Criminal Investigation
Agents of the Future

SAS Institute
Southeastern Association of Tax Administrators (SEATA)
July 2019
42 Years

42

World’s LARGEST
Privately held software company

94%
Of Fortune 100 Companies Use SAS

93%
Annual customer retention rate

14,021
SAS employees worldwide

#1
Advanced Analytics
“SAS leads in this space and it takes the next 8 competitors combined, to equal to SAS market share in advanced analytics” – IDC July 2016.

26%
SAS revenue reinvestment annually into R&D

>83,000
Customer sites in 146 countries

Leadership position in Analysts’ Validation

Gartner

IDC

FORRESTER

OVUM

US$3.24B
Revenue in 2017

Revenue growth for every year of operation

BUSINESS ANALYTICS

World’s LARGEST Privately held software company

94% Of Fortune 100 Companies Use SAS

93% Annual customer retention rate

14,021 SAS employees worldwide

#1 Advanced Analytics
“SAS leads in this space and it takes the next 8 competitors combined, to equal to SAS market share in advanced analytics” – IDC July 2016.

26% SAS revenue reinvestment annually into R&D

>83,000 Customer sites in 146 countries

Leadership position in Analysts’ Validation

Gartner

IDC

FORRESTER

OVUM

US$3.24B Revenue in 2017

Revenue growth for every year of operation
Current Agencies Using SAS Analytics for Tax and Revenue

SAS is used to process or protect

6 of every 10 tax dollars

collected by national governments globally

*Based on analysis of 2016 OECD annual revenue collections

1 Four additional countries who wish to remain anonymous
2 One additional state who wishes to remain anonymous
SAS is the de facto standard for tax analytics

Tax authority clients include...

Australia
Austria
Belgium
Canada
Canary Islands
Chile
Denmark
Finland
Hong Kong
Hungary
India
Ireland
Lithuania
New Zealand
Philippines
Poland
Punjab (India)
Saudi Arabia
Singapore
Slovakia
South Africa
Spain
United Kingdom
Japan
SAS is the de facto standard for tax analytics

Tax authority clients include...

U.S. Internal Revenue Service
U.S. Treasury Inspector General (TIGTA)
North Carolina
Iowa
Minnesota
Wake County (North Carolina)
Tennessee
New Mexico
Kansas
New York
Delaware
California
Oregon
District of Columbia
Wisconsin
Maryland
Rhode Island
Maine
Pennsylvania
Montana
Missouri
Massachusetts
Florida
Kansas
New Jersey Treasury
New York City
Oklahoma
Virginia
Positions Held at IRS Over 30 Years

Executive Director, International Operations (Retired)
IRS Criminal Investigation Division (CI)

- Executive Director, International Operations, CI
- Executive Director, Quality & Technical Support, SBSE Collection
- Executive Director, Rulings & Agreements, TEGE
- Deputy Director, International Operations, CI
- Deputy Director, Office of Strategy, CI
- Director, Office of Planning & Strategy, CI
- Director, Office of Refund Crimes
- Assistant Special Agent in Charge, Cincinnati Field Office, CI
- Supervisory Special Agent, Newark Field Office, CI
- Senior Analyst, HQ, Washington DC, CI
- Special Agent, Baltimore Field Office, CI
- Revenue Agent, SBSE Examination Division
The Data Tells The Story
INTERNATIONAL OPERATIONS

IRS CI INTERNATIONAL Operations (IO) educates foreign governments and agencies on crime detection, investigative techniques, case studies, emerging trends, and best practices. CI trains foreign governments through collaborative efforts with the International Law Enforcement Academies (ILEA) in Budapest, Bangkok, San Salvador, and Gaborone, Botswana. CI IO also presents training for foreign law enforcement and governments around the world sponsored by the Bureau of International Narcotics and Law Enforcement Affairs (INL), Overseas Prosecutorial Development, Assistance and Training (OPDAT), and the State Department. This included training for judges, prosecutors, and investigators in Nepal; International Financial Investigations courses in Colombia, Mexico, and Argentina; and cybercrime and virtual currency training sessions in Australia and New Zealand.

