McKinsey&Company

NOVEMBER 2017

© AF-studio/Getty Images

Risk

The new frontier in anti–money laundering

New analytical tools and surgical automation can help banks take the fight to fraudsters.

Stuart Breslow, Mikael Hagstroem, Daniel Mikkelsen, and Kate Robu

In recent years, three factors have heightened the risk banks face when combating financial crimes. First, the growth in volume of cross-border transactions and greater integration of the world's economies have made banks inherently more vulnerable. Second, regulators are continually revising rules as their focus expands from organized crime to terrorism. Finally, governments have expanded their use of economic sanctions, targeting individual countries and even specific entities as part of their foreign policies.

Banks have responded to these trends by investing heavily in people, manual controls ("checkers checking the checkers"), and systems addressing point-in-time needs. For example, in the United States, anti–money laundering (AML) compliance staff have increased up to tenfold at major banks over the past five years or so. Banks have typically used a piecemeal approach, adding staff to areas with the weakest controls. Often this has resulted in compliance programs built for individual countries, product lines, and customer segments—with all the duplication that suggests. Banks have also hired thousands of investigators to manually review high-risk transactions and accounts identified through inefficient, exception-based rules. For example, one big US bank expanded the ranks of its compliance team by one-third in recent years, including many people who work on "know your customer" (KYC) and AML compliance. Banks are also spending hundreds of millions of dollars to maintain the processes and systems they built in response to remediation needs.

As a result, second-line AML compliance programs now look more like operational utilities or, as one executive put it, "factories," and less like the independent oversight functions that banks first envisioned. These factories are expensive yet might be acceptable if the huge teams and manual processes were working well. But many are not. Most financial institutions continue to face challenges that erode the effectiveness and efficiency of their AML programs, including the following:

- Poor-quality data, nonstandard data structures, and fragmented sources make data aggregation by legal entities, subsidiaries, and vendors difficult. For example, many banks are still making tens of thousands of costly customer calls every month to refresh KYC documents, updating information that is incorrect or missing in their databases.

- Analytical approaches for customer risk scoring and transaction monitoring suffer from high rates of false positives, resulting in significant resources focused on investigating low-risk accounts and transactions. Adding new calibration tools and thresholds often leads to another spike in the number of false alerts.

- Inconsistent standards in processes such as customer identification, enhanced due diligence, and account monitoring and screening mean that businesses do not agree on what constitutes risk and violation of compliance requirements.

- Similarly, inconsistency in the reporting of suspicious activities and currency transactions means banks sometimes produce too many reports, and sometimes too few, exposing them to the twin dangers of regulatory sanctions and excessive cost.

- Fragmented systems and platforms limit the ability to automate transaction monitoring and due diligence. Instead, compliance teams spend the bulk of their time collecting data, and then on "stare and compare" sessions, instead of investigative work.

- Reliable quantitative metrics to assess risk across products, geographies, and processes are often not available.

- Ever-faster launches of new products and services, as well as instant fund transfers and mobile payments, add complexity to real-time detection and prevention. For example, "intelligent" ATMs allowing customers to anonymously deposit and transfer cash even when banks are closed certainly offer convenience but lack adequate KYC and AML safeguards.

Leading banks are trying to crack these problems by turning to new technologies. Machine learning, real-time data-aggregation platforms using fuzzy logic, rapid automation, and text and voice analytics offer a fundamentally new approach to managing compliance. Even better, they also offer an opportunity to simultaneously cut structural costs and improve the customer experience. As they take up these new tools, banks are shifting financial-crime compliance toward a more forward-looking and sustainable approach.

Traditional improvements in operations, governance, and management information systems will continue to be important elements in financial-crime-prevention programs. But technology and advanced analytics can raise these programs to much higher levels of effectiveness and efficiency. While there are many opportunities, our experience shows that banks should invest in three areas: efficient *data-aggregation* platforms, *advanced statistical modeling* (such as machine learning–based risk scoring and alert-generation engines), and *automation of processes* (such as investigator visualization tools).

Banks that invest strategically in these three areas, rather than tactically reacting to market and regulatory changes, can over time substantially reduce their risk exposure and capture other substantial benefits. For example, compliance-error rates measured through sample-based testing can be reduced from more than 30 percent to less than 5 percent. At the same time, false-positive alerts can be brought down from more than 90 percent to below 50 percent. These steps reduce the risk of regulatory fines and other penalties related to noncompliance, as well as help banks avoid potential reputational issues. The following discussions review ideas and techniques in the three areas and suggest ways banks can apply them.

Data aggregation

Banks in all markets struggle with the quality of data they keep on their customers, creating a significant obstacle to data aggregation. Long-time clients may have signed up when information standards weren't as rigorous and manual forms were prone to error. Most banks have established modern data-entry processes for new customers—yet these might be followed inconsistently across countries or even branches. The challenge can be especially daunting in some countries like the United States or the United Kingdom that have only partial nationwide identification systems.

