Leading Practices to Leverage Forensic Data Analytics in Compliance Monitoring and Investigation
Table of contents

1. Proactive Compliance Monitoring
2. EY’s First Annual Global Forensic Data Analytics Survey
3. Big Data and Forensic Data Analytics
4. EY’s Forensic Data Analytics Solution
5. Beyond Compliance and Investigation
6. Case Examples
Proactive Compliance Monitoring
Top Compliance Issues—What We are Seeing

► Bribery and corruption remain top risks
► Regulatory compliance and fraud & abuse
► Third-party integrity
► M&A due diligence
► Risk areas include:
  ► Integrity of vendors, suppliers and distributors
  ► Improper payments in the forms of bribes or kickbacks
  ► Travel and entertainment abuse
  ► Conflicts of interests (e.g., employee and supplier matches)
Now, more than ever, increased transparency is top-of-mind among our clients in...

► Internal Audit
► Compliance & Legal
► Investigations
► Business / Operations

What we hear:
1. Make my program more effective and measurable
2. Make my program more efficient (reduced sample sizes, risk based, cost savings)
50% by tip or accident demonstrates the need for improved analytics

Initial Detection of Occupational Frauds

<table>
<thead>
<tr>
<th>Detection Method</th>
<th>Percent of Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tip</td>
<td>43.3%</td>
</tr>
<tr>
<td>Management Review</td>
<td>14.6%</td>
</tr>
<tr>
<td>Internal Audit</td>
<td>14.4%</td>
</tr>
<tr>
<td>By Accident</td>
<td>7.0%</td>
</tr>
<tr>
<td>Account Reconciliation</td>
<td>4.8%</td>
</tr>
<tr>
<td>Document Examination</td>
<td>4.1%</td>
</tr>
<tr>
<td>External Audit</td>
<td>3.3%</td>
</tr>
<tr>
<td>Notified by Police</td>
<td>3.0%</td>
</tr>
<tr>
<td>Surveillance/Monitoring</td>
<td>1.9%</td>
</tr>
<tr>
<td>Confession</td>
<td>1.5%</td>
</tr>
<tr>
<td>IT Controls</td>
<td>1.1%</td>
</tr>
<tr>
<td>Other*</td>
<td></td>
</tr>
</tbody>
</table>

**Other* category was not included in the 2010 Report.

2012 ACFE Report to the Nation on Occupational Fraud
Compliance Monitoring Challenges

► A good compliance monitoring program needs to be flexible and adaptive
► Investigators, attorneys and regulators need information now
► Convergence of structured and unstructured data
► Organizational silo leads to redundant efforts to meet monitoring and reporting requirements
► The volume of business activities that should be monitored can overwhelm the resources of most organizations.
► Get the right FDA tools and the right people to operate FDA
► The data available for analysis are incomplete or inaccurate
► Need for innovative analytics to increase detection and reduce false positives
► Timely follow up and remediation
How Global Companies are Responding

► **Compliance** and **legal** are often teaming with **internal audit** to look beyond anti-corruption policies and training and into tests of books and records
► Integrating **new analytics** specifically targeting corruption – these aren’t your typical rules-based, process control SOX tests
► Integrating **“Big Data”** concepts including:
  ► Text mining (unstructured data)
  ► Statistical analyses and anomaly detection
  ► Visual analytics and interactive dashboards
  ► 100% data sampling, not just random sampling
► Analytics used to assess high fraud/corruption risk areas
Ernst & Young’s First Annual Global Forensic Data Analytics Survey

Telephone interviews conducted globally with leading companies

<table>
<thead>
<tr>
<th>Markets</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Americas</td>
<td>105</td>
</tr>
<tr>
<td>US</td>
<td>65</td>
</tr>
<tr>
<td>Brazil</td>
<td>40</td>
</tr>
<tr>
<td>EMEIA</td>
<td>200</td>
</tr>
<tr>
<td>France</td>
<td>40</td>
</tr>
<tr>
<td>Germany</td>
<td>41</td>
</tr>
<tr>
<td>India</td>
<td>40</td>
</tr>
<tr>
<td>Italy</td>
<td>40</td>
</tr>
<tr>
<td>UK</td>
<td>40</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>160</td>
</tr>
<tr>
<td>Australia</td>
<td>40</td>
</tr>
<tr>
<td>China / HK</td>
<td>40</td>
</tr>
<tr>
<td>Japan</td>
<td>40</td>
</tr>
<tr>
<td>Singapore</td>
<td>40</td>
</tr>
<tr>
<td>TOTAL</td>
<td>466</td>
</tr>
</tbody>
</table>