Internationally, IRS CI has special agent attachés strategically stationed in 10 foreign countries (Canada, Mexico, Colombia, Panama, Barbados, The Netherlands – Europol, England, Germany, China, and Australia). Attachés continuously build and maintain strong alliances with foreign governmental, law enforcement, and industry partners. These strong strategic alliances provide CI with the ability to develop international investigative leads and support domestic investigations that have an international connection. Special agent attachés are vigilant in uncovering emerging schemes that promoters and financial institutions perpetrate to help U.S. taxpayers evade their federal tax obligations.
NCITA provides training to foreign governments and agencies on crime detection, investigative techniques, case studies, developing trends and best practices. CI trains foreign governments through collaborative efforts with the International Law Enforcement Academies. These academies are in Budapest, Hungary; Bangkok, Thailand; San Salvador, El Salvador; and Gaborone, Botswana and at the International Academy for Tax Crime Investigation at Guardia di Finanza Economic and Financial Police School in Ostia, Italy sponsored by the Organization for Economic Cooperation and Development (OECD), as well as State Department-sponsored training.

IRS CI special agents attend quarterly firearms, defensive tactics and semi-annual building entry training. Through frequent use of force training, they maintain their skills and abilities so they always use good judgement and the appropriate degree of force necessary to safely carry out enforcement activities. These activities include search warrants, arrests, surveillance, armed escorts, dignitary protection, undercover activities and seizures.
EVERY INVESTIGATION CRIMINAL Investigation conducts involves digital and multimedia evidence of some type. Digital and multimedia evidence comes from many sources, such as witnesses, subpoenas, personal computers, mobile devices (phones, tablets, etc.), small and large business computers/servers, server farms, cloud storage or even the Dark net. The proper collection and review of digital evidence requires specialized skills to be admissible in court proceedings.

IRS–CI Electronic Crimes’ primary function is the forensic acquisition, analysis and testimony of the digital and multimedia evidence related to ongoing criminal investigations. Electronic Crimes has a presence in 60 posts-of-duty across the United States and consists of five computer investigative forensic analysts, six senior analyst computer investigative forensic analysts (CIFA), 80 special agent – computer Investigative specialists (SA-CIS), four forensic assistant contractors, and one administrative support employee. The SA-CISs and CIFAs receive advanced training in the proper procedures of collecting, acquiring, analyzing and testifying on digital and multimedia evidence. Additionally, they provide assistance in drafting search warrants, court orders and subpoenas. They travel across the country to participate on search warrants and conduct the digital and multimedia evidence seizures.

Types of forensic analysis the SA-CISs and CIFAs perform include:
- Dark web activity
- Encryption and password recovery
- Targeted searches
- Deduplication of large data sets (such as email)
- Recovery of hidden and deleted data
- Extraction of data from proprietary financial software (tax preparation, accounting, payroll, point of sale systems, custom database, etc.)
- Internet activity and history analysis
- Data conversion from proprietary formats
- Website preservations
- Taint reviews and segregation
- Recovery & repair of damaged disk drives

In FY 2018, Electronic Crimes personnel participated on 327 search warrants, performed at 553 locations; seized 1.76 petabytes of data from 1,918 computers/laptops/external devices and 1,845 mobile devices; and testified in 16 trials. Electronic Crimes is the premier source for digital and multimedia evidence in IRS–CI.*
Cyber Crimes

- Data intrusion, business email compromise, phishing schemes, bank account takeovers and data loss incidents.
- Selling, buying and compromise of PII via internet
- Virtual currency-based tax and money laundering schemes
- Dark web marketplace owners, admins, and large vendors
- Terrorist financing, incl. use of virtual currency, mesh networking, and other online means to raise and funnel funds
Crossing the Streams

Current State:

• Desktop based investigation methods

• Several different systems to get results. You use the one that you like, not the best one (might use 5 different computers)

• Agencies collect a lot of information, but not analyzing it all cohesively – e.g. cash flow analysis for a particular industry (e.g. roofers, restaurants)

Future State:

• Server based investigation methods. Agent data taken off desktop, moved to server, and merged (across cases).

