DATA ANALYTICS & FRAUD PREVENTION

Increased volumes of business data means corporations should consider leveraging data analytics for fraud prevention and detection, say Greg Mason and Frances McLeod...

/ INTERMEDIATE

We are all acutely aware that the amount of data that an organization uses and stores has been growing exponentially over the past few years. IDC projects that data volumes will reach 40,000 exabytes (40 trillion gigabytes) by 2020[1]. IDC makes an analogy for the lay person that helps us visualize the enormity of that amount, 40,000 exabytes represents more than 57 times the number of all the grains of sand on all the beaches of the earth. This means the digital universe doubles every two years between now and 2020. Much emphasis is being placed on the potential rewards of unlocking valuable information from such data volumes; the so-called promise of 'Big Data'. However, the associated potential risks for corporate bodies that may be buried in these voluminous datasets are clearly growing too.

Big Data accounts for much of this new information, that is, largely untagged file-based and unstructured data, about which little is known. This means not only that large quantities of potentially useful data is getting lost, for example data which could allow a consumer goods company to draw insights on consumer behaviour and generate greater value from marketing campaigns, but that fraud, bribery and corruption, money laundering and other white collar criminal activity may remain either very difficult to detect or, even worse, undetected because it lies buried in the morass of data.

The Association of Certified Fraud Examiners estimates that companies around the world lose $3.5 trillion to fraud annually [2]. However, in addition to the very real commercial implications of the costs of white-collar crime, many of these activities potentially expose corporate bodies and their management to serious compliance risks as well as reputational damage.

A company which does not have sufficiently robust internal security measures, controls and does not test them for efficacy, continuously updating them; is at risk of both internal and external perpetrators leveraging the proliferation of data to their advantage in order to hide fraudulent transactions, improper payments, layering of funds and even "forge" electronic documentation in support of illicit acts. Equally, we find that companies with systems which are either un-integrated or poorly integrated are unable to effectively identify compliance breaches such as the payment of bribes, procurement fraud or money laundering, as they have no means of analysing the complete transaction flow or to identify anomalous or unauthorized payments.

/ THE POWER OF DATA MINING

As forensic accountants and data analysts that leverage the strong IT skills required in order to tackle large amounts of data, we frequently identify anomalous, problematic payments in our clients' financial accounting systems. We also help our clients test the effectiveness of their controls by regularly monitoring high risk payments to ensure that they are appropriate, correctly authorized and supported by a full audit trail.

If we are conducting a review or investigation, we rely heavily on data mining. Data mining combines data analysis techniques with high-end technology for use within a process. We begin by defining the potential problem; what is the allegation or suspicion or specific risk and where is the associated data likely to be held? ···›

NEW EU DATA PROTECTION REGULATION

The Data Protection Regulation was proposed by the European Commission in January 2012, in order, in part, to update the 1995 EU Directive (95/46/EC) to bring it in line with challenges from technological advances and globalization and to strengthen online privacy rights. There are many changes proposed under the Regulation but from a data analytics perspective of note are:

- Change in territorial scope: The Regulation would bring confirmation that EU law would apply to EU citizens' personal data even if they are processed outside of the EU;
- Definitions & Conditions to Consent: The Regulation would introduce that data subjects will have to give explicit consent to any one processing (and that would include analysing) their personal data, the consent must be by affirmative action and fully informed. It must also be as easy to withdraw consent as it is to give it;
- Profiling: The Regulation could introduce new restrictions on the ability to profile and a right to object which the individual must be advised of in a highly visible manner.
- Right to Compensation: The Regulation could introduce the right for an individual to seek compensation for pecuniary and non-pecuniary damage as a result of the unlawful processing of their data;
- Sanctions: The regulation will introduce larger maximum fines of up to 100 million Euros or 5% of a company's annual worldwide turnover, whichever is greater.

"THE ASSOCIATION OF CERTIFIED FRAUD EXAMINERS ESTIMATES THAT COMPANIES AROUND THE WORLD LOSE $3.5 TRILLION TO FRAUD ANNUALLY."