Target audience/ recruitment criteria:

- Senior individuals in global companies: Finance (excluding audit), Internal audit & risk, Procurement, Legal, Compliance, Investigations, Business/management
- Companies with over US $100 million annual revenue, over 50% greater than US $1 billion
- Criteria: Must have decision-making responsibility around the company’s anti-fraud and anti-bribery efforts, compliance monitoring and investigations
- Respondents are users of forensic data analytics only
- Research period: November 2013 and January 2014

This work has been conducted in accordance with ISO 20252, the international standard for market and social research
FDA efforts are well aligned with perceived fraud risks
► Bribery and corruption is top fraud risk at 65%
► Further, 74% are using FDA to combat bribery and corruption

Big data has big potential
► 72% of respondents believe emerging big data technologies can play a key role in prevention and detection

Key benefits
► Enhance our risk assessment process (90%)
► Detect potential misconduct that we could not detect before (89%)
Need to do more

- 63% of respondents agree that they need to do more to improve their anti-fraud/anti-bribery procedures, including the use of FDA

Biggest challenge

- Getting the right tools or expertise for FDA (26%)

Data volumes analyzed are relative to company size

- 71% of companies over US$1 billion in revenues report working with data sizes less than or equal to 1 million records
- 42% of companies with revenues between US$100 million and US$1 billion report working with data sets under 10,000 records
Big Data and Forensic Data Analytics
What is Big Data?

Gartner:

Big Data is high **volume**, **velocity** and **variety** information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.
## IBM Survey - Big Data Sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transactions</td>
<td>88%</td>
</tr>
<tr>
<td>Log Data</td>
<td>73%</td>
</tr>
<tr>
<td>Events</td>
<td>59%</td>
</tr>
<tr>
<td>Emails</td>
<td>57%</td>
</tr>
<tr>
<td>Social Media</td>
<td>43%</td>
</tr>
<tr>
<td>External Feeds</td>
<td>42%</td>
</tr>
<tr>
<td>Sensors</td>
<td>42%</td>
</tr>
<tr>
<td>Free-form text</td>
<td>41%</td>
</tr>
<tr>
<td>RFID scans or POS data</td>
<td>41%</td>
</tr>
<tr>
<td>Geospatial</td>
<td>40%</td>
</tr>
<tr>
<td>Audio</td>
<td>38%</td>
</tr>
<tr>
<td>Still images/video</td>
<td>34%</td>
</tr>
</tbody>
</table>

Source: IBM Technology Outlook 2012
IBM Survey - Big Data Analytics Activities

- Query and reporting: 91%
- Data mining: 77%
- Data visualization: 71%
- Predictive modeling: 67%
- Optimization: 65%
- Simulation: 56%
- Natural language text: 52%
- Geospatial analytics: 43%
- Streaming analytics: 35%
- Video analytics: 26%
- Voice analytics: 25%

Source: IBM Technology Outlook 2012
What is Forensic Data Analytics?
Forensic data analytics maturity model
Beyond traditional “rules-based queries” - consider all four quadrants

Structured Data
- Low Detection Rate
  - Matching, Grouping, Ordering, Joining, Filtering

Unstructured Data
- High False Positive Rate
  - Keyword Search

“Traditional” rules-Based Queries & Analytics
- Low Detection Rate
  - Anomaly Detection, Clustering, Risk Ranking

Statistical-Based Analysis
- High False Positive Rate
  - Data visualization, Drill-down into data, Text Mining

Data Visualization & Text Mining
- Low Detection Rate
  - Traditional Keyword Searching

- High False Positive Rate
  - Data Visualization & Text Mining
## Unstructured data sources

<table>
<thead>
<tr>
<th>Who</th>
<th>What</th>
<th>When</th>
<th>Why</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Networking</td>
<td>Concept Clustering and Keywords</td>
<td>Communication Over Time</td>
<td>Sentiment Analysis</td>
</tr>
<tr>
<td>“Who is talking to whom?”</td>
<td>“about what?”</td>
<td>over which time period?</td>
<td>how do they feel?”</td>
</tr>
<tr>
<td>▶ People-to-people analysis</td>
<td>▶ Top words mentioned</td>
<td>▶ When communications occur</td>
<td>▶ Positive vs. Negative Sentiment</td>
</tr>
<tr>
<td>▶ Entity-to-entity analysis</td>
<td>▶ Key concepts / topics</td>
<td>▶ Communication spikes around key business events</td>
<td>▶ Top 10 angry or negative emails</td>
</tr>
<tr>
<td>▶ Map communication lines to organization chart</td>
<td>▶ Top or unusual dollar amounts</td>
<td>▶ Linked to T&amp;E or other journal payment related transactions</td>
<td>▶ Ethical tone: flag for secretive, harassing, confused, or derogatory communications</td>
</tr>
<tr>
<td></td>
<td>▶ Sensitive words / phrases</td>
<td></td>
<td>▶ Fraud Triangle analytics: link communications to an employee’s pressure, opportunity or rationalization risk score</td>
</tr>
<tr>
<td></td>
<td>▶ Potentially privilege flagging</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▶ Potentially non-responsive</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>▶ Predictive coding (more-like-this)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Emerging Trends: Information and Analytics