• One or two systems to get results. Data integrated, patterns identified, and surfaced in single user interface for agent.

• Self-service data analysis tools that don’t require coding/SQL; visual for non-technical agents
Data Extraction from Investigation Target

Current State:

- Search warrant; computers (cell phones, tablets, hard drives, thumb drives, Alexa, etc.)
- Agent do all manually: 1) Extract data 2) Make it analysis-ready 3) Find the connections and patterns
- CIs have typically done #1 well, but slowly. Still room for improvement in data extraction. #2-3 most room for improvement.
- Swiss Bank – IRS systems vs. DOJ system compatibility issues; slowed down intelligence gathering.
- Data goes in, but can’t get it out.

Future State:

- Data ingestion, linkage and analysis done on site vs back at office. Example = Scanning, text mining, entity extraction, and link analysis done all in real time.
- Make sure the company you engage with is cutting edge with active R&D to keep up with any data and any systems it may have to engage with. Tech keeps advancing.
- Make sure you own your data! (Data goes in some systems, but not allowed to come out!)
Tech vs Policy

Current State:
• Policy that governs global economy not keeping pace with technology (e.g. base erosion, profit shifting, crypto (BITCOIN) and virtual (VENMO)
• Criminals are outrunning our policy
• Lagging policy creates a problem with identifying the entity’s tax home by getting data from other countries
• Issue is national and international

Future State:
• Automatic Exchange of Information (EOI) – Exchange of information in real time with other tax departments worldwide
• Policy and tech need to be in place to support this. Need to be forward thinking.
State and Federal Collaboration

Current State:
• Compliance group within FTA
• Tax fraud info sharing in ISAC
• Otherwise, CIs are isolated – somewhat intentionally, but what are consequences of this?
• But what happens with the cases that fall under IRS CI threshold (e.g. $1M in tax liab)? There’s a gap – opportunity for improvement.

Future State:
• Collaboration with State police – Car stop; Person had cash and Treasury checks in car (possibly tax prep materials); often criminals move from violent crime to tax fraud (financial crimes)
• Do we create a CI-focused branch of the ISAC?
• Does FTA create a CI annual conference / focus group? Is there one today?
• What policy need to be created legally to support this information sharing and collaboration?
“Whole of Government”
Global approach to combatting financial fraud.

J5 Australia, Canada, Netherlands, UK and US

JITSC International Taskforce on Shared Intelligence and Collaboration

FATCA Foreign Account Tax Compliance Act

OECD Organization for Economic Cooperation and Development

CRS Common Reporting Standard
Agents of the Future: Using Math to Find Insights in Data

Taking complex math that looks like this...

\[
\begin{align*}
X_1 &= \mu_1 + \sigma_1 \left( \sqrt{\frac{1}{\rho U_1}} + \sqrt{\frac{1}{1-\rho} U_2} \right), \\
X_2 &= \mu_2 + \sigma_2 \left( \sqrt{\frac{1}{\rho U_1}} - \sqrt{\frac{1}{1-\rho} U_2} \right),
\end{align*}
\]

where \( U_1 \) and \( U_2 \) are independent and standard normal random variables, \( \sigma_1 \) and \( \sigma_2 \) and \( \mu_1 \) and \( \mu_2 \) are real numbers, and \(-1 < \rho < 1\).

(i) (10 points) Find \( E X_1 \), \( E X_2 \), \( \text{var}(X_1) \), \( \text{var}(X_2) \), and \( \text{cov}(X_1, X_2) \).

(ii) (10 points) Find the joint density function of \( (X_1, X_2)^\prime \), starting from the bivariate normal density and making a transformation.

