BEST PRACTICES TO LEVERAGE FORENSIC DATA ANALYTICS IN INVESTIGATION AND COMPLIANCE MONITORING

In the big data era, companies face the challenge of effectively leveraging forensic data analytics (FDA) to protect their interests, optimize their operations, and grow their businesses. This session will discuss the basics of FDA, its benefits, methodologies, best practices, and challenges. You will also learn about tools to use in your analysis, including the risk scoring framework, FDA maturity model, interactive dashboards, social network analysis, and predictive modeling.

You will learn how to:

- Explain what FDA is and how it can help you with your work.
- Incorporate FDA workflow and best practices.
- Use visualization and predictive analysis.

CHI CHEN, CFE, CISA
Director, Fraud Investigation & Dispute Services
EY
Shanghai

Chi Chen has more than 13 years of experience in financial and enterprise data analytics and has worked in the United States, Australia, and Greater China. He has extensive experience utilizing analytical software packages to join disparate data sets into a single repository, performing various analyses upon them, and identifying trends, patterns and outliers to detect fraud. He has led the design of a number of complex analytical engines leveraging predictive analysis, statistical modeling, data visualization, and text mining to help clients build continuous monitoring systems to identify fraudulent or high-risk transactions timely and efficiently. He is also experienced in providing clients leading forensic data analytics solutions, including AML analytics; anti-bribery and anti-corruption analytics; Fraud Triangle analytics; third-party due diligence; business intelligence; text mining; and statistical anomaly detection.

“Association of Certified Fraud Examiners,” “Certified Fraud Examiner,” “CFE,” “ACFE,” and the ACFE Logo are trademarks owned by the Association of Certified Fraud Examiners, Inc. The contents of this paper may not be transmitted, re-published, modified, reproduced, distributed, copied, or sold without the prior consent of the author.
Foreword

*We Live in a Data-Centric World*

Data is everywhere. From complex general ledger accounting systems to network security and social media, it is the all-encompassing backbone of the corporate world. If you are responding to litigation issues or regulatory demands, are engaged in an investigation or are trying to prevent fraud, you will need to understand complex data sets from multiple sources within your business. The key to making sense of and commanding this rapidly expanding frontier is to effectively leverage new technology in conjunction with subject matter professionals.

---

Today, an hour’s worth of business for a typical Fortune 1000 company can create millions of transactional records. The entirety of data from the private sector doubles every 14 months. Big data is no longer a phenomenon but a vastly expanding data footprint in our everyday course of business. This data footprint is at the core of most regulatory matters, investigations and complex business issues.

---

*The Bar Is Raised: Big Regulatory and Legal Risks Require Big Data Thinking*

In today’s regulatory environment, global companies are under greater pressure to improve their anti-fraud and anti-corruption compliance program. Compliance continues to be a top priority for boards of directors, audit committees, and senior management at many multinational companies.

It is clear from recent regulatory enforcement actions and settlement agreements across the globe that the bar has been raised. Simply having an anti-fraud and anti-corruption compliance program is not sufficient.
Traditional financial control tests are fundamentally different from the controls required for effective detection and monitoring of fraud, bribery, and corruption. To be successful, companies need to integrate forensic data analytic techniques that incorporate Big Data concepts across multiple data sources, third-party watch lists, transactional data, text mining, and even social media and email to prioritize and isolate areas of risk or rogue activity.

Integrating forensic data analytics (FDA) into a robust anti-fraud program or investigative process enables today’s chief compliance officers, general counsel, and chief audit executives to ask questions of their data that they have never been able to ask before. We are beyond traditional rules-based tests or random sampling that may miss important information or generate large amounts of false positive results. Executives can now pinpoint regions or business units where increased risks are surfacing based on multiple attributes in the data—not just from one source or isolated test. This yields significantly higher quality results. At the same time, companies can reduce overall internal audit or compliance field work costs by leveraging forensic data analytics as a component of the risk assessment process to determine which locations to audit, rather than running analytics after the fact.

**Forensic Data Analytics (FDA)**

*Defining Forensic Data Analytics*

We defined the term “data analytics” as the ability to collect and use data to generate insights that inform decision making. More specifically, we describe forensic data analytics in the context of managing fraud and bribery risk as the ability to collect and use electronically stored information, both structured and
unstructured data sources, to identify potentially improper payments, patterns of behavior, and trends. Forensic data analytics encompasses integrating continuous monitoring tools, analyzing data in real time, and allowing for immediate action to prevent suspicious or fraudulent payments.

