Data Quality: The Journey at IRS Research, Analysis, and Statistics (RAS)

Robin Rappaport
Senior Operations Research Analyst
Data Quality Team Leader
IAIDQ Webinar Facilitator
Background

• **Data Quality Team Leader** responsible for delivery of Data Quality Initiative for Research Databases at Internal Revenue Service (IRS). **Work of team contributed to IRS being awarded a The Data Warehousing Institute (TWDI) 2011 Best Practices Award for the Compliance Data Warehouse (CDW), a Computerworld Honor, and a Government Computer News (GCN) Gala Award.**

• **25 years experience** as Data Quality practitioner. Undergraduate degree in Economics with Computer Science. Graduate work in Operations Research with concentration in Mathematical Modeling in Information Systems. Worked in both **private (6 years) and public sectors (22 years).** Positions include Computer Programmer, Systems Analyst, and Operations Research Analyst.  

• International Association for Information & Data Quality (IAIDQ) **Webinar Facilitator**; a member of Institute for Operations Research and Management Science (INFORMS); Chairman, Individual Membership for Washington, D.C. chapter from 1987 – 1990; elected Secretary and served from 1990 - 1991.

• Lives in Baltimore, Maryland USA with her husband and their four children. She is also a Girl Scout Leader.

• **Specialties:** Data Quality, Metadata, SAS
Data Quality Initiative for IRS Research Databases

- **January 2005**, Temporary assignment to help with strategies and initiatives to improve the quality and delivery of data and information services to the Research community.
- **March 2005**, Discussion Paper by Director, Research Databases entitled: “Using the Compliance Data Warehouse to Improve Data Quality for Research”
- **September 30, 2005**, Permanent Position
Proof of Data Quality Value

- **Problem:** Why CDW numbers differ from another data source used by Projections & Forecasting Group (PFG).

- Response from IRS to Congress should be authoritative.
- Determined differences from CDW to PFG Data Source
- Compared CDW numbers to CDW Data Source

- **Actual Problem:** CDW numbers did not match CDW data source. Only affected certain states.

- **Resolution:** Three missing tapes identified and loaded.
**Why a Data Quality Initiative for IRS Research Databases?**

Data is fundamental to research.

- Data Quality should be applied wherever data is analyzed
- Analysis requires **data** to be **obtained**
- Analysis requires **assessment of quality** of data
- Analysis requires **knowledge of what the data represents**
- Analysis requires **knowledge of what fields should be selected**
- Analysis requires **knowledge of accuracy of the data**
- Analysis requires **knowledge of reliability of the data**
Findings and conclusions only as good as the data

Data Analysis Ladder

- Raw Data
- Static Reports
- Ad hoc Reports
- Descriptive Analysis
- Predictive Modeling
- Simulation & Optimization

What happened?
Why did it happen?
What will happen?

Return on Investment

Data and computing requirements
### Good research requires quality data

**IRS Strategic Foundations: Invest for High Performance**

*Use data and research across organization to make informed decisions and allocate resources*

- **RAS Goal:** Become our customer’s preferred source. Centralized source of *timely, relevant, accurate, accessible, interpretable, and coherent* Data; Metadata; Tools; System Infrastructure; and Training.

- **Ongoing Effort:** Improve quality of research databases; Increase number of users, and enhance online tools and web-enabled knowledge

- **Customer Comments:** “…should win … award for making such useful data available to the research community and working … to ensure data accuracy and consistency!”

- **Industry Recognition:** The Data Warehousing Institute (TDWI) 2011 Best Practices, IAC Excellence.gov Award Finalist 2008, ComputerWorld Honor 2007, and Award from Government Computer News (GCN) 2007 for IRS’ Update of Compliance Data Warehouse makes analysis less taxing

