Canada Should Take a Leading Position Integrating Big Data and Advanced Analytics into Key Control Processes

Author
Brian O'Donnell, Executive-in-Residence, Global Risk Institute

Foreword

Over the past 18 months the Global Risk Institute has dedicated significant focus to both emerging regulatory reforms and emerging technology innovation. Last July our seminar on Big Data and Advanced Analytics emphasized the emergence of Fintech in assisting established firms in the financial industry to innovate existing processes. And at the GRI conference last fall we emphasized the role technology is playing in transforming the financial services industry; indeed, Lowell Bryan's key note address highlighted unique opportunities for the Canadian Banks to utilize big data and advanced analytic solutions in order to leap ahead of other jurisdictions.

With this article we are kicking off a focus on the usage of Big Data and Advanced Analytics in the area of anti-money-laundering (AML). AML has been a particularly difficult solution for global banks as the evolving regulatory standards call for banks to be able to readily monitor all transactions across the firm, which requires an in-depth knowledge of their clients and their clients' counterparties (and often times the correspondent banks). This requirement lays bare a significant challenge facing most banks – with hundreds of systems and data bases, each with unique data models, unique customer identification protocols and differing data quality standards, AML monitoring and case work systems are overwhelmed. Indeed with an ever rising number of false positive alerts being triggered by such systems, banks are hiring thousands of case workers to manually resolve and close such alerts to the regulators satisfaction.

The GRI believes that finance technology, and specifically big data and advanced analytics, can be used to augment existing AML systems and significantly enhance both the effectiveness and efficiency of AML processes. By resolving client identities across the firm and using machine learning algorithms to monitor "unacceptable transactions," advanced analytics can significantly reduce the number of "false positives," and resolve those that do arise much more efficiently; by focusing on truly concerning patterns, advanced analytics is also more likely to identify truly concerning transactions.

Attached below is a white paper written by Abhi Mehta and Eliud Polanco of Tresata. They outline approaches that have had success with in helping large global banks drive greater efficiency and effectiveness in their AML process.

We plan to follow up this article with a seminar this fall, where we hope to bring the regulators and the Banks' AML Executives together to discuss how Canada can embrace innovative technologies and evolve a regulatory "innovation sandbox"; the goal is to drive greater efficiency and effectiveness across AML systems and processes, and to have Canada take a leadership role in integrating Big Data and Advanced Analytics into key control processes.

Breaking Bad Data & Solving for AML

Authors
Abhishek Mehta & Eliud Polanco

Mr. Abhishek Mehta and Mr. Eliud Polanco are independent contributors to the Global Risk Institute. They are solely responsible for the content of the article. Mr. Mehta and Mr. Polanco's biographies are at the end of the article.

"Here's to clean cars...and clean money."

-WALTER ("HEISENBERG") WHITE, BREAKING BAD, 2011

In season 4 of the hit television series Breaking Bad, the protagonist, Walter White, buys the A1A Car Wash as a front to hide the proceeds of his expanding drug empire. Using the structuring method to place "car wash" earnings into the financial system, Walter winds up laundering over $80 million in cash in one year!

Breaking Bad didn't just make for 'must watch' TV. To us, as financial industry participants, it also served as a stark reminder of the worst kept secret in banking: bad actors, when properly motivated, can circumvent controls in the financial system with ridiculous ease!

The effectiveness of anti-money laundering (AML) controls and regulatory compliance programs are one of the most significant issues facing the Financial Services industry today. In recent years, many countries have begun holding senior financial institution executives personally and criminally liable for failure to meet AML objectives. For many financial institutions, investments in AML have become the source of the single largest year-over-year cost increases in fiscal budgets.

And yet, despite the investment, bad actors are winning; thus, the total number of regulatory enforcement actions against financial institutions and the penalties being assessed continue to grow every year. One of the questions we look at as a predictive analytics company is how to resolve the following paradox:

"What's causing the diminishing returns on AML investments and is there a better way to find and combat fraud?"

What we have learned is that the problem is not as dire as it seems. **And data – bad data – seems to be at the heart of most, if not all, of the core issues.**

Tresata has been engaged with several financial institutions of varied size and scale that are deeply interested in breaking this impasse.

In this paper, we have summarized our findings and shared our perspective on possible solutions based on successful software implementations. We collapse them into the following three areas:

1. **Common issues in AML programs** that are impeding progress on transaction monitoring effectiveness,
2. **The root causes driving those issues**, and
3. **The three powerful innovations in technology –** bottomless storage / computing capacity, probabilistic data integration, and machine learning – that are powering a new and promising approach for detecting and mitigating AML fraud.