A General Framework for Forensic Data Analytics

BUSINESS UNDERSTANDING
This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives.

DATA UNDERSTANDING
The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or
to detect interesting subsets to form hypotheses for hidden information.

**DATA PREPARATION**
The data preparation phase covers all activities to construct the final data set (data that will be fed into the modeling tools) from the initial raw data. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modeling tools.

**MODELING**
In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to our estimated optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements for the form of data. Therefore, stepping back to the data preparation phase is often necessary.

**EVALUATION**
At this stage in the project you have built a model (or models) that appear to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.
DEPLOYMENT
Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the client can use. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases it will be the client, not the data analyst, who will carry out the deployment steps. However, even if the analyst will not carry out the deployment effort, it is important for the client to understand upfront the actions which will need to be carried out in order to actually make use of the created models.

Risk Ranking Framework
Risk ranking is a means of generating a targeted sample of transactions by calculating a normalized score across various attributes of the work plan. In a risk ranking framework, tests are assigned weights based on their significance in the model. The weights are applied to a normalized score by test.

Risk ranking can occur at either the entity level (i.e., vendor, employee, and customer) or at the transaction level.

The screenshot below illustrates how the weight of each test is incorporated into the final score assigned to the transaction or entity. Weights are applied to the score by attribute and then aggregated across all attributes to produce a final score.
The concept behind risk ranking is to isolate transactions or entities with the highest risk.

Three Components of Analytics Design

*Rules-Based*

Traditional SQL query execution is the basis of the rules-based testing. Rules-based tests can be taken from the library of fraud detection tests. Rules-based tests are the core to forensic data analytics testing. These tests drive the majority of the analysis and are commonly reflected in the visual deliverables. Common examples include:

- Large round payments
- Duplicative payments
- Payee not on vendor master

*Statistical and Visualization*

Data visualization presents an effective way to display our data and understand the results. We often use specific tools to visualize our data sets and analytics results, helping reviewers discover relationships, find patterns, locate outliers, determine structure, and summarize findings.
**Text Mining**

Text mining focuses on deriving patterns and trends from text fields with the help of statistical and linguistic tools. Text mining can be executed as a supervised approach (i.e., keyword searching) or as an unsupervised approach (i.e., entity extraction and concept extraction). Common text mining techniques used include:

- Concept analysis
- Link analysis
- Entity extraction
- Keyword hits

**Ernst & Young’s First Annual Global Forensic Data Analytics Survey**

**About the Survey**

This survey was conducted between October 2013 and December 2013 on behalf of EY’s Fraud Investigation & Dispute Services practice. We believe it is the first ever of its kind from a subject matter, size, global reach, and survey respondent perspective.

Here are the highlights of the survey design:

- 446 companies were surveyed, across 11 countries.
- Respondents are executive and senior management responsible for anti-fraud and anti-corruption programs.
- Forty-five percent of the companies generate $100 million to $1 billion in revenue, 55% over $1 billion.
- Over a dozen industries are represented, with the largest shares held by financial services, life sciences, oil and gas, utilities, and mining.
Key Findings
As a key focus, we sought to understand how companies are responding to global fraud and corruption risks from an analytics perspective as well as to gauge their overall maturity with respect to use of FDA technologies.

Here are some of our key findings:
- Bribery and corruption is the largest fraud risk area.
- Seventy-five percent of the companies surveyed use forensic data analytics (FDA).
- FDA includes a broad base of users, including corporate executive management (81%) and the board of directors (65%).
- Spreadsheets and database tools still dominate the technology landscape.
- There is relatively low awareness and use of the more sophisticated FDA technologies such as Big Data, natural language processing, statistical analysis, and data mining needed to fight today’s fraud and corruption risks.
- While 67% of the companies indicate that emerging Big Data technologies can play a key role in fraud prevention and detection, only 7% were aware of any specific Big Data technologies and only 2% were actually using Big Data technologies.
- Sixty-seven percent of respondents say their current anti-fraud and anti-bribery program is effective in preventing and detecting fraud and corruption; however, 64% say they need to do more to improve their current procedures, including the use of FDA.
- Sixty-two percent of respondents say they need to improve management’s awareness of the benefits of FDA and proactive transaction monitoring.
• Survey respondents reported the largest challenges were (1) getting the right FDA tools and (2) a lack of human resources or manpower to operate FDA.

