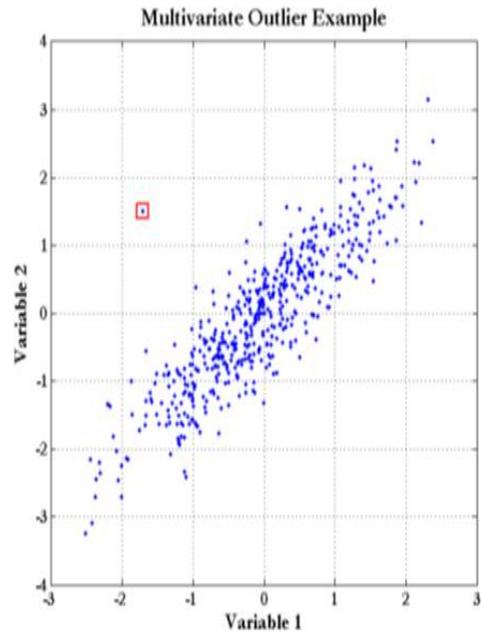# Treasury Inspector General Tax Administration (TIGTA)

Federal Audit Executive Council
Bi-Monthly Meeting & Training
October 24, 2012

Finding The "RED DOT" in Tax Administration's "BIG DATA"

OR

Reducing the Hay to Find a Needle

Agenda

- Background on IRS and TIGTA's Oversight Role
- Selected Audit Findings - Data Analysis Examples
- Data Mining & Predictive Analytics Project
  - Organizations Visited
  - Centers for Medicare & Medicaid
  - Postal Service Inspector General
  - Recovery Accountability and Transparency Board
- Lessons Learned
- Suggested Reference Materials
- Questions & Discussion

4

IRS Tax Statistics (BIG DATA)

Type of Return	Number of Returns	Gross Dollars
Individual Tax	143,607,800	$1,331,160,000,000
Corporation Tax	2,312,909	$242,848,000,000
Other Taxes	29,664,798	$825,922,000,000
Individual Refunds	110,900,000	$327,400,000,000
S Corporations	4,094,562	N/A
Partnerships	3,248,481	N/A
Nonprofits	406,034	N/A

Recent TIGTA Media Articles

- BNA Daily Tax Report
  - TIGTA finds IRS missing major "Red Flags" pointing to Billions in Identity Theft Fraud
- ABA Journal
  - TIGTA reports IRS sent refunds to ID thieves :
    - $3.3 Million Went to One Address Listed on 2,137 Returns
- South Florida Sun-Sentinel
  - TIGTA audit finds 741 Tax Returns filed from Belle Glade home
- Computerworld
  - Audit found the IRS paid refunds to criminals who filed false tax returns.

Selected TIGTA Reports

- Earned Income Tax Credit -$11-$13 Billion Improper
- Education Tax Credits-$3.2 Billion Erroneous
- Identify Theft-$5.2 Billion Lost
- Electric/Hybrid Car Credits-$33 Million in Error
- Adoption Tax Credit-$4 Million Questioned
- Prisoner Tax Fraud-$295 Million In Refund Fraud
- First Time Homebuyers Credits Questioned

$

Identity Thieves Move Into Tax Administration
Report Number: 2012-42-080

IRS Identified Cases

- 1 Million False Returns
- Blocked $6.5 Billion in Refunds

TIGTA Audit Findings

- Identified 1.5 Million Additional Returns
- $5.2 Billion in Fraudulent Refunds Escaped Detection

$21 Billion over 5 Years if not corrected

Identity Theft Address Analysis

City Address	Number of Returns To One Address	Refund Amounts To One Address
Lansing	2137	$3,316,051
Chicago	765	$ 903,084
Belle Glade	741	$1,004,897
Orlando	703	$1,088,691
Tampa	518	$1,791,770

Identity Theft
Returns by Taxpayer Category

Taxpayers Targeted	Number of Returns	Refund Amounts
Deceased	104,950	$ 415,047,568
Elderly	76,338	$ 374,419,730
U.S. Territories	67,789	$ 387,183,428
Students (16-22)	288,252	$ 695,343,022
Children (<14)	2,274	$ 3,960,327
Non-Filers	952,612	$3,345,064,109

