BIS Innovation Hub

Project Aurora

The power of data, technology and collaboration to combat money laundering across institutions and borders

May 2023

Preface: How to read this report

For an overview of the project, useful background and key findings read:

- Chapter 2: Executive summary
- Chapter 3: Introduction
- Chapter 4: Project Aurora – proof of concept
- Chapter 6: Conclusion

For more supporting details read:

- Chapter 5: Further considerations
- Annex A: Trends and opportunities

For other details read:

- Annexes B - E

Project Aurora was delivered in partnership with:

Publication date: May 2023.

Contents

1. Acronyms, abbreviations and definitions 7

2. Executive summary 10

1.1 Background 10

1.2 Data, technology and innovation 11

1.3 Project Aurora 12

1.4 Findings and key takeaways 12

1.4.1 A holistic view of payments data unveils money laundering networks 13

1.4.2 Behavioural monitoring and privacy enhancing technologies could be a game changer for AML efforts 13

1.4.3 Leveraging Project Aurora 14

3 Introduction 16

3.1 What is money laundering? 16

3.2 The money laundering process 17

3.3 AML monitoring and analysis today 18

3.3.1 Rule-based monitoring systems 18

3.3.2 Behavioural monitoring systems 19

3.3.3 Defensive reporting and de-risking consequences 19

3.4 Challenges facing AML efforts 20

3.5 Technology 22

3.5.1 Privacy-enhancing technologies (PETs) 22

3.5.2 Graph data structures 23

3.5.3 Machine learning 23

3.5.4 Network analysis 24

3.6 Summary of trends and opportunities 24

4. Project Aurora – proof of concept 26

4.1 Objectives and scope 26

4.1.1 Objectives 26

4.1.2 Scope 27

4.2 Part A: Synthetic data generation 29

4.2.1 Purpose of generating synthetic data 29

4.2.2 Generating the synthetic data 30

4.2.3 Leveraging a three-step approach to generate synthetic data 31

4.2.4 Constructing money laundering activities in the synthetic data set 34
4.3 Part B: Application of machine learning to the synthetic data set 39
4.3.1 Machine learning models 39
4.3.2 Testing the models 41
4.3.3 Results 42
4.3.4 Summary 46
4.4 Part C: Testing privacy-enhancing technologies for AML 48
4.4.1 Privacy-enhancing technologies explored 48
4.4.2 Testing a combination of privacy-enhancing technologies in four different collaborative analytics and learning arrangements 49
4.4.3 Results: applying machine learning models in combination with privacy-enhancing technologies 55
4.4.4 Privacy evaluation of PETs when encrypting transaction data. 61
4.4.5 Summary 62
5. Further considerations 65
5.1 Data 65
5.1.1 Additional data and money laundering typologies 65
5.1.2 Real-world data are crucial for understanding the feasibility and impact 66
5.1.3 Limitations of payments data and the need for other data 66
5.1.4 Data protection 67
5.2 Technology 67
5.2.1 Technical challenges with CAL arrangements 67
5.2.2 Machine readable typologies that facilitate information sharing 68
5.2.3 Explainability 71
5.3 Looking ahead 71
5.3.1 Instant payment systems, CBDC systems and financial crime 71
5.3.2 Legal and regulatory considerations 72
6. Conclusion 74
7. Annex A: Trends and opportunities 77
7.1 Standardisation, transparency and harmonisation in payments 77
7.1.1 G20 roadmap for enhancing cross-border payments 77
7.1.2 ISO 20022 harmonisation 77
7.1.3 Data standards for legal entity identification and beneficial ownership 78
7.1.4 The Wolfsberg Group Payment Transparency Standards 79

