Royal United Services Institute
for Defence and Security Studies

Occasional Paper

Sharpening the Money-Laundering Risk Picture

How Data Analytics Can Support Financial Intelligence, Supervision and Enforcement

Olivier Kraft

Sharpening the Money-Laundering Risk Picture

How Data Analytics Can Support Financial Intelligence, Supervision and Enforcement

Olivier Kraft

RUSI Occasional Paper, November 2018

Royal United Services Institute
for Defence and Security Studies

187 years of independent thinking on defence and security

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 187 years.

Published in 2018 by the Royal United Services Institute for Defence and Security Studies.

RUSI Occasional Paper, November 2018. ISSN 2397-0286 (Online); ISSN 2397-0278 (Print).

Royal United Services Institute
for Defence and Security Studies
Whitehall
London SW1A 2ET
United Kingdom
+44 (0)20 7747 2600
www.rusi.org

RUSI is a registered charity (No. 210639)

Contents

Acknowledgements v

Executive Summary vii

Introduction **1**
Objective 2
Methodology 2

I. Use of Data Analytics to Develop Intelligence on Money-Laundering Risks **5**
Section A: Objectives for the Use of Data Analytics by Public Authorities 5
Section B: Datasets Underpinning Data Analytics 6
Section C: Methods of Analysis 13
Chapter Summary 20

II. Devising and Implementing a Strategy for the Use of Data Analytics for AML Purposes **21**
Section A: Factors Shaping the Approach to Data Analytics 21
Section B: Key Steps for the Use of Data Analytics in AML Intelligence, Supervision and Enforcement 24
Chapter Summary 33

Conclusion and Recommendations **35**

About the Author 37

Acknowledgements

The author would like to thank EY, Lloyds Banking Group and Refinitiv for their financial support of the Financial Crime 2.0 programme,[1] and all those who generously shared their experience and expertise for this research. The author is also grateful to Malcolm Chalmers, Mario Gara, Emöke Jakab and Anagha Joshi for reviewing and commenting on earlier drafts of the paper, and to RUSI's Publications team for their support throughout the editing process.

1. The Financial Crime 2.0 programme, launched in March 2018, aims to determine how the anti-money laundering (AML) regime needs to be updated to be better aligned with today's technological landscape.

Executive Summary

THE USE CASES reviewed as part of this research illustrate the value of data analytics for anti-money-laundering (AML) supervision and enforcement. For the purposes of this research, the term 'data analytics' refers to methods allowing users to turn data into knowledge that would not be revealed through a human review of the data in question. This includes traditional statistical methods and more recent developments relating to 'big data' or machine learning.

If used strategically and at scale, data analytics can support a dynamic and more precise assessment of money-laundering risks, and thereby enhance the efficiency and effectiveness of authorities' AML efforts. This includes opportunities to monitor macroeconomic trends in real time, such as the evolution of financial flows in response to geopolitical or regulatory developments, or to more systematically analyse the financial activity of institutions or geographic areas to determine whether it is consistent with a range of contextual factors.

Notwithstanding the growing use of data analytics by AML authorities, the international dialogue on existing methods has been limited to date. This paper provides the first comparative study of the experience of national authorities in using data analytics for AML purposes and offers recommendations on how to maximise associated benefits.

The use cases reviewed for this research present a number of differences. First, they have been developed to serve different purposes. Some are primarily designed to detect or investigate criminal conduct, whereas others seek to enhance a sector's risk-mapping as a basis for better allocation of supervisory resources. Second, the datasets underpinning the analysis vary in nature (data relating to transactions versus data on contextual factors), scope (data on suspicious transactions only versus data on larger sets of transactions) and level of detail (aggregate versus granular data), with the most effective tools generally relying on a consolidated review of several datasets. Third, the umbrella term 'data analytics' includes a range of different methods that have been successfully used or are currently being explored by authorities, including trend analysis, network analytics and, more recently, machine learning.

Considering the diversity of potential models, the choice of an analytical model should be informed by several contextual and operational factors. Contextual factors include each jurisdiction's financial crime threat landscape, economy, data protection standards and non-AML policy objectives (such as facilitating business). Operational factors relate to the specific objectives and priorities of agencies, the availability of human and technological resources needed for the use of data analytics, and an agency's pre-existing access to data.

Taking into account both contextual and operational factors, the use of data analytics should begin with a determination of the priority questions to be answered through the analysis and

a clear understanding of how the results are relevant to AML efforts. For example, a method of analysis can be designed to determine if individuals use the services of several money service businesses to remit funds to the same destination, on the understanding that this behaviour indicates a higher risk of money laundering or terrorist financing.

As a next step, authorities should determine what data is needed to answer this question, whether such data is already available, and how additional data (if any) should be accessed, while keeping privacy considerations in mind. The selection of a dataset should be informed by consultations with other authorities, reporting entities and international partners. Irrespective of the specific method used, authorities need to develop a system to verify and ensure the quality of the data, including through routine verifications of consistency.

From the outset, each method should include an ethical framework clarifying how and for which purposes data may be used (for example, only for supervisory rather than investigative purposes, or only to support an ongoing investigation instead of revealing new leads), as well as a strategy for disseminating results to the intended end users. Finally, use cases should be subject to a periodic review of effectiveness, assessing if they meet the agreed objectives and how they contribute to overall AML efforts. This will inform any adjustments to the method and help policymakers assess whether resource and privacy implications are proportionate to the ultimate benefits.

In light of this, the paper concludes with the following recommendations:

- **Supervisors, financial intelligence units and relevant law enforcement agencies** should adopt a strategy to identify and harness opportunities for the use of data analytics in an AML context, taking into account available use cases and the factors discussed in this paper and in consultation with relevant private sector actors. This will provide a stronger basis for a transparent discussion with policymakers on necessary long-term investments and the balance with privacy considerations.
- **Regional organisations such as the EU** should foster a dialogue on the different approaches developed in this area by member states, to encourage a harmonised approach. Such coordination will reduce compliance costs for businesses operating in several countries and is likely to benefit the quality and usefulness of the data submitted under each of the national regimes.
- **Assessment bodies such as the Financial Action Task Force and the IMF** should more systematically review to what extent member countries have considered and harnessed the potential of data analytics for AML intelligence, supervision and enforcement purposes, which will allow for peer learning.

As the pace and complexity of the financial system continues to increase, the effectiveness of global AML efforts will be contingent on the readiness of governments to innovate in a strategic and coordinated manner.

Introduction

FOR THE PURPOSES of this research, the term 'data analytics' refers to methods allowing users to turn data into knowledge that would not be revealed through a human review of the data in question. This includes traditional methods of analysis (such as linear regression) and more recent developments relating to 'big data' (in other words, data that, due to its unstructured nature for example, is difficult to analyse using traditional data analysis methods)[1] or machine learning. The growing importance of data analytics in efforts to combat money laundering is not only a natural consequence of technological developments that make new approaches *possible*, it is also based on the recognition that traditional anti-money-laundering (AML) tools do not match the pace and complexity of today's international financial system, and that a different, technology-enabled response is therefore *necessary*.

As a result of these two factors, authorities in a number of jurisdictions have recently increased their use of data analytics to support AML prevention, supervision and enforcement objectives. Supervisory authorities analyse large amounts of transaction data to identify trends and patterns that may indicate a higher risk of money laundering (for example, a discrepancy between the volume of financial flows leaving a certain area and other characteristics of the same area). Network analytics allow financial intelligence units (FIUs) or law enforcement agencies to detect previously unknown connections between suspicious activity reports (SARs) – for instance, by clustering SARs involving seemingly unrelated companies that all have the same director.

While such initiatives are starting to produce results, research conducted for this paper indicates that the international dialogue on the use of data analytics by AML authorities has so far been limited. This reduces opportunities for peer learning despite a growing body of experience. In addition, the lack of dialogue also leads to a disparate landscape, with many authorities developing new systems based on different datasets. This fragmentation in turn increases the cost of compliance for financial institutions operating in several jurisdictions, and, to the extent it makes compliance more complicated, can also have an impact on the quality of the data obtained by authorities.

A dialogue on existing use cases of data analytics and their benefits, for example within the Financial Action Task Force (FATF), would enhance the coordination of initiatives in this area. Differences between methods of analysis will always exist, due to a series of factors that are discussed in this paper. However, such differences must be informed by an assessment of existing use cases and potential trade-offs. A shared understanding of the benefits of data analytics is also necessary to make further progress in the area of cross-border information sharing,

1. Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection', March 2017, p. 6, <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>, accessed 6 August 2018.

which requires a certain level of compatibility between national datasets. Finally, a review of the benefits of advanced analytics for AML purposes is essential to inform future discussions on the balance between these tools and privacy considerations.

Objective

The objective of this paper is to review and support the efforts of national authorities to harness the opportunities of data analytics for AML purposes. Specifically, the paper sets out potential methods and their benefits, as well as other factors that should be considered to ensure the proportionality and sustainability of innovative approaches.