We also share with you how our most popular product, **the 24-Day Data Diagnostic**, is a powerful tool to help you either design an effective AML program or successfully execute an AML modernization/transformation strategy.

THE AML PROBLEM AT A GLANCE

- It is estimated that fraud and money laundering makes up **~4%** of all global transactions, nearly **$2 trillion per year**.
- Investments in AML technology and operations worldwide have doubled between 2012 to 2016, **up to $8 billion**.
- Since 2012, there have been **over 40 regulatory enforcement actions** assessed on Financial institutions and Money Servicing Businesses.
- Financial Institutions and Money Servicing Businesses have paid **$5.3 billion in financial penalties** since 2012.

THE CURRENT AML DETECTION PROBLEM

In order to comply with the Bank Secrecy Act, the US Patriot Act, and other AML regulations, financial institutions must put in place monitoring systems that:

1. **Analyze transactions**
2. **Assess risk of potentially fraudulent behavior and**
3. **Generate risk alerts and aggregate them into cases.**

In addition, the financial institution's AML Operations team must review each case, make a determination as to whether the behavior is really fraudulent, and if so, issue Suspicious Activity Reports (SARs) to the Country's Financial Crime Enforcement unit.

For most financial institutions, the cost of executing the required activities has grown exponentially over the last five years, inhibiting their ability to invest in growth. While the financial impact may not be as extreme in some of the mid-market or smaller financial institutions we have observed, inefficiencies in AML systems and operations still threaten to keep regulatory risk levels above a comfortable threshold.

From our extensive work in this area,, the most significant driver behind inefficient AML performance is the lack of precision in current AML technologies.

This problem manifests itself in two ways:

1. **Generating an excessive number of false positives.**

   AML detection tools tend to generate significant numbers of false positives -- alerts for behaviors that are not actually fraudulent. In larger financial service institutions, **the false positive rates can fall within the range of 85% to 92% of all alerts generated**. Each false positive requires an AML analyst to conduct research activities and provide documentary evidence as to why the activity alerted is not fraudulent. Depending on the type and complexity of the alerts, it can take and even exceed a full man-hour to review and make a disposition for a single case.

2. **Inability to comprehensively detect true positives.**

   Even with robust monitoring and alert generation, AML systems still miss many true positives. According to the 2016 Pricewaterhouse Coopers Global Economic Crime Survey, only 50% of money laundering or terrorist financing incidents were detected by system alerts. Each time a financial institution finds a true positive that was not detected by the transaction monitoring system, the Firm's Compliance organization must (i) file a SAR and (ii) go back, tune the detection scenario to pick up the missed behavior, and retroactively look back at 6 months' to one-year's worth of historical transactions to see if there were any other missed incidents that match the same behavioral profile.

These "Lookbacks" can be tremendously expensive as they can create thousands of new alerts that have to be sent to Operations teams, which are already struggling to keep up with the volume of alerts that are coming in every month. The end result? More resources are needed to handle spikes in alert disposition just to keep pace.

Improving the detection precision of AML monitoring tools is critical. Unfortunately, AML Transaction Monitoring Systems (TMS) tend to be some of the most complex business applications that are deployed within financial institutions. Each system can have thousands of encoded business rules, detection scenarios, and complex probabilistic models all striving to find fraud needles in transaction system haystacks. And yet these solutions are still falling short, particularly in some of the largest financial instituions.

So, the logical question is: what is impeding AML monitoring tools from being as effective as possible?

WHY AML DETECTION IS SO DIFFICULT

Tresata's research, based on multiple implementations of our AML solution in financial institutions of various sizes, indicate three common issues that hamper the detection precision of AML systems:

1. **The quality of data being fed into AML transaction monitoring systems is extremely poor.**
2. **Tuning AML detection models to achieve optimum precision is more art than science.**
3. **Lack of "institutional memory" in alert disposition.**

ISSUE #1: THE QUALITY OF DATA BEING FED INTO AML TRANSACTION MONITORING SYSTEMS IS EXTREMELY POOR

The effectiveness of current AML transaction monitoring tools is dependent on the quality of the information about the counterparties and involved financial institutions in a transaction.

The more the transaction monitoring systems know about the originating entity (who initiated the transaction), the financial entities (which financial institutions were involved in processing the transaction), and the beneficiary entity (who is on the receiving end of the transaction), the better the tool's detection scenarios can work.

However, the quality of the counterparty and transaction chain of custody data being fed into AML transaction monitoring tools can be extremely poor.