70% Descriptive Analytics
What is happening?

30% Diagnostic Analytics
Why did it happen?

16% Predictive Analytics
What is likely to happen?

3% Prescriptive Analytics
What should I do about it?

Source: Gartner Research
Our Technology Ecosystem
Industry-recognized Technology Stack

- Structured data
- Unstructured data
- Early case assessment
- Dynamic reporting and visualization

Information Governance

Client presentation and reporting layer

Processing layer

Forensics and analytics core

Big data platform

Email and IM processing tools

Analytics and statistics tools

Forensic software

Big Data

Evidence Tracking System

Forensic software

Unstructured data

Structured data

Big Data
To drive better decisions, we must first ask the right business questions and then seek answers in the data. Thus, our work moves left to right, but our thinking must move from right to left.

Focus of many companies
- Many analytics companies in the marketplace today are dominated by data warehousing, enterprise dashboard and reporting solutions.
- Many clients, however, still struggle to embed analytics into operational decisions in a systematic and repeatable way, often resulting in clients not realizing the full value of analytics.

EY’s strategic focus
- Our focus is on becoming the leader in “value-driven analytics” by going to market through sectors and core competencies, supported by a centralized group providing market-leading, analytical and big data skills and technology.
- We also realize the importance of using change management skills to help our clients more effectively use analytics to create value.
Our Approach to Detective and Predictive Data Analytics

**Basic**
- Vendor/Customer list
- Contracts/agreements
- Purchase/Sales orders
- Receipts/invoices
- Vendor/Customer payment

**Advanced**
- All of data from basic
- Emails
- Chat logs
- E-files
- Call logs

**Predictive**
- All data from Basic and Advanced
- Statistical modeling
- External sources – blogs/posting, Twitter, Facebook

- Monitoring to identify irregular activities from numerical data sets using specified rules
- Linking the identified irregular activities with purchasing/sales reps and vendor/customer contact – “who’s talking to who”
- Identifying the “unknown” – using data sets to identify rules and to continuously redefine “basic”

**Remedial action identified**

**Sustainable and improved operating practices**
Below is an illustration of how a broad data collection exercise operates in practice. The objective is to gather data from a range of sources – and undertake initial processing to provide a central team with the ability to identify the higher risk activities. Following that review, targeted analytics would be deployed to identify the issues, transactions and relationships that need to be reviewed.

- **Gather**
  - Obtain data from all central systems and external sources.

- **Process**
  - Load, validate and transform data into define common model - independent of ERP.
  - Link sources to facilitate analysis.

- **Analysis**
  - Provide global dashboards to facilitate identification of risk issues.

- **Delivery/Follow up**
  - Deliver dashboards to be reviewed as part of the testing process.
Risk Indicator Framework Design

Tailored design with data analytic risk indicators

- Duplicate Payments: A%
- Meal Splitting: B%
- Travel Agents: C%
- Overbilling: D%

In-Scope Transactions ➔ High Risk Transactions

- Not every item bears the same risk level
- Define risk based on understanding of business process and potential control weaknesses
The dashboard tells you “who got paid what, where and what for”.

Data Visualization: Accounts Payable Monitoring
The 4W1H tell you “Who entertained who, where, what for, and for how much?”