Banks are turning to new tools to aggregate poor-quality data that can help them avoid hundreds of thousands of dollars in cost for manual data structuring and cleansing as well as hundreds of millions of dollars in investments required to build central "data lakes." For example, intelligent data platforms use machine learning or "fuzzy logic" (an approach to computing based on degrees of truth, rather than the more conventional binary true/false logic) on unstructured account and transaction data, to create a 360-degree view of suspected cases of money laundering. In practice, these new tools allow banks to automatically validate more customer identities, identify beneficial owners faster, and map how specific customers are connected to other individuals and legal entities, especially those earmarked as higher risks.

This can have significant implications on the volume of accounts and transactions that get escalated for manual reviews. For example, our analysis at one global institution showed that about half of the transactions flagged as "suspicious" would not have needed investigation if the bank had been able to connect the data held by its various divisions, some of which had identified and previously cleared the parties involved.

As another example, in a typical bank, data infrastructure and systems are not well positioned to quickly spot the connections among small cash deposits made by many different customers and wire transfers sent by those customers to the same recipient. The exhibit illustrates how a typical "smurfing" scheme works, in which cash deposits are broken down into amounts below the reporting threshold of $10,000. Analytics-driven data aggregation can help overcome these challenges by instantly connecting these individuals to the same geographic location, same behavioral pattern (for example, transaction types, frequency, and sequence), same destination account, and even block the wires from leaving the bank early in the process, before the laundered amount gets big.