We break down the causes of poor quality data into three categories:

1. The Complexity Problem
2. The Data Processing Problem
3. The "Not My Customer" Problem

1. THE COMPLEXITY PROBLEM

In most financial institutions, irrespective of size and scale, one of the main drivers behind the poor quality of information is the difficulty in complying with Know Your Customer (KYC) regulations. KYC regulations may seem simple and common sense – a financial institution should know who their clients are, what kind of business they are in, how they make their money, how they are transacting with the firm and how they are connected to other industries, buyers, and suppliers.

In theory, KYC information is current and is being fed into AML transaction systems. In practice, however, because of the variety of disaggregated processes in which customer lifecycle information can be collected and maintained, accurate KYC data is rarely collected and integrated into a universal KYC database. In many cases, even connecting the dots between entities with slight name variations can be challenging and very expensive to fix.

2. THE DATA PROCESSING PROBLEM

For mid-market or smaller sized financial institutions, the scale and complexity of curating quality customer and counterparty data may not be as extreme, but they share with large firms some of the same data processing technology challenges that are preventing quality data from making its way into AML Transaction Monitoring Systems. **The central obstacle is with a process called enrichment.**

Enrichment is the process by which contextual information about a transaction is physcially added to the transaction record before it reaches the AML Transaction Monitoring Systems. The slightest variation in data quality (missing attribute, added comma, misspelled name) reduces the effectiveness of the enrichment process, causing transaction records to lack context as they make their way into the TMS. This is illustrated in the insert on page 6.

3. THE "NOT MY CUSTOMER" PROBLEM

To complicate matters further, some of the data quality and integration challenges described thus far have only focused on the counterparties that are a financial institution's customer or client.

Financial Institutions that offer correspondent banking services have an even greater challenge of understanding the behaviors and risks of all the actors in the chain, including other originating, intermediary, and beneficiary banks involved in the transaction as well as any other companies that are conducting business with a financial insitution's customer but who are themselves not subject to their customer due diligence regime.

These "non-customer" counterparties - the buyers, affiliates, and other partners of a financial institition's customers are typically referred to as "pseudo-customers" or "customer's customers." Connecting the dots and compiling accurate data about the other financial institutions or pseudo-customers is even more complex and challenging than customers, leading to low quality information about those entities making its way into AML transaction monitoring tools.

ISSUE #2: TUNING AML DETECTION MODELS TO ACHIEVE OPTIMUM PRECISION IS MORE ART THAN SCIENCE

Even with proper counterparty and transaction chain of custody information, the degree of flexibility and configurability of AML transaction monitoring solutions can lead to a wide range of detection effectiveness. Most AML transaction monitoring tools work as follows:

1. SCENARIOS & RULES ARE DEFINED

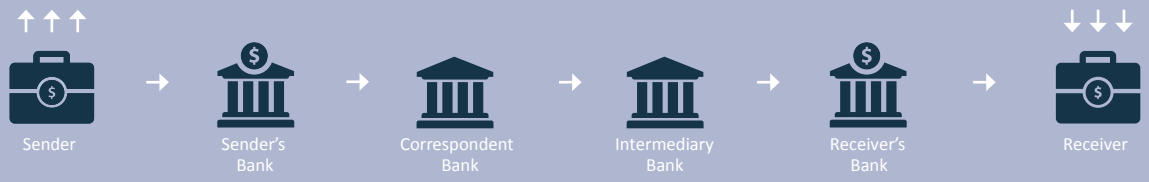
An observed fraudulent behavior is converted into what is called a "detection scenario," - a series of business rules or tests that can be run against each transaction. The variety of the types of transactions handled by a financial institution often determines how many scenarios and rules are required.

For example, either the threshold is fixed by an absolute rule (e.g., all cash withdrawals greater than $10,000 must trigger an alert), OR be relative to the type or size of the customer (e.g., cash withdrawals for high net worth individuals greater than $15,000 must trigger an alert, but for low net worth individuals the threshold is $5,000).

IMPACT OF POOR DATA QUALITY IN AML TRANSACTION MONITORING

What do you know about a customer and the transaction chain of custody is very different from what your Transaction Monitoring System (TMS) is seeing.

FINANCIAL TRANSACTION CHAIN



BREADTH OF AVAILABLE DATA FOR ENTITIES IN CHAIN

Sender
- Accountholder Name
- Location
- Type of Business
- Type & Purpose of Accounts
- Buyers & Suppliers
- Volumes
- KYC Risk

Sender's & Receiver's Bank
- Bank Name
- Bank Identification Code
- Account Number
- Branch Location
- Industries & Volumes they do business with

Receiver
- Accountholder Name
- Location
- Type of Business
- Type & Purpose of Accounts
- Buyers & Suppliers
- Volumes
- KYC Risk

DATA TRANSACTION MONITORING SYSTEMS ARE ABLE TO USE

Sender
- Accountholder Name
- Account Number

Bank Information
- Bank Name
- Bank Identification Code
- Transaction Date
- Transaction Amount

Receiver
- Accountholder Name
- Account Number

LOSS OF DATA...