7.2 Transaction monitoring utilities 79
7.2.1 TMU example: Transaction Monitoring Netherlands 81
7.3 Instant payment systems and potential CBDC systems 81
7.4 Public blockchains used for payments 82
8. Annex B: Machine learning in this PoC 83
8.1 Machine learning training, validation and evaluation 83
8.2 Machine learning model feature engineering 84
9. Annex C: Privacy-enhancing technologies 85
9.1 Overview of PETs 85
9.1.1 Homomorphic encryption 85
9.1.2 Local differential privacy 85
9.1.3 Federated learning 86
9.1.4 Other privacy-enhancing technologies 86
9.2 Application of PETs 87
10. Annex D: Questions to support real-world pilots 91
10.1 Objectives, performance monitoring and scope 91
10.2 Data and analysis 91
10.3 Post-pilot questions 93
11. Annex E: Additional acronyms and definitions 94
12. References 96
13. Acknowledgements 100

1. Acronyms, abbreviations and definitions

The project specific acronyms, abbreviations and definitions used in Project Aurora are listed in bold. Other useful acronyms, abbreviations and definitions are listed in Annex E.

AML	Anti-money laundering includes all kinds of actions – including sets of rules, legislation, principles, regulations, processes and tools specific to the financial sector – with the objective of tackling the laundering of illicitly obtained funds by criminals.
ANN	Artificial neural networks are a subset of machine learning, inspired by a simplification of neurons in a brain. They can be applied to model complicated network relationships and patterns.
BIS	Bank for International Settlements.
CAL	Collaborative analysis and learning. This collective term refers to approaches in which different parties either collaborate by sharing and analysing data in a centralised manner, or in which different parties collaborate using federated learning in a decentralised manner to train a machine learning model on local data. It then updates those learnings in a common model that is shared with all parties. Hybrid CAL comprises both centralised and decentralised approaches.
CFT	Countering the financing of terrorism is closely related to anti-money laundering and involves similar actions and forces. Such actions aim to tackle terrorist financing.
CLS	Complex layering schemes involve a complex network of multiple accounts across different financial institutions that are often used for money laundering.
CPMI	The Committee on Payments and Market Infrastructures.
FATF	Financial Action Task Force.
FL	Federated learning is a decentralised machine learning framework that enables multiple entities to train a shared model collaboratively without exchanging raw data.
FIU	Financial intelligence unit.
FSB	Financial Stability Board.
GNN	Graph neural networks are a class of artificial neural networks that operate on graph-structured data and contain a set of relationships between actors (edges and nodes).
HE	Homomorphic encryption is a specific kind of encryption that allows computation of data without revealing the underlying data.
IPS	Instant payment system.
ISO 20022	The ISO 20022 standard provides a common language and structure for financial messages that can be used within and across different payment systems and jurisdictions.
KYC	Know your customer/client are mandatory standards used in the investment and financial services industry to verify customers and understand their risk and financial profiles.
LEA	Law enforcement agency.
LEI	Legal entity identifier.
LDP	Local differential privacy.
Money mule networks	Money mules consist of individuals who intentionally, or unintentionally, launder money by receiving and transferring illegal profits on behalf of someone else.
Monitoring scenarios	Transaction data visible on three different levels for analysis, ie the view of each financial institution (siloed), the national view of a single country and the cross-border view across countries.
MPC	Secure multi-party computation is a cryptographic toolbox that allows multiple parties to make calculations using their combined data, without revealing their individual data.
PET	Privacy-enhancing technologies are a broad range of technologies that are designed to extract data value without risking fundamental data protection principles.

PII	Personally identifiable information.
PPP	Public-private partnership.
Smurfing	The act of breaking up the proceeds of illicit funds into small amounts that can easily be hidden amongst other small transactions.
Synthetic data	Information on real-world data that is artificially trained to reproduce the characteristics and structure of the original data.
TBML	Trade-based money laundering is a method used by criminals to launder the proceeds of illicit activities through the international trade system.
TMU	Transaction monitoring utility.
VASP	Virtual asset service providers are any natural or legal persons who exchange, hold, safekeep, sell, convert or otherwise transfer virtual assets eg cryptocurrency exchanges.