Big Risks Require Big Data Thinking

Current Regulatory Risk Landscape

It is clear that today’s regulatory environment plays an important role in shaping the FDA landscape. Given recent headlines and government enforcement actions, this should not come as a surprise.

Beyond country-specific anti-corruption efforts, other common regulatory oversight considerations include financial reporting requirements, environmental regulations, human rights, and anti-money laundering (AML), among a host of many others. In a recent 2013 EY business risk survey entitled, “Business Pulse: Exploring dual perspectives on the top 10 risks and opportunities in 2013 and beyond,” a survey of over 600 companies globally identified regulatory and compliance risks among the top ten business risks. As depicted in the graph below, at the center of the radar are the risks and opportunities that the survey respondents felt were having the biggest impact on major organizations worldwide. Arrows indicate the extent to which the ranking is expected to increase, decrease, or remain the same between 2013 and 2015.
Regulatory requirements, laws, and recent enforcement trends play a key role in design and use of FDA programs.

**Key Benefits of FDA**
The primary benefit reported from the use of FDA was the ability to “enhance our risk assessment process,” followed by “the ability to detect potential misconduct that we couldn’t detect before.” Referencing the previous discussion on the use of unstructured data, we also see that 62% of respondents included being “able to analyse non-}

*The top-level management of a commercial organisation (be it a board of directors, the owners or any other equivalent body or person) are committed to preventing bribery by persons associated with it. They foster a culture within the organisation in which bribery is never acceptable."

**UK Bribery Act of 2010 Guidance**
structured data formats, alongside structured data formats to identify potential misconduct,” which further demonstrates the need for text mining capabilities.

Who Owns the Anti-Fraud and Anti-Corruption Program?

Thirty-one percent of the companies surveyed pointed to legal and compliance functions as having overall responsibilities for the company’s anti-fraud and anti-corruption program. Twenty-two percent of respondents report the function lying within the internal audit and risk function and 32% within business management and C-suite.

Tone from the top is extremely important; therefore, it is very encouraging to see such high proportion of senior executives taking charge of anti-fraud and anti-corruption compliance.
As a compliance monitoring and investigative asset, FDA is an essential component to a robust and effective anti-fraud and anti-corruption program.

SIDEBAR: The following is a word cloud listing the answer to the question:

Q1. Who has overall responsibility for the company’s anti-fraud and anti-bribery program, in terms of compliance monitoring and investigations?
Both the Team and the Analytic Program Are in Place

Seventy-two percent of the companies we surveyed have an internal investigations group and 74% indicated that they conduct FDA.

Where the differences lie is in the combinations of the two. In the combined approach, 48% indicated that they have both an internal investigations team and also conduct FDA. Twenty-six percent reported conducting only FDA without a formal internal investigations team, while 24% reported having an internal investigations team but no FDA. Very few companies, only 2%, reported not having either an internal investigations team or FDA activities.

While financial services sector have the highest level of both investigative teams and the use of FDA (63%), the life sciences industry boasts the largest adoption of FDA with
81% either using it with internal investigations (46%) or in isolation (35%). FDA adoption is the lowest in the mining industry at 69%.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Mining</th>
<th>Oil &amp; Gas</th>
<th>Power &amp; Utilities</th>
<th>Financial Services</th>
<th>Life Sciences</th>
<th>Other Sectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>We have an internal investigatiosn team, AND ALSO conduct forensic data analytics</td>
<td>46%</td>
<td>41%</td>
<td>33%</td>
<td>45%</td>
<td>53%</td>
<td>46%</td>
<td>46%</td>
</tr>
<tr>
<td>We conduct forensic data analytics, but do not have an internal investigations team</td>
<td>26%</td>
<td>28%</td>
<td>37%</td>
<td>29%</td>
<td>14%</td>
<td>35%</td>
<td>27%</td>
</tr>
<tr>
<td>We have an internal investigations team, but do not conduct forensic data analytics</td>
<td>24%</td>
<td>31%</td>
<td>31%</td>
<td>21%</td>
<td>20%</td>
<td>17%</td>
<td>25%</td>
</tr>
<tr>
<td>None of the above</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>5%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
</tbody>
</table>

