Using “Big” and “Little” data to build a compliance analytics program

May 2017

PwC

Using “Big” and “Little” Data to Build a Compliance Analytics Program

Marko Kuzmanovic
Manager
Enterprise Systems Solutions
PwC
marko.kuzmanovic@us.pwc.com

Ken Kozakura
Manager
Advanced Risk and Compliance Analytics
PwC
kenneth.kozakura@us.pwc.com
**Objectives**

- Big opportunities for compliance professionals to use data to proactively manage compliance risk
- A pragmatic approach for compliance professionals to use data science
- **Case Study:** How Microsoft used data science in their Compliance Analytics Program (and how you can too!)

**The Journey**
“I want to report the following misconduct…”

“We have questions about a contract you entered into with…”

How did this happen?

How did we not know that it was happening?
Data. Lots of data.

Data sets stored separately.
Targeted business and finance reports

Different stakeholders use the data for different reasons

Perspective matters
Data science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured, which is a continuation of some of the data analysis fields such as statistics, data mining, and predictive analytics.

Data science brings together a number of disciplines and approaches.
Data science for compliance professionals

1. Gather
   internal & external data

2. Analyze
   using algorithms

3. Report
   actionable insights

A Compliance Analytics Framework
What are the compliance risks in your sales model?

Indirect sales models are becoming increasingly complex as organizations heavily leverage channel partners to reach emerging markets or increase market share. Additionally, various levels of employees are empowered to approve greater pricing discounts, credit extensions, and other non-pricing concessions to win business, further increasing operational and compliance risks.

This model limits visibility into final end customer pricing and introduces unique operational or compliance and legal risks. Few organizations have the capabilities to identify overarching trends and themes to quantify the impact of these risks.

DOJ – Evaluation of Corporate Compliance Programs*

- On February 8th, 2017 the DOJ issued guidance on its evaluation of corporate compliance programs in the context of criminal investigations of corporate entities
- Describes the factors that prosecutors should consider when determining whether to bring charges, negotiate pleas, or evaluate other agreements. It focuses on the existence and effectiveness of the corporation's pre-existing compliance program
- Lists eleven high-level topics that DOJ has frequently found to be relevant in evaluating corporate compliance programs
- A common theme throughout the eleven topics is identification, controls, and detection through analysis and information gathering

*United States Department of Justice, Criminal Division, Fraud Section
**Transaction Monitoring: What is the Value Add?**

**The Problem:**
- Form our experience and benchmarking data, we've found the front lines of compliance staff can spend upwards of 80% of their time on issues with low to medium materiality and only 20% of their time on critical or high risk issues
- Compliance often occurs in silos causing overlap where several teams are unknowingly reviewing the same issues
- The absence of an enterprise wide view creates risk, increases costs, and limits the value of the information to the organization

**The Value Proposition:**
- **Reduce cost:**
  - Focus on what matters: leverage data analytics to focus on the most critical risks by both weight and materiality
  - Eliminate inconsistencies and redundancies in data collection, analysis, and review by potentially disconnected business units such as risk, compliance, and regulatory
- **Increase revenues:**
  - Centralize management of key data sets across the organization
  - Establish one master record, eliminate inconsistencies that could require future reconciliation
  - Free up front line staff to work on other revenue generating activities
- **Improve operational efficiency**
  - Leverage analytics to identify patterns within key compliance areas: customers, vendors, expenses, etc.
  - A more complete view allows for strategic decision making: cross-sell, up-sell, customizations, etc.
  - Leverage data in ways it was never used before such as advanced analytics, data visualization, machine learning, and inclusion of new unstructured or external data sets

**Transaction Monitoring: Maturity Model**

**The challenge: Integrating technology and fraud risks**

As the volume of meaningful data continues to grow and the quantity and nature of valuable data sources expands, it becomes increasingly important to visualize links that indicate a potential threat to the organization.