2 Executive summary

2. Executive summary

2.1 Background

Money laundering is a global problem that undermines the integrity and safety of the global financial system. Currently, financial institutions monitor transactions for suspicious activities in a siloed way. However, this approach is ineffective as many payment transactions are complex and involve interconnected networks that span multiple financial institutions and borders. Criminals operate in networks and exploit this complexity.

Both legitimate and illicit transactions flow through payment systems. A network view of payments data is essential to combat money laundering.

Financial institutions are exposed to increasing levels of various types of financial crimes, with 67% exposed to financial crimes involving digital payments and over 60% to various forms of money laundering.[1] The amount of money laundered globally is estimated to be between 2 and 5% of global GDP, or between $2 trillion and $5 trillion.[2] However, the estimated total sum seized annually amounts to less than 1% of this – between $20 billion and $50 billion.[3]

A 2022 study found that financial institutions face compliance costs of approximately $274 billion globally,[4] an increase of approximately 28% on the 2020 figure of approximately $214 billion.[5] Between 2019 and 2022, the average costs of compliance grew by approximately 54% in the United States, 80% in Canada, 30% in Germany and 23% in France.[6, 7]

When asked about the major factors driving compliance costs, increasing anti-money laundering (AML) regulation was cited by 68% of financial institutions, while another 68% cited evolving criminal threats.[8] These factors combined with the risk of sanction, have contributed to a defensive approach to AML compliance being adopted by financial institutions. This defensive approach can lead to over-reporting to authorities, which can become a drain on public resources. It can also result in the

1 See Lexis Nexis (2022).

2 See UNODC. The source states $800 billion to $2 trillion based on global GDP at the time of writing. When the estimated value laundered is calculated based on global GDP in 2022, which was approximately $100 trillion (see Statista 2023), the estimated value laundered would be between $2 trillion - $5 trillion. It should be noted that it is difficult to estimate the total amount of money that goes through the laundering cycle.

3 See Statista (2023). The value for the total sum seized is calculated from the estimated value laundered based on global GDP in 2022, which was approximately $100 trillion.

4 See Lexis Nexis (2022).

5 See Lexis Nexis (2021).

6 See Lexis Nexis (2022).

7 See Lexis Nexis (2021).

8 See Lexis Nexis (2022).

termination of customer relationships to reduce overall exposure to financial crime risk (known as "de-risking").[9]

2.2 Data, technology and innovation

In response to some of these challenges, the Financial Action Task Force (FATF) has identified that data-sharing and collaborative analytics are critical for effective anti-money laundering and countering the financing of terrorism (CFT) efforts.[10] In its *Stocktake on data pooling, collaborative analytics and data protection*,[11] the FATF outlined several technologies and approaches that could be used to improve AML/CFT efforts, including different approaches to data-sharing,[12] privacy-enhancing technologies (PET),[13] advanced analytics,[14] data standardisation[15] and data protection.[16] Digital transformation to enhance AML/CFT efforts is a strategic priority of the FATF.[17]

Additionally, in 2020, the G20 leaders endorsed a *Roadmap for enhancing cross-border payments*. As part of this roadmap's prioritisation plan, the FATF is also considering updating its recommendation 16 (the travel rule)[18] to take into account developments in the architecture of payment systems, including the adoption of ISO 20022 messaging standards. This is to improve the consistency and usability of payment message data in cross-border payments and could also facilitate more efficient AML/CFT checks.

Technology and collaboration could support financial institutions, central banks, supervisory and other public authorities to address AML challenges through collaborative analytics and learning (CAL). Such initiatives could leverage payment system-level data and public-private collaborative approaches to analyse privacy protected data[19] to reveal suspicious networks and activities that may not be detected by financial institutions acting in isolation.

---

9 See FATF (2021a).

10 See FATF (2021a).

11 See FATF (2021b).

12 Sharing information could also support customer due diligence measures such as institutional risk assessment, customer onboarding, risk management of a business relationship, identification of the beneficial owner, and could help identify and share patterns and flows, such as typologies.