"FRAUD INHERENTLY REQUIRES EFFORTS AT CONCEALMENT, SO DETECTION MAY REQUIRE A DISCREET COLLECTION OF DATA BELONGING TO THE POTENTIAL PERPETRATOR, FOR EXAMPLE WHERE THE CORPORATE SUSPECTS THE THEFT OF COMMERCIALLY SENSITIVE DATA."

/ IS BIG DATA ANALYTICS COMBATING MONEY LAUNDERING?

The United Nations Office on Drugs and Crime has recently stated that $1.6 trillion "dirty" dollars are moving around the global economy and that this number is increasing every year. Some of it is being laundered through online payment systems such as Paypal and Bitcoin. US regulators have recently been focusing on guarding against this money laundering typology. For example, the US Treasury Department ruled in March of this year that firms issuing or exchanging online currency would have the same reporting thresholds ($10,000) under money laundering rules as traditional payment providers (e.g., Western Union). As we have discussed, both fraud and money laundering can sometimes be detected through Big Data searches for suspicious anomalies amongst millions of transactions. Companies have "taken it on themselves to spot fraudulent transactions". In part, because banks and online payment providers usually reimburse victims of fraud they have tended to invest most heavily in fraud detection analytics to help combat that type of abuse. However, according to security experts, the same efforts and expenditure is not going into data analytics systems to combat money laundering. In the opinion of Daniel Weitzner, a former White House deputy chief technology officer, now a researcher at the Massachusetts Institute of Technology, and security expert: "They have reporting obligations over a certain dollar amount and they report suspicious activity. But the understanding is the government will do the analysis to spot money laundering,"[6]

We can then undertake a targeted data collection and enhancement exercise. For example, we may need to pull data from more than one data source such as the financial accounting system, project management database, time and expense systems, or from disparate systems that perform the same function but at different locations (E.g. where accounting systems are not integrated across business units, which is frequently the case for companies that have grown rapidly through acquisition).

We then normalize/clean the data and where necessary enhance it, for example exclude irrelevant data fields or supplement empty fields from other sources, as necessary. We then begin to run queries, which we may need to test, refine and validate based on our knowledge of typologies, country specific practices, the nature of the business, etc. Finally, we begin to analyse the results by generating reports, using visualization tools, etc. This allows us to review relevant supporting documentation (e.g. purchase orders, authorizations, etc.), conduct interviews (as appropriate) and "fill in the gaps" to better understand what went wrong and why. We may also place the conduct in the context of the companies' own procedures. Once investigations are complete, we are frequently in a position to make recommendations as to how to tighten controls, improve processes and enhance record keeping in order to prevent future abuses and to allow the company to test the effectiveness of its systems and compliance with its programs and regulatory obligations.

/ PROACTIVE RISK MANAGEMENT

With enforcement efforts strong, particularly in the US, and the long reach of US jurisdiction (e.g. as at December 2013 nine of the top ten FCPA settlements involve non-US companies [3]), companies and their boards are becoming more sensitized to the risks of fraud, bribery and corruption, and other categories of white-collar crime. As a result they are committing more time and resources to combating these risks. We

/ WHAT IS BIG DATA?

Big Data describes data sets that are too large, too unstructured or too fast changing to allow for analysis using traditional relational or multidimensional databases. Analysing Big Data can require dozens, hundreds or even thousands of servers running parallel software. What truly distinguishes Big Data, however, aside from its volume, velocity (i.e., how quickly it is changing) and variety is the potential to analyse it in order to make more informed decisions. In the realm of security and risk management, corporate governance, and fraud detection and prevention, we expect in coming years to see the integration of Big Data analytics into risk management and security operations. In fact, we already see the big security and compliance software companies developing tools targeting the cyber security and fraud prevention space, particularly in insurance and banking. We also expect to see increased demand for analysts who combine IT and data science skills with great understanding of commercial and cyber risks, as these analysts are in high demand and difficult to find, we expect to see corporate bodies buying in expertise in consulting form and or partnering with third parties such as business intelligence software providers to develop industry specific tools.

see this trend being reflected in the types of mandates we as a company are winning. Namely, we are increasingly being asked to provide proactive advice from a detection, prevention and compliance perspective – and to test existing controls and or leave behind protocols for on-going, repetitive testing in order to identify and shut down improper activity – for example we can design continuous testing which focuses on narrow bands or transactions or areas which pose particularly strong risks for that particular client.

