Digging for Pay Dirt

HUD OIG's Journey From Data Mining to Predictive Analytics

Mid-America Intergovernmental Audit Forum

Overland Park, KS ~ December 2, 2016

Cliff Cole
Director, Integrated Data Analytics Division

"The Changing Face of Auditing"

Before – Reported ROI – Early 1990's

$800.00

Today – Reported ROI – 2016

$11,500,000,000

$E=MC^2$


Office of Inspector General | U.S. Department of Housing and Urban Development
OFFICE of
INSPECTOR GENERAL
UNITED STATES DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

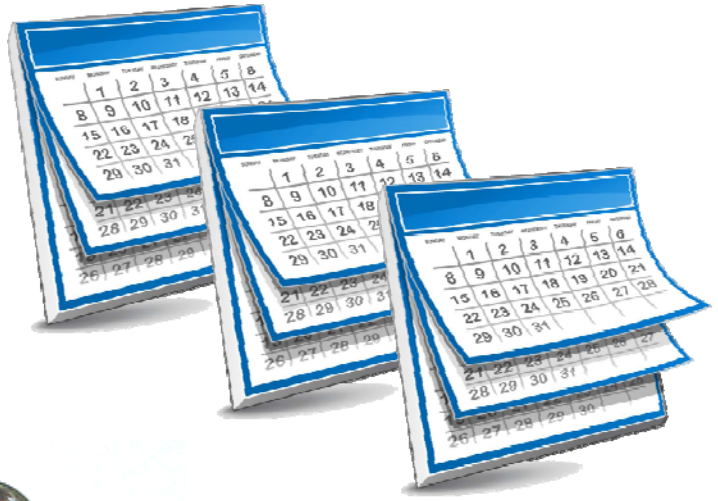
TCFOA=?


Office of Inspector General | U.S. Department of Housing and Urban Development
OFFICE of
INSPECTOR GENERAL
UNITED STATES DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

Lots of auditors …

+

Lots of time …



=

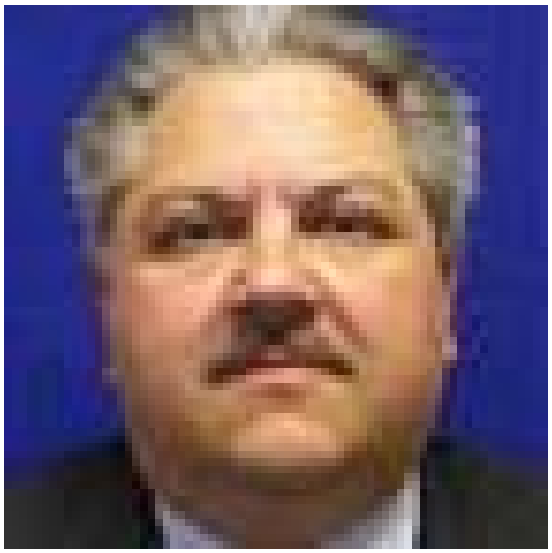
Fewer auditors …

+

Less time …

=

TCFOA=



Agenda

- Who we are …
- What we do …
- When did we start …
- Where we went …
- Why we do it …
- How we do it …
- Pay Dirt!
- Questions

Who are we? (The Team)

- iDAD - Integrated Data Analytics Division
  - ✓ Auditors
  - ✓ Statisticians
  - ✓ Computer Scientists
  - ✓ Data Analysts
  - ✓ Research Specialists
  - ✓ Program Managers

What do we do? (The Mission)

iDAD is responsible for the development and implementation of OIG's **Advanced Analytics Program** (AAP) which encompasses the collection and management of internal and external datasets; and implementation of data virtualization (integration of disparate systems and multiple databases)

What do we do? (The Mission)

The Goal

To create a data-driven and analytical environment for a proactive approach in carrying out the OIG mission by leveraging data for enhanced decision making and activity planning

What do we do? (The Mission)

Advanced Analytic Program (AAP)

The program is designed to:

- ✓ Facilitate thorough analysis of current and historical data
- ✓ Present an overview of the housing industry to apply insights and enhance OIG's strategic planning
- ✓ Create an accessible portal to housing program data needed for OIG day-to-day activities

What do we do? (The Mission)

iDAD will implement AAP by:

- ✓ Acquiring and managing direct access to HUD and internal databases
- ✓ Collecting and integrating structured and unstructured data to build OIG's knowledgebase
- ✓ Applying various analytic techniques (modeling, data mining, machine learning, etc.)

