers with *****2(f)*****

*****.

The Internal Revenue Service Does Not Have a Centralized Data Log for the Third-Party Data It Receives and Does Not Validate All Received Data

An IRS review²² of its third-party data completed during our audit found the following:

Currently external third-party data is collected from (a) variety of governmental and private industry sources but is not clearly defined or shared. It is unclear as to what data from those third parties is actually used. Same data may be purchased multiple times. It is unclear to the projects how it should be used within the enterprise target data architecture. The same data may be duplicated across organizations and platforms. The data quality is not fully understood. Frequency of data refresh is unclear.

There is no centralized control point for third-party data

The IRS does not have a centralized control point for third-party data requested or received from outside sources. To facilitate control and tracking of all third-party data, all data requested and received should be controlled on a centralized data log. This would help provide assurance that data are traceable and secure, and that a chain of control is established. Each business unit has its own system to control the data it receives from outside sources. Data received at the Detroit, Michigan; Memphis, Tennessee; and Martinsburg, West Virginia, locations have standard written procedures to log and control incoming data from third parties. Other electronic media come in at various places, sometimes directly to a business unit and are not logged into a single control system.

The IRS' own work in this area in the past has shown multiple payments made for the same data. A centralized control point could also be part of a system to re-request data that have not been received timely so that the IRS does not have to delay processing because of a missing file. The IRS has stated it has the "As-Built Architecture" database, which includes vendor names.

²² "IRS Third-Party Data Strategy" July 2008 PowerPoint presentation by the IRS Modernization and Information Technology Services organization.



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

However, this database is more application-based rather than control-based. It does not have specific purchased data file names, costing information, incoming date information, or data element information that is easily accessible. Also, there is no assurance the database is complete.

We requested that the IRS provide a list of all third-party data received from outside sources. It was unable to do so, stating it did not have a centralized control point for all of the data received from outside of the organization. In addition, the IRS could not provide us cost data for obtaining third-party data files. Consequently, we sampled specific data files. We provided the IRS with the names of four data files²³ received from third-party sources on which to base our questions and recommendations.

The IRS does not validate and correct all third-party data

The IRS does conduct limited checks of vendor-supplied data for validity and format before using it in various computer programs to authenticate what is reported on a tax return for income and deductions. However, the IRS does not use standardized procedures for validating the data or formatting data received from third-party vendors. For proper control, the data should be validated upon receipt to reduce delays and prevent incorrect data from being used later in processing. If good header information is not on file for the data, the computer systems are not always programmed to detect that the correct file is loaded, resulting in old data being used. Thus, the IRS relies on the computer operators to look for the information and decide if the data being loaded are correct, if they are able to make that determination. If the correct data are not available, the computer operators sometimes contact the third-party vendors directly without going through the IRS purchasing agent or being authorized in the contract to make such contact. We did not pursue this issue of unauthorized persons contacting vendors directly.

Combat Zone Service

According to the IRS, the data received on combat zone service from the Department of Defense are not logged into the normal IRS tape library where much of the data from third-party vendors are received and logged. The personnel handling the data stated the data are not validated or verified prior to being put into production. If some of the data are erroneous and do not post for various reasons, the IRS does not capture the unposted records for correction. The IRS did indicate that if the data file had more than a 10 percent error rate, the production runs could be stopped and the combat zone update information removed. When the data are corrected, they can be placed back into a production run at a different time.

²³ The Combat Zone Service data file, the Prisoner data file, the Social Security Administration updates file, and the Postal Address updates file are described in this report beginning on this page.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

Prisoner Data

The Criminal Investigation Division of the IRS annually requests and receives, on a voluntary basis, prisoner identification data from Federal and State prisons throughout the country. These data are not logged into the normal IRS library where much of the data from third-party vendors are received and logged. They build their own file and then give it to the programmers for the annual Electronic Fraud Detection System runs. It is used to examine both electronically and paper-filed returns to support the Revenue Protection Strategy and scheme development. The programmers verify the number of records processed, and in rare circumstances when the new data is not received, the program is run using that State's prior year data.