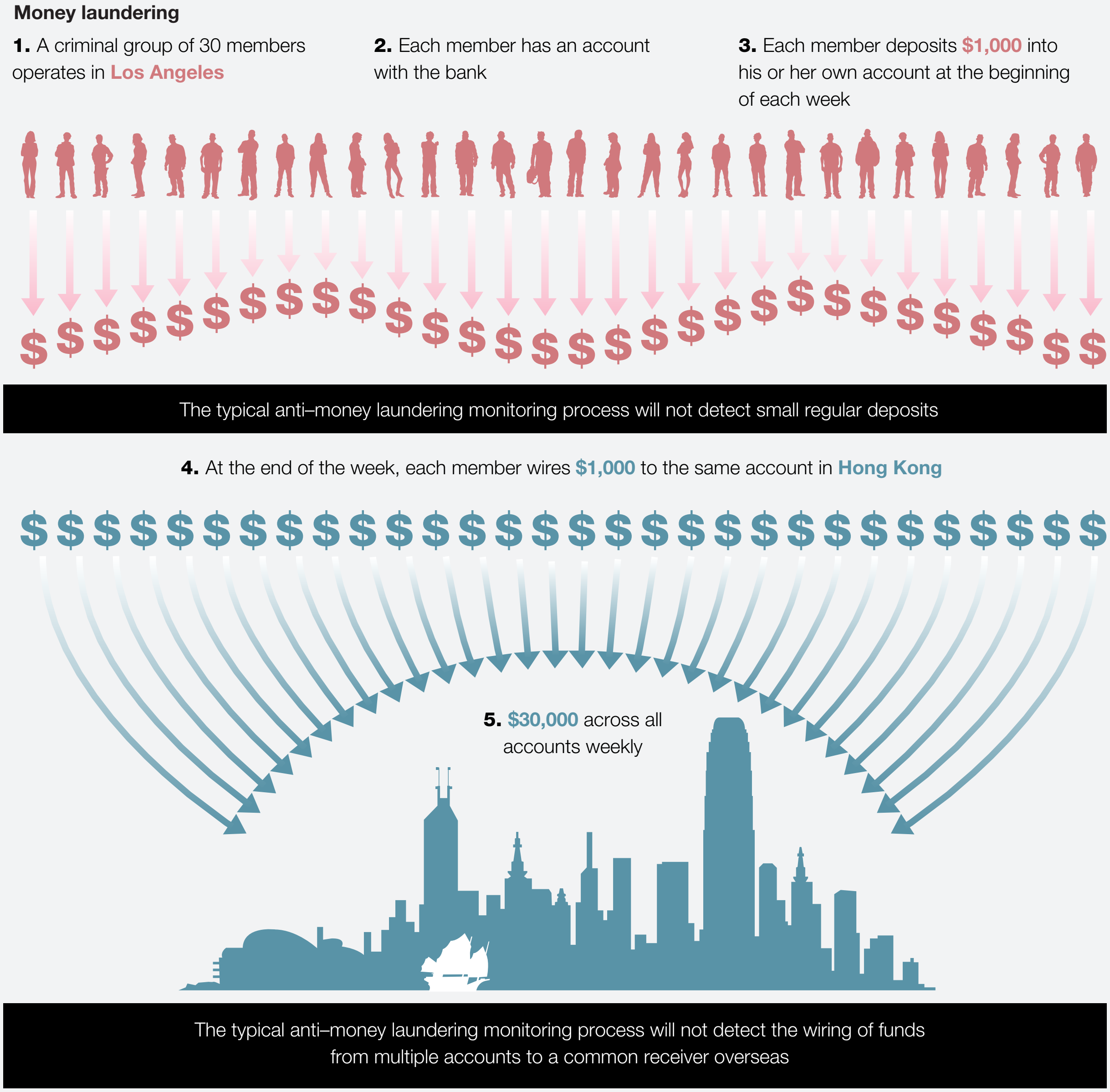
Advanced analytics

Intelligent data aggregation is not the only opportunity to apply advanced analytics in the AML space. Consider customer risk scoring and the tools used to generate alerts on suspicious transactions. Current tools are often not statistical models at all, but rather a series of linear rules based on an institution's experience, a typology of known money-laundering events, and explicit regulatory requirements (such as reporting any wire transfers of more than $10,000). Regrettably for banks, up to

90 percent of the alerts generated by these rules can be false positives, and should be quickly discarded by investigators (but often are not). Though rarer, false negatives (or criminal activity that goes unnoticed) also pose a significant risk to banks. It is relatively easy for criminals to understand the linear rules currently applied by many banks and then design approaches to circumvent them (like smurfing, including the use of dormant intermediary accounts before the funds converge into the target account).

Exhibit

How 'smurfs' tie together small deposits and wires, evading anti–money laundering detection.



Money laundering

1. A criminal group of 30 members operates in **Los Angeles**

2. Each member has an account with the bank

3. Each member deposits **$1,000** into his or her own account at the beginning of each week

The typical anti–money laundering monitoring process will not detect small regular deposits

4. At the end of the week, each member wires **$1,000** to the same account in **Hong Kong**

5. **$30,000** across all accounts weekly

The typical anti–money laundering monitoring process will not detect the wiring of funds from multiple accounts to a common receiver overseas

Source: McKinsey analysis

Statistical models based on machine learning and other forms of artificial intelligence can help banks raise their game. Such models review verified events to identify the often obscure combinations of predictive variables most likely to help minimize losses. Learning algorithms take advantage of the large pools of data and heightened computing power now available to detect patterns that might go unnoticed by data scientists. Systems using artificial intelligence can discern, for example, whether a series of transactions represents possible money laundering or a more innocent activity, such as a sudden wave of overseas expenses. In our experience, machine-learning algorithms can help reduce the number of false reports by 20 to 30 percent. As a result, investigators can spend more time on high-risk cases, and the manual work required can be reduced by as much 50 percent.

The impact of advanced statistical modeling is further increased when it's applied across a network of financial institutions. For example, one major European payments processor implemented machine-learning algorithms to follow the money across many banks and various entities, accounts, and locations. The approach allowed investigators to identify the paths used by "mule" accounts that are notoriously difficult to detect. Such accounts, spread across several financial institutions, bounce and "clean" the funds as they move from an illegal source into the formal financial system. Besides identifying the at-risk accounts in their network, investigators were able to develop powerful predictive variables to flag suspicious transactions and accounts newly entering the legal payments system. Forewarned is forearmed: banks that are on the alert for markers of increased money-laundering risk—such as use of Bitcoin services, prepaid cards, accounts opened by foreign students overseas—are able to stop transfers in real time.

The financial industry has been slow to adopt advanced tools such as machine learning, partly because the models are difficult to explain and validate to satisfy regulatory requirements. However, the techniques are becoming commonplace in other parts of the bank. Machine-learning algorithms are being used to offer better products and advice to customers, as well as to manage customer retention more effectively. Regulators are becoming more comfortable with validation approaches involving random forest and other such algorithms, which produce models that are relatively easy to explain and test for stability.

Banks can start with simple uses of analytics, like those involved in smart triage and microsegmentation of accounts and transactions to reduce false positives. For example, instead of making binary "file/do not file" decisions, some banks score each account and transaction that did not immediately require filing of suspicious-activity reports (SARs). They "hibernate" them until the cumulative view of triggers over time surpasses a predetermined threshold. Some institutions achieved a threefold improvement in SAR conversion rates through tighter segmentation of accounts and transactions based on behavioral and demographic characteristics, allowing them to distinguish between suspicious and nonsuspicious transactions the same way experienced investigators do.

Down the road, other tools might accelerate progress, given AML's heavy reliance on human judgment and expertise. Deep learning is an advanced form of machine learning that is already being used in image analysis and human language processing. It attempts to mimic human thought processes like those used by financial-crimes investigators and requires large amounts of data and fine-tuned models. Deep learning will likely start being deployed at scale in the next three to five years for banks to combat money laundering, fraud, and other financial crimes.