**Broad Business Application Across the Enterprise**

FDA is beneficial to a wide range of stakeholders. This includes senior management and board of directors as well as professionals in audit, compliance, and investigation roles. It is not surprising to see the results of FDA being shared among a vast group of business executives. However, it is surprising to see such a high level of senior leadership named as business users or beneficiaries of FDA. Eighty-one percent of respondents listed “corporate executive management” and 69% the “board of directors” as primary users or beneficiaries.
BEST PRACTICES TO LEVERAGE FORENSIC DATA ANALYTICS
IN INVESTIGATION AND COMPLIANCE MONITORING

FDA Efforts Are Aligned with the Key Areas of Risk
The risk of bribery and corruption was identified by our respondents as having the highest priority. Sixty-six percent of the companies indicated that it was of major or some concern. Notably, more respondents named it a “major concern” than any other category.

Asset misappropriation was a close second at 63%, followed by capital projects at 50%.

We were pleased to see that FDA efforts are well aligned with perceived risks. FDA is most often used in high risk areas such as bribery and corruption (75%) and asset misappropriation (76%). However, there is room for improvement with developing more FDA methodologies around monitoring for capital projects, where just over half of the respondents (54%) indicated they were using FDA.

From a sector perspective, the mining and life sciences industries stand out as being the most concerned (major/some/minor concern) with bribery and corruption—83% of mining companies and 71% of life sciences.
Geographically, Brazil, China/HK and the UK are leading the pack with 88%, 85%, and 85% of respondents indicating that bribery and corruption are of major or some concern.

Over half of the respondents in China and Brazil indicate that bribery and corruption are a major concern. On the opposite end of the spectrum, Japan was by far the least concerned about bribery and corruption with only 21% of respondents indicating major or some concern.

Where FDA Is Deployed
Asset misappropriation and bribery and corruption risks were the top areas of focus for FDA, followed by financial statement fraud and capital projects. Survey respondents could check multiple answers. Interestingly, more than 60% of all respondents indicated FDA being deployed in
combination with asset misappropriation, bribery and corruption, and financial statement fraud. This broad application is a strong indicator of robust adoption. Not surprisingly, the mining, life sciences, power and utilities, and oil and gas industries were above the overall average when it came to FDA focusing on bribery and corruption, while the financial services industry clearly dominated the AML response with 81% use.

Perhaps a new area of consideration is the use of FDA in capital projects. In the mining, oil and gas, power and utilities, and life sciences sectors, FDA was utilized more than 50% of the time. It is also worth noting that life sciences and mining led the responses with the use of FDA for mergers and acquisitions, which we feel will be an upward trend in the future.

By functional area, we see Investigations and Internal Audit and Risk functions representing the highest user base of asset misappropriation based FDA uses. Not surprisingly, Legal and Compliance functions stand out as the highest group for use of Bribery and Corruption risks with an 86% response rate, followed by Finance (76%) and Business/Management (77%). We believe this is an opportunity for Internal Audit and Risk to get more involved in anti-bribery and anti-corruption analytics as they fell behind Business/Management with 72% indicating use.

The Business/Management function was the clear leader in FDA use for financial statement fraud with 65%. Similar to bribery and corruption risks, analytics around money laundering primarily fell within the domain of Legal/Compliance at 53% reporting.
The Core Components of a Successful FDA Program

Need to Go Beyond Traditional Rules-Based, “Spreadsheet” Analytics

Seventy-nine percent of our survey respondents reported that they use descriptive analytics for historical look-back data analysis, trend analysis, matching, and sorting. However, when it comes to fraud detection and prevention, these traditional “rules-based” accounting tools are not enough!

Leading innovations in FDA incorporate targeted model-based mining and visual analytic tools that allow the data to “speak for itself.” When deployed over large data sets, they can pinpoint unusual transactions or anomalies derived from the multidimensional attributes within the data.

Unstructured (text-based) data sources can also provide a wealth of analytic insight from evaluating free-text descriptions for suspicious payments such as “incentive payment” or “friend fee” to email communications, indicating who is talking to whom around key business events. This is a key area of analysis when monitoring for bribery and corruption risks.

FDA Maturity Model

With the continued growth of enterprise data and ever-increasing complex business activities, real value and insight is derived when organizations incorporate multiple data analytic techniques, beyond traditional rules-based tests. As the model depicts below, it is important to consider all four quadrants, not just the rules-based structured data analytics (upper left quadrant) as typically common with most spreadsheet and database uses of FDA.
Companies need to integrate all four quadrants for effective
detection and few false positives. This involves using
statistical based tools (upper right) as well as unstructured
tools such as keyword search (bottom left) and data
visualization and text mining tools (bottom right).