Payment messages come with the very basic data elements required to execute a transaction, as depicted. TMS requires much more contextual information about the parties involved in order to assess the likelihood of fraud. This information typically comes from KYC or Customer Master Data systems. And it typically doesn't make its way into the TMS.

...COUPLED WITH LACK OF ENRICHMENT

In order to process contextual information, transactions need to go through an enrichment. Currently enrichment involves extracting, transforming, and loading (ETL) data from multiple sources into large relational databases or data warehouses. But the data in source systems may be missing, inconsistent, duplicative or have a variety of typographical erros from manual entry, making it difficult to integrate using traditional ETL techniques.

...RESULTS IN MASSIVE FALSE POSITIVES

Without enrichment, most of the transaction monitoring models miss bad actors or generate false positive alerts. This increases the need for humans to manually look up the source systems to connect the dots, increasing the time and cost of executing AML transaction monitoring. Any new approach to AML fraud detection must account for data integration issues inherent in using poor quality data.

2. THRESHOLDS ARE CUSTOMER "SEGMENT" DEPENDENT

Developing relative thresholds depends on:

1. Being able to group the actor in the transaction into a pool or demographic segment of entities most like it (e.g. define what it means for an individual to be high or low net worth),

2. Identify the range of behaviors within the segment, and

3. Set the threshold to the limit of what is most common or normal for that segment (so that outlier behavior can be flagged as an alert).

Each of the activities around defining relative thresholds for a detection scenario is heavily dependent on human judgment; thus, there is a lot of variability in defining what the segments are and which customers get grouped into which segment.

Because the factors in segmenting customers and setting thresholds are so sensitive, we have seen from our experience that tuning detection scenarios to achieve optimum precision is more art than science. If one variable is changed slightly in one direction, it is possible to open the flood-gates and generate too many alerts from normal activity. If the same variable is changed slightly in the other direction, then this may lead to the valve closing too tight – too many true positives will not be caught.

The process of setting the dials in the detection models just so, such that they can produce the optimal alerts, is typically only as good as the quant analysts doing the work. An experienced analyst with the right level of intuition can account for inconsistencies in the data and optimally tune the scenarios, segmentations and thresholds. A poor quant can create tremendous noise leading to more false positives and higher review and disposition costs.

ISSUE #3: LACK OF "INSTITUTIONAL MEMORY" IN ALERT DISPOSITION

Finally, there is a lack of "institutional memory" in alert disposition. As discussed earlier, each alert initiates a workflow of operational processes for resolution. The resolution workflow tends to follow the procedure below:

1. A Middle Office AML analyst conducts third-party or external research from sites like Google or Lexis/Nexis on the identities of counterparties in the alert.

2. The analyst also manually looks up the same involved counterparties in existing internal systems (like various KYC databases) to understand current products and expected or anticipated behavior.

3. The analyst applies best practice knowledge on whether the transaction makes sense given the fraud scenario that triggered the alert.

4. Finally, the analyst documents all of their research and evidence into a written rationale explaining why they chose to either close the alert as a false positive or escalate it into a SAR.

Based on how alerts and cases get assigned to investigators, it is very common for counterparty research and analysis to be conducted over and over again, as if it has never been researched before. And, it is possible for the same or very similar type of behavior to be alerted each month, assigned to two different analysts and dispositioned differently (meaning one analyst may choose to close it as a false positive while the same behavior might get escalated by a different analyst the next month).

SOLVING FOR THE 'ANTI' - FIXING MONEY LAUNDERING

Since the early 2000's, especially the last decade, we have seen an explosion in new capabilities to manage and monetize big data. The step change improvements in artificial intelligence applications, foreign language translation, driverless cars, speech recognition and other amazing innovations are powered by a new class of data processing technologies that typically get classified under the umbrella term "Big Data Analytics." We believe that this new class of technology offers a way out of the vicious cycle that is driving inefficiencies in AML performance, regardless of the size or scale of the problem today.

The technology behind Big Data Analytics is unique and game-changing in three ways:

1. **Unlimited storage / computing capacity:** It leverages commodity, very low-cost hardware including CPU processors, memory, and storage. The low unit cost and ease of pooling together more and more servers into an integrated computing environment enables practically limitless storage and computational power.

2. **Probabilistic data integration:** It is designed to deal with chaotic data by enabling the connection of data from many sources based on the likelihood that the information is related, even if there are no system keys that have been designed to link them beforehand.

3. **Machine Learning:** It uses predictive models that improve in accuracy the more iterative user feedback and inputs are provided.