We find that in a prophylactic context, too, data mining followed by in-depth transactional analysis (from either integrated or disparate systems) can frequently identify anomalies, outliers and suspicious patterns that warrant further investigation or highlight an avenue of potential abuse that needs to be closed off by enhancing financial and IT controls, introducing a change in policy or other remedial action, such as re-education and or training.

A variety of data sources, structured and unstructured, can be mined to yield potentially useful information from a fraud detection and prevention perspective. Obviously, much may be contained in accounting systems, but, depending on the nature of the issue (asset misappropriation, bribery, financial statement fraud, etc.), other systems can be fertile sources of information that can be used in detection and prevention.

For example, we may look at the payroll data and conduct data matching to see if there are deposits from two different people going into the same bank account. This may be an indicator of fraud where employment activity has been falsified and one employee was receiving several pay checks. In such an instance we would also look into HR systems to look at names, addresses, phone numbers, serial numbers, etc. We have also looked at data matching in money laundering and terror financing contexts. Equally payroll and attendance records might raise red flags if an employee has never taken a vacation (rogue traders, for example, tend not to take time off for fear of having unauthorized positions discovered) or conversely filed claims for expenses for a period when they were on vacation.

In external fraud investigations we may look at utilizing the "sounds like" function (e.g., SOUNDEX) to help identify variations of valid company employee names that may have been high jacked by a fraudster who understands that many internal controls will only be looking for exact name matches. We also use fuzzy matching in FCPA investigations or compliance reviews, where we are looking for illicit payments to government officials of relatives of government officials and there may be name variations or transliteration issues.

Within the financial accounting systems we will also test for duplicates both simple and complex, which can suggest not only fraud by inadvertent double paying due to inefficiency or error. We look for gaps in sequential data that might indicate an attempt to circumvent or abuse the system, for example, a gap in sequential purchase orders every so often. This might indicate that an employee is attempting to take a purchase order "off ledger" and introduce into the system later for pay back.

/ CASE STUDY

DATA FORENSICS MEETS FORENSIC ACCOUNTING – POWERFUL TOOLS IN CONDUCTING A CORRUPTION INVESTIGATION

Our data forensics and analytics team were engaged by a multi-national client with suspicions about potentially corrupt payments to a third party and therefore possible liability under anti-bribery and corruption legislation (FCPA and UK Bribery Act). The level of potential risk was high, so our data forensics experts were under time pressure to confirm or refute the suspicions.

We received a hard drive with an image of a desktop but no additional information about the nature of the data on the drive. We knew that the key to information about the payments would be to locate and analyse the relevant financial transactions, however, the drive contained thousands of Gigabytes of data, none of which was easily identifiable or classifiable. Our experts conducted exhaustive searches to identify file types that might contain relevant data types and found hundreds of spreadsheets located in a directory on the image, but they turned out to be only summaries of transactions and we needed to find a complete set of transactions at a more granular level. We went back to the image and found several hundred Microsoft Access databases all of which were password protected. We used specialist software tools to identify possible passphrases by looking at and indexing the image and thus cracked into the databases. A review of the databases identified links to a SQL database but the links were broken. In order to reconstitute the links, our experts went back to the image and pulled 47 SQL databases and restored them all. Analysis revealed that fifty percent of the databases had the information we were looking for but for various time periods and much of it overlapped. We therefore consolidated the information over the relevant time period, removing duplicate transactions, into a single SQL database. Our experts ran queries across the database and were able to identify one hundred percent of the relevant transactions including amounts, dates, recipient, bank account, etc. We generated a report that allowed our client to confirm their suspicions with absolute certainty and take remedial action quickly.

/ BIG DATA, CLOUD AND DATA PRIVACY

In the ever-increasing need to manage and host vast quantities of data, cloud services provide a viable and cost effective alternative to institutions. Many cloud providers offer a fast, fully managed, petabytes scale data warehouse for a reasonable fee that provides a platform to host and analyse Big Data.

However, using the cloud to manage and store data is not without problems, especially in the context of data privacy. Some key points to contemplate when considering the need for a cloud provider:

- Data is subject to the laws of the jurisdiction in which it is stored.
- Private data stored in the US is at a significantly higher risk to be accessed by government agencies. Formal requests in the form of subpoenas and warrants generally compel the provision of information.
- Private data in the US is protected by the fourth amendment (where there is a reasonable expectation of privacy) – however once information is shared with a third party (i.e., cloud provider) the expectation of privacy could potentially be forfeited.