What do we do? (The Mission)

iDAD will implement AAP by:

- ✓ Establishing and maintaining a comprehensive database of all relationships
  - Individuals to Individuals
  - Individuals to Entities
  - Entities to Entities
  - Entities to Individuals

What do we do? (The Mission)

iDAD will implement AAP by:

- ✓ Identifying patterns in current and historical data to identify risk and opportunities
- ✓ Capturing relationships among many factors to conduct risk assessments or potential risk associated with a particular set of conditions

What do we do? (The Mission)

iDAD will implement AAP by:

✓ Designing, developing, implementing, and managing a:

- Comprehensive program risk assessment and monitoring tool to generate actionable leads for the OIG community
- Data-driven scoring mechanism for all identified data anomalies

What do we do? (The Mission)

iDAD will implement AAP by:

- ✓ Making recommendations for the acquisition of state-of-the-art technology to facilitate a dynamic and robust environment to foster effective and efficient data capturing, processing, storage, searching, analysis, and visualization

What do we do? (The Mission)

Success Factors

- Transforming the OIG to take advantage of an **Advanced Analytics Program**
- Creative use of both internal and external data sources to obtain a broad view of the housing industry that is useful to end users
- Designing effective data models that help identify housing issues and optimize OIG operations

What do we do? (The Mission)

Desired Outcomes

- Increase availability and accessibility of HUD program data to the OIG community
- Increase OIG's usage of data for strategic planning purposes
- Increase quantity and quality of actionable leads for OIG Auditors, Evaluators, and Investigators

When did we start … Where did we go?

1990's

Moon

HUD OIG Data Analytics Timeline

1. The Journey Begins

- Individual Office of Audit staff leveraging existing technology and HUD system access to accomplish the work
- Just a few data champions within the Office of Audit
- Data analyses use and results generally limited to individual audits and reviews
- Basic data mining of HUD and external data associated with the tenant based Section 8 program

Early 1990's

2. Informal Group

- National data mining group informally established within the Office of Audit
- Data champions given data analytic training and support responsibilities
- Duties are supplemental to existing work requirements
- Data focus centered on acquiring, converting, and providing FHA and rent subsidy data to support audit and investigative efforts
- ACL adopted as the standard data analyses tool

1996

3. Formal Group Established

- Data centric Computer Audit Specialist positions created
- Separate group formally established within Office of Audit to provide national data analyses support for on-demand services
- Data acquisition from additional HUD systems increases
- Became involved in nearly all national audits and external reviews by providing HUD data and advanced analyses expertise

2001

4. Data Matching

- In collaboration with the Office of Investigations, computer matching agreements established with FBI, NCMEC, USPS, and OPM
- On-demand data analyses support steadily increases to unprecedented levels
- Adoption of basic statistical sampling techniques
- Office of Audit data server created to house national HUD FHA and PIH and MF rent subsidy data sets for analyses and sharing with field-level data champions
- OIG data analyses journey chronicled in CIGIE's Journal of Public Inquiry

2006

5. Statistical Expertise

- Developed a higher level statistical analyses program, creating innovative designs and ground breaking methodologies
- Data analyses starts to become entrenched within the audit workload and strategic planning process
- Data analyses used to spearhead some of the largest, most successful joint multi-agency initiatives within audit and investigative operations

2010

6. Predictive Analytics

- Established a separate Predictive Analytics group
- Staffed by technical experts in computer science and data modeling and analyses
- Invested in creating and deploying an advanced system hard/software infrastructure
- Near-live connectivity established with key HUD PIH systems
- Invested in data visualization, and dash boarding technology

2013

7. Organizational Changes

- Merged the on-demand support group with the predictive analytic division, bringing all agency data analyses resources under one consolidated division
- Better to serve entire OIG field and HQs operations and to become a major contributor to agency strategic workload planning
- Leveraged skill sets to develop FHA predictive analytic models to assess mortgage industry trends

2015

★★ OFFICE of ★★
INSPECTOR GENERAL
UNITED STATES DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

When did we start … Where did we go?

1. The Journey Begins

- Individual Office of Audit staff leveraging existing technology and HUD system access to accomplish the work
- Just a few data champions within the Office of Audit
- Data analyses use and results generally limited to individual audits and reviews
- Basic data mining of HUD and external data associated with the tenant based Section 8 program

Early 1990's

When did we start … Where did we go?

2. Informal Group

1996

- National data mining group informally established within the Office of Audit
- Data champions given data analytic training and support responsibilities
- Duties are supplemental to existing work requirements
- Data focus centered on acquiring, converting, and providing FHA and rent subsidy data to support audit and investigative efforts
- ACL adopted as the standard data analyses tool

When did we start … Where did we go?

3. Formal Group Established

- Data centric Computer Audit Specialist positions created
- Separate group formally established within Office of Audit to provide national data analyses support for on-demand services
- Data acquisition from additional HUD systems increases
- Became involved in nearly all national audits and external reviews by providing HUD data and advanced analyses expertise

2001

When did we start … Where did we go?

4. Data Matching

2006

- In collaboration with the Office of Investigations, computer matching agreements established with FBI, NCMEC, USPS, and OPM
- On-demand data analyses support steadily increases to unprecedented levels
- Adoption of basic statistical sampling techniques
- Office of Audit data server created to house national HUD FHA and PIH and MF rent subsidy data sets for analyses and sharing with field-level data champions
- OIG data analyses journey chronicled in CIGIE's Journal of Public Inquiry

When did we start … Where did we go?