Identity Theft
In Tax Administration

- 1.5 Million Returns Questionable
  - $5.2 Billion in Refunds
- Delayed access to 3rd Party data
- All 3rd party data not utilized
- Pay & chase does not work
- $21 Billion over 5 years projected

American Opportunity Tax Credit Education Tax Credit
Report Number: 2011-41-083

- **2.1 Million Taxpayers submitted questionable returns claiming $2.6 Billion in erroneous credits**
  - **$2.1 Billion – No attendance documentation**
  - **$550 Million - Students not eligible**
  - **$88.4 Million –Student claimed on other tax return**
  - **$255,879 – Prisoners received credits**

$12.8 Billion Over Life of Tax Credit

TIGTA Data Mining and Predictive Analytics Project

November, 2011 - June, 2012

What a Fraud Detection Model Gets Us

- Use data mining to remove 90% of the hay to **focus on the 10% with the most needles**

- **Not relying on** tips, hotline, debarred lists, or random samples of data

- **Find highest ROI leads** that humans are unable to find with basic techniques

Selected Government Organizations

- **Centers for Medicare & Medicaid Services**
- **Justice Department**
- **Veterans' Affairs**
- **Coast Guard**
- **Postal Service Inspector General**
- **Education Inspector General**
- **Recovery Accountability & Transparency Board**
  - **The Recovery Operations Center**

Centers for Medicare & Medicaid Services (CMS)

- **CMS is largest single payer for health care in US**
- **1.5 billion claims submitted annually**
- **Additional data sources on the way (ACA)**
- **Billions of "non-claim" data points**
- **Expected to drive new innovation in Health Care**
- **Trusted to protect privacy**
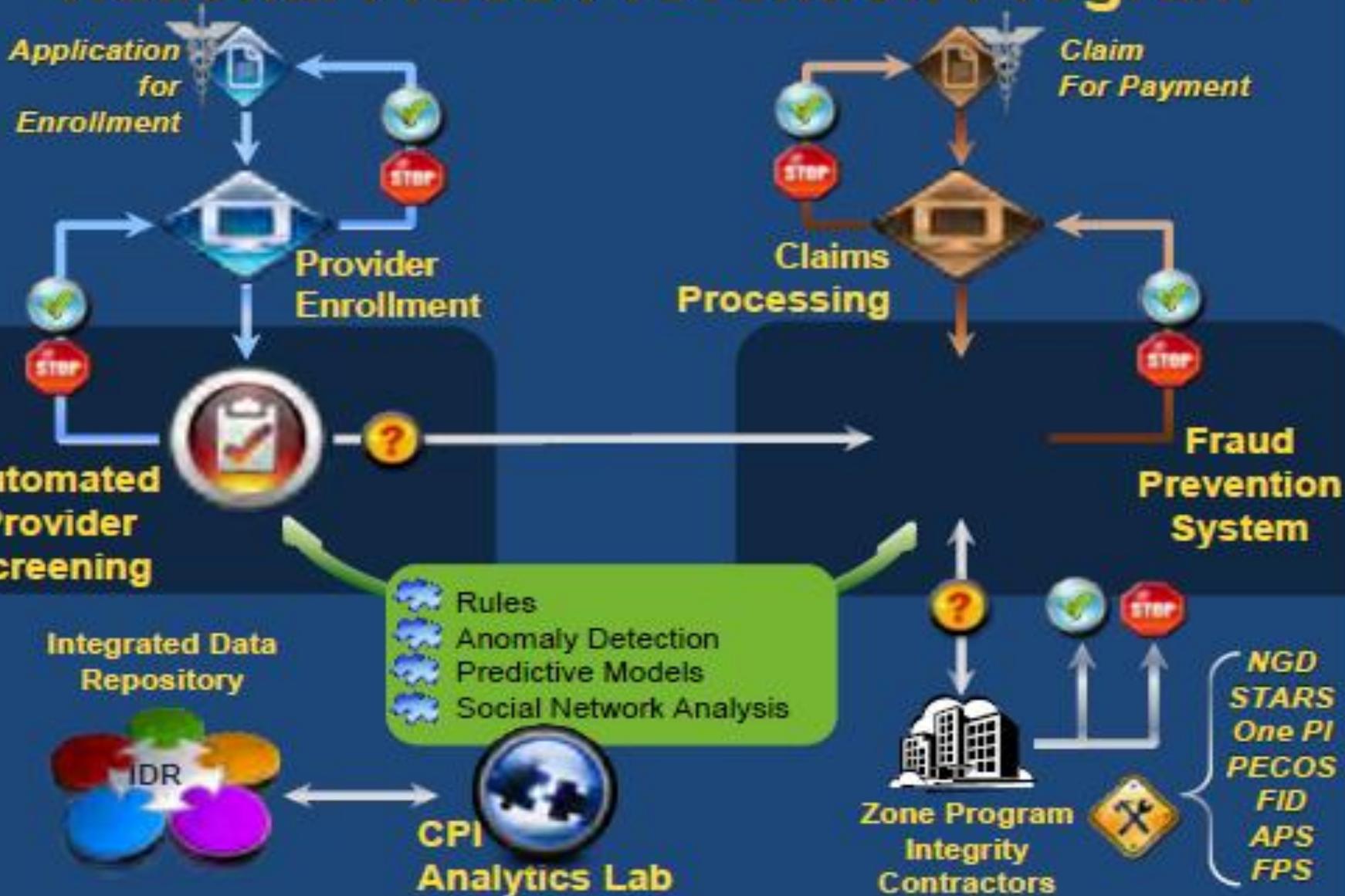
Centers for Medicare & Medicaid Services (CMS)