Social Security Administration Updates

The IRS receives weekly updates from the Social Security Administration. The update file contains the week's date, which helps identify the most recent file and is used to update IRS entity information data. IRS officials informed us they contact the Social Security Administration when records with errors are found, but we identified 419,000 records with missing data that could have been used to identify and stop potentially erroneous refunds.

Postal Address Updates

The Postal change file operator stated that if the file was close to the normal size, it is not checked. IRS employees indicated that if the number of total records processed is abnormally high or low, they will question the validity of the information. The IRS also indicated not all of the data received has header information containing the date of the data or the size of the file. When a computer system makes a data call, the data are loaded into the system. If there is a problem with the data, the system will stop and create an error ticket. The computer operators then must stop the process and try to obtain the correct file or take it completely out of the process.

Recommendations

The Deputy Commissioner for Operations Support should:

Recommendation 4: Institute procedures to create and maintain a centralized database that contains all third-party data. Procedures should ensure all data are included in the database, the data are received timely, the data elements and cost of the file are listed for each file, and the database is researchable so employees can query the database to identify files or data elements that may be useful to their programs.

Management's Response: IRS management disagreed with this recommendation to centralize the third-party database. The IRS has established procedures and processes to receive and maintain third-party data across the organization. The IRS will continue to review and update procedures and processes as needed to improve its management of third-party data.



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Office of Audit Comment: In discussions with IRS management subsequent to the issuance of our draft report, they indicated that third-party data are controlled in the IRS Government Liaison Division. We verified that the Government Liaison Division does have a list of third-party files; however the list is not complete, does not contain all relevant information mentioned in our report, and did not appear to be known by the necessary IRS personnel. At a minimum, the IRS should ensure that all employees with a need to know are aware of third-party data received and compiled by the IRS. Without a centralized data log containing information on all third-party data, different IRS functions may not be aware of information already being received by the IRS. As has happened in the past, this may result in different IRS functions requesting and paying for duplicate information. It may also result in beneficial information going unused. As such, we continue to believe that the IRS needs to implement corrective action to address this recommendation.

Recommendation 5: Institute procedures to ensure that all data files received from outside sources are validated for format upon receipt and documented with at least the date and size of the data received on the file, and that only the contract administrators contact vendors when there are problems with data being submitted or missing.

Management's Response: IRS management agreed with this recommendation. Data validation occurs as the Modernization and Information Technology Services organization processes the data received and loads the data to the various databases. If there are any problems, the Contracting Officer's Technical Representative contacts the vendor who submitted the data and alerts Procurement if contractual obligations are not met.

Office of Audit Comment: While IRS management agreed with our recommendation, they did not provide an implementation date or a responsible official. In addition, their response is not consistent with the facts reported. As stated in the report, the IRS does not use standardized procedures for validating the data or formatting data received from third-party vendors. If good header information is not on file for the data, the computer systems are not always programmed to detect that the correct file is loaded, which results in old data being used. Further, if the correct data are not available, the computer operators sometimes contact the third-party vendors directly without going through the IRS purchasing agent.



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Appendix I

Detailed Objective, Scope, and Methodology

Our overall objective was to determine how well the IRS verifies and controls information provided by third parties and whether the information is effectively used where possible to verify information on individual income tax returns during and after processing.

We relied on tax return information at the Treasury Inspector General for Tax Administration's Data Center Warehouse¹ because this audit dealt with third-party data received from outside vendors and because of where the IRS uses this information. This information was obtained from the IRS Master File² and is documented in IRS records where the information was generated. We believe the validation performed by the Data Center Warehouse staff is sufficient for this audit, and the information is considered adequate and reliable for our purposes.

We excluded primary SSNs with an invalid designation (name does not match Social Security Administration records). This would cause the numbers to be understated, but not significantly. The IRS indicated its error handling on some files would cause the records to be dropped, but we do not believe this would have significantly affected the outcomes of this report. There are no estimates or projections in this report. All numbers are from IRS data.