- **Government Recognition:** IRS Enterprise Data Management Office, Canada Revenue Agency, Department of Treasury, Pennsylvania Department of Taxation, National Security Agency (NSA), and Federal Aviation Administration (FAA)
<table>
<thead>
<tr>
<th>Area</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timeliness</td>
<td>- Minimize amount of time to capture, move, and release data to users</td>
</tr>
<tr>
<td></td>
<td>- Leverage new technologies or processes, where appropriate to increase efficiencies throughout the data supply chain</td>
</tr>
<tr>
<td>Relevance</td>
<td>- Ensure data gaps are filled</td>
</tr>
<tr>
<td></td>
<td>- Make new investments in data that reflect expected future priorities</td>
</tr>
<tr>
<td>Accuracy</td>
<td>- Create processes to <strong>routinely assess fitness of data</strong></td>
</tr>
<tr>
<td></td>
<td>- Report <strong>quality assessment findings</strong></td>
</tr>
<tr>
<td></td>
<td>- Publish <strong>statistical metadata</strong> through the CDW website</td>
</tr>
<tr>
<td></td>
<td>- <strong>Cross-validate</strong> release data against source and other system data</td>
</tr>
<tr>
<td>Accessibility</td>
<td>- <strong>Improve</strong> organization and delivery of metadata</td>
</tr>
<tr>
<td></td>
<td>- Provide online knowledge base</td>
</tr>
<tr>
<td></td>
<td>- Facilitate more efficient <strong>searching</strong></td>
</tr>
<tr>
<td></td>
<td>- Invest in third-party tools to <strong>enhance access and analysis</strong> of data</td>
</tr>
<tr>
<td>Interpretability</td>
<td>- <strong>Standardize</strong> naming and typing conventions</td>
</tr>
<tr>
<td></td>
<td>- Create <strong>clear, concise, easy to understand</strong> data definitions</td>
</tr>
<tr>
<td>Coherence</td>
<td>- Develop <strong>common key structures</strong> to improve record matching across database tables</td>
</tr>
<tr>
<td></td>
<td>- Ensure key fields have <strong>common names and data types</strong></td>
</tr>
</tbody>
</table>

## Focus on Improvements

- **Frequency of Updates:** Annual, Semi-Annual, Quarterly, Monthly, Weekly
- **Data Augmentation:** Derived fields, Pre-joining tables, Summary tables
- **Data Profiling, Rule Development, Quality Assessment**
- **Website Redesign and Search Capabilities**
- **Metadata Development and Maintenance** *(Data Definitions, Lookup Tables, Column Profile)*
- **Standard Key Structures, Naming Conventions, and Type Designations**
What is the Compliance Data Warehouse (CDW)?

The **Compliance Data Warehouse (CDW)** is an analytical data environment that is specifically designed to support research activities in the IRS. It is managed and administered by RAS. Key features of the CDW environment include:

**Data**: Taxpayer-level data from over 18 different legacy systems, including Master File, Returns Transaction File, Information Returns, case management systems, and third-party data. Total database storage is **460 terabyte**. Total disk storage is **1.2 petabyte**, CDW is the largest database in the IRS. There are over 41,000 unique data elements (columns) with over 550,000 attributes across 1,680 tables from 32 key data sources. Data sources are added regularly;

**Metadata**: Web-based metadata and dynamic data profiling for all databases, including definitions, lookup tables, cross-references, and other artifacts for over 32,000 data elements (columns); Average daily database queries: 1,200;

**Tools**: Software licenses for SQL clients, SAS, SAS Enterprise Miner/Text Miner, Hyperion Intelligence, ArcGIS, and support for any ODBC- or JDBC-compliant application;

**Computing**: Server-based computing environment for remote processing of complex and high-volume jobs; flexible storage management solutions for both temporary and permanent user files;

**Training and Support**: In-house training for SQL and Hyperion; group rate solutions for SAS and ArcGIS; and general support for data, tools, and account management services;

**Security**: FISMA Certification & Accreditation, Online 5081 Authorization, TIN Masking, Database Audits, System Logging, and other Security Controls.

---

**What is CDW?** It is an environment that provides data, tools, and computing services to the IRS Research community. It offers a high-performance analytical environment for conducting research studies.

**Why is CDW important?** It is the preferred environment for over 900 research analysts and other business users to perform analytical and high-volume data processing. Users include IRS, Treasury, GAO, and others. Average daily concurrent connections: **840**

**Who can benefit from CDW?**

Those who require access to a comprehensive set of compliance data for research and analysis in a secure environment. Average daily database queries: **6,500**.
What types of data are available in CDW?