Chapter I offers an overview of potential methods of data analysis in the context of AML supervision and enforcement, illustrated by international use cases. In addition to informing the decisions of relevant authorities, the proposed taxonomy is also meant to facilitate future discussions on this topic. Chapter II discusses factors (such as data protection, the regulatory burden for reporting entities, and the assessment of effectiveness) to be considered in developing and deploying a data management strategy for AML purposes. This paper complements existing work on the use of quantitative data in periodic national risk assessments,[2] the use of data to assess the effectiveness of AML regimes,[3] the relevance of individual datasets,[4] and the ongoing use of regulatory technologies within individual financial institutions.[5]

Methodology

The paper is based on a desk-based review of publicly available information, as well as information provided during 13 semi-structured interviews with authorities and former officials from seven jurisdictions.[6] It also builds on the outcomes of an international workshop held in London on 20 June 2018 with 60 representatives from FIUs, law enforcement, supervisory authorities (including representatives from additional jurisdictions[7]) and from the private sector.[8]

2. See, for example, Financial Action Task Force (FATF), 'National Money Laundering and Terrorist Financing Risk Assessment', FATF Guidance, 2013, p. 17.
3. See, for example, FATF, 'Guidance on AML/CFT-Related Data and Statistics', FATF Guidance, 2015; Organization for Security and Co-operation in Europe, 'OSCE Handbook on Data Collection in Support of Money Laundering and Terrorism Financing/National Risk Assessments', 2012.
4. Anagha Joshi, 'In Pursuit of Big Data: An Analysis of International Funds Transfer Reporting', *RUSI Occasional Paper* (April 2017).
5. Bart van Liebergen, 'Machine Learning: A Revolution in Risk Management and Compliance?', *Capco Institute Journal of Financial Transformation,* Institute of International Finance, April 2017.
6. Austria, France, Italy, the Netherlands, Singapore, Spain, and the UK.
7. Belgium, Germany, Gibraltar, Ireland, San Marino, Switzerland, and the US.
8. Workshop on 'Contribution of Advanced Analytics to AML Supervision and Enforcement', RUSI, London, 20 June 2018.

The publicly available information reviewed for this project includes, but is not limited to: risk assessments issued by public authorities in the UK and other jurisdictions; annual reports of FIUs and supervisors; guidance from international bodies on the use of data and statistics in the context of AML efforts; and academic and policy papers on data protection and governance.

I. Use of Data Analytics to Develop Intelligence on Money-Laundering Risks

THIS CHAPTER GIVES an overview of international experience in using data analytics for AML purposes. While many authorities were consulted during the research for this paper, the use cases presented here are not exhaustive. Rather, the paper seeks to offer an initial framework for further discussions that will reveal more approaches as technological opportunities evolve. The chapter first clarifies the objectives (Section A) and the datasets (Section B) underpinning data analytics, before discussing use cases among national authorities (Section C).

Section A: Objectives for the Use of Data Analytics by Public Authorities

The use of data analytics is not an end in itself. Rather, the use cases surveyed for this research are designed to: (1) identify suspicious activities that warrant a review by the national FIU or a law enforcement intervention; or (2) assess risks associated with individual financial institutions, economic sectors or financial products so as to inform supervisory priorities under the risk-based approach.

Identification of Suspicious Activities

Consistent with the FATF Recommendations,[9] nearly all jurisdictions require financial institutions and several other sectors to file suspicious activity – or transaction – reports (which are commonly abbreviated to SARs) if they suspect or have reasonable grounds to suspect that funds are the proceeds of a criminal activity or are related to terrorist financing. Data analytics may be designed to help the FIU determine which of these reported activities should be analysed as a matter of priority. In addition, authorities may develop tools to detect activities that are not reported but may nevertheless warrant further review. Such activities may remain unreported either because they simply cannot be detected at the level of an individual institution, or because the institution involved deliberately fails to comply with its reporting obligations.

9. FATF, 'Recommendation 20: Reporting of Suspicious Transactions', The FATF Recommendations, October 2018, p. 17; FATF, 'Recommendation 23: DNFBPs: Other Measures', The FATF Recommendations, p. 18.

Risk-Mapping for Supervisory Purposes

Under the recommendations of the FATF, AML supervisors are expected to adopt a risk-based approach. In other words, they should allocate resources considering a risk-mapping of the financial institutions they supervise. Risk-mapping tools generally distinguish between inherent risk, which exists prior to mitigation measures, and residual risk, which follows mitigation measures.

As an initial step, the evaluation of an institution's inherent risk requires an assessment of relevant factors, such as an entity's governance structure, the reputation of senior managers, the profile of customers, the geography of services, or the type of products offered. However, depending on a sector's characteristics, data analytics may be used to complement a qualitative risk-mapping. For example, the evolution of an institution's transaction volume on a certain geographic corridor over time, and/or a discrepancy with peers, may point to a higher inherent risk and therefore warrant closer supervision to ensure that the institution has taken appropriate mitigating measures. The outcomes of the analysis may also allow supervisors to provide additional guidance to supervised entities.

A targeted supervision of regulated entities may indirectly contribute to the detection of criminal conduct. However, this is not the primary purpose of supervision, which is designed to ensure that the AML efforts of financial institutions are consistent with their risks.

Section B: Datasets Underpinning Data Analytics

Aside from varying objectives, another reason explaining the diversity of methods of analysis applied by authorities is the range of datasets that may be available among jurisdictions. Accordingly, this section provides a taxonomy of potential datasets used, recognising that a given method of analysis may use more than one dataset. Transaction data is discussed first, followed by other types of data.

Transaction Data

Given the characteristics of money laundering, transaction data is currently the most common basis for conducting data analytics in the AML context. However, this category includes a number of different datasets, which can be grouped according to three criteria: the level of granularity (do authorities hold data about individual transactions, or about groups of transactions?); the relevant transaction characteristics (what type of data do authorities hold for a given transaction or group of transactions?); and the relevant universe of transactions (which transactions are included?).

These three criteria are independent of each other and the number of potential combinations is therefore considerable. Datasets are discussed here in two broad categories based on the first criterion: (a) granular data and (b) aggregate data.

Granular Transaction Data

Granular transaction data relates to individual transactions. It often includes all the characteristics of a transaction that a financial institution is required to collect under applicable record-keeping standards, including amount, currency, parties involved, and country of destination. Granular data generally allows an analyst to establish a link between transactions, either through names or through other data points such as a credit card number, a telephone number (for mobile money) and/or a public key (for virtual currency transactions). However, granular data can also be anonymised. For example, supervisors may request samples of anonymised transactions to decide where to set regulatory thresholds.

Depending on the jurisdiction, granular data is reported to the FIU either for suspicious transactions only, or for additional categories of transactions.

- Suspicious transactions

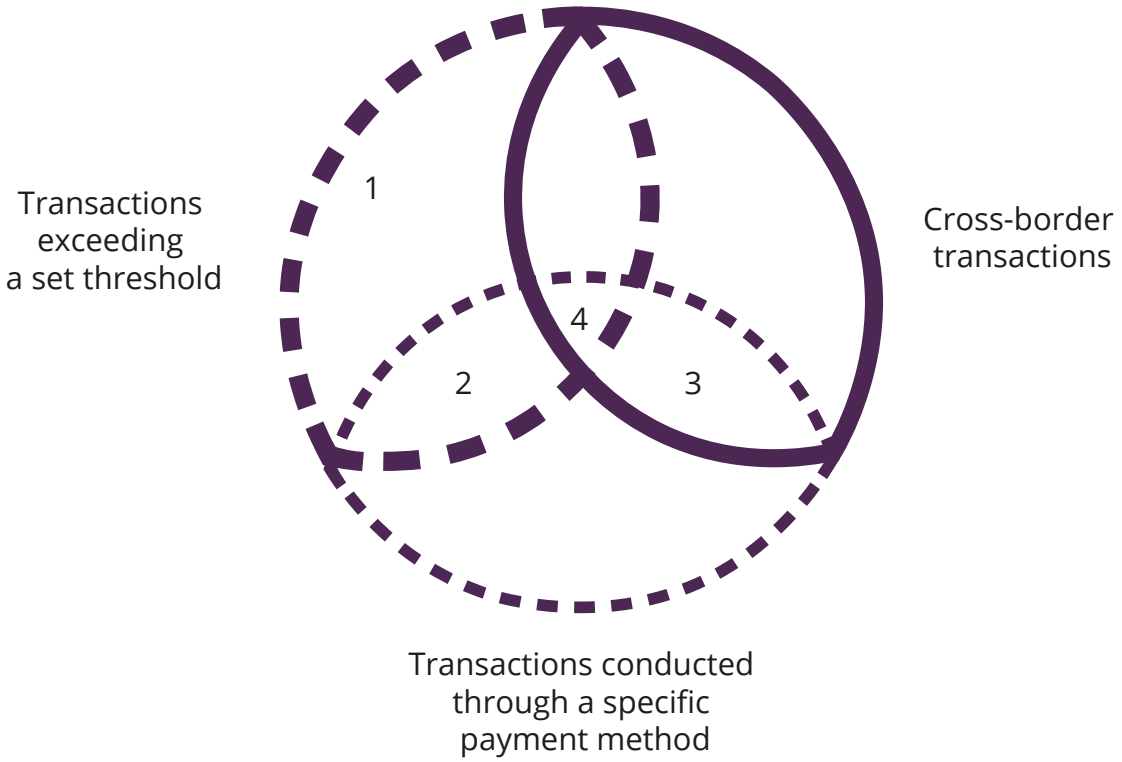
By definition, SARs are meant to provide FIUs with intelligence leads on potential cases of money laundering and terrorist financing. As a result, the effectiveness of the SARs regime is often assessed based on the ratio of SARs that lead to investigations, or the 'conversion rate'. While the conversion rate is a key indicator of the quality of SARs and of the overall effectiveness of the AML regime, the role of SARs in AML efforts is not necessarily limited to their individual contribution to an investigation. Rather, irrespective of that contribution, the consolidated analysis of SARs can reveal networks or broader trends. This is particularly true when the analysis also draws on other databases such as police records or company registries. Research for this paper indicates that this additional value of SARs is increasingly used to focus analytical and investigative resources.[10] While the use for supervisory purposes has been more limited to date, various use cases are described further in Section C.