13 Privacy-enhancing technologies (also referred to as cryptography/encryption technologies) such as homomorphic encryption, secure multi-party computation, differential privacy and zero-knowledge proofs can facilitate secure and privacy-protected information-sharing and analysis.

14 Advanced analytics such as machine learning, federated learning, deep learning, network analysis and natural language processing can be applied to analyse large amounts of structured and unstructured data more efficiently, and identify patterns and trends more effectively.

15 See FATF (2021a).

16 See FATF (2021b).

17 See FATF (2022).

18 See FSB (2022).

19 Collaborative approaches to analysis include centralised, decentralised or hybrid (centralised and decentralised) at a national and cross-border level. These are discussed further in section 4.4.

The protection of individual and fundamental rights to privacy can be a concern when considering the use of data and technologies to fight financial crime. Data privacy and protection, and countering financial crime are important public interests that are not opposed to each other. They should be supported by the right technological tools and by a balanced legal framework.

2.3 Project Aurora

Project Aurora builds upon the above-mentioned initiatives and challenges in a proof of concept (PoC). The PoC investigates the use of privacy-enhancing technologies and advanced analytics for different CAL approaches for detecting money laundering activities. The PoC contains the following parts:

- Generation of a synthetic data set that represents transactions between financial institutions, individuals and businesses within a country and across borders. The data set also reflects complex money laundering events that are embedded in the data. The data set consists of a minimum set of data fields, which are common to different payment ecosystems, such as instant payment systems (IPS) and potential CBDC systems, as well as data fields required in any CAL arrangements.

- Testing three different simulated monitoring scenarios (views the synthetic data at the single financial institution level, at the national level and at the cross-border level) with machine learning models[20] and network analysis to compare the performance[21] of the scenarios in detecting money launderers and suspicious networks.

- Testing and comparing the performance of different CAL approaches – such as centralised, decentralised or hybrid at the national and cross-border levels – in detecting money launderers and suspicious. Privacy-enhancing technologies were applied to the data in each approach and analysed using advanced analytical methods to examine how privacy-enhancing technologies could support privacy and data protection.

2.4 Key findings and takeaways

Project Aurora demonstrates the advantages and potential of using payments data in combination with privacy-enhancing technologies, machine learning models and network analysis for the detection of complex money laundering schemes. The project also simulates how these data and technologies could be brought together to enable public-private collaborative analysis and learning (CAL) arrangements, both nationally and internationally, to counter money-laundering. The project demonstrates that **CAL**

20 Machine learning is a subset of artificial intelligence. It enables a machine to learn from insights from the data. Machine learning is used in this report refers to "artificial intelligence and machine learning".

21 Performance is made up of two parts: effectiveness and efficiency. These refer to the fraction of money laundering activities detected in the data while keeping the number of false positives low.

approaches are more effective in detecting money laundering networks than the current siloed approach (in which financial institutions carry out analysis in isolation).

2.4.1 A holistic view of payments data unveils money laundering networks

A **holistic view of payments** data is essential to effectively identify and combat suspicious activities that take place beyond the bounds of single financial institutions and national borders. Leveraging these data could lead to improvements in monitoring by opening up a holistic view on transaction networks that unveil money laundering networks.

At a national level, the analysis approaches explored in this project could be performed via transaction monitoring utilities or CAL arrangements in which different ecosystems of payments data (eg financial institutions, fintech, virtual asset service providers (VASPs), card schemes, e-money or others) are connected. At a cross-border level, similar analysis could take place in a CAL arrangement.

Similarly these approaches could be used by operators (eg central banks or private sector entities) of **IPS or potential CBDC** systems that include AML monitoring and analysis capabilities. Operators of these systems could provide participants with additional tools and support to enhance their monitoring efficiency.[22]

2.4.2 Behavioural monitoring and privacy enhancing technologies could be a game changer for AML efforts

Utilising network analysis for detecting anomalous and suspicious networks shifts the focus from individual behaviour to the overall behaviour of suspicious networks, resulting in improved detection capabilities.