**Major Categories of Data in CDW**

- **Customer Account**: Payments, abatements, credits, adjustments, freeze conditions, additional assessments, adjustments, reversals, bankruptcies, claims, penalties, offer in compromise, etc.

- **Tax Return**: Federal tax returns filed by individuals, businesses, exempt organizations, and government entities. Includes Forms 1040, 6251, 3520, 1120, 1065, 1041, 990, and others.

- **Information Reporting**: Information filed by a financial institution, employer, partnership, or other party on behalf of a taxpayer. Includes Forms W-2, 1098, 1099-B, 1099-MISC, Schedule K-1, and others.

- **Compliance/Case Management**: Case management systems containing information on examinations, delinquent accounts, underreporter activity, enforcement revenue, or other compliance-based data.

- **Third-Party**: Taxpayer data from federal-state sharing agreements, treaty partners, and other federal agencies, including Social Security Administration, Department of Justice, and State Department.

- **Other**: Customer surveys, national statistical samples, credit bureau data, publically available data, fee-based financial data, and other sources.
CDW: Infrastructure

Database Server (Sybase IQ)  

Shared Storage (1.2 Petabyte)  

Application Server (SAS, Hyperion)  

IRS Network  

SPSS  
Access  
SQL  
Web  
SAS
How are CDW metadata created and published?

**What are metadata?**

Metadata is sometimes defined as “data about data”. Metadata is ideally maintained as a repository of searchable information about available data.

**Why are metadata important?**

Metadata provide the key to turn data into information. They can be used as a tool to select appropriate data for a study and then understand and interpret findings. They can also be used as a basis for developing business rules for data validation and augmentation.

**Who can access CDW metadata?**

Any IRS employee with access to the intranet can view CDW metadata on the CDW website.

CDW maintains a web-based repository of metadata for over 32,000 columns of data. Metadata are available at the database, table, and column level, and are created and updated based on some of the following sources:

- Internal Revenue Manual (IRM)
- Document 6209 (*IRS Processing Codes and Information*)
- Functional Specification Packages (FSPs), Computer Programming Handbooks, Core Record Layouts (CRLs), Program Requirement Packages (PRPs)
- Tax Returns and Tax Return Instructions
- Other official documents and materials

**Examples of CDW Metadata at the Database, Table, and Column Level**

<table>
<thead>
<tr>
<th>Level</th>
<th>Types of Metadata Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>Name, Source, First Year, Last Year, Last Updated</td>
</tr>
<tr>
<td>Table</td>
<td>Name, Database, Frequency, Frequency Type, First Year, Last Year, Last Updated, Number of Columns, Legacy Source</td>
</tr>
<tr>
<td>Column</td>
<td>Name, Legacy Name, Definition, Table, First Year, Last Year, Last Updated, Data Type, Distribution Type, Range Type, Nulls Allowed, Has Lookup, Minimum Length, Maximum Length, Primary Key, Last Updated</td>
</tr>
</tbody>
</table>
The <legacy name> ....

Choose all that apply:

was added in <extract cycle>.  
(It) has data through <extract cycle>.  
(It) is <short description: clear, concise, and easy to maintain>.

It is reported on <Form #, Line#>.  (It is (transferred to OR included on <Form #, Line#> (notated '<notation>').)  (The format is (<word for number> character(s) OR numeric).  OR It is reported in (positive, negative, or positive and negative) (whole dollars OR dollars and cents).)  Valid values (if known) are ....  Values are (if valid not known) .....  It is (zero, blank, null) (if not present OR if not applicable).  (Values (other than valid) also appear.)  (See <related fields>.)
SMEs Needed for Metadata Development: Value Statement

If you provide SMEs for metadata development, you can expect better understanding of data fields due to more complete definitions which should result in improved and more reliable research.

If you do not, then you can expect continued difficulty in identifying fields for research which could result in flawed research and misinformed decisions.

Developed two years ago while attending Danette McGilvray’s tutorial at DGIQ 2010 Conference.

Presently looking for hard-workers capable of synthesizing information from various sources into clear, concise definitions.
What metadata are available on the CDW website?

Database and Table-Level Metadata

Names, descriptions, sources, availability, update status, and links to other internal websites for program or operational information.

Metadata Availability

- Database-Level
- Table-Level
- Column-Level
- Lookup Tables
- Reviews
What metadata are available on the CDW website (cont’d)?