To accomplish our objective, we:

- I. Determined who in the IRS has control over all third-party data and whether the process for controlling this data is effective.
 - A. Attempted to obtain a list of all third-party data, when and where data are received, who sends data to the IRS, and the costs associated with obtaining the data. The IRS was unable to provide a complete list of all the data received from third-party sources and the associated costs. This represented a significant scope limitation. We used a sample of four data files³ to make our determination regarding the accuracy and completeness of the various data files received by the IRS.

¹ The Data Center Warehouse provides data and data access services in a centralized storage facility with security and administration of the files. It is an architecture used to maintain critical historical data that has been extracted from operational data storage and transformed into formats accessible to our organization.

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

³ The Combat Zone Service data file, the Prisoner data file, the Social Security Administration updates file, and the Postal Address updates file are described in this report beginning on page 13.



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

- B. Determined what steps are taken to verify and control third-party data. We requested copies of procedures and manuals indicating any checks the IRS performs prior to using the data.
 - C. Determined how and when the availability of third-party data is communicated to IRS functions.
 - 1. Interviewed IRS personnel about how the third-party data are received and tracked.
 - 2. Checked with IRS programmers and in manuals on how the third-party data are validated and used in the various computer systems.
 - 3. Analyzed the process for IRS functions to gain access to current data received on a set schedule or newly requested third-party data.
 - D. Interviewed other large agencies and corporations on their use and control of third-party data.
 - E. Identified the third-party data available to validate the dollar amounts claimed on the tax return.
 - 1. Used Data Center Warehouse information to obtain data to perform validation of dollar amounts.
 - 2. Used Data Center Warehouse information to obtain any prior year data (TYs 2002 through 2007) for EITC amounts claimed by taxpayers ***2(f)*****

- II. Determined when the IRS uses the third-party data to validate the information claimed on U.S. Individual Income Tax Return (Form 1040) series.
- A. Requested pertinent computer programming procedures and manuals that identify checks or verifications done during computer processing of Forms 1040.
 - B. Followed up on corrective actions to recommendations made in our prior reports.⁵
 - 1. For the recommendation to establish procedures requiring use of the National Account Profile database date of death information to identify and correct during processing those returns filed using the SSN of a deceased individual, we:

*****2(f)*****

*****2(f)*****
***** and *Better Use of the National Account Profile During Returns Processing Can Eliminate Millions of Dollars in Erroneous Payments* (Reference Number 2004-40-098, dated May 12, 2004).



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

- a) Obtained programming documentation to ensure corrective action was included for returns processing.
 - b) Discussed corrective action, if any, with the IRS.
 - c) Reviewed procedures and reports to help ensure corrective action was effectively taken.
2. For the recommendation to ensure all manuals and procedures are updated to include research for a date of death when a taxpayer contacts the IRS to request a tax benefit, we verified that Internal Revenue Manual sections include proper research for this issue.
 3. For the recommendation to conduct studies on the accuracy of EITC claims on tax returns containing individuals claimed for EITC purposes that are 20 or more years older than the primary taxpayers and individuals listed as children that are up to 19 years older than the primary taxpayer, we discussed study results with the IRS and determined what corrective action was taken, if any.
- C. Reviewed Internal Revenue Manual sections for processing steps.
 - D. Determined if there are differences in the validation for paper returns versus electronic returns.
 - E. Obtained information on IRS authority to change Form 1040 information during initial processing, including processes for math error adjustments, unallowables, and unpostables.
 - F. Reviewed the Automated Underreporter function process for use of third-party data.
 - G. Reviewed the examination process for use of third-party data.
- III. Determined whether the IRS could increase use of tax return data and third-party data.
- A. Interviewed IRS management regarding current use and possible increased use of third-party data, including potential new uses and barriers faced in using information currently available.
 - B. Reviewed information gathered in Steps I., II., and III.A. to determine current use of tax return and third-party data.
 - C. Obtained pertinent reports and studies done by the IRS regarding third-party data use.

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



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

for measuring, reporting, and monitoring program performance. We determined the following internal controls were relevant to our audit objective: the office of the Chief Technology Officer and the Wage and Investment Division policies, procedures, and practices for processing selected work streams in campus⁶ operations. We evaluated these controls by interviewing management and reviewing case files.