/ REPETITIVE ANALYSIS FOR ANTI-CORRUPTION & FRAUD MONITORING

Bad actors can and will exploit weaknesses in controls wherever they can find them. We have leveraged data analysis in a wide range of fraud detection and anti-corruption compliance audits and have found it an invaluable resource. We frequently detect inappropriate or illegitimate payments made on corporate credit cards or claimed for in employee expenses. We have even found an instance of an employee claiming for reimbursement of $75,000 of taxidermy charges (obscuring a bribe). Increasingly, we find repetitive analysis of particular value where the area under scrutiny is identified as high risk for fraud or corruption; where we are reviewing very large volumes of transactions regardless of amount, and where transactions are on-going over a period of time. From a compliance testing perspective too, on-going monitoring is an essential component, repetitive testing allows for patterns to be identified over periods of time, for benchmarking, and, from a prevention perspective. If staff knows on-going fraud and anticorruption analytics are being undertaken, this can prove an effective deterrent.

In addition ERM systems, materials management and inventory control databases, file systems and project management systems can be fertile data sources in bribery and corruption and fraud detection, as costs can be misallocated or bribes (and kickbacks) accounted for. This means that at management reporting time, irregular payments are often consolidated or miscategorised into low risk cost categories to avoid raising red flags.

The more data sources can be mined, analysed and tested, the more likely anomalies will be detected and/or alerts raised. Further, we find that companies that take a holistic approach and are able to integrate and consistently test and analyse data from systems across geography, business unit, etc. can identify schemes (even if they are cross-border) more quickly and effectively. We have clients who, for example, from an ethics and FCPA compliance perspective keep globally integrated gifts and entertainment data base and review these expenses across business units, locations, etc. to ensure that individual clients are not collectively the recipients of lavish gifts and entertainment.

/ THE POWER OF FORENSICS

Data forensics is also part of our arsenal in the detection of illicit activity. Fraud inherently requires efforts at concealment, so detection may require a discreet collection of data belonging to the potential perpetrator, for example where the corporate suspects the theft of commercially sensitive data, collusion or misappropriation of funds. Frequently, when we find a problematic transaction or series of transactions we will leverage our data forensics skills to identify the relevant sources of data, such as email or instant messaging, in support of the investigation.

Performing forensic analysis on a device is often critical to identify information buried in various forms of electronically stored information (ESI), such as emails, spreadsheets or word documents and files. In many cases, the relevant information (such as an accounting database) may have been deleted. Digital forensics can help to recover that information. Further, it may allow for the identification of accomplices who may not have been specifically linked to the transactions in question.

Banks and insurance companies suffer significant exposure to fraud. The Coalition against Insurance Fraud estimates that approximately $80 billion of fraudulent claims are made in the US annually [4]. The Association of Certified Fraud Examiners estimated the cost of credit card fraud at $5.5 billion[5]. Banks are increasing the victims of attacks launched from the Cloud. For example, when a fraudster is looking to set up a phishing scheme to gain access to victims' bank accounts, the built-in redundancy, scalability and automation capabilities of cloud servers are very effective. In fact, all it takes to procure cloud services is a working credit card; without ever needing to deal with a live salesperson, the cloud becomes an even more viable base from which criminals can commit fraud. Only recently, JPMorgan Chase & Co had to warn some 465,000 holders of prepaid cash cards issued by the bank that hackers who attacked its network in July may have accessed personal information. As a result, these at-risk industry sectors are increasingly testing Big Data fraud prevention and detection software that is designed to detect anomalous behaviour that might signal attempted cyber-attack and or fraud.

Financial and insurance companies are also increasingly looking to aggregate data from multiple sources, including mobile data, along with social media from Facebook to Twitter, to profile customers' behaviour and to detect fraud as well as to minimize disruption to their customer's banking activity. Solutions include anomaly detection programs that monitor the online and mobile banking activity of each account holder from login to logout, and use it to distinguish fraudulent activity from established legitimate patterns.