### Data Visualization: Travel & Entertainment Monitoring

#### WHERE | Expense Location

<table>
<thead>
<tr>
<th>Location</th>
<th>Amount by Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>$2,965,477.04</td>
</tr>
<tr>
<td>Sa</td>
<td>1,219.91</td>
</tr>
<tr>
<td>Austria</td>
<td>$161,147.81</td>
</tr>
<tr>
<td>Switzerland</td>
<td>$155,656.48</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>$100,121.12</td>
</tr>
<tr>
<td>Germany</td>
<td>$100,300.47</td>
</tr>
<tr>
<td>Spain</td>
<td>$567,106.73</td>
</tr>
<tr>
<td>Canada</td>
<td>$567,003.77</td>
</tr>
<tr>
<td>France</td>
<td>$223,424.62</td>
</tr>
<tr>
<td>Vietnam</td>
<td>$333,393.68</td>
</tr>
<tr>
<td>Others</td>
<td>$380,880.00</td>
</tr>
</tbody>
</table>

#### HOW | Top Concepts

- **CONCEPT**
  - [No Concept] $90,247.64
  - [None] $93,669.63
  - [Do] $150,485.50
  - [Dinner] $188,147.63
  - [Events] $373,347.42
  - [Conference] $73,024.26
  - [Club] $569,661.14
  - [Meals] $680,235.78
  - [Business events] $60,049.07
  - [Business] $60,049.07

#### WHY | Expense Purpose

<table>
<thead>
<tr>
<th>Purpose_Detail</th>
<th>Entry_Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airfare for trip to Vietnam and Cambodia to participate in business meetings</td>
<td>$49,696.40</td>
</tr>
<tr>
<td>Property Loss Control - Site Visits: Vienna, Austria</td>
<td>$46,196.10</td>
</tr>
<tr>
<td>Airfare for trip scheduled to India, Chicago, Ottawa</td>
<td>$42,631.32</td>
</tr>
<tr>
<td>Business meetings, meals and events in VA and DC</td>
<td>$33,095.82</td>
</tr>
<tr>
<td>Travel to European Business Units to discuss tax issues</td>
<td>$379,908.99</td>
</tr>
<tr>
<td>Grow member onboard GHQ aircraft</td>
<td>$34,712.74</td>
</tr>
<tr>
<td>Meetings at Jet Aviation</td>
<td>$34,712.74</td>
</tr>
<tr>
<td>Business meetings, meals and events in VA, DC</td>
<td>$34,642.60</td>
</tr>
<tr>
<td>Club Membership</td>
<td>$33,960.00</td>
</tr>
<tr>
<td>Travel to Cambodia and Vietnam as a participant</td>
<td>$33,393.88</td>
</tr>
<tr>
<td>Receipt for fees to participate in the US ASEAN bus</td>
<td>$32,600.00</td>
</tr>
</tbody>
</table>

#### WHAT | Expense Type

<table>
<thead>
<tr>
<th>Expense_Type</th>
<th>Entry_Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airfare</td>
<td>$1,505,514.87</td>
</tr>
<tr>
<td>Lodging</td>
<td>$869,341.68</td>
</tr>
<tr>
<td>Seminar/Training/Conference Fees</td>
<td>$369,465.27</td>
</tr>
<tr>
<td>Meals/Meetings, Local Business Meals</td>
<td>$252,923.06</td>
</tr>
<tr>
<td>Other Expenses</td>
<td>$229,292.31</td>
</tr>
<tr>
<td>Taxi/Subway/Bus/Limo/Car Service</td>
<td>$209,010.49</td>
</tr>
<tr>
<td>Meals/Meetings, Business Meals</td>
<td>$174,799.22</td>
</tr>
<tr>
<td>Airfare - Fare Difference</td>
<td>$131,948.02</td>
</tr>
<tr>
<td>Airfare - Group Fares</td>
<td>$133,206.79</td>
</tr>
<tr>
<td>Car Rental</td>
<td>$133,206.79</td>
</tr>
<tr>
<td>Dues - Other Club Dues</td>
<td>$120,317.90</td>
</tr>
</tbody>
</table>

#### WHO | Employee

<table>
<thead>
<tr>
<th>Employee</th>
<th>Entry_Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dean Bunch</td>
<td>$2,474,047.53</td>
</tr>
<tr>
<td>Loris Almeasides G</td>
<td>$34,616.00</td>
</tr>
<tr>
<td>Maurice Raymond M</td>
<td>$33,618.65</td>
</tr>
<tr>
<td>Maher, Sean William J</td>
<td>$56,555.46</td>
</tr>
<tr>
<td>Hayduk, Kenneth R</td>
<td>$54,405.98</td>
</tr>
<tr>
<td>Ohto, Yutaka A</td>
<td>$52,117.69</td>
</tr>
<tr>
<td>Jordan, Mauri T</td>
<td>$53,355.62</td>
</tr>
<tr>
<td>Lamb, Gerard F</td>
<td>$54,161.97</td>
</tr>
<tr>
<td>Andersen, Brian D</td>
<td>$50,239.14</td>
</tr>
<tr>
<td>Michael Lestavy</td>
<td>$50,078.48</td>
</tr>
<tr>
<td>Ritter, Douglas D</td>
<td>$49,645.99</td>
</tr>
</tbody>
</table>