⁶ The data processing arm of the IRS. The campuses process paper and electronic submissions, correct errors, and forward data to the Computing Centers for analysis and posting to taxpayer accounts.



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Appendix II

Major Contributors to This Report

Michael E. McKenney, Assistant Inspector General for Audit (Returns Processing and Account Services)
Kyle Andersen, Director
Richard J. Calderon, Audit Manager
Glory Jampetero, Lead Auditor
George Burleigh, Senior Auditor



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

Appendix III

Report Distribution List

Commissioner C
Office of the Commissioner – Attn: Chief of Staff C
Assistant Secretary of the Treasury for Tax Policy
Chief, Criminal Investigation Division SE:CI
Commissioner, Small Business/Self-Employed Division SE:S
Commissioner, Wage and Investment Division SE:W
Deputy Commissioner, Small Business/Self-Employed Division SE:S
Deputy Commissioner, Wage and Investment Division SE:W
Director, Customer Account Services, Wage and Investment Division SE:W:CAS
Director, Electronic Tax Administration and Refundable Credits, Wage and Investment Division
SE:W:ETARC
Director, Earned Income Tax, Wage and Investment Division SE:W:ETARC:E
Director, Strategy and Finance, Wage and Investment Division SE:W:S
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 Liaison: Chief, Program Evaluation and Improvement, Wage and Investment Division
SE:W:S:PRA:PEI



Better Use of Available Third-Party Data Could Identify and Prevent More Than One Billion Dollars in Potentially Erroneous Refunds

Appendix IV

Outcome Measures

This appendix presents detailed information on the measurable impact that our recommended corrective actions will have on tax administration. These benefits will be incorporated into our Semiannual Report to Congress.

Type and Value of Outcome Measure:

- Funds Put to Better Use – Potential; *****2(f)*****

Methodology Used to Measure the Reported Benefit:

Our data were originally derived from the approximately 154 million TY 2007 individual tax returns filed as of cycle 200853 that contained a valid SSN. The National Account Profile database was used to determine the citizenship code associated with each SSN to obtain the number of SSNs being used by individuals *****2(f)*****

Type and Value of Outcome Measure:

- Revenue Protection – Potential; \$17 million for 5-year period (***2(f)***** is \$13 million and taxpayers 120 years old and older is \$4 million) affecting 6,406 taxpayers (see page 4).
Funds Put to Better Use – Potential; \$30 million for 5-year period (***2(f)***** is \$27.5 million and taxpayers 120 years old and older is \$2.5 million) (see page 4).

Methodology Used to Measure the Reported Benefit:

Our data were originally derived from the approximately 154 million TY 2007 individual tax returns filed as of cycle 200853 that contained a valid SSN. From the 154 million individual tax returns, we used the ages as calculated by the IRS from information provided by the Social Security Administration. The ages were then sorted into a file for those ***2(f)***** and another file for those 120 years old or older. The outcome measure applies to the credits claimed



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

by 4,454 taxpayers ***2(f)***** using a filing status other than Single (\$5.5 million refundable and \$2.6 million other credits) and all 1,952 taxpayers 120 years old or older (\$5 million refundable and \$.8 million other credits) on TY 2007 returns. At the time of the audit, the oldest person in the world was 115 years old, so it is assumed anyone claiming to be 120 years old or older is in error or fraudulent. The current year credit totals were multiplied by 5 to obtain a 5-year projection.



**Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds**

Appendix V

Management's Response to the Draft Report

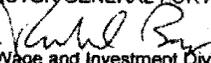


DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
ATLANTA, GA 30308

RECEIVED
MAY - 5 2010
BY: MAS

APR 29 2010

MEMORANDUM FOR MICHAEL R. PHILLIPS
DEPUTY INSPECTOR GENERAL FOR AUDIT

FROM: Richard Byrd, Jr. 
Commissioner, Wage and Investment Division

SUBJECT: Draft Audit Report – Better Use of Available Third-Party Data
Could Identify and Prevent More Than One Billion Dollars in
Potentially Erroneous Refunds (Audit #200840021)

Thank you for the opportunity to review and respond to the subject draft report. As indicated in your report, the IRS received almost 1.8 billion information documents from third-party providers for Tax Year 2006 and we were able to process and use 88 percent of that information in gauging compliance with the tax laws. This is a substantial undertaking and it presents a unique set of challenges. Accordingly, I would like to comment on three topics addressed in the report:

- o Social Security Numbers (SSNs) which are 
 used to claim the Earned Income Tax Credit (EITC);
- o 
- o A centralized control point for all third-party data.