(iii) (5 points) Make a surface plot of the joint density function of \( (X_1, X_2)^\prime \), with \( x_1 \) and \( x_2 \) varying from \(-\frac{\pi}{2}\) to \( \frac{\pi}{2} \), and \( \rho \) generated from \( \text{rand}(5) \).

(iv) (5 points) Generate bivariate normal random numbers via those equations.

(v) (5 points) Construct a contour plot for your simulated data in Part (iv).

(vi) (5 points) Make a 3d histogram of your simulated data in Part (iv).

...and turning into easy to use data analysis tools that look like this.

Investigators doing same thing today with Excel and sticky pads, just much more slowly. How can we help our agents find (and work) better cases faster?
4 Main Types of Analytics Used in Tax Investigations

- **Predictive Modeling**
  - Identifies characteristics of bad tax returns

- **Outlier Detection**
  - Peer groups/outliers

- **Network Analysis**
  - Connections

- **Text Mining**
  - Unstructured data
Predictive Modeling – Generates a Profile of Bad Actor

- Uses prior case data to teach the computer - “Labeled”
  - Cat / Not Cat
  - Hot Dog / Not Hot Dog
  - Tax Evader / Not Tax Evader

- That resulting list of characteristics is called a “model”.

- That model is then used to evaluate new data for that same infraction (e.g. tax fraud, insider fraud, money laundering, cyber breach)

- Model gets smarter with new case closures – More labeled data to learn from!
Computer ingests all of the prior cases that were successful and unsuccessful.

- Adjusted Gross Income Decile: 0.80
- Tax liability due: 0.92
- Charitable Contributions: 0.01
- Medical expenses: 0.75
- Filing Status (Married, Single): 0
- Geography (e.g. region): 0.64

Computer tells us which of the characteristics have the most “predictive power”. If > .7, considered highly predictive.

New CI Case Leads Generated

Runs on new data

Feedback loop

Model gets smarter over time if case outcomes are fed back in. Feedback loop not automatic, needs to be configured to do this.
4 Main Types of Analytics Used in Tax Investigations

- Predictive Modeling: Identifies characteristics of bad tax returns
- Outlier Detection: Peer groups/outliers
- Network Analysis: Connections
- Text Mining: Unstructured data
Outlier Detection

- Computer groups taxpayers by like characteristics ("peer groups") and looks for taxpayers who fall outside

- Used when past case data not available or reliable

- Helpful to find new segments to investigate that were previously ignored (e.g. nail salons vs restaurants), or for forensics purposes to find financial transactions or activities that are aberrant.
4 Main Types of Analytics Used in Tax Investigations

- Predictive Modeling: Identifies characteristics of bad tax returns
- Outlier Detection: Peer groups/outliers
- Network Analysis: Connections
- Text Mining: Unstructured data
Network Analytics (Connections)

Networks are built on the implicit relationships between individuals and companies (for example, people who share addresses, phone numbers, or bank accounts). Not all fraud is networked, but much is.

- Particularly useful for complex cross border activities and to see “social networks” of suspect individuals.
- Take down the network not just the individual.
- Corporate structure easily revealed and visualized.
- Complex flow of financial transactions can be followed.
- Structures and transaction flow can be identified as “high risk”
Fraudulent Tax Preparer

Tax preparer associated with flagged return.

9 other tax returns associated with same preparer. They have also been flagged for fraud, for the same reason.

Original tax return flagged for alert due to unusual deductions.