National Fraud Prevention Program

Application for Enrollment

Claim For Payment

Provider Enrollment

Claims Processing

Automated Provider Screening

Fraud Prevention System

Rules
Anomaly Detection
Predictive Models
Social Network Analysis

Integrated Data Repository

IDR

CPI Analytics Lab

Zone Program Integrity Contractors

NGD
STARS
One PI
PECOS
FID
APS
FPS

The New Approach to Combating Fraud, Waste and Abuse

Yesterday	Today & Future State
• Providers suspected of fraudulent activity are put on prepay review, sometimes indefinitely	• CMS will deny individual claims
	• CMS and its contractors will use prepay review as an investigative technique
• CMS initiates overpayment recovery	• CMS will revoke providers for improper practices
• Law enforcement determines if an arrest is appropriate	• CMS and Law Enforcement collaborate before, during and after case development
	• CMS will address the root cause of identified vulnerabilities

28

Postal Service Inspector General

- **Lease Building Risk Model**
  - **Select facilities with greatest impact for Audit**
  - **Identify risk of fraud**
  - **Predict future space requirements**
  - **Calculate market value of properties**

MAP   LIST

Hawaii

Alaska

Puerto Rico
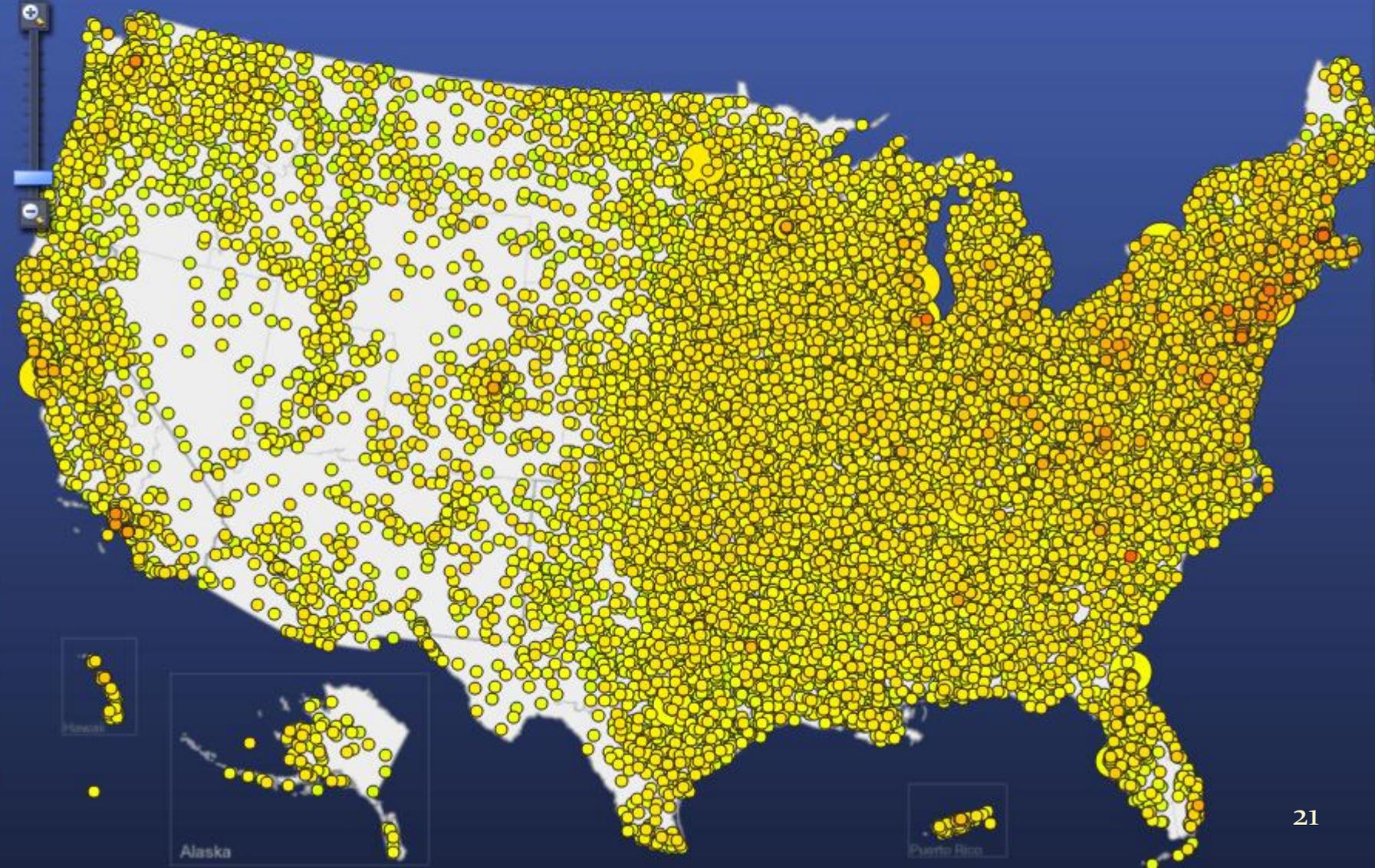
Metrics in Lease Renewal Model

- Metrics
  - Safety (Building Age)
  - Risk Ranking
  - Worth (Revenue, Lease Expenses, Labor Costs)
  - Capacity (Volume, Revenue, Hours Worked)
  - Lease Rate (Current rate, Suggested rate, Growth)

- Each scored from 0 – 10 points

Postal Service Inspector General

- Claimant Risk Analysis Model
  - Postal Service Workers' Compensation Claims
    - Totaled more than $1.3 Billion
    - Future Costs: $15.1 billion
- Model Description
  - Predictive Statistical Model
  - Identifies claimants with risk of fraud
  - Statistical model used 25 characteristics

Finding Healthcare Fraud

Healthcare Fraud: Map View

Use filters to help narrow down your search.