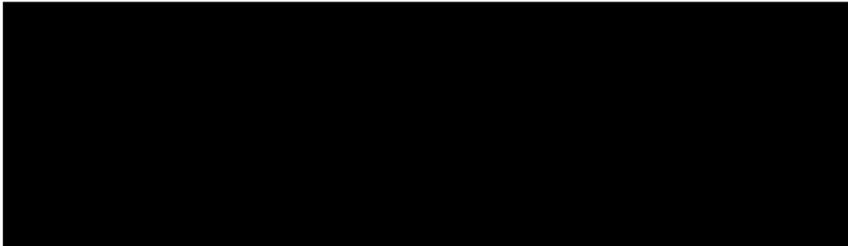




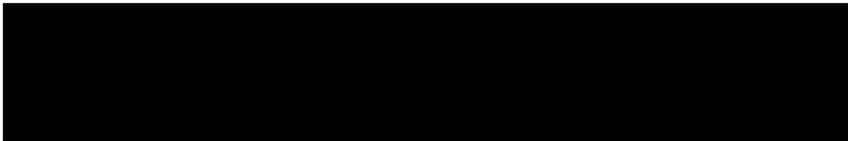


***Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds***

2



We employ a sophisticated risk-based scoring model to identify and select the most likely EITC noncompliant returns for examination. Because our resources are finite, we carefully weigh all areas of potential EITC noncompliance against burden to the taxpayer, return on investment, and a balanced EITC compliance program. Through these efforts, we protect \$2 billion, annually.



This conclusion is based on your sample of tax return data during the five year period from Tax Year 2003 through Tax Year 2007. We recognized this issue and beginning with the 2010 Filing Season, we addressed the matter of [redacted] of the taxpayer by adding a check through the Error Resolution System (ERS) for any primary taxpayer who meets the following conditions:

- [redacted]
- is not another taxpayer's dependent;
- reports wages or income from self-employment; and
- requests a refund of \$300 or more.



**Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds**

3

The ERS function refers returns meeting the above criteria to the Wage and Investment Division, Accounts Management, Identity Protection Specialized Unit, which reviews the returns and provides instructions for processing. Although the ERS age verification check is not directly linked to EITC, it does cause a closer review of the entire return

with a filing status other than single, are detected by the age check above. Notwithstanding ERS efforts to address these age anomalies up front, the IRS also has other processes in place to filter these returns for indications of noncompliance and select them for examination. The extension of math error authority outlined in Recommendation 3 of your report would enable ERS to adjust claims for EITC.

Tax examiners request official documents to establish taxpayer identity, if missing. The ERS tax examiners freeze taxpayers' accounts if they do not reply to the request, or if the reply is inadequate. The ERS tax examiners allow claims for EITC only for children who died in, or after, the Tax Year of the claim.

Centralized Control Point for Third-Party Data

Your report states that the IRS does not have a centralized control point for third-party data requested or received from outside sources. However, IRS Procurement tracks contracts awarded to third parties with Central Contractor Registration (CCR). The CCR records include basic procurement and financial information, Data Universal Numbering System (DUNS) number, and banking information. This data is interfaced into the vendor database contained in the web-based Integrated Procurement System and the Integrated Financial System. Furthermore, all IRS third-party data files are maintained by Modernized Information Technology Services (MITS) in the IRS Computing Centers and made available to the various Business Operating Divisions (BODs), as needed. This process is designed to meet different law and regulation requirements, and business data needs of the BODs.