- Transactions meeting objective criteria

In addition to SARs, financial institutions in certain jurisdictions are required to report additional transactions based on objective criteria defined by competent authorities.[11] Criteria are generally defined following an assessment of the types of transactions that are most commonly used in money-laundering schemes, including high-value, cross-border and cash transactions. The main difference with SARs is that, in the case of strictly objective criteria, transactions must be reported regardless of a case-specific suspicion. As reflected in Figure 1, criteria may be applied on their own or in combination. Table 1 lists relevant examples by category.

10. Workshop on 'Contribution of Advanced Analytics to AML Supervision and Enforcement'. See also, National Crime Agency, 'Suspicious Activity Reports (SARs) – Annual Report 2017', p. 36.
11. For a more detailed discussion on the difference between suspicion-based and other forms of reporting, see Law Commission, 'Anti-Money Laundering: The SARs Regime', Consultation Paper No. 236, 2018, pp. 179–84.

Figure 1: Potential Subsets of Transactions According to the Three Most Common Criteria



Source: Author's research.

Table 1: Examples of Jurisdictions Requiring Transaction Reports Based on Objective Criteria (Non-Exhaustive List)

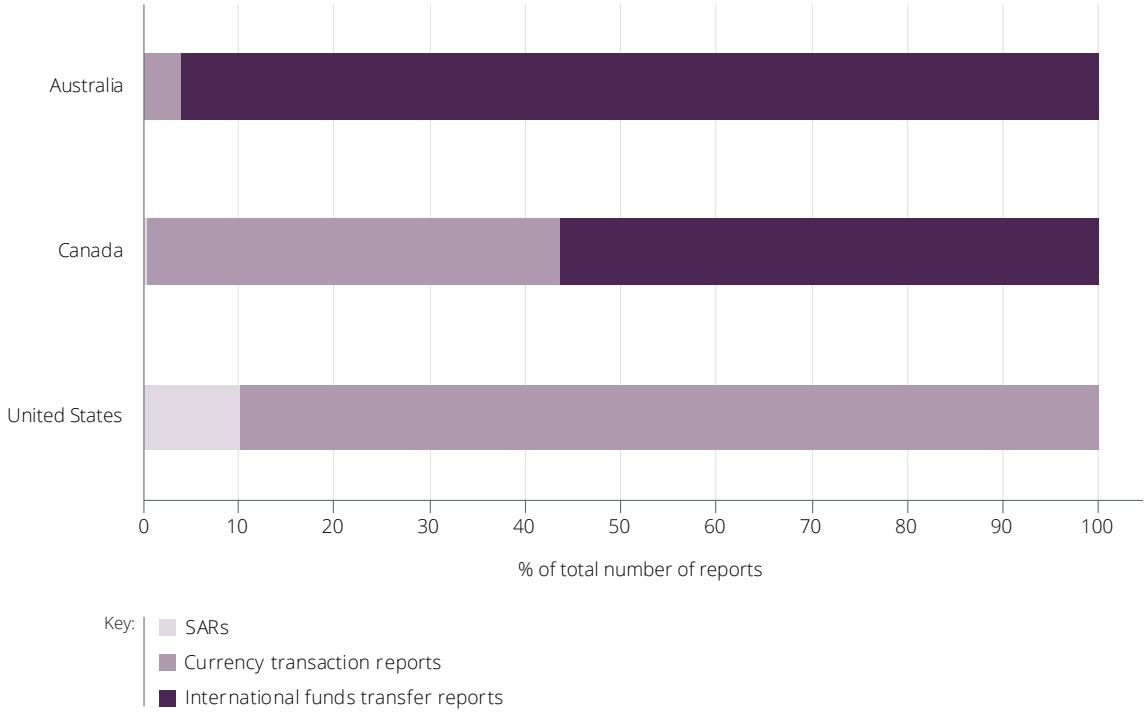
Subset	Examples
1	**The Netherlands**: Financial institutions and designated non-financial businesses and professions must report all transactions above a certain threshold (which varies depending on the sector) to the FIU.
2*	**US**: Financial institutions must report each cash deposit, withdrawal, exchange of currency or other payment or transfer above $10,000. **France**: Financial institutions must report all cash or e-money transfers above €1,000 (or linked transfers exceeding €2,000 per calendar month), as well as cash withdrawals above €10,000. **Spain**: Money service businesses must report cash transfers above €1,500.
3	**Australia**: Financial institutions must report all cross-border wire transfers.
4	**Canada**: Financial institutions must report all cross-border wire transfers above CAD 10,000. **India**: Financial institutions must report all cross-border wire transfers above INR 500,000.
Other	**Italy**: Professional gold traders and financial institutions must report all transactions in gold exceeding €12,500.

** Subset 2 is the only category of systematic transaction reports explicitly foreseen by the FATF Recommendations, although their adoption is left to countries' discretion. Specifically, Interpretative Note to FATF Recommendation 29 states that '[c]ountries should consider the feasibility and utility of a system where financial institutions and DNFBPs [Designated Non-Financial Businesses and Professions] would report all domestic and international currency transactions above a fixed amount'.*

Sources: Implementation Decree for the Money Laundering and Terrorism Financing (Prevention) Act (the Netherlands); 31 CFR 1010.311 (United States); Code monétaire et financier 2018, Article L561-15-1, D561-31-1, D561-31-2 (France); Reglamento de la Ley 10/2010, de 28 de abril, aprobado por Real Decreto 304/2014, de 5 de mayo, Article 27 (Spain); Anti-Money Laundering and Counter-Terrorism Financing Act 2006, Section 45 (Australia); Proceeds of Crime (Money Laundering) and Terrorist Financing (PCMLTF) Regulations, Section 12(1)(b) and 12(1)(c) (Canada); Master Circular RBI/2015-16/42, Section 8.b.4 (India); Legge del 17 gennaio 2000, n.7, Article 1.2 (Italy); FATF, 'Interpretive Note to Recommendation 29 (Financial Intelligence Units)', The FATF Recommendations, October 2018, p. 97.

The size of the dataset resulting from objective transaction reporting depends on the applicable criteria (for example, the lower the threshold for currency transaction reports, the larger the dataset), but is generally larger than the SARs database. Figure 2 indicates the ratio of SARs, currency transaction reports and international funds transfer reports (where applicable) in Australia (2016/17), Canada (2016/17) and the US (annual average between 2012–14).

Figure 2: Ratio of Different Types of Reports to FIUs in Selected Jurisdictions



Source: AUSTRAC, 'AUSTRAC Annual Report 2016/17', 2017; Financial Transactions and Reports Analysis Centre of Canada, 'FINTRAC Annual Report', 2017; FATF, 'FATF Mutual Evaluation Report of the United States', 2016.

The geographic targeting orders (GTOs) foreseen under US legislation constitute an alternative model for the collection of granular transaction data. Unlike the permanent reporting requirements described above, GTOs are used on an ad hoc basis and, as illustrated below, are limited to a specific area.

> [GTOs] require any domestic financial institution or group of domestic financial institutions in a geographic area and any other person participating in a given type of transaction to file a report in the manner and to the extent specified in such order. Four GTOs have been publicly issued in the last two years (previously, they were not public). Examples include: a 2015 GTO requiring trades or businesses that export electronics located near Miami to record and report to FinCEN [Financial Crimes Enforcement Network] information on certain transactions in excess of USD $3000; and a 2016 GTO on title insurance.[12]

12. FATF, 'Anti-Money Laundering and Counter-Terrorist Financing Measures – United States, Fourth Round Mutual Evaluation Report', December 2016, p. 57.

Aggregate Transaction Data

Aggregate data does not reflect the characteristics of individual transactions, but the collective characteristics (for example, the total amount) of groups of transactions. Similar to granular data, aggregate transaction data may be distinguished depending on the characteristics that are captured, and the universe of transactions covered:

- Characteristics

For any group of transactions, characteristics that may be captured include: the total volume (amount); the total number of operations; and the number of customers involved.

- Universe of relevant transactions

Aggregate data may cover different subsets of transactions defined by their amount, their cross-border nature, the method of payment, the type of customer, the nature of the transaction (for instance, cash withdrawal), and/or other characteristics. For example, the aggregate data collected by one authority covers transactions above a threshold of €15,000 (both domestic and cross-border), whereas the aggregate data collected by another authority covers all transactions on certain corridors considered to be high risk (without any monetary threshold). As part of the supervision of money service businesses (MSBs), certain supervisors collect aggregate data that reflects all transactions conducted.

Data on suspicious transactions can also be aggregated. The key advantage of SARs compared to objective transaction reports is that, in addition to basic transaction characteristics, they generally also include a more detailed analysis of the transaction. This makes it possible, for example, to aggregate SARs based on keywords that are frequently associated with a specific predicate offence. On the other hand, from a statistical point of view, objective criteria allow for a more consistent benchmark, both across institutions and over time (assuming that any monetary thresholds are adjusted based on the inflation rate). The main reason is that the policy for filing SARs varies significantly across sectors and over time. In other terms, a financial institution with a low suspicion threshold may file twice as many SARs as another institution that has a similar risk exposure but conducts more robust verifications before deciding whether to file a SAR. Similarly, an increase or decrease in the number of SARs filed by a given financial institution does not necessarily reveal a change in the threat landscape, but could instead be due to an adjustment of that institution's reporting policy. Statistical methods that rely on SARs should account for these extraneous factors. One FIU reported risk-scoring SARs into several categories based on the FIU's own methodology, whereby only SARs that meet a certain threshold will be included in the dataset underpinning the analysis. The use of a uniform classification for all SARs enhances the consistency of the dataset across the sector and over time but assumes the FIU's capacity to review and risk-score all SARs.