Highlights of Claimant Model

- Reactive to Proactive
- Return $$ to Postal Service
- Risk Scores – Work the "best" cases
- Shortens time to work cases
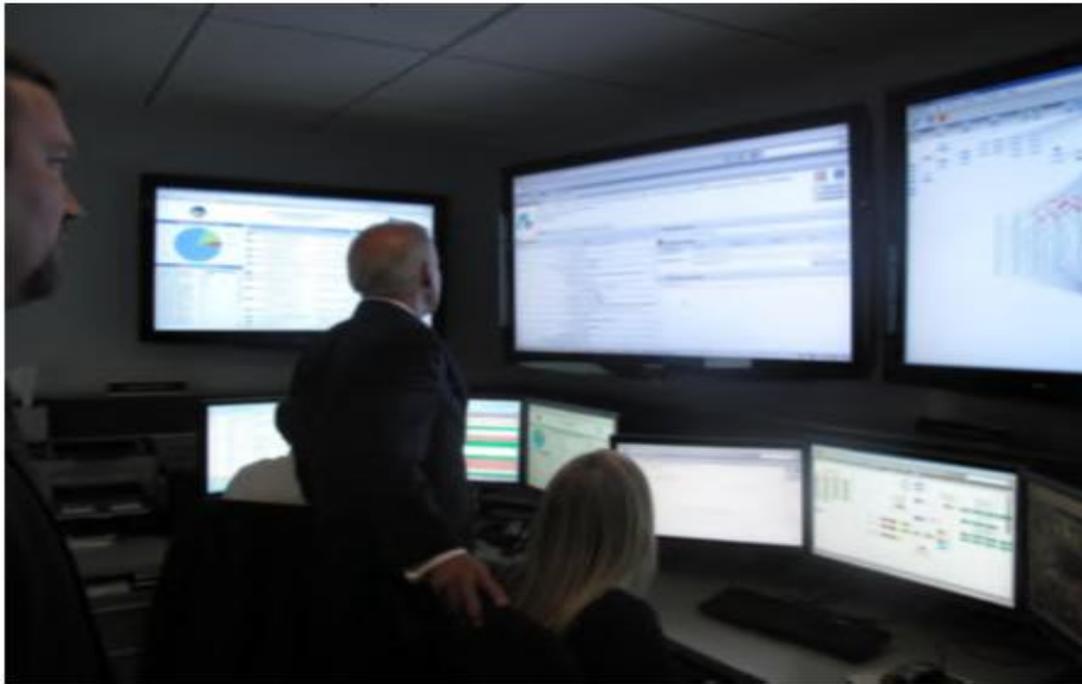- Discover relevant and timely leads early

RETURN ON INVESTMENT IS KEY

Recovery Accountability & Transparency Board

New Monitoring Levels

Recipients are put through a risk model using variables that are constantly being monitored for relevancy

RECOVERY.gov

ROC Data Mining

- Sifting through massive amounts of data "rapidly":
  - Find hidden patterns and information
  - Identify fraudulent behavior
  - Dissect fraud exposing operational factors of crimes

RECOVERY.gov

Data Challenges

- Large volumes of data
- Data systems not integrated
- Numerous legacy systems to plow through
- Unstructured data
- Noise, uncertainty, quality, complexity

RECOVERY.gov

Lessons Learned

- Obtain top executive sponsorship
- Identify goal(s) and start small
- Identify the data needed and location
- Identify data cleansing requirements
- Identify software, staffing, and consulting
- Construct the business model
- Interpret the results visually
- Document return on investment
- Share results with interested parties
- Plan to refine your models often

Reference Material

- Treasury Inspector General Tax Admin.
  - www.treasury.gov/tigta
- Journal of Advanced Analytics
  - 3rd Quarter, 2012, Fraud Prevention
- TechAmerica Foundation
  - Demystifying Big Data
- Association of Government Accountants
  - Leveraging Data Analytics in Federal Organizations (May, 2012)

Reference Material

- Government Business Council
  - Technology Tackles Waste, Fraud, and Abuse
- Rexer Analytics
  - 5th Annual Date Miner Survey
- Predictive Analytics World Government Conference
- Interagency Fraud and Risk Data Mining Group
  - www.va.gov/oig/ifrdmg

Questions

Contact Information:

Alan Duncan
202-622-5894
alan.duncan@tigta.treas.gov