Project Aurora demonstrates the potential to improve the detection of money laundering while reducing the number of incorrect alerts. Furthermore, the project shows that the optimal performance of machine learning models is observed in a simulated cross-border scenario in which sensitive transaction data are protected and secured (using encryption or a combination of privacy preserving methods), consolidated into a centralised system[23] and where network analysis is utilised.

Moreover, a centralised approach that consolidates privacy-enhanced transaction data at a national level and collaborates with other countries to collectively train a machine learning model (in a decentralised approach using federated learning) that

---

22 There could be limitations on the types of money laundering activities and actors that could be detected depending on design choices and data available. Section 5.1.1 discusses this further.

23 While Project Aurora simulated a centralised cross-border CAL approach in the experiment, it should be noted that in reality the challenges associated with data protection, data localisation, legal, regulatory and other factors would be complex. A decentralized CAL approach using federated learning at a cross-border level may offer an alternative solution, however there could be a trade-off with effectiveness in detecting money laundering activities.

can be applied locally, could support potential cross-border collaboration on AML efforts, while upholding the data sovereignty of individual countries. [24]

2.4.3 Leveraging Project Aurora

To leverage Project Aurora, three aspects should be kept in mind:

First, the specific data fields and sources required for detection of financial crime may vary depending on the techniques and methods used by criminals. It is essential to have a thorough understanding of different types of financial crimes and identify the data fields and sources that may help indicate their occurrence. Project Aurora shows that the performance of such analysis is only as good as the breadth of data available. Data quality and standardisation of data identifiers and fields are important factors.

The adoption of ISO 20022 could be an opportune moment to catalyse greater international consistency in the use of data identifiers and fields, and their shared definitions (available in machine readable form), that could be used for financial crime detection and to enable CAL arrangements. For example, the inclusion of the legal entity identifier (LEI) in ISO 20022 payments messages. When combined with the additional data fields available in these messages, the LEI could help identify a greater range of money laundering activities involving legal entities.[25] Similarly, standards being developed for identification of beneficial owners, would be important too. These standards are further discussed in Annex A.

Second, further discussion on the public benefits of CAL arrangements for AML followed by legislative clarity to support such arrangements would be necessary. Data protection agencies should be engaged at an early stage to be part of co-design processes with other stakeholders in CAL arrangements to identify and mitigate risks and address uncertainties, for example considering the role and application of PETs.

Third, effective CAL arrangements, as a public-interest tool for financial crime detection, as part of a broader strategic framework for financial crime prevention and disruption, would require collaboration between the public and private sectors to contribute to and deliver such a strategic approach. Any such approach would need to consider the prioritisation of risks and responses to them, the data required, trust between participants and the legal certainty needed to enable a CAL arrangement. National strategies for AML monitoring and analysis could also include the appropriate cross-border CAL initiatives to gain a broader view of money laundering networks and the flow of funds.

It takes a network to defeat a network.

---

[24] See The White House (2021): the US and UK prize challenge to advance privacy-enhancing technologies as they present an opportunity to leverage the power of data while protecting privacy and intellectual property, and enable cross-border and cross-sector collaboration.

[25] The legal entity identifier (LEI) is the global standard for legal entity identification. It could enable data associated to legal entities to be linked to transaction data within and across borders. The LEI could connect a greater range of data sets and capture different relationships which could be useful in AML efforts. It could also address the challenges faced by monitoring systems from inconsistencies in how entities are identified.

3 Introduction

3 Introduction

3.1 What is money laundering?

Money laundering is the act of hiding the origin of illegal assets, often involving a series of transactions that may appear legitimate on the surface. During the course of these transactions, the nominal owner of the funds and the form of the assets can change in order to hide any connection with the original assets.

Money laundering fuels corruption, organised crime, terrorism, modern slavery and environmental crime.[26] These illicit activities may discourage foreign investment, distort international capital flows and may also result in welfare losses by draining resources from other economic activities, as well as undermining the integrity and stability of the financial system and the broader economy.[27]

The consequences are far-reaching and in today's interconnected world the impact of these activities is not confined to individual financial institutions (FIs) or individual countries.

