

Data Analytics' Lessons Learned/Best Practices



OIG Data Analytic System

“An independent source of mission critical knowledge.... formulated through timely and reliable analysis, resulting in added value and insight.”

Midwestern Intergovernmental Audit Forum

September 8th 2017



U.S. Department of Education, Office of Inspector General

ODAS Presentation AGENDA

- **CAATs team skills, roles and responsibilities.**
- **Strategic approach to establishing a data analytic presence within the OIG.**
- **ODAS Data Warehouse Foundation Systems – Immediate ROI**
- **Data Mining in the 21st Century and Beyond – What the heck is Data Mining?**
- **Student Fraud Ring Project.**
- **Purchase Card Data Visualization demonstration.**

Provided Handouts

- **Memorandum of Understanding (MOU) Exhibit** - This agreement governs the relationship between the two organizations regarding the interconnecting information technology systems, including designated managerial and technical staff. OIG and FSA will abide by the policies and procedures set forth in this agreement and the information technology policies and procedures of the Department of Education.
- **How to Select a Successful Data Analytic Project** - In order to select and stand up a successful data analytical project, what are the fundamental procedures and/or guidelines to follow?
- **Student Fraud Ring ~Data Analytical Project Summary Guide Points** - I have selected this project for this discussion because I believe the lessons learned and best practices can be readily understood and mimicked in other data analytic projects.
- **Top Strategies for Productive Data Analytics Shop** -12 key points reflecting lessons learned/best practice that you may find valuable.

CAATs Support

Process Audit / Investigation
Adhoc Data Requests.
Mainframe Access
Oracle Access
Medium/Complex
programming.
On-site Audit Support
Technical Consultant
Statistical Sampling Support
Close team support /onsite
participation
ODAS System Project Lead

ODAS Support

Develop/Maintain
Foundation Systems
Risk Models /
Analytical Systems
Support Production System
Network
Security
Backup/Recovery
Hardware/Software
FISMA Compliance/
Continuous Monitoring
Big Data Analytics

DATA → **INFORMATION** → **KNOWLEDGE** → **ACTIONS** → **OUTCOMES**

Predictive Analytics

Ad-Hoc Report Writing

2020

Uses data mining and statistical techniques to examine patterns in data to infer causal relationships. They go beyond what has occurred in the past to seek why it occurred and quantify probability of future occurrences.

Reduces the time and labor associated with exploratory data analysis using prepackaged data mining algorithms in conjunction with intelligent software agents to find and present interesting patterns, correlations and anomalies.

Learning What You Did Not Know

Risk Models - Trends

Data Mining - Patterns

2018

Early warning indicators identifying emerging risks allowing auditors to fully understand critical control points, rules and exceptions; perform control and risk assessments in near real-time; and integrate analysis results into all aspects of the audit process.

Ability to uncover hidden patterns, relationships and anomalies by sifting through large volumes of data to discover that which is not apparent and obvious for further investigation, begin the process of learning what was previously unknown.

Learning Where You should Investigate

Data Visualization Scorecards

Dimensional Drill Down

2017

Representing abstract data as images allowing Decision Making 'At a Glance'. Displays what is happening now with limited view of where and why. With trend lines, they may offer hints about the future.

Opportunity for real insight and is focused on the what, is capable of where with a strong spatial dimension, but is still anchored in the past. This is a analytical tool that can help to determine next questions and what you need to know next.

Seeing What You Expect To See

School Summary

Student Look-up

2016

Allows authorized users to search for schools using OPEID number, school name, or DUNS number and view a comprehensive school profile.

Provides general information about the selected school to include Title IV funding, Dun & Bradstreet, and a list of all students receiving Title IV funding for last 3 years.

Allows authorized searches of financial aid and grant information for students using SSN, last/first name, school name, DOB and view a comprehensive student profile.