Automated processes

Automation and standardization of critical portions of the due diligence and investigation processes can make expert staff more effective and significantly reduce their caseload. Robots can be used to automate certain activities, including the population of case files for investigators, the closing of level-one alerts, and the population of SAR forms. These measures can reduce the investigation time for alerts and allow for workforce optimization.

For example, a leading North American wealth manager used many techniques to move from a largely manual process for customer identification and due diligence to a reengineered and tech-enabled process. The solution included case-management work flow to guide due-diligence analysts faster and more effectively through the process; an integrated interface to bring all the data and third-party applications that analysts typically need into a single screen; rules-driven pipeline management to ensure priority-based resolution of cases; and so on. Under the new processes, staff were able to make decisions on low- to medium-risk customers almost instantly, and within 24 hours on most high-risk clients. The initiative also enhanced the customer experience by speeding up decisions and eliminating unnecessary follow-ups for missing information. All told, the firm was able to improve operational efficiencies in KYC by up to 50 percent.

The integrated interface is particularly important for speeding up the alert-investigation process and can be quickly acquired and deployed from a number of third-party vendors. This type of tool automatically gathers information through online searches, internal data, and third-party databases and highlights concerns such as relevant sanctions, negative media, and political exposure. This information is visualized into a clear, on-screen report that helps an investigator quickly assess the case and make a decision.

Filing of SARs with regulators is another area that presents high potential for automation. Natural-language-processing software converts data into text and can replace most of the work that investigators are traditionally putting into writing the reports that support their decisions on a case when it's filed. Integrated with the case-management work flows, nonperforming-loan applications can be really powerful tools that automatically generate the SARs as soon as the investigator pushes the "generate report" button—all it takes is a quick review and edit, followed by pushing the "file report" button.

How to get there

Our experience suggests that analytics and technology are important, but they alone will not provide a silver-bullet solution to all AML challenges. The key to impact is being able to deploy analytics and technology in a business-specific way and to embed them organically into business processes, which in turn often have to be fundamentally reshaped to take advantage of new tools.

With this context in mind, leading institutions are focusing on four key initiatives to both generate substantial value in the near term and course-correct their in-flight efforts to achieve a more sustainable target operating model:

1. Develop a truly end-to-end view of an optimized, tech-enabled KYC and AML process, from new standards for customer-data intake to customer identification to risk-based due diligence to monitoring. The design of this "north star" process should cover complexity-based triage, rules-based routing of files to investigators, standards of work, quality tollgates, and so on. Currently no single third party supports the entire process; hence the bar is very high for up-front system-architecture design and integration of internal and third-party point-to-point solutions.

2. Define a strategy for data quality and aggregation, including linking KYC and AML data closer together. Consider application of analytical tools such as fuzzy logic and machine learning to connect the dots in the known KYC/AML reference data—such as customer investments

and associated entities and individuals—and sanctions data. This will address a significant share of customer due-diligence escalations. It will also end many of the sanctions and transaction-monitoring alerts that result from gaps and problems with matching reference data, including known intracompany and intercompany transfers and customer identification data from another business unit.

3. Fold simple analytical approaches like micro-segmentation into current systems and model-validation processes. In this way, added rules on top of existing rule engines for customer risk scoring and alert generation can be consolidated. In parallel, banks can invest in longer-term solutions such as training neural networks through supervised learning, which will further reduce false positives and false negatives. These tools can be brought up to high performance and be ready to go once regulatory approval is completed.

4. Implement a set of metrics and practices to measure effectiveness of the KYC/AML processes and assess impact from operational and system improvements. Potential metrics could include the following:

    - Establish the expected volume and quantity of alerts. For example, set targets in 90-day intervals to reduce false alerts as new controls are launched.

    - Set rate of conversion of alerts to cases: for example, aim to reduce the SAR conversion rate by 1 or 2 percent every 90 days.

    - Reduce time per case: for example, set a target to reduce the investigation time by case type.

    - Set targets to reduce false positives and negatives, rather than focusing on the number of SARs filed or overall transaction volumes.

The industry is at a turning point. Not only are many banks reconsidering their approach to KYC and AML, but many regulatory-technology start-ups are launching products to support and sometimes supplant their efforts. Every new technology reaches a point when the hurdles fall away, and the benefits become too numerous to ignore any longer. As pioneering banks are finding out, automation and analytics for AML are at that point.

Stuart Breslow is a partner in McKinsey's New York office, **Mikael Hagstroem** is a partner in the Charlotte office, **Daniel Mikkelsen** is a senior partner in the London office, and **Kate Robu** is a partner in the Chicago office.