Our survey found a striking distinction between companies
that rely solely on spreadsheets and database tools as
compared to the companies that use more sophisticated
 technologies.

**In more analytically mature companies, FDA results are a lot more visible.** Seventy-one percent of respondents list
board members as users or beneficiaries of FDA in more
mature organization as compared to only 58% in less
mature ones. Similarly, executives are listed by 82% versus
76% of respondents, legal and compliance—by 78% versus
64%.

**Respondents are less uncertain about necessary program improvements.** When asked if they agree with
the statement that they “need to do more to improve,” a
higher percentage of survey participants from the mature
organization say “yes” or “no,” as opposed to remain neutral or decline to answer.

**Analytically mature companies are more receptive toward technologies.** For example, 68% of respondents in analytically mature group see potential of Big Data versus 62% in the remaining population.

**Mature organizations see much higher benefits from FDA.** Eighty-three percent of respondents in more analytically advanced organizations believe that FDA leads to earlier detection of misconduct as compared to 70% in less mature organizations. Eighty percent of members in more mature organizations think that FDA allows them to review large amounts of data in short period of time, while only 73% of members in the rest of companies share this belief.

**Analytically mature companies view FDA as more cost-effective:** 64% versus 55% for less mature organizations.

**Effectiveness of anti-fraud and anti-corruption program overall is viewed a lot more positively.** Sixty-three percent of respondents in more mature organizations agree with the statement “our current anti-fraud and anti-bribery program is effective in preventing and detecting fraud and corruption” versus only 48% in less analytically mature companies.

Finally, and perhaps most importantly, **FDA shows better results/recoveries.** Sixty-three percent of respondents in the more mature group said that they “get positive results and/or recoveries from the forensic data analytics tools” as compared to 56% in the less analytically mature group.
## Five Success Factors for FDA Integration

To build a successful FDA program and progress toward integrating all four quadrants in the FDA Maturity Model with high ROI, companies should consider the following five success factors:

1. **Low-Hanging Fruit:** The priority of the initial projects matters.
   - Start with the “low-hanging fruit.” Since the first project normally incurs the largest cost associated with setting up the analytics infrastructure, it is important that the first project yield tangible recoveries or risk mitigations.
     - Example: T&E of sales force or the procure-to-pay process.
   - Pick the first project that has a big demand, where data resides in multiple sources and a compelling, real ROI can be measured.
     - Example: T&E + credit card data, plus rental car data, plus travel agency data to provide “risk score” by employee.
     - Another example: vendor risk ranking by payments + third party watch lists + vendor master/employee master

2. **Go beyond the “rules-based,” descriptive analytics**
   - One of the key goals of FDA is to increase the detection rate of non-compliance fraud and corruption, while reducing the risk of false positives.
   - From a technology perspective, companies need to move beyond rules-based tests and embrace both structured and unstructured data sources considering all four quadrants.

3. **Communicate:** Demonstrate the first success story, then share it widely through the organization.
• Success stories will generate more of a pull versus push strategy for expansion.
• Involve a multidisciplinary team: IT, Business Users, and Functional Experts
• Communicate across multiple departments.
• Seek to increase transparency in your user day-to-day operational data.
• Don’t just audit to report the noncompliance; audit to improve the business.
• Seek investment in incremental chunks, not the entire enterprise at once.

4. Leadership support gets it funded, but heavy end-user involvement makes the program successful.
• Leadership sets the tone and gets the doors open, but people need to adopt it in their daily operations to make it sustainable.
• The team should not develop in a vacuum; every project needs a business champion.
• Insist on process improvement, standardization, and simplicity.
• Invest in automation, not manual refreshes, to make it sustainable.

5. Enterprise-wide deployment takes time; don’t expect overnight adoption
• Analytics integration is more of a journey, not a single destination.
• While quick-hit projects may take four to six weeks, the program and integration will take one to two years, or more.
• Programs need to be refreshed as new risks and business activities change

Companies need to look at a broader set of risks, incorporate more data sources, use better tools (move away from end-user tools), and move to real-time or near-time analysis and increase data volumes. By embracing these
potential areas for improvement, organizations will deliver a more effective and efficient anti-fraud and anti-corruption compliance program that is highly focused on the key fraud risk areas and improved internal audit, compliance, or investigative quality.