Money launderers exploit the complexity of the global financial system, information asymmetries due to gaps in regulatory data visibility and the ability to share information, and differences in national laws. According to the FATF, all large-scale money laundering schemes invariably contain cross-border elements.[28] Therefore, money laundering is a challenge that requires coordinated global efforts, which are essential to protect the integrity and stability of the financial system.

Financial institutions play a critical role in detecting and preventing these activities, since they serve as the "first line of defence". Despite global efforts, most FIs rely on siloed data and isolated systems for their suspicious transaction monitoring, thus limiting their ability to detect complex cross-border and cross-institutional money laundering networks.

Combatting financial crime is only as strong as the weakest link and is a continuous process requiring the ability of human and technological processes to adapt to evolving risks. New money laundering techniques continue to constantly emerge as criminal elements use creative ways to obfuscate the source(s), destination(s) and the flow of funds between them. Criminals exploit complexity in the financial system and quickly adapt to find new opportunities to evade rules and regulations or to make detection of their activities as difficult as possible.

---

26 See IFAC (2022).

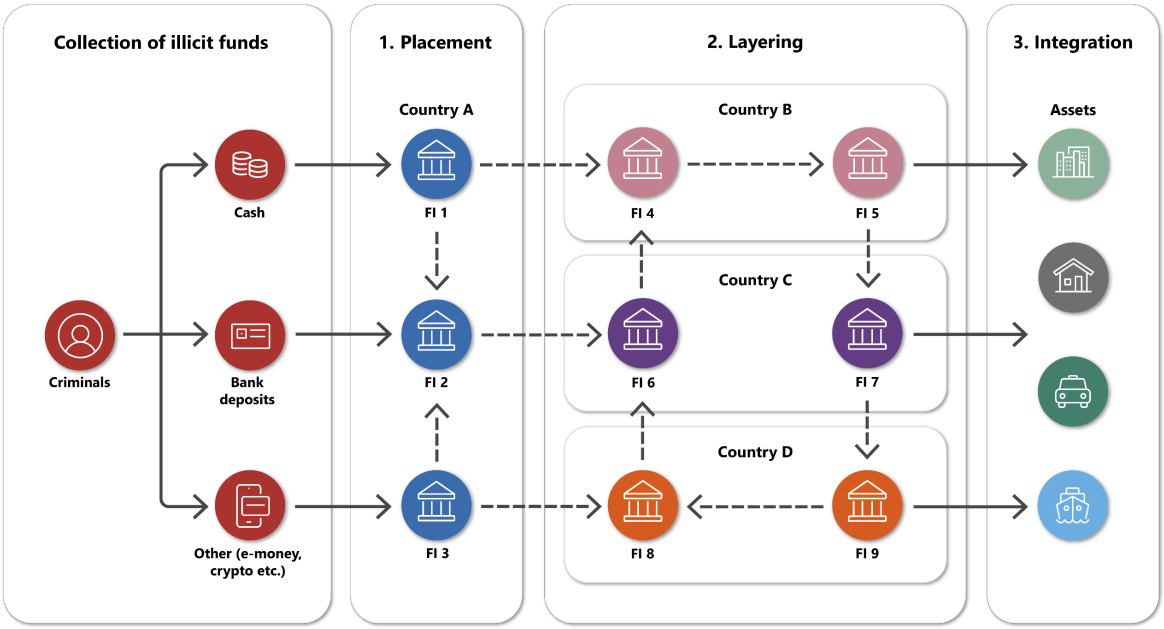
27 See IMF (2021).

28 See FATF (2023).

3.2 The money laundering process

Money laundering typically involves the following simplified three stages as illustrated in Graph 1.[29]

Graph 1: A simplified view of the three stages of money laundering



In reality, this process is not as linear as shown in this diagram.