**Tier 1: Ad Hoc Reports**
- Utilizes standard reporting from source systems
- Manual review of random samples or results
- Tools: Standard ERP reports

**Tier 2: Structured Reports**
- Extract data from source system for custom analysis
- Analytical review using filtering, sorting, aggregation, etc.
- Tools: ACL, Access, Excel

**Tier 3: Advanced Analytics**
- Leverage prior experience or risk assessments to create custom analytics
- Identify unusual transactions using risk based approach and perform a targeted review
- Tools: Oracle, SQL Server, Data Visualization Tools, DQ Tools, Watch Lists

**Tier 4: Continuous Monitoring**
- Ongoing monitoring using proven technology, methodologies, and tools
- Tools: Oracle CCM Solutions, SAP CCM Solutions, Approva T&E, Greenlight Technologies, Infogix
**Transaction Monitoring: Approach**

Mature red flag detection approaches proactively use datasets to scrutinize for trends and relational outputs, and establish common and custom rules within an analytics library to guide detection and monitoring efforts toward areas of vulnerability. The process captures the continuous flow of lessons learned and is scalable to any magnitude of data and rules-based complexity.

1) **Acquire.** Request data from disparate data sources and load to an enterprise analysis environment

2) **Consolidate.** Consolidate to facilitate and identify anomalies and data quality issues

3) **Profile.** Enhance the data by supplementing or remediating issues

4) **Analyze.** Perform detailed analytics to highlight potential issues and identify rules that should be used for monitoring efforts

5) **Enhance Library.** Create, maintain, and enhance the ranking and scoring system by appending new rules and removing false positives

6) **Report.** Use data visualization tools to generate custom reporting and dashboards

---

**Transaction Monitoring: Ranking & Scoring**

Each step works towards improving the search for entities and transactions that may pose a threat to the organization:

1. Data is gathered from multiple sources and is standardized.
2. Intelligent decision systems capture expert knowledge and previous experiences.
3. Exception based & statistical analysis are used to evaluate transactions in context to identify abnormal events.
4. Behavioral models establish potential links with known high-risk entities through current events and third party databases of known offenders.
5. Data Analysis and field analysis are combined to select a sample set of transactions for deeper investigation.
6. Visualization and reporting allow for advanced visual discovery.
### Potential Data Sources & Tools

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Where to Look</th>
<th>Tools</th>
<th>Red Flag Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unstructured</strong></td>
<td>• Paper documents&lt;br&gt;• Emails&lt;br&gt;• Cell Phones&lt;br&gt;• Computers&lt;br&gt;• External Storage (hard drives, flash drives)&lt;br&gt;• Server/Access Logs&lt;br&gt;• Contract Management System&lt;br&gt;• Social Media</td>
<td>• Preservation/Collection: EnCase, FTK, NIUIX, Cellebrite, ABBYY&lt;br&gt;• Processing: NIUIX, Trident, Clearwell, OutIndex, CDFSrtch&lt;br&gt;• Review Platforms: Relativity, Stratify, Recommind, iConect, Concordance, Brainspace&lt;br&gt;• Social Media/Sentiment: Crimson Hexagon, Clarabridge</td>
<td>• Bribery: kickbacks, reciprocal agreements, bid rigging/collusion&lt;br&gt;• Insider Threat: Unsecured sensitive data (PII, PHI), unused access, identify and review user exceptions</td>
</tr>
<tr>
<td><strong>Structured</strong></td>
<td>• Transactional and General Ledger Data&lt;br&gt;• Inventory&lt;br&gt;• Fixed Assets&lt;br&gt;• Vendor, Customer, and Employee lookup tables&lt;br&gt;• Bank transactions&lt;br&gt;• WorldCompliance, WorldCheck, Dow Jones, Lexis/Nexis, D&amp;B, OFAC, EPLS, Local listing of debarred entities</td>
<td>• Database Platforms: Oracle, SQL Server, MS Access, MySQL&lt;br&gt;• Data Cleansing/Text Matching Tools: BusinessObjects, Datamatics, Global Name Recognition, Google Rfina, Monarch, Informatica&lt;br&gt;• Data Visualization: Tableau, MicroStrategy, Power BI, Qlik, Datawatch&lt;br&gt;• Continuous Monitoring: Oversight, ACL, Approva (more CCM vs CTM), SAP &amp; Oracle CM Solutions&lt;br&gt;• Case Management: SharePoint, QuickPlace, Oversight (Case Mgmt module)&lt;br&gt;• Advanced Analytics: SAS, R, Palantir, Python</td>
<td>• Payments: round dollar, cash, manual, unusual 3rd parties, inaccurate descriptions, many/large payments to single entity, duplicates, same requestor/approver&lt;br&gt;• Invoices: One time vendors, sequential invoices to same vendor, $0 invoices, substantial activity for new vendors</td>
</tr>
</tbody>
</table>
**Case Study**