5. Statistical Expertise

- Developed a higher level statistical analyses program, creating innovative designs and ground breaking methodologies
- Data analyses starts to become entrenched within the audit workload and strategic planning process
- Data analyses used to spearhead some of the largest, most successful joint multi-agency initiatives within audit and investigative operations

2010

When did we start … Where did we go?

6. Predictive Analytics

2013

- Established a separate Predictive Analytics group
- Staffed by technical experts in computer science and data modeling and analyses
- Invested in creating and deploying an advanced system hard/software infrastructure
- Near-live connectivity established with key HUD PIH systems
- Invested in data visualization, and dash boarding technology

When did we start … Where did we go?

7. Organizational Change

- Merged the on-demand support group with the predictive analytic division, bringing all agency data analyses resources under one consolidated division
- Better to serve entire OIG field and HQs operations and to become a major contributor to agency strategic workload planning
- Leveraged skill sets to develop FHA predictive analytic models to assess mortgage industry trends

2015

HUD OIG Data Analytics Timeline

2016 Forward

8. The Future

- Invest in geospatial data analyses resources and develop robust in-house geographic information systems and expertise
- Continue to connect the OIG data analytic infrastructure to an increasing number of HUD data systems
- Invest in acquiring uniquely qualified staff with highly specialized skill sets
- Incorporate housing-related publically available national data sets within the data infrastructure
- Continue to provide data based HUD program risk assessment systems and make significant contributions to the OIG's audit , investigative, and evaluation strategic workload planning processes
- And most importantly, continue to develop a proactive predictive analytics capability to measure future trends affecting HUD operations and the housing industry

Where did we go? Where are going?

Moon

STARS!

Why do we do it?

FY	Number of Reports Using CAATs and Data Analytics					
	External	Reported ROI	External ROI/Report Ratio	Internal	Reported ROI	Internal ROI/Report Ratio
2015	29	$634,969,383	$21,895,496	7	$110,366,806	$15,766,687
2016	30	$130,492,196	$4,349,740	5	$3,246,697,766	$649,339,553
TOTAL	**59**	**$765,461,579**	**$12,973,925**	**12**	**$3,357,064,572**	**$279,755,381**

FY	Number of Reports Not Using CAATs Data Analytics					
	External	Reported ROI	External ROI/Report Ratio	Internal	Reported ROI	Internal ROI/Report Ratio
2015	35	$45,720,892	$1,[illegible]311	4	$0	$0
2016	11	$24,297,[illegible]	$2,208,841	6	$255,972	$42,662
TOTAL	**46**	**$70,018,146**	**$1,522,134**	**10**	**$255,972**	**$25,597**

Why do we do it?

How do we do it?

- Obtain Executive Level Buy-in and Support
- Promote/Hire the Right People
- Research, test, and acquire your “tools”
- Invest in the Technology
- Get Some Wins
- Do What You’ve Never Done and What No One Else is Doing
- Break Your Paradigms
- Build On The Successes

Pay Dirt! - iDAD FY16 Workload – Summary Highlights

- Continued to enhance the OIG's effectiveness in carrying out its mission by providing timely and efficient response to over **500** region, headquarters, and outside agency requests for data, analyses, statistical sampling, and other services.
- Used advanced statistical methodologies to quantify over **$13 billion** in estimated monetary impact associated with work performed for the annual audit of HUD's financial statements and other OA initiatives.
- Continued timely data analytic support to the joint OIG/HUD/DOJ BLI task force, contributing to the **$1.3 billion** in civil remedy settlements negotiated so far in FY 2016.

Pay Dirt! - iDAD FY16 Workload – Summary Highlights

- Developed HUD program assessment and risk-based targeting systems designed to identify high risk multifamily non-health care project operations.
- Developed a data mining tool to search SF FHA claims and default history data sources and target loans that have significantly exceeded title conveyance timelines. The algorithm identified over 250,000 delayed conveyance loans since October 2010 and about $4.1 billion in ineligible claim payments.