Another caveat regarding the use of SARs as a basis for data analytics is that, by definition, the dataset is limited to those instances in which a suspicion of money laundering or terrorist

financing has already been noticed by a financial institution. The SARs database is therefore most useful for the analysis of trends on established risks but may not reveal 'unknown unknowns'.

In certain cases, data is aggregated by authorities that hold granular transaction data. For example, an FIU that receives cross-border transaction reports can aggregate those reports according to the amount (for example, x% of transaction reports involve an amount between y and z). The FIU may also generate aggregate data based on criteria submitted by another authority that does not have full access to the underlying data.

In other cases, authorities may decide that collecting granular data is not necessary and therefore require financial institutions to aggregate data prior to submission. In order to obtain a consistent dataset, reporting requirements therefore need to also address the following aspects:

- Level of aggregation

Once a subset of transactions has been defined, one option would be to require financial institutions to report the aggregated characteristics of that subset (such as the total value and number of transactions). For example, an institution would indicate how many transactions above a certain threshold it has conducted during a given time period. While potentially useful for certain purposes, data aggregated in such a way is too general to provide a basis for a detailed analysis. Most authorities therefore require reporting entities to break down the data according to predefined criteria. The most common criterion is based on the origin or destination of funds (in other words, the 'corridor' in the case of cross-border transactions). In other terms, in the example above, financial institutions would be required to state the total number of relevant transactions conducted per corridor.

While almost all models surveyed for this research require data to be aggregated by corridor, other criteria vary. For example, certain supervisors require MSBs to aggregate data at the national level, whereas others ask for data to be aggregated at the level of individual agents. Authorities may also request data to be aggregated by the sector of the financial institution's customer, or develop more bespoke criteria for specific sectors, for example by requiring MSBs to aggregate transactions depending on whether the sender of a money transfer is a national of the country of destination.

- Number of reported data points

Reporting models also differ in the number of data points that financial institutions are required to report once they have aggregated the relevant data. While in certain countries institutions are required to report all data points (for example, the total value of relevant transactions on each corridor), other supervisors collect only the data points with the highest value (for example, the top five corridors for each institution).

- Frequency

The models surveyed for this research relied on monthly, quarterly or annual reporting requirements.

Non-Transactional Data

Considering that transaction data needs to be analysed in context, most methods of analysis developed by authorities also draw on additional datasets. One of the key benefits of using non-transactional data is the ability to establish links or clusters that may not be apparent from transaction data. Such links may be revealed through beneficial ownership and other data contained in company registers, information on business relationships with certain service providers,[13] real-estate ownership, media references or sanctions data (such as UN lists). Non-transactional data may also serve as a point of reference in identifying unusual financial flows at the macro- or microeconomic level. Such combined analysis may, for example, rely on international trade or shipping data, demographic data, criminal justice data, or indicators of an area's economic activity.

Other types of non-transactional data may be used to better assess the business model of regulated entities and adapt supervisory measures. This includes the number of customers from specific jurisdictions, the number of politically exposed persons holding accounts with a given institution (see for example the Financial Crime Report required by the UK's Financial Conduct Authority),[14] or the number of corporate customers versus individual customers. Analysing non-transactional data is particularly important with respect to forms of money laundering, such as trade-based money laundering, that do not rely on financial transactions, or to terrorist-financing cases that involve very small amounts. Where non-transactional data is obtained in text form, its analysis often requires the use of text-mining analytics that turn text into numeric representations that can be processed at scale.

Section C: Methods of Analysis

In all jurisdictions considered for this research, a significant amount of relevant data is already available. While the datasets can be enhanced to provide a more effective basis for data analytics, the main challenge for authorities is to make full use of the available data through appropriate methods of analysis. Some of these methods may be intended to better identify indicators of known risks. Other methods are exploratory and may reveal previously unknown patterns, such as correlations between two seemingly unrelated factors.

Consistent with the definition stated in the introduction, this section also includes use cases of data analytics that do not necessarily rely on machine learning, but on more traditional statistical methods. While techniques such as machine learning may lead to additional insights in the future, traditional methods are more accessible in the short term and allow for the development of datasets that can pave the way for more advanced techniques. Most methods are illustrated

13. For example, the information leaked through the 'Panama Papers' contained data on the activities and clients of the law firm Mossack Fonseca. The data on offshore firms incorporated by Mossack Fonseca was used by law enforcement agencies to investigate cases of alleged money laundering involving complex corporate structures.
14. Financial Conduct Authority, 'Annual Financial Crime Report', SUP 16 Annex 42A, 2016.

by a brief description of relevant use cases, keeping in mind that disclosing too many operational details in a public paper would compromise the effectiveness of the methods used.

Comparative Review of Institutions

As a minimum, an automated processing of aggregate data allows public authorities to compare indicators within a given sector or a peer group of comparable institutions and identify potential outliers. For example, one supervision model analysed during the research relies on the ratio of higher-risk transactions (such as cash transactions) over the total number of transactions conducted by each bank branch. For each peer group of banks, outliers can be identified, first at the local and then at the national level. A bank's unusually high ratio of high-risk transactions does not mean that the bank poses a higher risk overall, as it may have taken sufficient mitigation measures. Rather, the risk score allows supervisors to enhance their assessment of banks' inherent risk and adjust their priorities accordingly.

Another model relies on the ratio of activities reported as suspicious and high-risk transactions meeting predefined criteria. The number of SARs cannot be expected to be proportional to the overall volume of transactions. However, the authority using this approach indicated that an unusually low ratio (at the level of a region or an institution) would generally warrant additional verifications and has led to the successful detection of money-laundering schemes involving complicit or unwitting financial and non-financial institutions. By contrast, an unusually high ratio may point to defensive reporting.

Trend Analysis

A trend analysis is subject to at least two conditions. First, by definition, the analysis of a dataset over time presupposes the availability of historic data and, unlike the comparative review of institutions described above, is therefore not available immediately upon the introduction of a data collection system. Second, to provide a solid basis for a trend analysis, a dataset needs to be stable over time. For example, the criteria for the relevant universe of transactions should remain unchanged, barring any adjustments to monetary thresholds based on inflation. Assuming the two conditions described above are met, a trend analysis can reveal unusual developments pointing to potential cases of non-compliance, as illustrated in Box 1.

Box 1: Detection of Unusual Evolution of Remittances Through Trend Analysis of Aggregate Data

Using aggregate data submitted by financial institutions, the Italian FIU noticed a drop in reported remittances to China from €2.674 billion in 2012 to €557 million in 2015, during a period when remittances to other countries were broadly stable. Additional verifications revealed that a high proportion of remittance agents had moved to payment institutions that did not have sufficient controls in place and did not submit the required aggregate data to the FIU.

Source: Oral evidence by Claudio Clemente, Head of the Italian FIU (Unità di Informazione Finanziaria) at the Finance Committee of the Chamber of Deputies, 'Money transfer e prevenzione del riciclaggio e del finanziamento del terrorismo' ['Money Transfer and the Prevention of Money Laundering and Terrorist Financing'], 19 April 2016, p. 10.

Tracking transaction data over time can also prove valuable in assessing risk displacement following legislative or regulatory developments. For example, aggregate data may allow authorities to determine if the imposition of sanctions on country A is followed by an increase of financial flows to country B, which could point to evasion mechanisms.

A trend analysis can also be combined with the comparative analysis described above, to determine to what extent the evolution of a certain institution's operations on a particular corridor is consistent with broader trends within the relevant peer group.

Review of Consistency Between Datasets

Considering that transaction data may be difficult to analyse without additional context, a number of authorities review whether reported financial flows are consistent with the level of activity that can be expected, based on independent variables. This is based on the understanding that a discrepancy may point to higher risks of illicit financial flows and therefore inform a supervisor's risk-based approach. Table 2 provides examples of datasets that have been matched to detect such discrepancies.

Table 2: Examples of Combined Analysis of Transaction Data and Non-Transactional Data (Based on Case Studies)

Dataset 1 (dependent variable)	Dataset 2 (independent variables)	Example of Discrepancy
Aggregate data on remittances, broken down by location of sending agent and country of receiver	Demographic data	A high amount of remittances is sent to country A from a region in country B. However, only few nationals of country A live in the relevant region, and few nationals of country B live in country A. The level of financial flows suggest that the services of money remitters may be used for other purposes that remitting legitimate funds.
Aggregate data on number of cash transactions per municipality	Socio-economic characteristics of relevant municipalities	The application of an econometric model at the national level may indicate where, taking into account factors such as the level of financial inclusion and the number of bank branches, the number of cash transactions is unusually high, which can be an indicator of the size of the grey economy.

Source: Telephone interview with representatives of a national FIU, 8 March 2018; Guerino Ardizzi, Pierpaolo De Franceschis and Michele Giammatteo, 'Cash Payment Anomalies and Money Laundering: An Econometric Analysis of Italian Municipalities' (Rome: UIF-Banca D'Italia, Quaderni dell'Antiriciclaggio, 2016; forthcoming in the International Review of Law and Economics*).*

In addition to verifying the consistency between financial data and non-financial data, another approach relies on the comparison of complementary datasets, for example, through mirror statistics. Mirror statistics are defined as 'bilateral comparisons of two basic measures of a trade flow'.[15] Subject to a sufficient level of detail (for instance, sector or type of goods), foreign-trade data has been used to compare transactions reported in both the exporting and the importing countries. While small discrepancies may be explained through technical factors, larger or systematic discrepancies may point to illegal activities and trade-based money laundering. If a discrepancy is noticed for a particular type of goods, customs data can be used to identify the main importers and exporters of the relevant goods for further verifications.