How Microsoft used data science in their Compliance Analytics Program (and how you can too!).

Unlock insights from our data

Designing and implementing a Compliance Analytics Program (CAP) that will create an early warning and monitoring system for compliance risks
Microsoft’s Compliance Analytics Program

**Global Compliance Data Analytics Program**
Implement an integrated program to provide prioritized, risk-based analytics at the corporate and subsidiary level.

**Robust Field Readiness & Clear Accountabilities**
Land focused readiness processes to ensure that analytics are appropriately prioritized and in-depth training lands with stakeholders. Establish clear roles and responsibilities, and accountabilities.

**Sustainable and Proactive Model**
Use rapid analytic development methodologies and leverage machine learning techniques.

**Showcase as “Leading Practice”**
Showcase as "leading practice" for other companies to enhance their compliance programs through data analytics and visualization.
Microsoft’s road map

Creating new analytics or perspectives on existing analytics to identify sales transactions for further review (i.e. after-the-fact reviews)

Incorporate sales pipeline data to provide a similar type of review before sales transactions are approved (i.e. proactive reviews)

Compliance Analytics Building Blocks

1. Proof of Concept
2. Detective Processes
3. Integrate Feedback & Machine Learning
Microsoft’s High Risk Deal (HRD) Dashboard Toolkit

1. HRD Landing Dashboard
   Summary view of riskiest transactions & review status

2. High Risk Deals Report
   Analytics that identify which deals require additional review

3. Action Framework
   Recommended next steps based on risk flag and risk attribute

4. Review Tool
   High risk deals threshold feed into this tool which tracks the review status of each

High Risk Deals Analytic Demo
Becoming proactive and predictive
What does proactive look like?

**Past** ➔ **Present** ➔ **Future**

- Fragmented and siloed deal approval
- Targeted, post-deal review using analytics
- Targeted, pre-deal review using analytics

---

**The roadmap**

**People** ➔ **Process** ➔ **Tools**

01. Locate a pre-deal review and approval function in the existing sales process and add the analytic to their toolkit

02. Upgrade the analytic to allow real-time risk scoring as deals move through the quoting or sales process

03. Change the pre-deal sales review and approval processes to use the analytic as a control, instead of a monitoring solution

04. Connect the analytic to deal quoting systems to support business processes and force reviews before deal closure
Microsoft's Compliance Analytics Program will provide an early warning and monitoring system for a defined set of compliance risks.

1) Use data analytics to reactively and proactively cast a wider net over compliance risks.

2) Use probability driven analytics along with trending analysis to offer predictive insights and increased self-monitoring through early warning signals.

3) Provide interactive, real-time dashboards and reporting with automatic refresh and limited manual intervention.

4) Offer some level of user customization to tailor reports for specific needs.

5) Include a feedback loop to incorporate learnings and support continuous improvement.