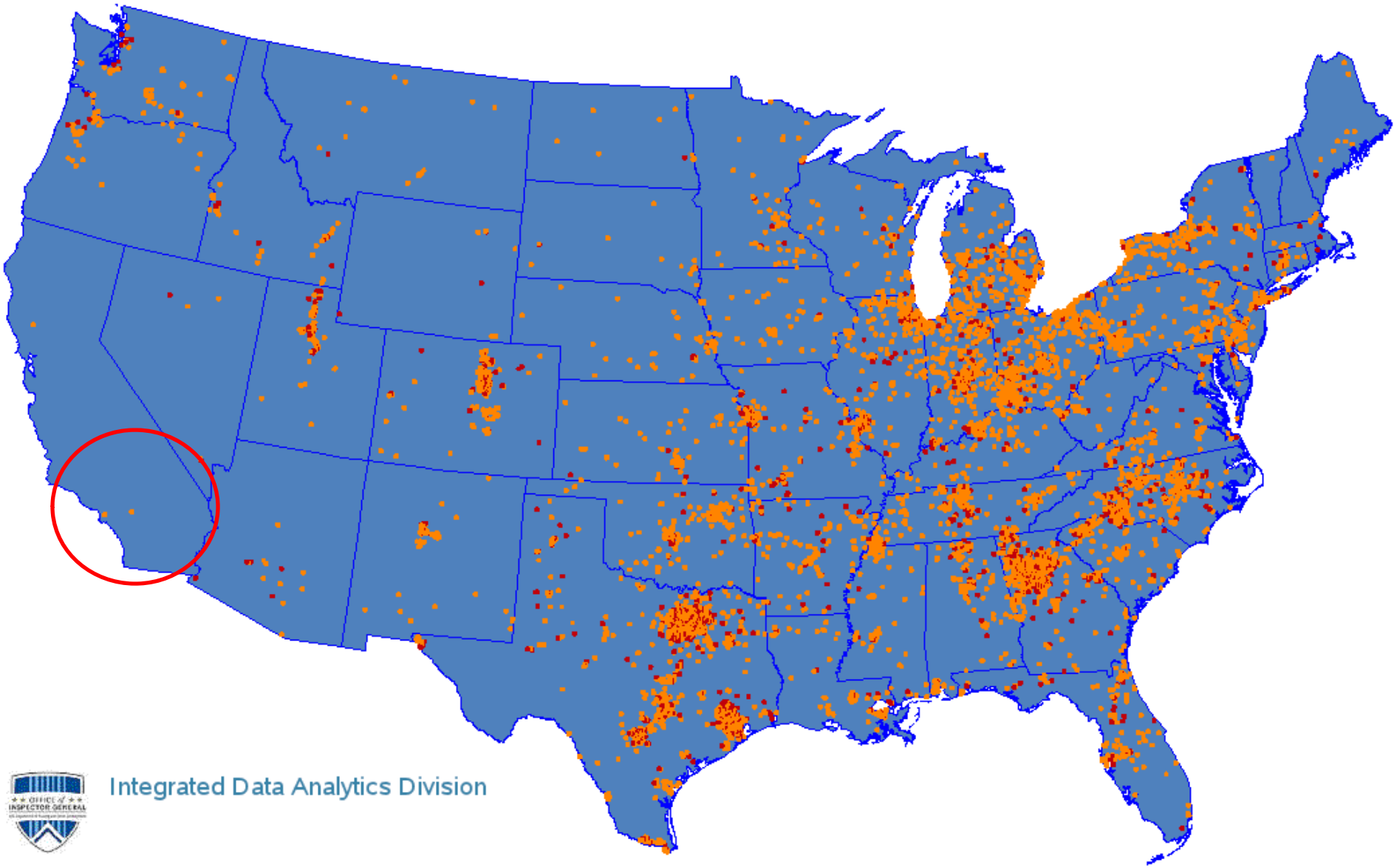
Pay Dirt! - iDAD FY16 Workload – Summary Highlights

- Developed, tested, prototyped, and fielded a comprehensive FHA mortgage claims predictive model:
  - ✓ Successfully applied it to 13 banking sectors, several at-risk watch-list banks, and included localized built-in analyses algorithms to highlight different regions of the country according to price volatility in the housing market.
  - ✓ Developed a website template/visualization dashboard with several drill-down layers showing historical rates of loan failures to display and showcase the results.

RETURN

Historical Claim Patterns by Region: 2006_Q1

Adjusted to represent an equal number of active FHA loans in each quarter.