15. OECD Glossary of Statistical Terms, 'Mirror Statistics', <https://stats.oecd.org/glossary/detail.asp?ID=6635>, accessed 22 October 2018.

Assuming a certain level of harmonisation in the collection of aggregate data on cross-border financial flows, a similar comparison could be conducted between financial flows from country A to B reported in country A, and the same financial flows reported in country B. A significant discrepancy between the two could, for example, be indicative of undeclared financial services in either of the two countries.

Network Analytics

Network analytics refer to a range of tools used to identify, analyse and visualise links and patterns within and across datasets. While network analytics are routinely used within financial institutions, applying those tools to datasets available to FIUs or supervisors makes it possible to identify and analyse networks across several financial institutions.

Network analytics are not a recent concept but have received wider application because of technological developments. As the examples in Boxes 2 and 3 show, network analytics have been applied to different datasets, ranging from SARs to larger datasets of objective transaction reports (where available).

Consistent with the general distinction discussed in Section A, network analytics have been used for both law enforcement and supervisory purposes. In the law enforcement context, the main objective is to detect or better understand networks around individuals or entities suspected of illegal conduct. In the case of supervision, the primary purpose of network analytics is to better assess financial institutions' exposure to networks posing a higher risk.

Box 2: Application of Network Analytics for the Detection of Undeclared Remittance Services

Undeclared money transfer services are not subject to regulatory oversight and pose a high risk of being misused for terrorist-financing or money-laundering purposes. Network analytics have been used to detect such activities by establishing a connection between multiple transactions that may involve separate financial institutions or branches, and that do not raise a suspicion when evaluated individually.

According to a typologies report published by the Australian FIU, indicators for undeclared money transfer services include:

- Large cash deposits made at different bank branches on the same day.
- Large international funds transfers to common beneficiaries.
- Multiple customers conducting international funds transfers to the same overseas beneficiary.
- Multiple international funds transfers to different beneficiaries sharing the same overseas address.
- Multiple international funds transfers undertaken through different remitters on the same day.

Source: AUSTRAC, 'Typologies and Case Studies Report 2014', Commonwealth of Australia, 2014.

Box 3: Application of Network Analytics for the Risk Mapping of Supervised Entities

The Netherlands: Based on the transaction data it receives from MSBs every quarter, the Dutch Central Bank (DNB) has been able to detect networks of individuals sending funds to the same counterparties in high-risk countries via several businesses. This information can then be used to identify which businesses and agents are most exposed to these networks, which is not necessarily indicative of non-compliance but may warrant more intensive supervision. As part of its effort to strengthen the sector's own preventive measures, the DNB has informed relevant businesses of some of its findings (for example, the network portion involving a specific business).

Singapore: The Monetary Authority of Singapore (MAS) applies network analysis to Suspicious Transaction Reports (STRs) to identify groups of related STRs filed across financial institutions, and over time. This analysis, combined with other information and data sources, allows MAS to identify and focus supervisory attention on financial institutions affected by networks of higher-risk accounts, entities or behaviour. In some cases, this analysis could also uncover potentially illicit activities which would be referred to law enforcement agencies for investigation.

Source: Presentation by a representative of the DNB at RUSI's public conference on 'Financial Crime 2.0', London, 27 November 2017; Speech by Chua Kim Leng, Special Advisor, Financial Supervision Group, Monetary Authority of Singapore, at the AML/CFT Industry Partnership Dialogues, Singapore, 14 May 2018, <http://www.mas.gov.sg/News-and-Publications/Speeches-and-Monetary-Policy-Statements/Speeches/2018/Speech-at-the-AMLCFT-Industry-Partnership-Dialogues.aspx>, accessed 23 September 2018.

In any case, the datasets need to include criteria that are detailed enough to allow for links between transactions to be established. For example, the names of the parties to a transaction may be used to determine if two transactions involve the same individual. However, a system that relies solely on names to connect data points is vulnerable to false negatives based on differences in spelling. Network analytics therefore often use additional criteria (such as ID number, date of birth/registration, account number) for purposes of entity resolution. Based on the number of criteria that data points have in common, probabilistic matching algorithms assess the likelihood that those data points are identical.

Transactions may be linked through criteria other than identifiable information on the sender or recipient of a transaction, allowing for the use of network analytics beyond the use cases reviewed for this research. For example, a connection may be established between a set of credit cards that have been issued by different institutions to different customers but are repeatedly used at the same cash machines within a few minutes. Such a connection can be an indicator of a money-laundering typology where cards issued to victims of human trafficking are controlled by a single person.

Even in the absence of transaction data, network analytics may help supervisors and law enforcement gain a clearer understanding of large sectors with a high turnover, such as the sector of MSBs. The MSB sector is characterised by a high number of registered businesses and

an even higher number of agents, who may offer financial services on behalf of more than one principal. The complexity of the sector is compounded by its turnover (with businesses being set up and shut down within a short period of time) and its interconnectedness (with several businesses sharing the same director or the same address, for example).[16] Applying network analytics to information submitted by MSBs upon registration (for example, name of director or address) would reveal clusters of businesses that can otherwise only be identified through a manual and reactive case-by-case review.

Outlook: Big-Data and Machine Learning Techniques

Consistent with a broader trend across sectors, big-data and machine learning techniques have the potential to significantly enhance the outcomes of all methods described above. Clusters, patterns and anomalies can often only be revealed through the combination of a range of data points, including both transaction data and non-transactional data. While the use cases discussed in this chapter are designed to identify inconsistencies between a limited number of indicators, big-data techniques can enrich such analysis with unstructured data. For example, the big-data platform recently developed by an FIU that was visited for this research draws on press articles to more accurately assess ties between third countries and high-risk jurisdictions as another factor contributing to the contextual monitoring of financial flows.[17]

By allowing for a more flexible approach than traditional statistical methods, machine learning is also likely to further increase the ability of supervisors, FIUs and law enforcement agencies to detect anomalies beyond the patterns revealed through the methods described above. In turn, a better understanding of anomalies will allow for a more effective use of supervisory resources, and for the more effective and efficient investigation of money-laundering cases.

For the time being, one of the challenges to deploying machine learning in the context of AML supervision and enforcement is the need to validate data, or confirm whether an individual pattern was indicative of criminal conduct. While critical to the machine learning process, validation can be difficult in the context of money laundering, where investigations may take several years before leading to an outcome. More general challenges relate to the availability of qualified staff and of long-term funding, to the ability of law enforcement and supervisors to extract relevant information from unstructured data, and to the uncertainty over legislative and ethical barriers to the use of big-data techniques.[18]

16. See Olivier Kraft, 'Money Service Businesses in the UK: Improving the Conditions for Effective Financial Crime Supervision and Investigations', *RUSI Occasional Paper* (January 2018).
17. Author visit to FIU in EU member state, 21 March 2018.
18. John Coyne and Amelia Meurant-Tompkinson, 'I Can See Clearly Now! Technological Innovation in Australian Law Enforcement: A Case Study of Anti-Money Laundering', Australian Strategic Policy Institute, 19 July 2018.

Notwithstanding these challenges, supervisors and law enforcement in certain jurisdictions are currently making initial attempts at using machine learning for AML supervisory purposes.[19]

Chapter Summary

The use cases discussed in this chapter reflect the many ways in which data analytics have been – or in the future could be – used to strengthen AML efforts. While the use cases surveyed for this research are all intended to help authorities gain a more granular and real-time understanding of risks, a more detailed taxonomy is necessary for a more structured and informed discussion on the role of data analytics. As explained above, several sub-categories may be distinguished according to the following criteria:

- **Objectives:** Analytics may be used to detect, analyse and investigate criminal conduct, or to enhance a sector's risk-mapping and inform the supervisor's risk-based approach.
- **Underlying data:** Analytics rely on a multitude of datasets that vary in nature, in scope (for example, data on suspicious transactions versus data on sets of transaction defined by objective criteria) and level of detail (for example, aggregate versus granular data).
- **Methods of analysis:** Subject to its scope and level of detail, each dataset may provide the basis for a range of different methods of analysis, including trend analysis and network analytics, which are likely to be strengthened by big-data and machine learning techniques.

As data analytics pursue different objectives and have been deployed in different legal and economic contexts, a general conclusion on the most effective approach is not advisable. However, existing use cases illustrate the various degrees to which methods of analysis, when used with a clear objective, have, for example, allowed FIUs to proactively identify unusual financial flows at the macro- and microeconomic levels, and supervisors to better identify high-risk institutions and strengthen AML prevention efforts. On that basis, the next chapter identifies factors to consider when selecting and implementing an approach.

19. Financial Stability Board (FSB), 'Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications', 1 November 2017, p. 23.

II. Devising and Implementing a Strategy for the Use of Data Analytics for AML Purposes

BUILDING ON THE overview provided in the previous chapter, this chapter offers guidance for authorities when setting up, implementing and assessing strategies to leverage data analytics for AML purposes. Section A sets out several factors that should be considered at the outset, and which may vary from one jurisdiction to another. Section B then discusses the practical process that may be followed by authorities seeking to make (greater) use of data analytics.