These three powerful forces create exciting opportunities for a new approach to tackling fraud, whether it is anti-money laundering fraud, credit card transaction fraud or even health insurance fraud.

Tresata has pioneered the use of the aforementioned technology forces to address the common issues that are preventing maximum effectiveness of AML detection processes:

1. Improve the quality, and quantity, of data being fed into AML transaction monitoring systems

2. Refine AML segmentation to achieve optimum precision, using probabilistic techniques, and

3. Create "institutional memory" in alert systems, leveraging advanced machine learning techniques.

FIX #1: IMPROVE THE QUALITY OF DATA BEING FED INTO AML TRANSACTION MONITORING SYSTEMS

The difficulty in creating and maintaining accurate and trusted KYC and pseudo-customer data is, in fact, an at-scale data integration challenge. All of the core information required to know and understand a financial institution's customers, counterparties and dependent financial parties is distributed among tens or hundreds of product processors, local KYC databases, reference data systems, etc. - it would be impossible for financial institutions to complete transactions, maintain balances or collect on loans without some accurate data somewhere in the ecosystem.

Tresata's approach to AML introduces the opportunity of having largely machine-driven processes make sense out of data. "Smartbots" are deployed to cycle through a pool of all of the transaction, KYC and reference data that can be collected at an organization, sift through and look for connections, weed out inconsistencies and conflicting information, and effectively organize information around every unique counterparty that is found (customers, pseudo-customers and involved financial institutions). Human subject expertise provides the last mile, providing feedback and business judgment to continuously improve the quality of the machine-driven data integration process.

The end result is the creation of a Data Asset which fully integrates an entity's demographic information, using inputs from many different KYC and referential data sources, with their behaviors, based on connecting every transaction that an entity has engaged in across any product processor. This data asset can be extracted from the Big Data pool and fed as an input into AML transaction monitoring tools which should immediately improve the precision of fraud detection.

FIX #2: REFINE AML SEGMENTATION TO ACHIEVE OPTIMUM PRECISION

AML scenarios depend on the quality of the counterparty segmentation used to define the thresholds. This creates a big challenge - if segments are too broad or narrow, it will impact the ability of the transaction monitoring system to precisely determine whether a specific activity for a customer is suspicious or not.

Ideally, a better approach would be to not segment at all.

Financial institutions have years of transaction data in their archives. Rather than looking at a broader population of people, if a financial institution wanted to determine if a transaction a Customer recently made is suspicious, it should look at all of the Customer's history, assess what is 'normal' for that Customer and then rate how the current transaction compares to that behavioral profile. The thresholds for alert generation would be specific to each and every customer, as well as to every other unique entity within their role or context, whether they are the originators, beneficiaries or involved parties in a transaction.

Tresata calls this approach a "Segment of One," where products, pricing and offers are highly personalized at an individual customer level based on having a deep and rich understanding of how the customer has historically behaved.

Because of Tresata's ability to deliver very low cost super-compute capability and power with machine learning algorithms, Tresata's Segment of One model can now be practically applied to AML fraud detection.

THE POWER OF ONE

How Segment of One impacts on existing segmentation approaches.

Segment of One refers to the approach where a single entity is observed based on its historical behavior, as compared to aggregating in pools of similar entities. Segment of One enables:

Behavior based fingerprinting.

The behavioral fingerprint that is unique to each counterparty can be made up of an endless number of dimensions and factors, making for a very precise profile. For example, it is absolutely possible to discover through behavioral fingerprinting that a customer makes deposits typically on Sundays during the third week of a month in branch locations within a 20-mile radius of a certain zip code. Further, it can be ascertained that those specific features of that profile make that customer distinct compared to anyone else. (Another customer may be differentiated through different thresholds or entirely different features.)

Adaptive modeling.

Our model can intelligently adapt over time as customers, pseudo-customers and the dependent financial institutions change their behavior profiles. Rather than being driven by a static set of business rules that apply broadly across customer segments, it is possible to have machines continuously adapt the features and thresholds for each individual entity as changes to behavioral patterns are detected and reviewed. As a result, it is possible to test each transaction against a range of behaviors that are specific to each counterparty, enabling a more precise identification and isolation of non-normal behavior. The more historical transactional data is thrown into the Big Data pool, the more precise the fingerprint and ability to create thresholds to identify outlier behaviors.

The emerging best practice approach is to apply BOTH: fraud detection scenarios and business rules driven by known money laundering or terrorist financing behaviors AND with Segment of One techniques for identifying non-normal behavior at the individual counterparty level. Together, they offer a powerful vehicle for truly finding the suspicious needles in the transaction haystack.