---

Note: The table and diagram represent data visualization and analysis of travel and entertainment expenses, focusing on identifying key concepts and trends for anti-bribery and corruption monitoring in XYZ Corporation.
Payment Risk Scoring
Key Component to Educing False Positives and Focusing Risk Assessment
FDA - In-depth Overview of Big Data
Distributor or 3rd Party Analytics

Analytics include:

► Stratify revenues by top and bottom distributors
► Analysis of credits and returns
► Analysis of free goods and samples
► Analysis of write-offs
► Analysis of payment cycles
► Commissions analysis
► Analysis of price variances, by distributor and by product
► Zero dollar shipments
► Analysis of “debits” to revenue
► Compare order quantities with product and receipt quantities
► Profit margin analysis
► Anomalous or zero shipment weights
► Vendor due diligence and address verifications, sanction list check, etc.
Analytics include:

- Employee stratification
- Stratify by expense type
- Sensitive keywords
- Term frequency analysis (concept analysis)
- Round payments
- One-time payments
- Potential “gross ups”
- Potentially duplicative
- Out-of-policy spend
- Weekend or personal use
- High risk venues (e.g., adult entertainment, check cashing, etc.)
- Meal splitting
- HCP spend
- Spending over time/ trending
The dashboard tells you relationships identified through the analysis of structured and unstructured data sources.
Rather than simply comparing watch-list names to a vendor table in a spreadsheet, this example links accounts payable data to third-party watch-list data to identify potentially improper payments to sanctioned or high-risk entities and displays the results in an interactive dashboard.

Demonstrate Management Oversight & Intent
Linking Payment Data to Sanctions and Watch List Databases
These three variables were this highest drivers of suspicious transactions.

These variables were less important when predicting suspicious transactions. Client should focus resources on monitoring efforts for the three leading drivers, which accounts for 80% of the predictive value.
Merging Electronic Discovery Management Services and Forensic Data Analytics

FIDS investigated alleged improper payments through employee travel and entertainment expenses.

Using forensic data analytics and enhancements to our review platform, investigators used data visualization to select certain expenses for further review in Relativity.
EY investigated alleged improper healthcare provider (HCP) payments through employee travel and entertainment expenses.

Using forensic data analytics and enhancements to our review platform, investigators used data visualization to select certain expenses for further review in Relativity.
We identified several thousand wire transfers that were the subject of alleged use of customer funds in the company’s own investments.

We developed an algorithm that associated related email messages to the wire transfers based upon minimally available wire transfer information.

The wire transfers and corresponding email chatter were presented in our file review platform for further analysis.
Approach to Continuous Transaction Monitoring

Why Continuous Monitoring?
► Executive visibility and transparency
► Drive process improvements
► More advanced anti-fraud control
► Improved audit effectiveness

Enables:
► Proactively identify and remediate transaction-related issues and challenges
► Generate advanced analytics/insights
► Timely, accurate, complete reporting
Beyond Compliance and Investigation
Leverage Data Analytics to Optimize Operations and Grow Business

Operations/
Business Process

Data reflecting
the Business
strategies and
operations?
Case Example: Predictive Modeling

**Challenge:** Analyze 400,000 transactions for suspected bribery payments per DOJ subpoena

- Team reviewed 2,000 transactions from ledger data (text comments, amounts, dates, etc.)
  - Identified 400 suspicious and 1,600 non-suspicious entries
- Created statistical model: “Is Suspicious” / “Is Not Suspicious”
- Applied model to remaining 398,000 additional transactions
- Identified 14,000 new suspicious transactions
  - With confidence over 95% similar to “Is Suspicious”
  - Identified over $8 million in highly suspicious payments
  - Methodology accepted by the DOJ for this case
Case Example: Business Growth

Goal:
► Identify new customers most likely to buy Product A

Proposed Approach:
► Look at profile of physicians currently buying Product A and use predictive model to identify causal factors differentiating high potential customers from low potential customers
► Use causal success factors to score entire list of physicians to identify and target high potential customers

Potential model for identifying new customers most likely to buy

\[
\frac{1}{e^{7.7 + 0.8 \times \text{(Prod Count)\,\text{ln} - 1) + 0.004}}}
\]

Stats Equation

Model Results

75% results can be achieved by targeting 30% of customers
Thanks You