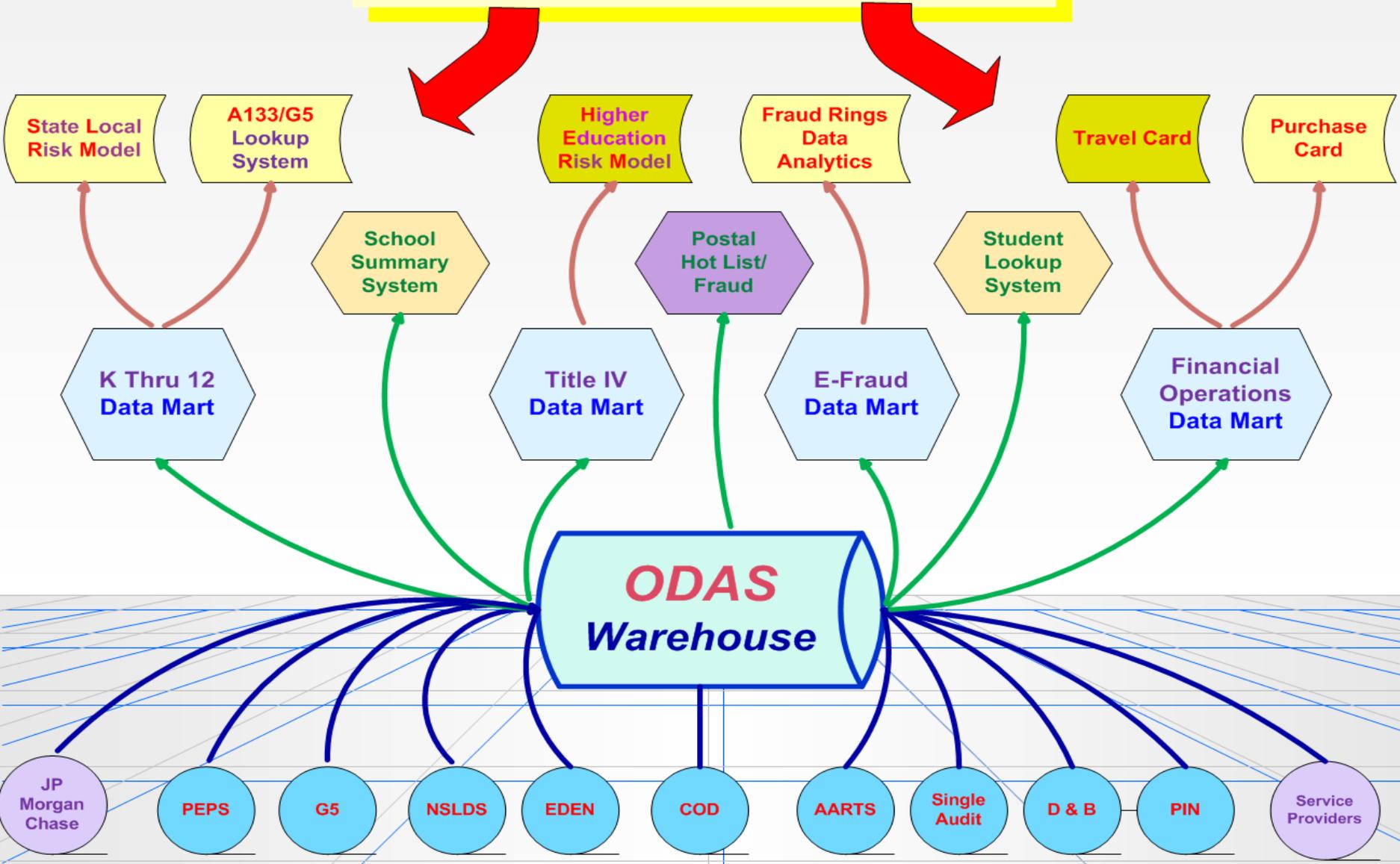
Provides data about individual students for such areas as FAFSA*; ISIR**; Loan data; Grant data; and miscellaneous data, such as e-mail and phone number.

FOUNDATION ~ SYSTEMS

**ODAS
Warehouse**

*FAFSA - Free Application for Federal Student Aid
**ISIR - Institutional Student Information Record

U.S. Dept ED - OIG Auditors - Investigators



In the beginning....

GAO Data Mining Reports

DODIG Purchase Card
Data Mining Project

Government Wide Report
June 2003

False Positives

The erroneous identification of a fraudulent event or dangerous condition that turns out to be harmless.

False positives **often occur** in Rule Based detection systems.