We disagree with the proposed outcome measure "Funds Put to Better Use – Potential;

We also disagree with the proposed outcome measures: "Revenue Protection - Potential; \$17 million for 5-year period is \$13 million and taxpayers 120 years old and older is \$4 million) impacting 6,406 taxpayers; and Funds Put to Better Use - Potential: \$30 million for 5 year period (is \$27.5 million and taxpayers 120 years old and older \$2.5 million)". Taxpayers are now validated as previously indicated.



**Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds**

Attachment

RECOMMENDATION 1



CORRECTIVE ACTION

The IRS is not commenting on this recommendation. Matters of tax policy are within the jurisdiction of the Office of Tax Policy at the Department of Treasury.

IMPLEMENTATION DATE

N/A

RESPONSIBLE OFFICIAL

N/A

CORRECTIVE ACTION MONITORING PLAN

N/A

RECOMMENDATION 2

Until clarifying legislation is passed, the Commissioner, Wage and Investment Division, should use the authority already provided in the law to 1) freeze refunds while contacting those taxpayers with potentially invalid EITC claims or questionable information on their tax return, 2) require a valid response from the taxpayers before allowing the EITC, and 3) adjust the return if the taxpayer does not respond within a specific time period.

CORRECTIVE ACTION

We do not agree with this recommendation. The existing authority of the IRS to address this issue, can only be exercised through an examination. Our Earned Income Tax Credit (EITC) examination strategy already focuses on selection of those cases most likely to be noncompliant and least likely to burden eligible taxpayers. We do not believe that refocusing other resources on this issue will meet our goal of a balanced compliance program which strategically addresses noncompliance with the resources available.

IMPLEMENTATION DATE

N/A

RESPONSIBLE OFFICIAL

N/A



***Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds***

2

CORRECTIVE ACTION MONITORING PLAN

N/A

RECOMMENDATION 3

The Commissioner, Wage and Investment Division, should work with the Department of the Treasury, Office of Tax Policy, to obtain limited math error authority so that the IRS can freeze refunds while contacting taxpayers [REDACTED]

This authority should allow the IRS to adjust the return if the taxpayer does not respond within a specific time period.

CORRECTIVE ACTION

The Wage and Investment Division, with the cooperation of the IRS Office of Chief Counsel, will discuss with the Department of the Treasury the merits of an administrative proposal to the Internal Revenue Code to obtain limited math error authority to freeze refunds, while contacting taxpayers [REDACTED]

IMPLEMENTATION DATE

January 15, 2012

RESPONSIBLE OFFICIAL

Director, Submission Processing, Wage and Investment Division

CORRECTIVE ACTION MONITORING PLAN

We will monitor this corrective action as part of our internal management system of controls.

The Deputy Commissioner for Operations Support should:

RECOMMENDATION 4

Institute procedures to create and maintain a centralized database that contains all third-party data. Procedures should ensure all data are included in the database, the data are received timely, the data elements are listed for each file, the cost of the file is listed, and the database is searchable so employees can query the database to identify files or data elements that may be useful to their programs.

CORRECTIVE ACTION

We do not agree with the recommendation to centralize the third party database. The IRS has established procedures and processes to receive and maintain third party data



*Better Use of Available Third-Party Data Could
Identify and Prevent More Than One Billion Dollars
in Potentially Erroneous Refunds*

3

across the IRS. The Service will continue to review and update procedures and processes as needed to improve its management of third party data.

IMPLEMENTATION DATE

N/A

RESPONSIBLE OFFICIAL

N/A

CORRECTIVE ACTION MONITORING PLAN

N/A

RECOMMENDATION 5

Institute procedures to ensure all data files received from outside sources are validated for format upon receipt and documented with at least the date and size of the data received on the file, and should also institute procedures to ensure only the contract administrators contact vendors when there are problems with data being submitted or missing.

CORRECTIVE ACTION

We agree with this recommendation. Data validation occurs as Modernization Information Technology Services (MITS) processes the data received and loads it to the various databases. If there are any problems, the MITS Contracting Officer's Technical Representative contacts the vendor who submitted the data and alerts Procurement if contractual obligations are not met.

IMPLEMENTATION DATE

N/A

RESPONSIBLE OFFICIAL

N/A

CORRECTIVE ACTION MONITORING PLAN

N/A