Section A: Factors Shaping the Approach to Data Analytics

The specific approach adopted by each jurisdiction is a function of a number of factors that vary from one jurisdiction to another, and which therefore contribute to international differences. For ease of consideration, these factors can be grouped into two categories, depending on whether they relate to the broader national context in which AML efforts are implemented (contextual factors) or to the relevant authorities seeking to use data analytics (operational factors).

Contextual Factors

Under the FATF Recommendations, AML efforts are meant to be risk based, taking into account the context of the relevant jurisdiction. This approach also applies to the use of data analytics.

Threat Landscape and Structure of Economy

According to the FATF, contextual factors include the nature and extent of money-laundering risks, as well as the characteristics of a given jurisdiction's economy and of its financial sector.[20] While these factors' impact on the design and implementation of data analytics is discussed in more detail in the next section, Table 3 provides some examples for clarity.

20. FATF, 'Methodology for Assessing Compliance with the FATF Recommendations and the Effectiveness of AML/CFT Systems', 2018, p. 5.

Table 3: Link Between Contextual Factors and Relevant Method of Data Analysis

Example of Contextual Factor	Impact on Method of Data Analysis
Level of cash usage in everyday life	The threshold for currency transaction reports may need to be higher in a country where cash is regularly used for large, legitimate purchases.
Size of financial sector	Assuming the same criteria are used for systematic reporting obligations in two countries, the amount of transactions meeting those criteria (and therefore the data to be processed) is likely to be significantly larger in a jurisdiction with a larger financial sector.
Predicate crimes	The predicate crimes generating the most illicit proceeds in a given jurisdiction (and the related money-laundering typologies) will determine which transaction data is most relevant. For example, a country faced with high levels of tax evasion is more likely to collect information on cross-border transfers.
Participation in international economic area	Where economies are integrated, for example in the EEA where financial institutions authorised in one member state can offer services in other member states, the effectiveness of data analytics may depend on the level of international harmonisation.

Source: Author's research.

Standards Governing the Collection and Handling of Data

The use of data analytics for AML purposes needs to comply with applicable data standards. Under the EU's new General Data Protection Regulation, for example, the collection of data must be necessary and proportionate to the objective pursued. Additional requirements apply in the case of data transfers to a third country outside the EU. Ethical standards may also determine how data can be lawfully used, for example by requiring that the analytical process be sufficiently transparent and give sufficient weight to a human review. This is particularly important when, as in the case of big-data techniques, personal data is extracted from several sources to develop a consolidated risk assessment. The implications of ethical standards are discussed further in Section B.

Other Policy Objectives

AML objectives need to be reconciled with other policy objectives, often represented by other authorities or departments. In particular, given that the use of data analytics is often predicated on the collection of additional data from financial institutions, these initiatives are generally balanced against the need to limit the regulatory burden on the sector. The weight given to this consideration varies from one jurisdiction to another.

Operational Factors

In addition to the contextual factors listed above, the research for this paper revealed several differences between the specific authorities that make use of data analytics for AML purposes.

Objectives and Priorities

The use of data analytics should always serve clear objectives. Consistent with the distinction described in Chapter I (Section A), the use of data analytics may serve a number of purposes. In particular, it may be targeted towards individual cases of suspected criminal conduct, or focused on the risk level of regulated entities. Within each of these categories, objectives may be informed by law enforcement or supervisory priorities. It should, for example, be clarified whether a method is intended to address money-laundering risks, terrorist-financing risks, or both. If the method is focused on money-laundering risks, it may prioritise the threat associated with a specific predicate crime (such as tax evasion, corruption, or drug trafficking) or a specific sector for which the risk mapping needs to be enhanced through data analytics.

Human and Technological Capabilities

During the research for this paper, significant differences were identified in the level of capabilities available for the analysis of data, in terms of both available human and technological capabilities. In particular, the use of data analytics for law enforcement and supervisory purposes requires not only investment in relevant tools, but also a more diverse skill set among staff, including qualified statisticians or mathematicians. If personal data is collected and processed, additional resources may be required to ensure adequate protection from internal leaks and external hacking. Certain countries are currently seeking to maximise capabilities through public–private collaboration, such as the Fintel Alliance in Australia or the National Economic Crime Centre in the UK.

The availability of human and technological capabilities has a direct impact on the balance with data protection standards. For example, if data cannot be effectively analysed and used for AML purposes, the benefits of collecting more data may be insufficient to justify privacy implications.

Access to Data

The diversity of methods developed across jurisdictions also reflects the different institutional set-ups of AML regimes, which are often the result of historic or legal considerations specific to each jurisdiction. For example, the FATF Recommendations acknowledge that FIUs may either be established as a stand-alone agency, or as part of an existing authority.[21] The FIU function is fulfilled by the same authority in charge of AML supervision in certain jurisdictions, or by a law enforcement agency in others. Subject to data protection rules governing the purposes for which specific data may be used, the FIU model has a direct impact on the types of data that

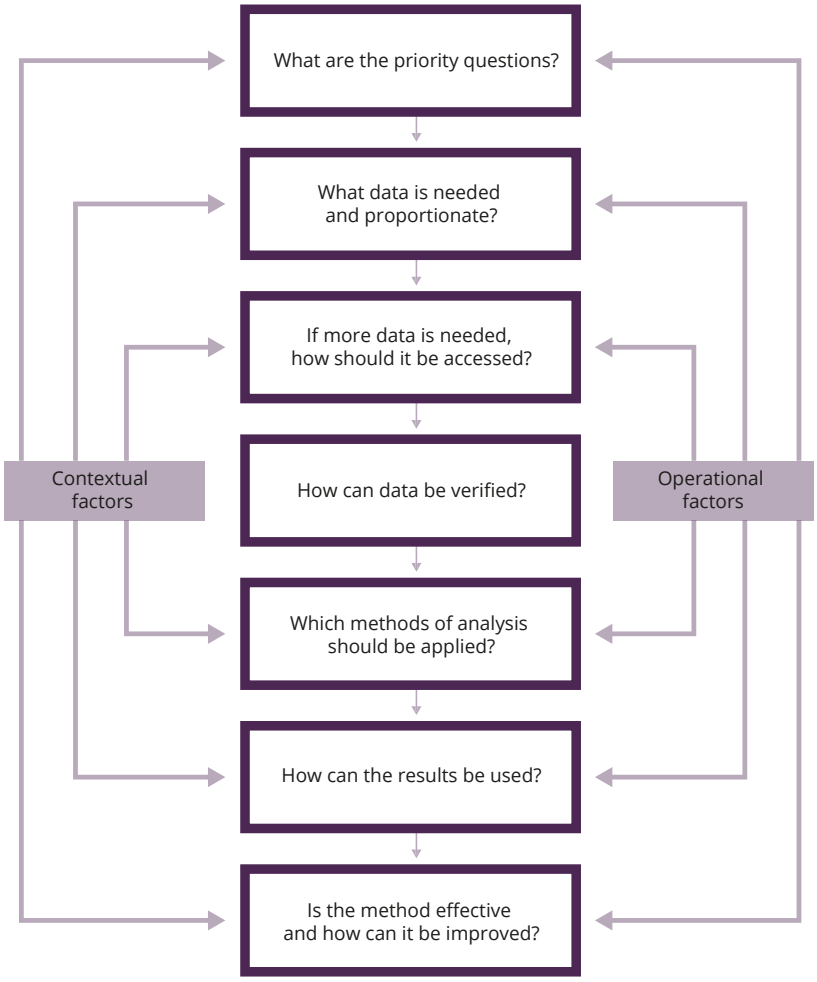
21. FATF, 'Interpretative Note to Recommendation 29'.

can be accessed by the FIU and other authorities. Where the FIU is distinct from other AML authorities, access to data depends on the level of inter-agency coordination.

Section B: Key Steps for the Use of Data Analytics in AML Intelligence, Supervision and Enforcement

The process for developing and implementing a data analysis for AML purposes may be broken down into seven steps: (1) determining specific questions; (2) identifying relevant data, starting with existing data; (3) obtaining access if needed; (4) ensuring data quality; (5) conducting the analysis; (6) using the results; and (7) assessing effectiveness. For ease of reference, these steps are summarised in Figure 3.

Figure 3: Key Steps for the Use of Data Analytics in AML Supervision and Enforcement



Source: Author's research.

Determining Priority Questions

Data analytics are not a panacea that can solve the issue of money laundering. As a first step, it is essential for authorities to clarify the priority questions that they would like to answer through data analytics and the underlying rationale.

The questions are generally based on contextual factors as well as operational factors such as law enforcement and supervisory needs. It is therefore essential for data analysts to work with the end users of the analysis (such as investigators or supervisors) from the outset. This will ensure that the questions are relevant, and that the outcomes are used in practice. Table 4 provides examples of potential questions to be answered through the use of data analytics, together with their underlying rationale.

Table 4: Examples of Priority Questions and Underlying Rationale

Research Question	Rationale
How is a specific institution exposed to a set of high-risk jurisdictions?	A high exposure to certain jurisdictions should be matched by commensurate risk mitigation measures.
Which geographic areas register financial flows that are not commensurate with local characteristics?	A discrepancy between reported financial flows and other socio-economic characteristics of a particular region may be an indicator of the size of the illegal or grey economy.
What is the impact of financial sanctions or conflicts on financial flows to specific countries?	Analysing the correlation between financial flows and sanctions may reveal corridors used to circumvent regulatory hurdles.
Are there any clusters of (active and past) MSBs that are connected despite seemingly unrelated names?	A pattern of repeated registrations and de-registrations of MSBs at the same address or by the same person may indicate an attempt to reduce the supervisor's visibility over a business's long-term activity.