FIX #3: CREATE "INSTITUTIONAL MEMORY" IN ALERT DISPOSITIONING, LEVERAGING ADVANCED MACHINE LEARNING TECHNIQUES

Because of the 'rote' nature of the alert or case disposition processes, financial institutions have finally begun investing in machine based systems to improve productivity, increase accuracy, and save costs.

Two clear options have gained traction: Robotic Process Automation (RPA) and Machine Learning. RPA systems emulate the repetitive tasks of the investigative analyst's workflows (e.g., Lexis-Nexis searches, screen-scraping and saving screenshots to PDFs), only with the ability to execute them faster, at significantly larger scale and consistently.

Machine Learning systems are also referred to as 'Selflearning Systems', where not only can machines be trained to "read" and effectively make recommendations based on prior cases, but they can improve their recommendations as human subject matter experts continuously provide feedback. Over time, such recommendations will mimic the knowledge and behavior of the best analyst.

In a new approach to tackling AML fraud with Big Data, a small team of the best AML investigators would use Tresata's human-machine learning interface to "train" machines on what factors to look for and how to write the narratives used to disposition cases. Then, they would sample and score one-month's worth of machine-generated recommended dispositions and narratives providing feedback inputs on which recommendations were correct or incorrect.

There are four key benefits to an automated AML workflow approach:

1. **It unleashes investigators to spend their time and brainpower actually investigating rather than doing the administrative research and documentation** required to complete narratives (many replete with mind-numbing tasks like copying and pasting Google search result or Lexis-Nexis screenshots into their documents).

2. **It powers each human investigator with machine assisted knowledge to enable them to be as smart, intelligent and productive as the best analyst.**

3. **It institutionalizes "muscle memory" in the case management process, as machines can immediately scan and process information about every prior case dispositioned** (even the ones just completed yesterday) and incorporate the findings and results in future recommendations and narratives.

4. **It focuses assignment of the toughest cases to humans, with the obvious cases** (where there are clear indicators of false positives or suspicious behavior) handled by machines that can be scaled up infinitely at nearly no cost.

In summary, an ability to augment human AML investigative systems with automated behavior based learning systems is no longer science fiction, but a reality every financial institution needs.

At Tresata, we realize that improving and automating AML systems is not just a technical challenge, but also a process one. Financial institutions need to understand how to potentially evolve into this new approach while simultaneously improving the performance and effectiveness of existing AML compliance efforts already underway.

Given the issues financial institutions of all shapes and sizes are facing with currently meeting BSA/AML requirements, many under the unfortunate spotlight of regulatory "Matters Requiring Attention" (MRAs), the massive IT infrastructure investments and projects in-flight and the human resource management of an ever-growing Middle and Back Office, even thinking about a new approach can seem daunting. In some financial institutions, this transformation might be too risky or expensive to consider as this involves new and to them, unproven technology where there may not be enough internal expertise present in the IT organization.

However, we are seeing early adopters willing to invest in transformation initiatives to show early wins and break the doubting cycle of turning good money over to massive, inflexible, and weak AML investigations. They have already proven that implementing a transformation does not have to be done in one fell swoop and best practices for how to transition into the new approach are well known.

AS IT SHOULD SAY - HERE IS TO CLEAN CARS AND ALWAYS CLEAN MONEY!

(DON'T) FEAR THE MACHINES

We want to demystify some of the hype around "machine learning." Machine learning simply means the ability of sophisticated computer programs to observe, understand and learn from behaviors or patterns that weren't explicitly defined by a human. Features of smart "machines" include:

AUTOMATION

Machine learning can run repetitive workflows millions of times in a consistent way without getting tired. The only resources it consumes are CPU and memory, which can be scaled at a low cost. At its most basic, RPA tools simply emulate human tasks and keyboard presses. Newer approaches skip the human tasks altogether and run more efficiently by working directly with the data sources behind business applications.

AUGMENTATION

Smart machines can not only find patterns, but raise the interesting findings to a level of attention. In other words, machines help prioritize what to focus on given all of the possibilities.

ADVANCED ANALYTICS

As an example of advanced analytics, machine learning is good at comparing billions of different data points together to find correlations. This enables machines to find and reveal hidden relationships beyond human perception.

OPEN-BOX MODEL

Most importantly, Tresata is pioneering an "open box model". All of our software (especially those using machine learning capabilities to make better predictions) have been given semantic capabilities. (Think Alexa.) What it means for you is that the predictions from our software will always be explained in plain English (a feature we call "STORYBOARD"). This will explain to an end user (or regulator) why the machine made the recommendation it did.