Integrated Data Analytics Division

Pay Dirt! - iDAD FY16 Workload – Summary Highlights

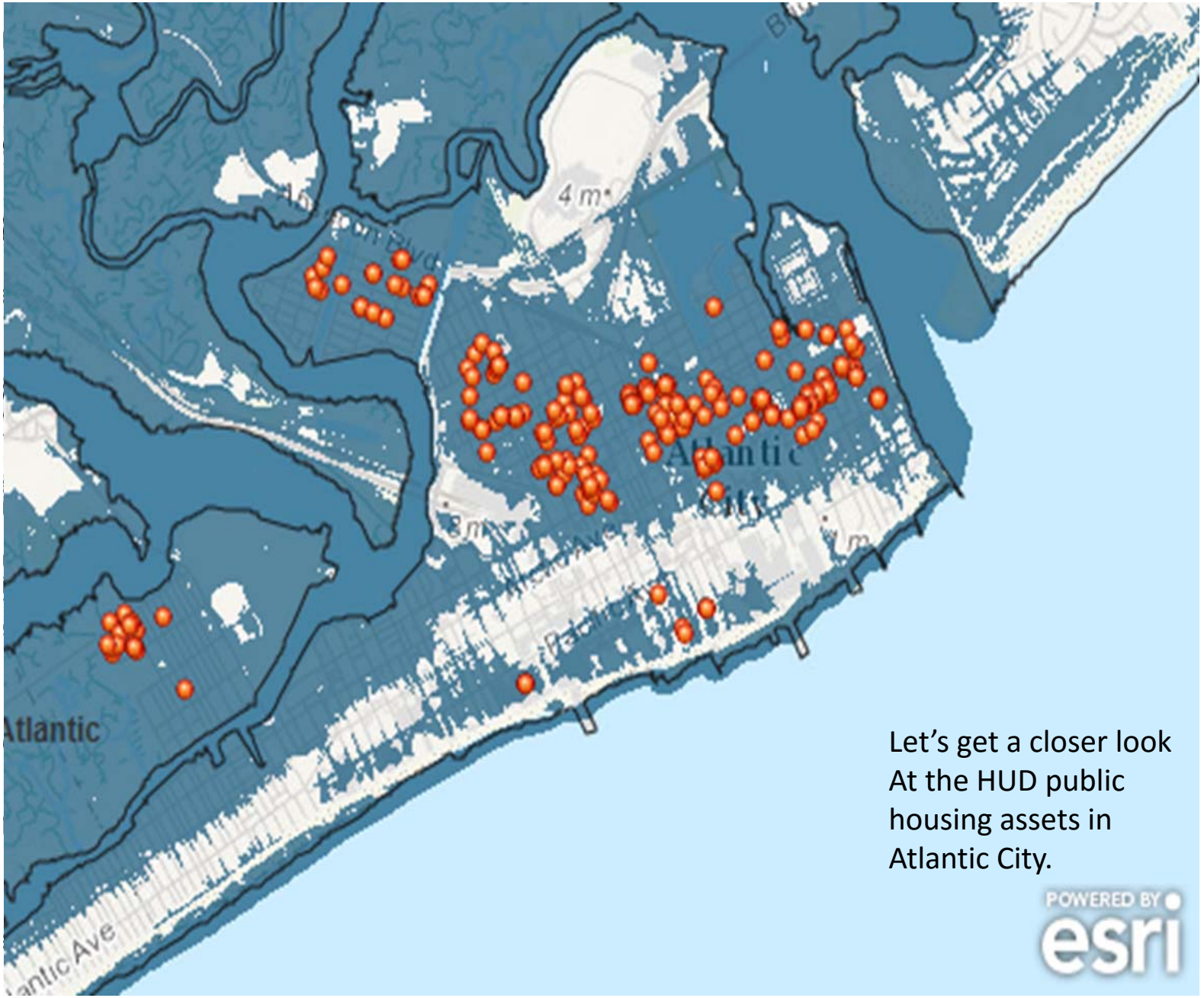
- Developed and fielded multi-family project and single family loan servicing risk assessment and targeting systems designed to rank potential audit, investigative, and evaluation workload and to facilitate strategic planning.
- Conducted focused, in-depth outreach to Region 4 Office of Investigations to better understand specific case workload, the factors contributing to successful prosecutions, and established unprecedented collaborative working relationships to foster a deeper understanding of field level data and analyses support needs.

iDAD FY16 Workload – Summary Highlights

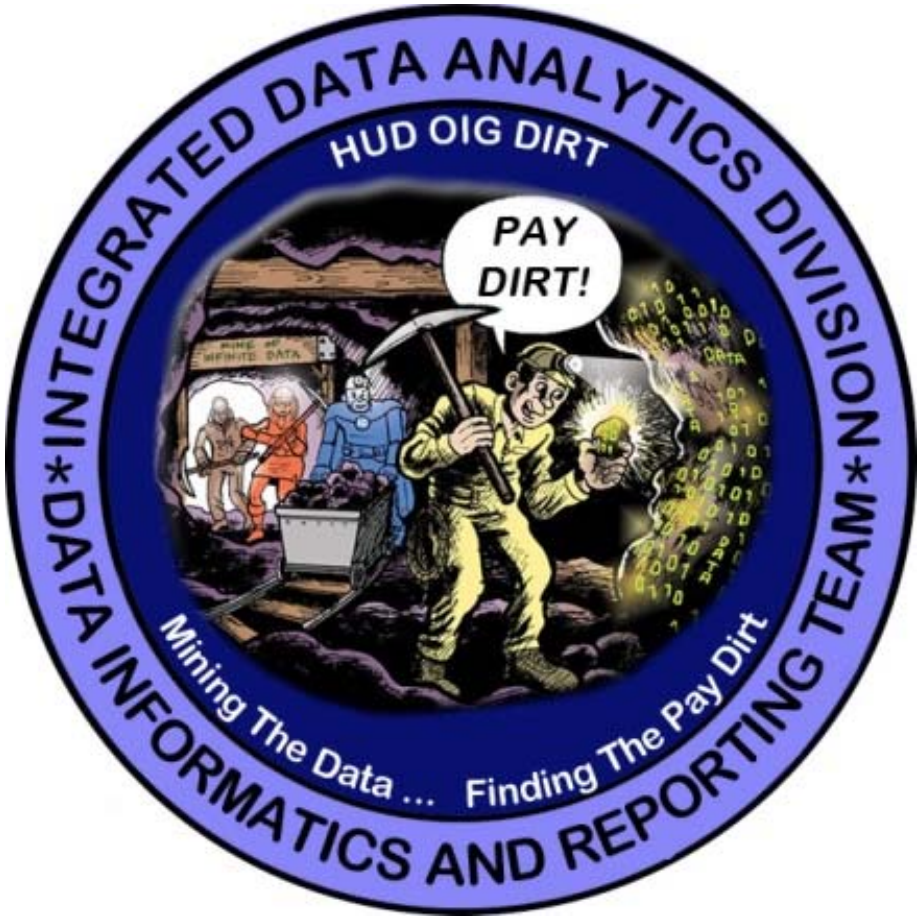
- Enhanced the predictive analytics infrastructure by ingesting HUD and third party data sources related to multi-family and PIH income verification and benefits eligibility determinations and oversight.
- Developed a HECM fraud targeting interactive geospatial dashboard that allows interested agents and other users to target reverse mortgages across the nation that appear to have significantly over appraised market values when compared to it's neighbors within local proximity.