**Metadata Availability**

Database-Level

Table-Level

Column-Level

Lookup Tables

Reviews

---

**Column-Level Metadata**

Definitions, legacy references, availability, release frequency, data types, primary key candidates, Nulls, distribution type, range type, and other attributes.
What metadata are available on the CDW website (cont’d)?

Lookup Tables and Column Reviews

Definitions for unique values of discrete (categorical) fields, and ability to view and submit comments about anomalies or other features.
Standardized Search Results on the CDW Website

Compliance Data Warehouse
Research, Analysis, and Statistics

Resources
- Getting Started
- Databases
- Knowledge Base
- Data Analysis Tools
- Data Reviews

Services
- Training
- Data Analysis
- Record Matching
- Data Transfer
- Storage Services

Support
- Data Alerts
- News and Events
- Contact Us
- About CDW

Related Links
- Select Related Links

32 Search Results - page 1 of 4 for adjusted gross income

Adjusted Gross Income Amount Per Return
Database: Automated Underreporter (AUR)
Table: AUR • AGI

The Adjusted Gross Income Amount Per Return is the amount claimed as Adjusted Gross Income (AGI) after subtracting all adjustments (Form 1040, Line 36) from Total Income (Form 1040, Line 22). It is reported on Form 1040, Line 37. It is reported in positive and negative dollars and cents.

Adjusted Gross Income Prior Amount Per Return
Database: Automated Underreporter (AUR)
Table: AUR • AGI PRIOR

The Adjusted Gross Income Prior Amount Per Return is the amount claimed as Adjusted Gross Income (AGI) after subtracting all adjustments from Total Income for the prior year. It is reported in positive and negative dollars and cents.

Underreported Earned Income Tax Credit (EITC) Adjusted Gross Income (AGI)
Database: Automated Underreporter (AUR)
Table: AUR • UREITCAGI

The Underreported Earned Income Tax Credit (EITC) Adjusted Gross Income (AGI) was added in 2004S2. It is the amount of underreported EITC AGI. It was changed in 2005S2 to remove "Modified" from the legacy name. It is reported in positive and negative dollars and cents. Values are zero in 2003S2.
Identifying patterns (and problems) in data with profiling

**What is Data Profiling?**
Data profiling involves a standardized analysis of data to determine its completeness and suitability for use.

**Why is profiling important?**
Analyzing basic patterns in data can reveal both insights and anomalies that are useful in describing the underlying structure and improving data quality.

**Who can profile CDW data?**
As of March 2010, those with access to the IRS intranet can profile data via the CDW website. One does not need to be a CDW user to view the CDW website and perform basic profiling.

Data profiling is a process of analyzing data to gauge overall suitability for use. Common profiling tasks can help identify:

- Invalid values in fields (values out of range)
- Missing values or empty fields (fields containing no data at all)
- Inconsistent methods of representing the same value
- Data elements used for purposes other than expected
- Violation of business rules
- Unrealistic frequencies or percentages of specific values in a column
- Violations of referential integrity
- Misspelled text values

**The Role of Metadata in Data Profiling**
Metadata includes valid values, when known. The process of data profiling is more informative when valid values are known. The actual values identified as part of the data profiling can be compared to the valid values. This allows for Data Validation of a specific data field using the data in that one field. Metadata can also enable cross-validation to what should be the same field on other data tables. Metadata can also include more complex business rules that can be used for validation.
What data profiling features are available on the CDW website?

Table Statistics

Row counts for a given table by State, County, and Zip Code. Row counts represent original, unaltered data from the authoritative source.
What data profiling features are available on the CDW website (cont’d)?

Frequency Table

Frequencies (row counts) for unique values of a discrete (categorical) field for a specific time period. Cumulative statistics are generated.
What data profiling features are available on the CDW website (cont’d)?

**Data Profiling Features**

- **Table Statistics**
- **Frequency Table**
- **Column Statistics**
- **Trend Analysis**
- **Geographic Maps**
- **Reviews**

**Column Statistics**

Basic distributional statistics for a given time period and filter condition. Users can drill down on unique values of a discrete (categorical) field.
What data profiling features are available on the CDW website (cont’d)?