Case turned out to be 10 tax evaders and 1 criminal tax preparer working in collaboration.
Untangling the Knot
Super Clusters
Network Analytics
Much More Than a Neat Diagram

• **Strength or frequency of interaction**
  • Modeled as weights (strong or weak link) on the links of the resulting network

• **Network Metrics**
  • Degree – How many transactions; size of transaction
  • Influence – Large transaction in sea of small ones
  • Closeness – 1 hop away vs. 5 hops
  • Centrality – Identifies the most important entities in a network
  • Hub – What is central point
  • Betweenness – Who has the ability to control info flow (“gatekeepers”)
  • Eigenvector – How many high influence nodes are you connected to? (guilt by association)

• Perform network optimization based on a set of algorithms such as clique, cycle, shortest path, minimum-cost network flow, linear assignment, minimum spanning tree, minimum link-weight cut, transitive closure, and traveling salesman problem
# Network Analysis Techniques

## Networks visualization and ego-centric analysis
- Displays relationship between selected alert and any other related alerts through link and nodes
- Identify links with known (blacklisted) entities

## Entity link analysis / entity resolution
- Detecting rings in two mode networks of people and attributes
- Identify rings in first party fraud collusions

## Graph walking to identify rings
- Walk through the graph to identify rings
- Identify paper rings of fraudulent collusions

## Centrality measures
- Ranking nodes based on various graph centrality parameters
- Identify leaders in fraud networks

## Snowball method
- Identify suspects and recursively expand their connections using snowball method
- Identify linkages to known (blacklisted) entities

## Peer group analysis
- This technique detects abnormal behavior of a target by comparing it with its peer group and measuring the deviation of its behavior from that of its peers
- Abnormal changes when compared to the peer group

## Network topologies - cliques and stars
- Any quantitative or qualitative features of a user behavior in online social networks that are inconsistent with the rest of users can be considered anomalies
- Anomaly Detection - Identify outliers in networks

## Page-rank
- The Page-Rank algorithm can be used to discover the critical accounts of the groups
- Collusive fraud groups

## Combining user level and network level features
- Combine user level attributes with network level attributes
- Identify conspired groups
4 Main Types of Analytics Used in Tax Investigations

- Predictive Modeling
- Outlier Detection
- Network Analysis
- Text Mining

- Identifies characteristics of bad tax returns
- Peer groups/outliers
- Connections
- Unstructured data
Data comes in native language and translated, then linked to tax authority data (and other external data).
UK Land Overseas Company Ownership Data (OCOD)
- Properties owned by offshore and foreign companies

UK Land Commercial and Corporate Ownership Data (CCOD)
- Properties owned by UK companies

Brazil STJ Court of Appeals
- Provides basic descriptions of most recent case decisions for high profile appeals, party names and individuals who run the case

China Foreign Trade Enterprise Directory
- Provides company name, in Chinese and English

China Liaoning Quarantine Bureau
- Official English Trade names for Chinese companies

China MOFCOM Foreign Investment Directory
- Official English Trade names for Chinese companies

China SAIC Corporate Registry
- Directorship, shareholding, subsidiaries, change notices, and standard company information

DPRK Maritime Authority Seafarers Directory
- Name, DOB, licenses, and certification for all registered Seafarers, in English

DPRK Maritime Authority Vessels Directory
- Vessel ownership and identifying information

China-DPRK Trade Directory 2013-2016 & 2017
- Chinese companies that export or import to and from the DPRK, with contact personnel and information

Unique personal identifiers are necessary to disambiguate between individuals with the same name or prevent falsely linking possibly same as entities. Fortunately, some of these identifiers are hidden in official public records. Chinese ID numbers, for instance, provide additional information beyond just the ID number itself, including:
- Date of Birth
- Place of Birth
- Gender Information
- Job Status

LG Public Procurement Network
- Standard company information and banking information for registered government traders

Copyright © SAS Institute Inc. All rights reserved.
Agora Marketplace

- HTML scraped and parsed
- Intensive ETL/data Cleansing process
- Loaded into investigative user interface for exploration
- Text analytics, network analytics, risk scoring

CI Use Case: Dark Net Monitoring
Global Collaboration is Increasing
Joint International Take Down of “Wall Street (DarkNet) Market”

• U.S. Department of Justice (DOJ), Internal Revenue Service (IRS), Federal Bureau of Investigation (FBI), U.S. Drug Enforcement Agency (DEA), HSI, U.S. Postal Service, Interpol, Europol

• Germany, Netherlands, Brazil, U.S. and Eurojust.