Stage 1: Placement

In the placement stage a money launderer introduces illicit funds into the financial system. This could occur by transferring funds to accounts or wallets controlled by money launderers. The transfer method used would depend on the source and type of illicit funds, for example cash, bank deposits or cryptoassets.

Stage 2: Layering

In the layering stage, a money launderer moves the funds from the placement stage through different layers, involving transfers between different financial institutions or conversion to different assets, to obfuscate the source(s).

Techniques used by money launderers depend on the type of asset, but large-scale money laundering schemes consist of illegal funds being moved through the financial system using an array of complex layering structures. For example:

[29] See FATF (2018) page 18.

- **Money mule networks** consist of individuals who, intentionally or unintentionally, are used to launder money by receiving and transferring illegal profits on behalf of someone else. The goal is to add layers between the source of the funds and the criminals by obfuscating the fund's route. Professional money mule networks often include multiple accounts across different financial institutions and jurisdictions, making it difficult to trace the funds.
- **Smurfing** is the act of breaking up larger transactions into smaller ones to avoid detection. This can be done by using multiple bank accounts, credit cards and shell companies, sometimes across different countries.

Stage 3: Integration

In the third stage, the laundered funds are reintegrated into the economy by being transfered to accounts controlled by the criminal or invested in assets such as real estate, valuable or luxury items, or business ventures.

3.3 AML monitoring and analysis today

Financial institutions are required to implement a range of AML controls and monitoring processes, including assigning customers a risk value, implementing transaction monitoring systems, labelling and reviewing suspicious transactions, and reporting any suspected money laundering activities to public authorities, typically national financial intelligence units (FIUs).

The monitoring of transactions for suspicious activities is typically based on one or other of two approaches: rule-based systems and behavioural monitoring systems.

The risk of sanction from ineffective compliance procedures, increasing costs of regulation and compliance uncertainty (due to risks arising from an evolving financial crime threat landscape) can lead to financial institutions adopting a defensive approach to compliance.

3.3.1 Rule-based monitoring systems

Rule-based monitoring systems use a set of pre-defined rules or algorithms to identify suspicious transactions. These rules can be based on historical data, and on other factors such as size, frequency and the origin of transactions.

The rule-based approach is effective when detecting common money laundering typologies, for example in cases in which a customer deposits funds in a manner designed to avoid reporting requirements.

However, the rule-based system can result in a high number of:

- **false positives** – legitimate transactions flagged as potentially suspicious; and
- **false negatives** – illicit transactions not flagged as suspicious.

The occurrence of false positives or false negatives is often due to the use of broad, generic rules that trigger alerts for a wide range of transactions. It is estimated that 90–95% of all alerts generated are false positives.[30, 31] Significant resources are spent examining and reviewing these false positives.[32]

3.3.2 Behavioural monitoring systems

While rule-based monitoring relies on pre-defined rules and triggers to flag transactions, behavioural monitoring examines patterns of human behaviour associated with transactions, using machine learning models to understand typical behaviour and relationships between accounts and transactions.

This approach can identify deviations from typical patterns of anticipated normal behaviour, detect more sophisticated money laundering techniques and potentially reduce the volume of false positives. This is because the system is able to analyse large sets of transaction data, and is capable of learning and adjusting its model over time.[33]

Behavioural monitoring could be a valuable tool in detecting suspicious behaviour, but it is limited in its effectiveness in detecting money laundering across financial institutions and borders due to siloed views of data and the inability to share information.

3.3.3 Defensive reporting and de-risking consequences

Financial institutions can be sanctioned if they are deemed not to have necessary and effective AML compliance procedures in place and process transactions that turn out to be illegal. This risk, as well as the costs associated to increasing AML regulation and exposure to new financial crime threats can result financial institutions adopting a defensive approach to AML compliance.