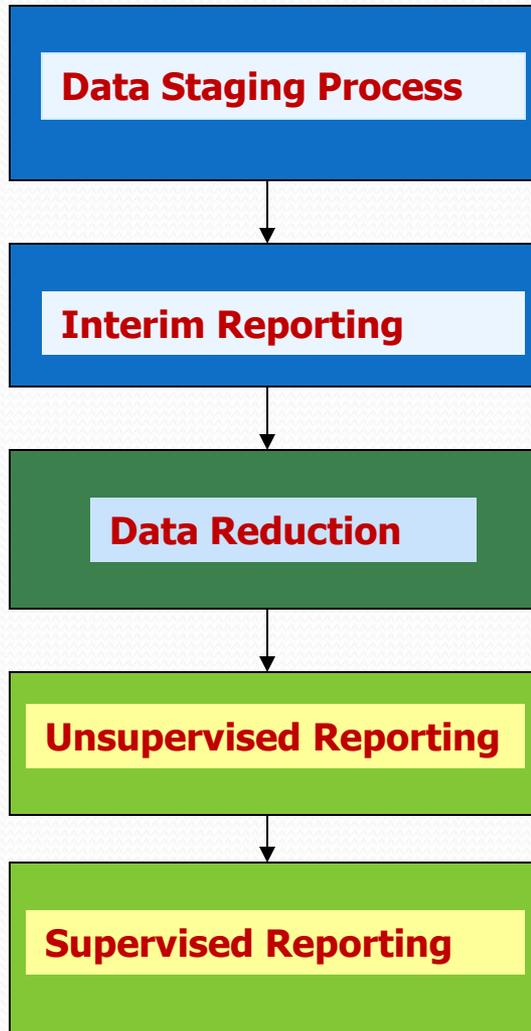




Data Mining Defined

“Data Mining is the ability to *predict* with a high degree of probability, anomalies where fraudulent or inaccurate activity is likely using statistical and mathematical techniques.”

Data Analytics Development Process



Cleaning, standardizing data elements and tagging fraudulent indicators.

Develop Trend and Rule Based Exception Reporting using fraud indicators.

Aggregate or amalgamate the information contained in large datasets into manageable (smaller) information nuggets to **filter out false positives**.

Run transactions thru Statistical and AI modules to identify possible fraudulent outliers.

Using Knowledge Database, run transactions thru Statistical and AI modules to *predict* fraudulent anomalies.

**E-Fraud
Data Analytical
Model**

~

**Detecting Fraud Rings
Within Title IV
Student Loan Arena**

~

Ed

Student Fraud Ring

Proof Of Concept Project

Objective - develop a Proof of Concept module that successfully detects with a fairly high degree of probability fraudulent or suspicious activities relating to student fraud rings. The activities would include identifying suspicious matching data of the students enrolling at a school. Then, risk measure this activity by identifying potential loss or Federal Student Aid (FSA) dollars at risk.

Background - Technology can be viewed as a two edged sword. The increased efficiencies and web-based systems allow for prompt user services that promote customer satisfaction but also open up the process for abusive and fraudulent activities.

- Management controls of these new automated processes are being challenged as never before.
- Identifying the increased potential for theft and abuse has also been a challenging pursuit.
- Student fraud rings have become a rapidly growing crime activity that now have targeted the U.S. Department of Education (ED) FSA programs.
- ED processed over 19 million applications for student financial aid and disbursed over \$90 Billion in FSA funds in SY2010.

Fraud Indicators

Home Address

Phone Number

**Multiple IP
Addresses**

E-Mail

**PIN
Number**

**PIN
Question/Answer**

False Positive Filter and Scoring

	Risk Indicator	Weight	Weight Factor/Value	Comment
1	Residential Address	3	0.13043	Count > 3, or
2	Phone Number	3	0.13043	Count > 3, or
3	Email Address	3	0.13043	Count > 3
4	IP 1	2	0.08696	Count <= 100
	IP 1	0	0	Count > 100
5	IP 2	1	0.04348	Count <= 100
	IP 2	0	0	Count > 100
6	IP 3	1	0.04348	Count <= 100
	IP 3	0	0	Count > 100
7	PIN	1	0.04348	
8	PIN Q&A - 1,5,6,7,8	1	0.04348	
	PIN Q &A - 2,3,4,9	0	0	
9	CMRA	3	.13043	

Student Fraud Ring

Actual Immediate Filtered Results

15 Post Secondary Schools selected as part of the Proof of Concept Project.

School Code	Total Student Population	Filtered Student Population
999999	21,035	89
999999	4,171	26
999999	12,020	183
999999	65,457	129
999999	19,193	8
999999	5,566	4
999999	5,332	5
999999	3,572	28
999999	44,130	366
999999	11,727	51
999999	6,106	11
999999	7,787	86
999999	40,441	100
999999	1,701	19
999999	48,598	62

Investigation Agents

Assessment/Feedback

- Results of running this system against 15 schools were disseminated to **Investigations** for their **assessment and feedback** of the effectiveness of this model.
- Conclusion - model had **identified all known fraud rings** from SY2010. Statistically this is very rare, which further gave us a sense of the value generated.
- **Identified new** previously unknown **fraud rings**.
- **Added additional students** to fraud rings under investigation.