Program Success Factors

- Solve for a clear, well-defined risk
- Start with a manageable data set
- Leverage the user community
- Evolve with the risk landscape
Illustrative applications beyond technology sales

<table>
<thead>
<tr>
<th>Industry</th>
<th>Compliance Challenges</th>
<th>Data Analytics Use</th>
<th>Insight</th>
</tr>
</thead>
</table>
| Healthcare Provider | • Non-compliance of physician compensation with fair market value; physicians cannot give themselves ongoing undue benefit. Non-compliance penalties are severe.  
• Up-coding during patient visits, coding the patient’s condition at higher levels of acuity or severity, allowing the physician to charge more and increase reimbursement. | • Total compensation data aggregation and comparison against fair market value trends by specialization, years of experience, location, hospital network, etc., and identification and analysis of outliers.  
• Correlate national and historic coding data for a broad set of services establishing a baseline cost across a variety of factors, enabling easy identification of outliers and exceptions. | • Early warning and monitoring system for total physician compensation and per-visit coding exceptions, allowing for quick remediation and significant cost savings. |
| Retail         | • Supply chain includes unauthorized or unapproved suppliers in a higher risk market.  
• Distributors and customers are fraudulently returning products that may have been stolen, broken, used, or purchased to hit sales targets. | • Review transactions that are being transacted through “higher risk” distributors or resellers. Layer in other risk attributes such as if they are also receiving incentives or discounts.  
• Monitor the occurrence of high inventory shrink rate over time or in comparison to other stores  
• Monitor stores that have dramatic increase in number of returns or markdowns. | • High risk deals review of transactions in higher risk markets, with unknown distributors, and/or with multiple incentives.  
• Early warning system to determine if there are stores or employees correlated with high returns rates or markdowns. Implement stronger controls and trainings for stores or employees. Revisit returns policies if necessary. |
| Banking        | • Adherence to the Bank Secrecy Act and Anti-money Laundering Act (BSA/AML)  
• Bank employees may open new accounts with no activity to hit sales targets | • Automate the review of transactions to correlate AML red flags and review high risk deals. Once matured, bring in text and web analytics.  
• Monitor the activity of bank accounts opened by employees. | • Implement a high risk transaction review to perform detailed review of transactions to determine if money laundering is occurring.  
• Flag employees for review who have hit sales targets with a significant number of accounts that do not have activity. |
Questions?

Thank you!

**Marko Kuzmanovic**
Manager
Enterprise Systems Solutions
PwC
marko.kuzmanovic@us.pwc.com

**Ken Kozakura**
Manager
Advanced Risk and Compliance Analytics
PwC
kenneth.kozakura@us.pwc.com
Appendix

Sample Vendor Risk Dashboard
**Sample Time & Expense Dashboard**

**Executive level summary**

1. **Summary Statistics**
   At the top of the report, there are high level summary statistics around volume and status of reviews.

2. **Country Level Metrics**
   Drill down capabilities at the country level.

3. **Aging**
   Understand how quickly issues are getting resolved and click-down to outcome of review.
Identify transactions for review

1. Summary Stats
   At the top of the report, there are high-level summary statistics around total agreements, sales, amounts, and customers.

2. Country Level Metrics
   For each of the countries in the selected region/area, you can see which countries have the highest aggregated risk.
   Looking at Brazil, we can see that there's a 62 score (out of 100) but a high geographical risk as a Tier 1 country and a medium risk partner and trend score.

Identify transactions for review

3. Agreement Hover
   Hovering over the top red circles in the bubble chart, you can see the highest risk transactions colored red. In this example, agreement 850378-1 has a high risk score of 83 (out of 100)
   On the bottom left pane we can see transactional details associated with this agreement update with hover.
Agreement level

4 Agreement Score
Overall this agreement has been identified as having a high risk geography, partner, government, and large discount risk with a "Tier 3" medium risk associated with trend.

5 Agreement Details
This particular agreement happened in Brazil through the reseller Contoso, Ltd and sold to Government de Estado.

Brazil is a Tier 1 risky country and Contoso, Ltd generally does 37% of total government business in Brazil, however, in Q1 FY16 there was a significant transaction spike.

Additionally this enrollment has a 54% discount which is large compared to similar deals of 18%.

Follow-up
We can scroll through the "details and transactions" text box to see which specific products were sold and export this information to Excel for further follow-up.
Partner level

Partner Level Summary

Look at the same transaction, to see how it compares to how this partner generally sells to see if it fits the trend. In the bar chart/line graph, there are deals that closed in Q3 FY16 that had a relatively low list amount but very high in terms of discount percent.