Pay Dirt! - Status of HUD OIG Geo-Spatial Journey

- Mid-2016 – OIG acquires ESRI ARC GIS license
  - ✓ iDAD appoints a geo-spatial guru – may later hire a degreed analyst/specialist
  - ✓ Worked with IT to setup a server-based infrastructure
  - ✓ Developing map-based visuals and dashboards


New Jersey Shore
Atlantic City
- Storm Surge
- PHA Asset
Let's get a closer look At the HUD public housing assets in Atlantic City.
POWERED BY esri

Pay Dirt!


INTEGRATED DATA ANALYTICS DIVISION
HUD OIG DIRT
PAY DIRT!
Mining The Data ... Finding The Pay Dirt
DATA INFORMATICS AND REPORTING TEAM

Office of Evaluation - Integrated Data Analytics Division - DIRT

DIRTY Fact Sheet – 2016-001-R7-MF-011516

Data Informatics and Reporting Team Yield Dispatch

Prepared For Ron Hosking, Region 7 Inspector General for Audit
Date: January 15, 2016 **Page 1 of 2**

PROGRAM AREA: Multi Family (MF)

PURPOSE: To present the Office of Evaluation iDAD DIRT MF non-health care project risk assessment system results for Region 7.

BACKGROUND: Various HUD system data points designed to measure MF project/property post-endorsement performance risk were linked, scored, ranked, and organized by OIG region in an attempt to identify the currently active, poorest performing non-health care related projects (as determined by HUD implemented quality assurance measures) for potential audit targeting and strategic planning purposes.

SCOPE: Data refreshed as of October 19, 2015 from the MF Data Mart Property *active_property* and *active financing* tables (34,368 and 27,597 records, respectively).

QUICK FACTS: The risk factors used by the system and the points assigned are described as follows:

HUD OIG MF Non-Nursing Home Assessment - Description of Risk Factors and Assigned Point Values Ranked by OIG Risk Score and Number of Assisted Units (Higher Points/Units = Higher Risk)

Risk Factor	Description	Points Assignment
1. Troubled	Project flagged in OPIS as in a "TROUBLED" or "POTENTIALLY TROUBLED" status	T = 15; P = 10
2. OPIS Risk Category	OPIS integrated risk score computed using annual financial statement data submitted through the Multifamily Financial Assessment Subsystem (FASS-MF), physical inspection data collected through the Physical Assessment Subsystem (PASS), and selected contract, loan and profile data from the REMS. Calculated using the financial, physical, loan payment status history, management review, and other data for an active project or contract to assist in identifying at-risk properties.	Low = 5; Med = 10; High = 15
3. Default	Project flagged in OPIS as currently in default	Y = 15
4. DEC	Department Enforcement Center referral status	15 = Currently active DEC case; 10 = Not Current DEC case, but previously referred
5. Financial	FASS last performance rating color code	RED = 15; YELLOW = 10
6. Watch List	Project has been flagged as being on the OPPIS watch list	15 = put on list within the last year; 10 = was put on the list more than a year ago
7. Physical Inspection	Last REAC PASS inspection score	15 = Inspected within the last 2 years, scored less than 60, and had exigent deficiencies; 10 = Inspected within the last 2 years and scored less than 60
8. OIG Risk Score	Accumulation of the above risk scores	105 Maximum Possible Points

Office of Evaluation - iDAD DIRTY Fact Sheet – 2016-001-R7-MF-011516

Data Informatics and Reporting Team Yield Dispatch

Prepared For: Ron Hosking, Region 7 Inspector General for Audit
Date: January 15, 2016 **Page 2 of 2**

RESULTS (PAY DIRT): The top 25 highest risk MF projects scored by the system are:

Region Rank	OIG Risk Score	Property ID	Property Name	Address	City	State	Zip Code
1	80	800007192	CROSSLINES RETIREMENT CENTER I	3030 POWELL AVE	KANSAS CITY	KS	66106
2	70	800011758	New Horizons	1715 E Linwood Blvd	Kansas City	MO	64109
3	70	800022630	MEADOWS - THE	1415 S 6TH Avenue	SUNNYSIDE	WA	98944
4	70	800220089	Soulard Market Apartments	1535 S. 8th Street	St. Louis	MO	63104
5	70	800235087	University Village Apartments	1270 South Sandhill Road	Orem	UT	84058
6	65	800210409	HILLSIDE II APARTMENTS	160-198 KINGSTON	SAINT LOUIS	MO	63125
7	65	800210499	Hillside 1 Apartments	200 -232 Kingston Drive	SAINT LOUIS	MO	63125
8	65	800044820	HANIGAN TERRACE	1421 W. 35th Avenue	DENVER	CO	80211
9	60	800020027	TAMARAC APTS	1818 Arizona AVE SW	HURON	SD	57350
10	60	800011506	CHEVY CHASE APARTMENTS	1224 W BRECKENRIDGE	MEXICO	MO	65265
11	60	800022390	CAM-BEY SENIOR APARTMENTS	50 N Main St	Coupeville	WA	98239
12	60	800023746	CHIEF WASHAKIE	82 E HAYDEN	EVANSTON	WY	82930
13	60	800073064	WHITCOMB APARTMENTS	1507 FILLMORE ST	STERLING	CO	80751
14	60	800022569	JACKSON APTS	670 S JACKSON ST	SEATTLE	WA	98104
15	60	800236615	Four J Apartments	24530 Southside Road	Waynesville	MO	65583
16	60	800073050	FOUNTAIN RIDGE APARTMENTS	370 Comanche Village Dr	Fountain	CO	80817
17	60	800011650	JAMES VALLEY GROUP HOMES	651 W NICHOLS ST	SPRINGFIELD	MO	65802
18	55	800011863	ROOSEVELT TOWNE APARTMENTS	711 N Euclid Ave	Saint Louis	MO	63108
19	55	800003106	MORNINGSIDE HEIGHTS	1516 E 6th St	La Junta	CO	81050
20	55	800017878	FANNO CREEK	4205 SW Beaverton Hillsdale Hwy	Portland	OR	97221
21	55	800003244	WISE AND HARRIS ARMS	605 26TH ST	DENVER	CO	80205
22	55	800011971	Wayne Apartments	3112-3114 BENTON BLVD	KANSAS CITY	MO	64128
23	55	800233032	Riverpointe Apartments	2550 Duportail Street	Richland	WA	99352
24	55	800073552	CEDARWOOD APARTMENTS	1875 SOUTH 75TH STREET	OMAHA	NE	68124
25	55	800220114	Jadwin Stevens Apartments	1851 Jadwin Avenue	Richland	WA	99352

Please refer to the accompanying Excel file, *"Region 7 – MF Property Risk Assessment – Non-Health Care Projects.xlsx"* for the data point details, individual risk factor scores - for the region and each state within the region.