A 24 DAY FIX TO AML

A (DATA) DIAGNOSTIC SOLUTION TO BREAKING BAD (DATA)

Many organizations today do not yet know or understand exactly what level of investment is required to undertake an AML transformation:

- Will this require a complete "rip and replace" of the entire data infrastructure?
- Does it involve hiring an army of application developers, data scientists and machine learning specialists (all of whom are much sought after and incredibly highpriced) to tune or build new tools?
- Do I need to re-train or re-hire hundreds or thousands of investigators?

In nearly all cases, the correct answer to those questions will be "NO", but the best way to measure exactly how much change and transformation is required, and in what sequence, is to begin with what we refer to as a Data Diagnostic.

A Data Diagnostic is a time-boxed exercise (up to 24 business days) where a financial institution leverages Tresata's advanced AML software on its existing infrastructure or a private secure cloud and uses the unique capabilities to rapidly build a one-time "snapshot" Diagnostic Data Asset.

Once created, this Diagnostic Data Asset will be used to compare results with existing AML transaction monitoring and modeling outputs to quantify:

1. **Product (or Technology) Issues:** The data quality of source systems feeding AML systems; the quality of the enrichment process integrating context to transactions.
2. **Process Issues:** The quality of KYC risk ratings and customer segmentations, based on assessing quality of inputs into models; targets for false positive reduction.
3. **People Issues:** Worker productivity optimization opportunities, based on identifying where robotic and machine learning processes can reduce rote and repetitive tasks.

This information can then be incorporated into a specific roadmap and plan of attack. Because the diagnostic activity is time-bound, the level of effort, risk and investment can be managed and promises guaranteed ROI.

WHAT WE DO IN A 24 DAY DATA DIAGNOSTIC

PROFLE:
Profile all source data from customer, transactions, reference, and compliance systems to quantitively measure their baseline quality.

LINK:
Clean and integrate disparate datasets across all entities, counterparties, customers, non-customers matched at a segment of one with context.

ENRICH:
Enrich and organize all source data for all unique entities across all values to produce a Diagnostic Data Asset.

ANALYZE:
Asses impact of improved data quality on existing transaction monitoring systems, processes and models.

IDENTIFY:

1. False positive optimization opportunities
2. True positive identification opportunites
3. Target state data, technology and software architectures
4. Potential operational cost saving opportunities

"A DATA DIAGNOSTIC IS A TIME-BOXED EXERCISE WHERE A FINANCIAL INSTITUTION LEVERAGES TRESATA'S ADVANCED AML SOFTWARE AND USES THE UNIQUE CAPABILITIES TO RAPIDLY BUILD A ONE-TIME 'SNAPSHOT' DATA ASSET."

SUMMARY

There is a new way to both meet the continuously evolving AML requirements and reduce the rate of growth in AML compliance technology. This new approach takes advantage of **new and innovative data processing technologies – unlimited storage & computing capacity, probabilistic data linkage and machine intelligence to:**

1. Transform how customer and counterparty data is collected, cleaned, organized and used,
2. Detect non-normal behavior more precisely through segment of one techniques, and
3. Improve investigator productivity by augmenting them with machine intelligence.

Financial institutions will have to pull the trick of defining how to take advantage of this new approach while sustaining and improving the current AML compliance infrastructure, **but a Tresata Data Diagnostic is a low-cost and low-risk tool that can help define the scope of an actionable and achievable AML transformation plan**. Financial institutions should explore including a data diagnostic into their overall AML compliance strategy and roadmap today.

TRESATA'S KYC AND AML RISK MANAGEMENT APPLICATION

Tresata's AML software is powered by a new class of technologies that enable smarter processing of high volumes of data. This allows companies to address the root causes of KYC and AML issues that in turn are responsible for high incidences of false positives.

Tresata's intelligent AML system delivers:

1. Holistic View of Customer & Counterparty Risk

We give existing risk and Customer Due Diligence processes superpowers by:

- Cleaning records from customer, product, transaction and compliance data sources into a single 360 view of customer and counterparty activity across all product relationships
- Applying complex technologies to identify, resolve and link unique entities across the value chain
- Optimizing name and adverse media screening case management by reducing false positives
- Enabling "single pane" viewing of full transaction history, screening history and AML history without having to open and swivel between multiple windows
- Assessing risks and enabling identification and traversal of "relationship networks" for each customer, such as their buyer-supplier or beneficial owner relationships
- Applying innovative machine learning techniques to process trillions of bytes of information and uncover hidden risk factors for customers and counterparties, enabling a thorough picture of risk

2. Predicting True & False Positives

Our AML software complements and enhances existing transaction monitoring solutions by:

- Building unique behavioral fingerprints for each party in a transaction – customers, beneficiaries and involved financial parties – from years of historical activity
- Assessing risk events, such as alerts from your transaction monitoring systems, against the unique fingerprints to isolate truly abnormal or suspicious behavior
- Quickly identifying obvious false positives and clear true positives

3. Optimized Case & Alert Dispositioning

Our AML software optimizes dispositioning by:

- Integrating findings from KYC risk rating and AML fraud likelihood predictions to recommend how to dispose open AML cases and alerts
- Auto-generating investigator narratives that describe key research findings about each counterparty in a case and the risk factors driving the case disposition recommendation
- Applying machine learning and AI to train machines how to manage and disposition cases the way your best investigators do

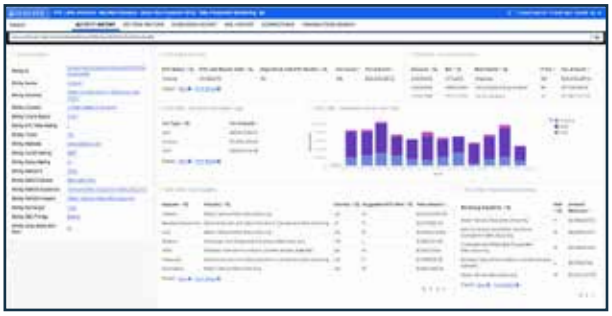


Figure 1. A Customer 360 profile with a single pane view of all transaction activity at a Segment of One

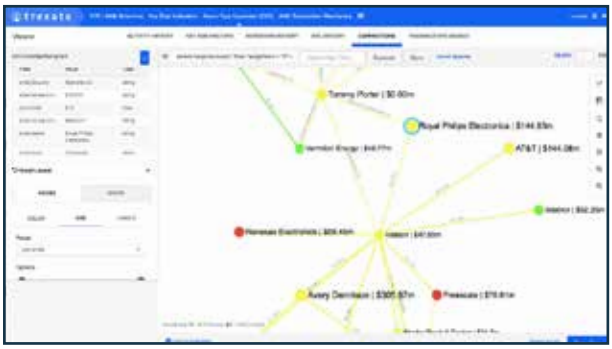


Figure 2. Interactive discovery and traversal of relationships between unique entities (companies and people) across all transactional data

TANGIBLE ROI

We have seen dramatic opportunities & improvements from use of our software:

- **Data quality improvement:** From 40% cleansed data to 98% cleansed and enriched with KYC factors & risk ratings
- **Entity resolution:** 100% of all customers, counterparties and involved financial parties enriched
- **False positive reduction:** Able to enable 50% suppression of false positives on Day One, just by improving the quality of data fed to AML Transaction Monitoring System
- **Improved workforce productivity:** Full automation of first level alert handling, 200-400% lift in second level (FIU) cases handled per person
- **Backlog reduction:** AML cases reduced by 50%, sanctions screening cases by up to 80% within first three months of implementation

BIOGRAPHIES

Abhishek Mehta

Abhishek Mehta is the CEO & Co-founder of Tresata, a predictive intelligence software company that in a short span of 4 years, he has built into one of the most innovative big data companies in the world.

Abhishek is recognized as one of the most influential thinkers, visionaries, and practitioners in the world of Big Data. His history is a rich combination of stints as a radical technology expert and a practical, in-the-trenches business leader. His experience includes Executive in Residence at MIT Media Lab, Managing Director at Financial institution of America, and various leadership positions at Cognizant Technology Solutions and Arthur Andersen.

A passionate supporter of entrepreneurship in the Southeast, Abhishek has been included in numerous lists of the top innovators, leaders, and disruptors of our generation. He is a highly sought after speaker on the topics of big data analytics, emerging business models, and all customary intersections of the two.

Eliud Polanco

Eliud Polanco is the Chief Scientist at Tresata Money, with over 15 years' experience in data analytics, business and technology strategy for financial institutions. Eliud has served as Global Head of Analytics and Big Data Strategy at multiple global systemically important financial institutions with responsibilities covering IT strategy and architecture, software partner management and build out of Data Science teams and analytics organizations. Analyses ranged from revenue and growth-oriented (Retail and Wholesale prospecting) to risk management and cost control (fraud intelligence, cybersecurity, financial stress test modeling and risk aggregation reporting). These experiences help shape Tresata's approach to building real-world, innovative solutions that will truly deliver realizable and tangible business benefit to its Financial Service customers.

ABOUT TRESATA

Tresata is the leading predictive analytics platform for understanding and monetizing customer behaviors with a singular goal – to enrich life™. This is achieved with great purity, precision and personalization by Tresata's analytics engines that have automated the discovery of knowledge™ from raw data to actionable insight.

For more information visit tresata.com or contact curious@tresata.com