• Significant assistance from Criminal Division’s Office of International Affairs and Organized Crime and Drug Enforcement Task Force Program.
Tax authorities are not just going after tax evaders themselves.

As of September 2018, over 80 banks had made such agreements with the United States under Swiss Bank Program.

Data can be used for modeling!
• **4 people charged** – Tens of millions; scheming for decades

• Former investment manager, a former U.S. resident, an American accountant and a Panamanian lawyer who once worked for the firm

• **11 million documents**, which a consortium of journalists made public in April 2016
Florida Keys Preparer Sentenced to Five Years for Filing Fraudulent Tax Returns That Resulted in an Aggregate Loss to the IRS of $14,500,000

A Monroe County, Florida tax return preparer was sentenced today to 60 months imprisonment after he pled guilty to filing fraudulent tax returns with the Internal Revenue Service (IRS).

Ariana Fajardo Orshan, U.S. Attorney for the Southern District of Florida, Michael J. De Palma, Special Agent in Charge, Internal Revenue Service, Criminal Investigation (IRS-CI), and George L. Piro, Special Agent in Charge, Federal Bureau of Investigation (FBI), Miami Field Office, made the announcement.

According to documents filed with the court, Pedro C. Rodriguez, 51, of Marathon, owned and operated the JC Mar Tax Services tax preparation business, located in Marathon, Florida. From approximately 2007 through 2017, Rodriguez filed fraudulent returns for his clients seeking refunds to which the clients were not entitled by reporting fictitious itemized deductions and fraudulent education and residential energy credits. For each of the years 2013 through 2017, Rodriguez submitted between 1,900 and 2,200 tax returns on behalf of his clients. At the sentencing hearing, the government provided evidence that established Rodriguez’s submission of fraudulent tax returns resulted in a loss amount to the IRS of $14,500,000.

U.S. District Judge K. Michael Moore also sentenced Rodriguez to one year of supervised release and ordered him to pay $14,569,171 in restitution.

U.S. Attorney Fajardo Orshan commended the investigative efforts of the IRS-CI and FBI in this matter. Assistant U.S. Attorney Daniel J. Marcet prosecuted this case.

• From 2007-17, Rodriguez submitted ~2,000 returns per year for his clients.
• Fictitious deductions and fraudulent education and residential energy credits.
• For one client, he claimed $6K in solar water heating, $9K in mortgage deductions, $13K on job expenses and others totaling $31K – none of which was true.
• Would give clients paper copies of final returns which did not show the deductions/credits.
• Charged between $80-$120 for preparation fees to his clients.
The best part. Internet review in 2015. Could it have been stopped earlier by scraping this data, and connecting it back to the tax returns filed?
In 2014, two North Carolina-based retail business owners found with $1.789M in cash in vehicle by Arkansas State Police.

- Found business ledger in their home consistent with illegal tax prep services.
- Officers seized 9 foreign ID documents, 300 ITIN letters, 113 tax related records, 179 W2s, and other ledgers indicating returns filed and refund status.
- Computer seized had TaxWise software files for 229 “clients” for 2008-11.
- 5 common employers and mailing addresses across many (75) of the tax returns
In Summary

- **DO Cross the Streams** - It is no longer enough to work in parallel on separate cases, but rather move towards a vision where both the agents (and their case data) are connected, thereby revealing overlaps and insights not previously realized.

- **Everyone Can (and Must) Contribute**—Enable your entire organization to do data analysis and intelligence gathering regardless of their technical skillset (or lack thereof). The line between “agent” and “data analyst” is growing thinner by the day. Fit-for-purpose user interfaces and capabilities must remain a focus for the organization.

- **New Data Sources** - PDF, Word, Excel, text, images, voice...resting in a database or on a PC...or even streaming in real-time from the edge. The clues aren’t just in DB2 and SQL Server anymore.