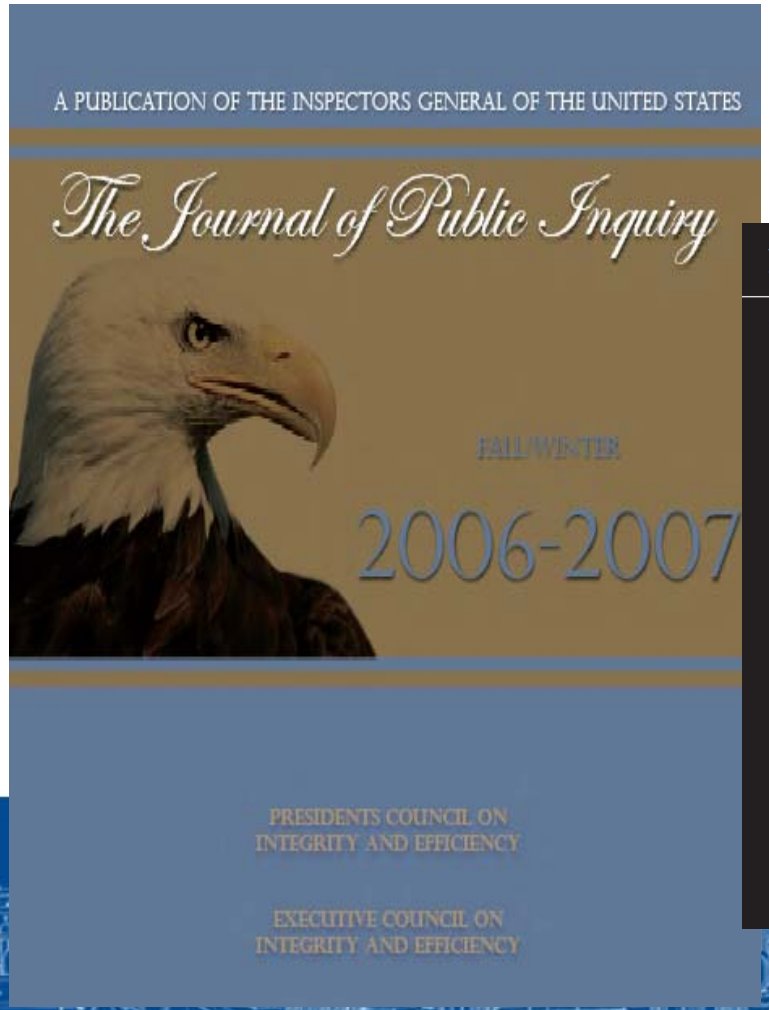
DIRT Lead: Cliff Cole, (202) 441-1605

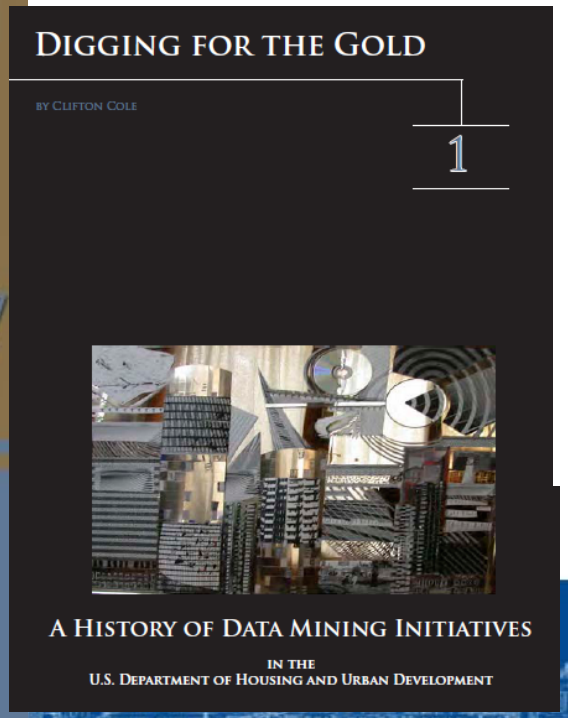
Pay Dirt!

Pay Dirt!

Journal of Public Inquiry Article Highlights the HUD OIG Data Mining Journey at

http://www.ignet.gov/randp/fw06jpi.pdf





INTRODUCTION

The inspector general community has long been in search of the audit and investigative Holy Grail–to identify, develop, and report results faster with fewer man hours and smaller budgets. While those agencies that have embraced and invested in the technological advances in computer hardware and software and related training have achieved significant operational efficiencies in attaining agency goals, it is the agencies that have incorporated the use of data mining and its ancillary techniques that have come closest to maximizing automation's productivity-enhancing potential. This article will focus on the U.S. Department of Housing and Urban Development, Office of Inspector General's (HUD OIG) data mining journey from its sporadic use by a selected, tech-savvy few to an agency wide implementation designed to produce more significant results with fewer resources.

DATA MINING DEFINED AND ITS USE

Because the uses of data mining techniques have proliferated over recent years among the inspector general community and other professional organizations, defining the term and its associated processes remains somewhat of an enigma. At HUD OIG, data mining and the use of computerized advanced audit techniques have become institutionalized in the day-to-day conduct of audit and investigative field operations. Accordingly, the definition that most aptly describes data mining at HUD OIG is:

The process of analyzing data from a variety of perspectives and summarizing the data into useful information that can be used to maximize audit and investigative resources.

Specific examples include analyzing a variety of HUD and other federal, state, and corporate databases to

- Perform data matching routines;
- Identify specific transactional universes;
- Isolate leads and potentially fraudulent transactions;
- Use as the foundation for developing statistical sampling plans; and
- Make sample selections and identify trends.

Data mining and associated computer-based analytical techniques have become an integral component of HUD OIG's effort to maximize operational efficiency in accordance with the mandate dictated by the President's Council on Integrity and Efficiency.

IN THE BEGINNING

Before the 1990s, computer-based auditing at HUD OIG was sparse, generally limited to a select few who were technically oriented and driven to do the best job possible with available hardware and software resources. This scenario likely described most federal audit and investigative organizations during that period. As the availability and use of personal computing, the Internet, and computer networks exploded through the '90s, so too did the role played by computers in the conduct of inspector general audit and investigative activities.

DATA ANALYSIS NETWORK ESTABLISHED

In 1996, HUD OIG data mining activities were formally organized with the creation of a core group of seasoned, tech-savvy auditors from each of the 10 HUD OIG regions. Group members represented their respective regions as the champion, resident expert, and support asset for advanced data mining and analysis. In addition to performing their day-to-day audit responsibilities as managers and auditors-in-charge, group members provided a variety of computer-assisted analytical services in support of their regional audit mission. Through regularly scheduled meetings and conferences and routine contact and networking with each other, this core group dramatically increased the use of advanced auditing techniques by sharing proven analytical processes and, if necessary, adapting them to work on current and future audits and investigations throughout the agency.

HIGHLY SPECIALIZED SUPPORT POSITIONS CREATED

With vigorous top management support, use of data mining and automated analytical techniques flourished. To further increase the use of these techniques and advance the level of assistance and support, HUD OIG created new computer audit specialist (CAS) positions in 2001 and established a headquarters-level CAS support branch. The positions were filled with proven

2 Journal of Public Inquiry

Questions - Comments

Contact Information

Digging for Pay Dirt

HUD OIG's Journey From Data Mining to Predictive Analytics

Mid-America Intergovernmental Audit Forum

Overland Park, KS ~ December 2, 2016

Cliff Cole
Director, Integrated Data Analytics Division