**Trend Analysis**

Basic statistics over time and by unique values of a discrete (categorical) field. Up to five years of data can be displayed.
What data profiling features are available on the CDW website (cont’d)?

**Geographic Maps:** U.S., State, and County level

**Data Profiling Features**

- **Table Statistics**
- **Frequency Table**
- **Column Statistics**
- **Trend Analysis**
- **Geographic Maps**
- **Reviews**
What data profiling features are available on the CDW website (cont’d)?

Data Profiling Features

- Data Reviews: Comments on specific data fields (columns) posted by users
- Table Statistics
- Frequency Table
- Column Statistics
- Trend Analysis
- Geographic Maps
- Reviews

Sample Data Reviews:

**Preparer Taxpayer Identification Number (TIN)**, May 03, 2012
By Marques Uma R
**Database**: Individual Master File (IMF)
**Table**: INF_TINS_HISTOERY • PREP_TIN
What is the significance of prep_on having a value of 0?

**Treatment Code**, April 04, 2012
By Svrilna Jennifer R
**Database**: Collection Information (COLL)
**Table**: COLL_TREAT_RESP • TREATNT_CD
Some treatment codes (treatnt_cd) appear to have dropped the 4th character in the loading process in August & September 2008. Research from ACS production server indicates the 4th character appeared in the production database, for those observations that were not already dropped off archive database, but did not make it to the CDW database. Affected treatment codes appear in CDW as: ‘L11’, ‘LV1’, ‘FM1’, ‘MC1’, ‘MC0’. These are levy and letter actions but when levy and letter actions cannot be known now that the data has dropped off the ACS production servers. Affected dates (actn_d) mostly distributed in August and September 2008 but span 22May2007-27SEP2008. 50% of the dropped treatment codes appear on 20SEP2008, 30AUG2008, 13SEP2008, 27SEP2008, and 26AUG2008, in descending order.

**Schedule C1 EIN**, March 22, 2012
By Canada Wanda H
**Database**: Electronic Tax Administration Marketing Database (ETA MDB)
**Table**: ETAatee • SCHD_C1 EIN
When will Processing year 2012 data be available.
What types of alerts are available on the CDW website?

### Data Alert Types

- **New Metadata**
- **Updated Metadata**
- **Data Augmentation**
- **Standardization**
- **Accuracy Issues**
- **Data Corrections**

#### Data Alerts

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr 2012</td>
<td>New BRTF Metadata for Form 3520&lt;br&gt;New Business Returns Transaction File (BRTF) metadata are available for Form 3520. The table is named BRTF_F3520. The updated metadata corresponds to Tax Year (TY) 2011/Processing Year (PY) 2012. <a href="#">View the data definitions.</a></td>
</tr>
<tr>
<td>Apr 2012</td>
<td>New TC420_AIMSSRCCD Lookup Table for IMF_TRANS_HISTORY&lt;br&gt;New lookup table is available for TC 420 AIMS Sourc Code. The table is named TC420_AIMSSRCCD. It is associated with four tables including IMF_TRANS_HISTORY. The new metadata corresponds to PY 2012. <a href="#">View the lookup table.</a></td>
</tr>
<tr>
<td>Apr 2012</td>
<td>New Collection Location Code (CLC) Lookup Table for Individual and Business&lt;br&gt;New lookup table is available for Collection Location Code (CLC). The table is named CLC. It is associated with CLC for both individual and business returns. The new metadata corresponds to PY 2012. <a href="#">View the lookup table for business returns.</a></td>
</tr>
<tr>
<td>Apr 2012</td>
<td>New TDA Closing Code Lookup Table for TC 470, 500, 520, 530&lt;br&gt;New lookup table is available for TDA Closing Code. The table is named TDA_CLS_CD. It is associated with closing codes for TC 470, 500, 520, and 530. The new metadata corresponds to PY 2012. <a href="#">View the lookup table.</a></td>
</tr>
<tr>
<td>Feb 2012</td>
<td>New TDI Closing Code Lookup Table for TC 59X&lt;br&gt;New lookup table is available for TDI Closing Code. The table is named TDI_CLS_CD. It is associated with nine fields across eight tables on databases including ARDI, BMF, COLL, ERIS, and IMF. The new metadata corresponds to PY 2012. <a href="#">View the lookup table.</a></td>
</tr>
<tr>
<td>Feb 2012</td>
<td>Updated IRTF Metadata for Form 1040 (Schedule A)&lt;br&gt;Updated Individual Returns Transaction File (IRTF) metadata are available for Form 1040 (Schedule A). The table is named IRTF_SCHED_A. The updated metadata corresponds to Tax Year (TY) 2012. <a href="#">View the data definitions.</a></td>
</tr>
</tbody>
</table>
CDW: Data Analysis Tools