Source: Author's research.

Identifying Relevant Data

As a next step, the relevant authority should determine what data is needed to answer a specific question. This will be based on the existing knowledge of the threat under consideration (based on SARs, for example) and can be tested on a data sample before being used at scale. For example, Table 5 reflects how the nature of a threat may inform the definition of relevant transactions.

Table 5: Link Between Priority Threats and Criteria That May be Used to Define Relevant Subsets of Transactions

Transaction Characteristic	Role in Defining Datasets Relevant to Specific Threats
Amount	For money laundering related to certain types of predicate offences (for example, high-level corruption or tax evasion), transactions above a certain threshold may be more relevant. By contrast, given that terrorist financing often involves small amounts, a dataset designed to address terrorist financing should generally include transactions involving smaller amounts. Data on transactions involving an unusually small amount (such as €1) may also be relevant for money-laundering purposes, as such transactions can be used by criminals to send a signal to their counterparts.
Destination or origin	Depending on the research questions, the relevant set of transactions may focus on jurisdictions with a high level of financial secrecy that are often used for tax evasion, on jurisdictions in or bordering a conflict zone, on jurisdictions with close ties to a sanctioned country, or on jurisdictions known to be the origin of illegal commodities, such as narcotic drugs.
Method of payment	The role of this criterion depends on whether the predicate offence is known to generate proceeds in a particular format (such as cash proceeds or cryptocurrencies).

Source: Author's research.

The definition of the relevant datasets should be carefully considered, given that changing criteria too frequently has an impact on the consistency of the dataset over time and therefore on the opportunities for monitoring trends. Regardless of the dataset, the relevant criteria must be clearly defined and explained to ensure consistency across institutions and over time. For example, the term 'cash transactions' may refer to a withdrawal, a deposit, a payment, or other transactions.

In addition, the choice of a dataset should take into account the risk that criminal actors may adapt their methods. For example, the collection of granular transaction data based on monetary thresholds may lead to a higher number of transactions under that threshold (known as 'smurfing'). One way to address the risk of circumvention is to complement granular data on a subset of transactions with aggregate data on a larger set. For example, if cash transactions have to be reported above a certain threshold, the risk of 'smurfing' can be assessed based on the mean value of all cash transactions. There will generally be a range of datasets potentially relevant to any given objective, though with varying degrees of usefulness. For example, data aggregated at the level of each bank branch/agent naturally allows for a more granular analysis than data aggregated at the level of a jurisdiction. Similarly, monthly data may reveal more insights than annual data. In any event, the most effective analytical methods generally draw on multiple datasets, including non-transactional data. Consistent with the factors described above, the choice of a relevant datasets will depend on the following.

Data Protection Standards

In most countries surveyed for this research, the collection of data is subject to a legal assessment that the data in question is necessary to a certain objective (for instance, AML), and that the impact on privacy is proportionate to the objective. The concept of proportionality means that, even if data is deemed potentially relevant, a determination needs to be made as to whether its value is such that it justifies the related privacy restrictions. This determination should take into account whether the same objective can be achieved with a dataset that would be less intrusive.

For example, if an agency has identified cross-border transactions on a specific corridor as posing a higher risk, it could collect individual data on all transactions on this corridor, individual data on transactions above a certain threshold, aggregate data on relevant transactions, or other datasets. Each of these datasets has different degrees of operational relevance and intrusiveness. Proportionality requires a balance between these two factors.

Types of Data Already Available

Making greater use of data does not necessarily imply collecting additional data. Rather, priority should be given to relevant data that is already available, sometimes to other national agencies or to other departments of the same agency. For example, central banks often collect data on cross-border flows to monitor the balance of payments. Using existing data for a new purpose (such as AML) may require a legal assessment under data protection standards to ensure that such use is necessary and proportionate to AML objectives. Assuming these conditions are met, coordination among and within agencies avoids the unnecessary duplication of data collection initiatives. Even if additional data is collected, coordination with existing mechanisms increases the opportunities for a consolidated analysis of multiple datasets.

Economic Factors

Various economic factors such as the level of financial inclusion or the level of cash use should be considered. The threshold for currency transaction reports would generally be higher in countries in which cash is regularly used for large, legitimate purchases.

Sectors and Business Models

The most relevant dataset is likely to be different from one sector to another. For example, an approach developed for the MSB sector may not work for credit institutions. In addition, traditional datasets may need to be adjusted to account for the emergence and business model of new sectors, such as electronic payment providers. Within each sector, different approaches may be justified depending on institutions' business lines (for instance, retail banking versus investment banking) or size.

Consultations with the Private Sector

Any reporting requirements calls for additional resources from regulated entities. As it is in the interest of the AML system that available resources (both in the public and private sectors) are used efficiently, engaging with the private sector allows authorities to determine how data is currently structured and to take those arrangements into account when deciding what data to use. Drawing on existing systems rather than creating a new structure not only creates efficiencies, but is also likely to lead to better data than having two parallel systems. For example, most authorities that collect and analyse data on cross-border transactions rely on the Society for Worldwide Interbank Financial Telecommunication (SWIFT) format that has been used by the financial sector for many years, so this pre-existing format may be effectively adopted for reporting purposes.

International Dialogue

While national characteristics may justify different approaches, authorities should also consider engaging with their international counterparts when deciding what data to use for AML purposes. Such dialogue may initially be beneficial for facilitating peer learning. In addition, early consultations might also avoid unnecessary differences and facilitate a more consolidated analysis of risks. This is particularly important for countries that are part of an economic block in which financial institutions are allowed to provide services across borders (such as the EEA), but often have to comply with different reporting regimes in each country. Consultations may also be possible in global organisations such as the IMF, which provides guidance on other statistical datasets, or the FATF, which has the mandate to lead international efforts against money laundering.

Obtaining and Regulating Access

Authorities may have different options for obtaining access to the data identified as relevant. The most common form for accessing the data is for relevant entities to file reports with a designated agency. The agency acts as the central repository and holds the data. Alternatively, new technologies may allow authorities to analyse encrypted data without storing it in a central location, thus limiting privacy implications. For example, an agency would be able to run algorithms on a large dataset and see the results of the process without reviewing all the underlying data. This is subject to the development of compatible systems between financial institutions and public authorities, and is currently considered in the context of various financial intelligence-sharing partnerships.[22] If a new reporting system is seen as necessary, the following factors should be considered.

22. See Nick J Maxwell and David Artingstall, 'The Role of Financial Information-Sharing Partnerships in the Disruption of Crime', *RUSI Occasional Papers* (October 2017).

Format

One common obstacle to the use of transaction reports (including SARs) as a basis for data analytics is their format. A SAR submitted in hard copy, or as a scanned PDF, may contain useful information, but cannot be used unless its contents are first translated into a machine-readable format. The use of online forms for the submission of reports will help structure data to be more effectively processed.

Coordination with Existing Reporting Regimes

To encourage effective and accurate reporting, authorities should consider how additional reporting requirements may be consolidated with existing regimes. The coexistence of separate reporting regimes serving related objectives can have an impact on the willingness and ability of regulated entities to meet their various obligations.

AML authorities should also engage with their counterparts in other areas of regulation to identify lessons learnt from other reporting systems (for example, the disclosure and analysis of trading data in the context of market abuse supervision)[23] and potential synergies with existing standards (such as the legal entity identifier, which facilitates entity resolution across datasets, meaning that identical data points can be more easily recognised and linked).

Capacity of Reporting Entities

Any reporting regime should consider the capacity of various sectors or various parts of a given sector to submit required information. In certain cases, considering the amount of data to be provided and/or the available resources of various institutions, tailored reporting models may be used to collect similar information from different sectors. The Australian AML regime, for example, foresees two different models for the reporting of international funds transfers, depending on whether the regulated entity is a bank or a money remitter.[24] Offering a free or low-cost IT solution to smaller financial institutions can limit the costs associated with additional reporting requirements.

Engagement with Reporting Entities

Engaging with all reporting entities is critical when introducing a new reporting regime. First, engagement is necessary for the successful and consistent implementation of any reporting requirements. Questions may relate to definitions, but also to the scope of reporting requirements (for example, in the case of domestic institutions with foreign branches, or foreign institutions operating in the relevant jurisdiction through agents). Second, interventions made during the workshop held for this research confirmed the importance of communicating the

23. See, for example, the data submitted under the EU's Markets in Financial Instruments Directive.
24. See Attorney-General's Department, Australian Government, 'Statutory Review of the Anti-Money Laundering and Counter-Terrorism Financing Act 2006', April 2016, p. 71.

rationale and expected benefits of the data collected. Such transparency contributes to the private sector's commitment to the regime, and therefore to the quality of the data submitted.[25] Third, authorities should clarify how any new reporting requirements relate to existing ones. This is particularly true in cases in which reporting entities are required to submit data on non-suspicious transactions. In those cases, it is essential for financial institutions to recognise that the new requirements do not limit the general obligation to have systems in place to detect suspicious transactions and to report those to the FIU.[26]

Controlling Access to Data

In addition to considering the proportionality of additional data collection, data protection standards also require that access to any personal data should be granted on a need-to-know basis. This is particularly true for models requiring the reporting of transactions that are not deemed suspicious. If such a dataset is meant to serve exclusively as a basis for data analytics (as opposed to screening for specific names), full access may be limited to those data analysts who work directly with the data. To maximise the usefulness of available data without compromising privacy, authorities holding larger amounts of data can monitor trends on behalf of other authorities. For example, one FIU consulted during this research provides aggregate data on reported transactions to the supervisor, without sharing the underlying data.