Some banks are also using another innovative approach; using social data to 'cluster' information. For example, if the cardholder's social media contacts are tagged as co-workers and their social profiles indicate their current geographical location, a bank can use this information again to correlate recent charges made by the cardholder.

We fully acknowledge the possibilities of the high performance Big Data analytics in this context, but we do not see it as a panacea. We also recognize

REFERENCES
1. IDC Digital Universe Study, Sponsored by EMC, December 2012
2. http://www.acfe.com/uploadedFiles/ACFE_Website/Content/2012-report-to-nations.pdf
3. http://www.fcpablog.com/blog/tag/top-ten
4. http://www.insurancefraud.org/the-impact-of-insurance-fraud.htm
5. http://www.acfe.com/uploadedFiles/ACFE_Website/Content/rttn/2012-report-to-nations.pdf
6. http://blogs.wsj.com/cio/2013/06/03/anti-money-laundering-tools-lag/
7. http://www.acfe.com/uploadedFiles/ACFE_Website/Content/rttn/2012-report-to-nations.pdf

that many companies are struggling with how and where to start using it for fraud and other white-collar crime detection. Ultimately, the quality of analytics driven by software is impacted by the quality of the data itself and as humans create data, it is subject to human error (typos, confirmation bias, etc.).

So in conclusion we would recommend the following best practices for those thinking of taking a foray into Big Data fraud prevention analytics:

1. We recommend starting with small and specific uses for big data. For example, identify one or two business problems or risk areas that can be resolved by improving fraud detection, and create a task force comprised of compliance, legal, IT and business unit representatives to devise an outcome based plan. Get senior level buy in; this is part of a risk management strategy.

2. It is critical to ensure that the company is working with high quality data. Impress upon your team (and external consultants, as necessary) the need to ensure the proper data is being collected and the signal is separated from the noise to allow for effective, meaningful data analysis.

3. Assess and plan for the relevant regulatory environment. It is critical to understand the boundaries for using customer data and the relevant privacy laws. Obtain legal and other expert advice on this area and acknowledge the added complexity for companies that operate globally, particularly in Europe. This will, in our opinion be a, if not the, major challenge.

4. Finally, we would stress that in our opinion this technology will be most effective if used by experienced forensic accountants, data analysts and both internal and external compliance and security professionals as part of a strategic approach to this aspect of risk management. In our experience over reliance on IT solutions without sufficient human expertise and analysis may lead to at best too many false positives (and a diversion of valuable resource to resolve them) and at worst missed inappropriate activity.

AUTHOR BIOGRAPHY

Greg Mason is a partner and one of the co-founders of FRA. His expertise lies in database architecture and programming, software design, mass data analysis and data mining for the purposes of investigations, disputes and litigation. Greg has recently worked on a high-profile FCPA matter where he analysed a global oil services company's internal financial database, comprising over 21 million transactions made in over 25 countries, for presentation to SEC investigators. He has also developed a tailored database and eDiscovery platform to review documents and capture electronic financial information for the forensic audit of monies related to over 500 bank accounts in an investigation of bribery allegations in connection with a Central Asian government's privatisation of its national oil company.

AUTHOR BIOGRAPHY

Frances McLeod is one of the co-founders of FRA and the Managing Partner. She advises diverse clients on anti-corruption (FCPA/OECD/Bribery Act) issues in terms of response to internal and external investigations, in a compliance context, and in related civil and criminal litigation, in a variety of jurisdictions. She also advises a number of European clients on data protection and privacy-related matters in the context of US-driven eDiscovery requests, with an emphasis on providing practical solutions that balance potential conflicts of law. Frances has been involved in advising companies responding to Oil-for-Food investigations, contributing to the formulation of strategy, presenting complex financial analysis and data, electronic and paper discovery, conducting a comprehensive evaluation of the database system which underpinned the whole Oil-for-Food Program, and preparing witnesses for interview.

"A COMPANY WHICH DOES NOT HAVE SUFFICIENTLY ROBUST INTERNAL CONTROLS AND SECURITY MEASURES IS AT RISK OF BOTH INTERNAL AND EXTERNAL PERPETRATORS LEVERAGING THE PROLIFERATION OF DATA TO THEIR ADVANTAGE."