Large data volume means choosing the right tool for the right job

<table>
<thead>
<tr>
<th>Database Greater Than</th>
<th>Data Warehouse Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^6$ (1 MB)</td>
<td>Tiny</td>
</tr>
<tr>
<td>$10^9$ (1 GB)</td>
<td>Small</td>
</tr>
<tr>
<td>$10^{10}$ (10 GB)</td>
<td>Big</td>
</tr>
<tr>
<td>$10^{11}$ (100 GB)</td>
<td>Large</td>
</tr>
<tr>
<td>$10^{12}$ (1 TB)</td>
<td>Very Large</td>
</tr>
<tr>
<td>$10^{13}$ (10 TB)</td>
<td>Huge</td>
</tr>
<tr>
<td>$10^{14}$ (100 TB)</td>
<td>Massive</td>
</tr>
<tr>
<td>$10^{15}$ (1 PB)</td>
<td>Ridiculous*</td>
</tr>
</tbody>
</table>

* Attributable to Ed Wegman, Center for Computational Statistics, George Mason University

- The larger the amount of data being analyzed, the greater the efficiencies from remote computing and using SQL or SQL-based products
## Tools for Data Quality and Analysis

<table>
<thead>
<tr>
<th>Analysis Tool</th>
<th>Functional Characteristic</th>
</tr>
</thead>
</table>
| SQL           | ▶ Easy-to-use database language consisting of a few basic commands that provides fast and efficient retrieval, summarization, and processing of data  
▶ Virtually all ad-hoc or descriptive queries performed by Research users are conducive to SQL |
| Hyperion      | ▶ Web-enabled, point-and-click tool for query, analysis, and report-writing activities  
▶ Ability to interactively create pivots, charts, and user-defined table views  
▶ Server-side implementation for maximum performance |
| SAS           | ▶ End-to-end data management and business intelligence software package supporting a wide range of functionality for data management, statistical analysis, econometrics and time series, operations research, GIS, data visualization, data mining, database connectivity, web and application development, and other business and analytical needs  
▶ Server-side processing available to support analysis of very large databases  
▶ Most widely used all-purpose analysis software among federal statistical agencies  
▶ **DataFlux** product purchased to assist with Data Quality Assessment |
| SPSS          | ▶ Data analysis tool widely used in the Research community that can be used to retrieve data directly from CDW databases without any intermediate steps |
| Other         | ▶ Microsoft Access, Excel, and other third-party tools can be used for data retrieval, but are not suited for processing data on the server |
# Quality Assessment

## Rule Failures Reported by Column

<table>
<thead>
<tr>
<th>Column Name</th>
<th>YEAR</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Rules</th>
<th># Errors</th>
<th># Rows</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule 1: (Valid Values)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule 2: (Sign Test)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule 3: (Range Test)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule 4: (Legislative)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rule n: (Others)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Current Status and Future Plans

- Data Integration
- Data Profiling
- Rule Development
- Quality Assessment
- Continuous Monitoring
- Data Stewardship
Acknowledgements

• Jeff Butler, Director, Research Databases, IRS RAS

• CDW Data Quality Team (Headquarters, Field, and Contractors)

• Lwanga Yonke, Information Quality Process Manager, Aera Energy LLC and Advisor to Board of Directors, IAIDQ
Data Quality: The Journey at IRS Research, Analysis, and Statistics (RAS)

Robin Rappaport
Senior Operations Research Analyst
Internal Revenue Service (IRS)
Research, Analysis, and Statistics (RAS)
RAS:DM:RDD:TRD
1111 Constitution Avenue, N.W. K
Washington, D.C. 20001

202-874-0578
Robin.Rappaport@irs.gov