Ensuring Data Quality

Considering that the usefulness of data analytics depends largely on the quality of the underlying data, the development of a new method should include a process to ensure the quality of data and, where several authorities or departments are involved, assign clear responsibilities to this end. Authorities consulted during this research reported investing staff resources into quality assurance. In other words, the data collected is generally reviewed upon receipt to ensure that it is complete and to detect any apparent inconsistencies (for example, by comparing the amount of reported transactions with previous reports filed by the same institution). If needed, the reporting institution may be asked to verify and confirm the data. The burden on staff may be reduced by adding data validation in templates or providing software to reporting entities.

Data Analysis

This is the stage at which methods of analysis are applied to the datasets. For examples of such methods, please refer to Chapter I.

25. Workshop on 'Contribution of Advanced Analytics to AML Supervision and Enforcement'.
26. 'Code monétaire et financier 2018 (France)', Article L561-15-1, Section II, states that the submission of a currency transaction report is without prejudice to the obligation to file a SAR in the case of a suspicion.

Using Results

Data analytics can only be effective if the results of the analysis are subsequently used and acted upon. Conversely, the lack of such use can affect the assessment of whether the collection and processing of data is proportionate to the benefits. Irrespective of the methods used to analyse data, the following observations should be considered from the outset.

Disseminating Results

One of the challenges cited by authorities interviewed for this research who had invested in data analytics was to make sure that the results of the analysis were effectively taken into account by the intended 'users', including front line supervisors or investigators. This is particularly true when those users are not familiar with data analytics and their potential value.

In addition to involving operational staff when defining the priority questions, authorities should ensure that those staff are made aware of the availability and results of data analytics. For example, if a specific analytical product can be generated by an FIU upon request from police forces (for example, a 'financial image' of a region based on aggregate transaction data), this possibility needs to be proactively communicated. Similarly, if data analytics are meant to inform supervisory priorities, their results should be integrated into standard processes. Where appropriate, insights drawn from analytical tools should also be shared with relevant regulated entities. In any case, the results of data analytics should be 'translated' from scientific language into standard language to encourage their wider use.

Anticipating and Mitigating Risks

In addition to the benefits of data analytics, potential risks should also be considered and addressed. This includes, for example, setting parameters for the use of analytical tools.

- Permitted uses

By definition, objective transaction reports relate to transactions that meet a certain set of criteria but are not necessarily deemed suspicious. The use of such reports to identify new leads for law enforcement (rather than assisting in an existing investigation) may therefore raise concerns of surveillance. As a result, certain jurisdictions have set restrictions on the purposes for which these reports may be used. In France, for example, the FIU cannot open an investigation based solely on currency transaction reports.[27] In the Netherlands, the

27. Tracfin, 'Communication systématique d'informations relative aux transmissions de fonds', <https://www.economie.gouv.fr/tracfin/communication-systematique-dinformations-cosi-relative-aux-transmissions-fonds>, accessed 30 July 2018: 'La communication systématique d'informations (COSI) ne nécessitera aucune analyse et ne sera la manifestation d'aucun soupçon. Elle ne permettra pas de fonder la conduite d'investigations et n'entraînera pas d'exonération de responsabilité pénale, civile et professionnelle du déclarant'.

systematic transaction data submitted by MSBs is used solely for the purposes of supervision (see Box 3). Any supervisory action based on the analysis may subsequently lead to a referral to law enforcement, if appropriate. However, the underlying data cannot be shared in bulk.

- Ethical framework

As noted in the aforementioned report of the FSB, the reliance on AI and machine learning in supervision may introduce new risks if those technologies are 'used without a full understanding of the underlying methods and limitations'.[28] As in other areas, the use of data analytics – and in the future big data and machine learning – for the purposes of AML supervision and enforcement should therefore be subject to a clear and robust ethical framework. This includes, for example, ensuring the transparency of the analysis, clarifying how results should be treated, and maintaining accountability for the subsequent decision-making process.

Assessing Effectiveness

Assessing effectiveness is critical for improving data analytics, but also for confirming whether the benefits are proportionate to the privacy implications and the resource implications for public and private stakeholders. The assessment will naturally depend on the priority questions and the underlying objectives set at the beginning of the exercise.

In the law enforcement or FIU context, it is not always possible to measure the value of a specific source of intelligence in isolation. The assessment of effectiveness is therefore likely to be based on qualitative data about the role of data analytics for both strategic and operational purposes, particularly considering to what extent data analytics have contributed to the detection of previously unknown money-laundering cases, patterns or typologies.[29] The assessment can also consider to what extent the use of data analytics has increased the efficiency with which the FIU can process and prioritise incoming SARs.

From a supervisory perspective, the main objective is to determine to what extent data analytics contribute to an improved risk-mapping. Two supervisors indicated, for example, that the results of risk-scoring tools using data analytics had been subsequently matched against the results of on-site visits to the relevant institutions, to determine whether both were consistent.[30] More generally, the ratio of inspections that conclude with a request for corrective measures can serve as an indicator of a risk-scoring tool's accuracy. An increase in that ratio following the introduction of an analytical tool may therefore also be a sign of effectiveness.

28. FSB, 'Artificial Intelligence and Machine Learning in Financial Services', p. 32.
29. See Financial Crimes Enforcement Network (FinCEN), 'Implications and Benefits of Cross-Border Funds Transmittal Reporting', January 2009, Appendix B: Financial Intelligence Unit Letters of Support, Letter from Neil J Jensen, CEO of AUSTRAC.
30. Telephone interview with AML supervisor in EU member state, 8 March 2018; Workshop on 'Contribution of Advanced Analytics to AML Supervision and Enforcement'.

Assessment methodologies should avoid 'self-fulfilling' performance indicators based on spurious correlations. For example, conducting more inspections in financial institutions may reveal more cases of non-compliance in those institutions, but this increase may only be a function of increased scrutiny rather than an indication of an effective allocation of supervisory resources.

Over time, it should be assessed to what extent the use of data analytics has contributed to the quality of SARs, in particular by reducing the number of false positives (such as SARs that do not relate to any illegal activity) and the number of false negatives (such as activities that are not reported despite involving criminal proceeds).[31] This presupposes that the results of the analysis are fed back to the private sector, through guidance and more targeted risk assessments. The assessment of effectiveness should also inform improvements to the analytical tool, including by adjusting the underlying questions and the criteria to define the most relevant and proportionate datasets.

Chapter Summary

For data analytics to meet the high expectations that are often voiced, the underlying method should be based on a clear strategy. As discussed in this chapter, the strategy should be informed by several contextual factors (such as the threat landscape, data protection standards, and other policy objectives) and operational factors (such as priorities, human and technological resources, and access to data).

Based on these factors, the implementation should include the following steps:

- Determining the questions to be answered and the underlying objective
- Identifying the relevant data
- Obtaining and regulating access
- Ensuring data quality
- Analysing the data
- Using the results (within an ethical framework)
- Assessing the effectiveness of the model

Over time, the development of data analytics should also benefit from the experience of a growing number of AML authorities in using data analytics, including the use cases described in the first part of this paper.

31. Donato Masciandaro (with Lucia Dalla Pellegrina, Giorgio Di Maio and Margherita Saraceno), 'I-Type and II-Type Errors in Reporting Suspicious Transactions', 2018, unpublished.

Conclusion and Recommendations

DATA ANALYTICS ARE not a panacea for the challenges of today's AML system. However, as this paper shows, they have a significant potential to enhance the effectiveness of AML intelligence analysis, enforcement and supervision.

Harnessing opportunities associated with data analytics will require a more comprehensive dialogue on this topic, both at the national and the international level. To support this dialogue, this paper has taken stock of the initiatives undertaken by authorities in several jurisdictions. Existing use cases vary in many respects, including their purpose, the underlying data, and the method of analysis. The choice of an approach should be informed by several contextual and operational factors. In particular, clarifying specific objectives and verifying the potential value of existing data before collecting additional data can help ensure that the use of data analytics is aligned with data protection principles. In light of this, it is recommended that:

- **Supervisors, financial intelligence units and relevant law enforcement agencies** should adopt a strategy to identify and harness opportunities for the use of data analytics in an AML context, taking into account available use cases and the factors discussed in this paper and in consultation with relevant private sector actors. This will provide a stronger basis for a transparent discussion with policymakers on necessary long-term investments and the balance with privacy considerations.
- **Regional organisations such as the EU** should foster a dialogue on the different approaches developed in this area by member states, to encourage a harmonised approach. Such coordination will reduce compliance costs for businesses operating in several countries and is likely to benefit the quality and usefulness of the data submitted under each of the national regimes.
- **Assessment bodies such as the FATF and the IMF** should more systematically review to what extent member countries have considered and harnessed the potential of data analytics for AML intelligence, supervision and enforcement purposes, which will allow for peer learning.

As the pace, complexity and size of the financial system continue to increase, the effectiveness of global AML efforts will be contingent on the readiness of governments to innovate in a strategic and coordinated manner.

About the Author

Olivier Kraft is a Research Fellow at RUSI's Centre for Financial Crime and Security Studies. His research interests include the opportunities of new technologies for anti-money laundering (AML) efforts, and the synergies between AML and cyber-security measures. Prior to joining RUSI in 2017, Olivier worked with the Financial Action Task Force, the global standard-setter in the areas of AML and counterterrorist finance (CTF), where he focused on evaluating the effectiveness of countries' AML/CTF efforts. From 2011 to 2015, Olivier advised the World Bank Group Sanctions Board on allegations of fraud and corruption in development projects co-financed by the World Bank Group.