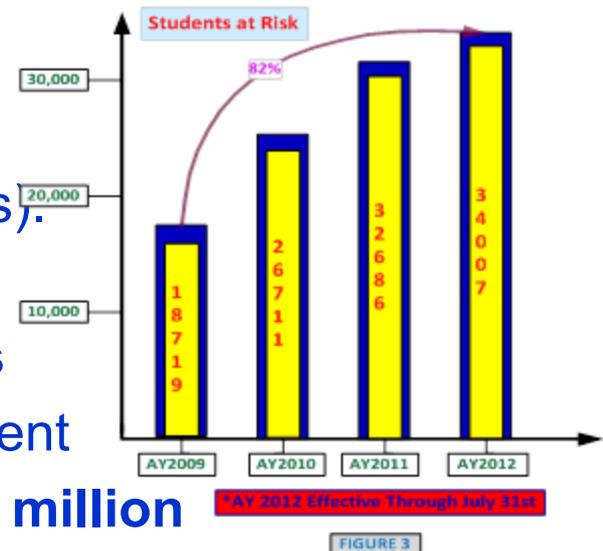
Unintended Consequences

Though our project objective was to identify fraud rings in direct support to our investigative staff, we were now also able to:

- Identify what the big picture most likely is. Up to this point, no one really knew how large this problem is.
- Impact on the Financial Statements – As the primary lender in the student loan arena, financial auditors now want to know what is the potential impact on the financial statement?
- Are we working the right cases – Effective resource management of limited investigation assets? Up till now, our cases have been primarily from hot line or school referrals. There may well be very large rings out there that were up to now hidden in the background.

Data Analytical Project Results

- Student aid fraud ring activity has increased 82 percent from award year (AY) 2009 (18,719 students) to AY 2012 (34,007 students).



- We identified a total of over **85,000 recipients** who may have participated in this type of student aid fraud ring activity, who received over **\$874 million** in Federal student financial aid.
- Applying a statistical model, we estimated that **\$187 million** of this \$874 million in Title IV funds are probable fraud loss.
- For FY2016 we have referred over **\$2.5 million** in referrals to both Investigations and the Department, in addition to cases developed by agents individually using this system.

Lessons Learned

- Every effective analytical system has an embedded **risk model** built within it.
- Beware of **the wall**...you will hit it, but do not get discouraged.
- **Vetting Process** – we discovered that this is the final crucial step in the Phase III – filtering out false/positives.
- Unintended Consequences - is a frequent **byproduct** of analytical projects if you look for it.

Data Visualization System

~

Purchase Card

Project Objective

Develop an automated system using data analytic techniques to identify p-card transactions having a high risk of fraud or abuse as well as to assess and identify emerging risk. More specifically, the system aims to provide an automated solution that satisfies the defined oversight responsibilities of the following OIG components:

- **Audit Services (AS)** - the Government Charge Card Abuse Prevention Act of 2012 requires agencies' IG to conduct periodic assessments and audits of agency card programs to identify and analyze risks of fraudulent, improper, or erroneous purchases and payments in order to develop an audit plan.
- **Investigation Services (IS)** - conducts, supervises, and coordinates investigations of suspected fraudulent activity by Departmental employees.
- **Quality and Integrity Group (QIG)** - conducts fair, impartial and independent reviews of OIG's processes and business areas. It also coordinates with IS on allegations of misconduct, including criminal misconduct, by OIG employees; and as such access to travel and purchase card transactions is necessary to support QIG's mission.

Assessment Rules/Indicators

Transaction Level

- Billing Amount > 3 Standard Deviation
- Billing Amount > Single Purchase Limit
- Closed Account Activities
- PayPal & Amazon Transactions
- Potential Splits
- Prohibited MCCs (merchant category code)
- Questionable MCCs
- Transactions with Sales Tax
- Weekend/Holiday Transactions

Assessment Rules/Indicators (Cont.)

Account Level

- Frequency of Purchasing from Same Vendor
 - a tool for identifying spending patterns of cardholders and potential fictitious vendors.
- Frequency of Purchasing under Same MCC
 - a tool for identifying spending patterns of cardholders and potential “problem” cardholders.
- Percentage of Whole Dollar Transactions
 - a tool for identifying potential suspicious & fictitious transactions.

Demonstration

Interagency Fraud Risk Data Mining Group

Mission Statement

The Interagency Fraud Risk Data Mining Group (IFRDMG) is a collection of investigators and auditors within the federal community that has been formed for the purpose of sharing best practices, raising awareness and offering a forum for the evaluation of data mining and risk modeling tools and techniques to detect fraudulent patterns and emerging risks.

<http://www4.va.gov/oig/ifrdmg/>