This can lead to the overreporting of transactions to authorities, which becomes a drain on public resources. In some cases when increased compliance costs and additional risks outweigh business benefits, financial institutions can terminate relationships with particular customers to reduce their overall exposure to financial crime risk (also known as de-risking). This can have high social costs such as financial exclusion, and could ultimately create distrust in risk assessment tools and regulatory frameworks in both the public and private sectors.[34]

---

30 See Oracle (2019).

31 See BIS (2019).

32 See Finanstilsynet (2019) at page 69 - The director of a systemically important financial institution (SIFI) has stated to a financial supervisor that if an ordinary customer is selected just once for a manual review, the economic benefit of having that customer will disappear.

33 FATF (2021b) at p 22.

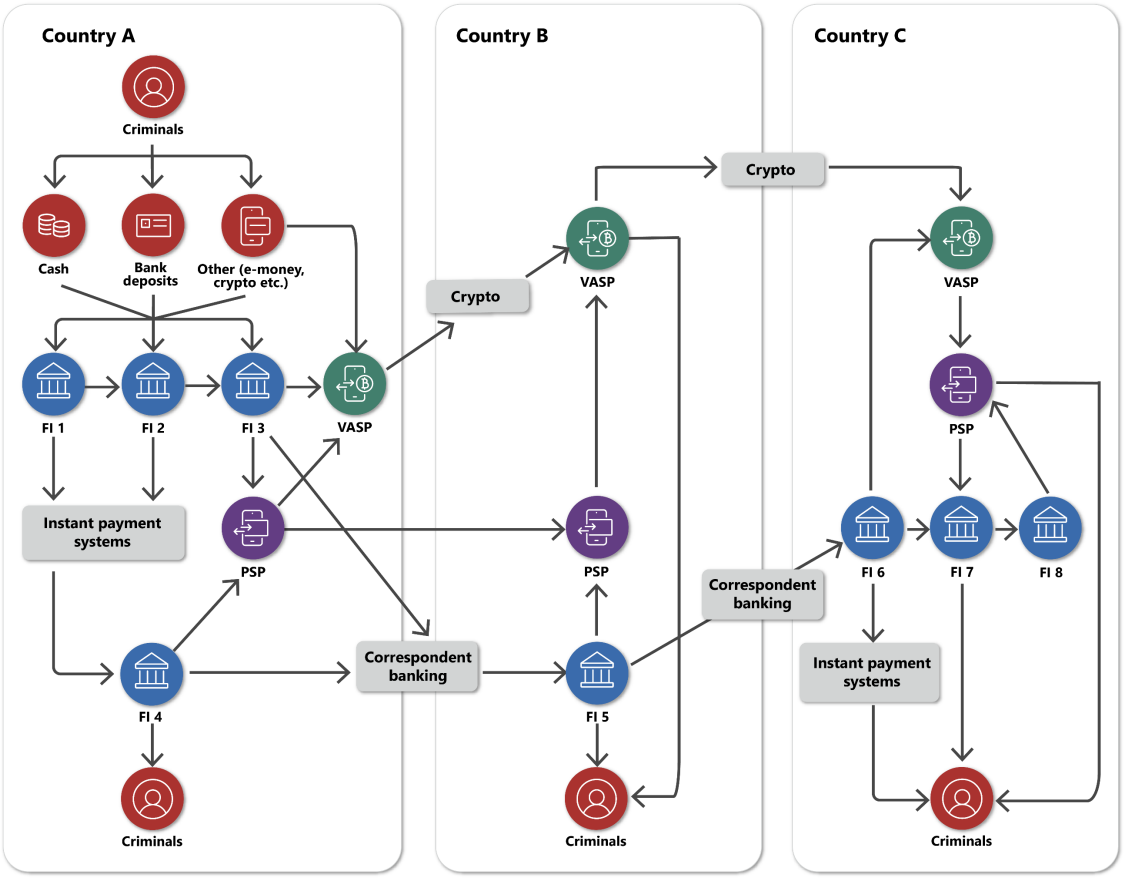
34 FATF (2021a).

3.4 Challenges facing AML efforts

Money launderers create a complex network of transactions across financial institutions and borders, while in contrast, these institutions themselves often have a limited and siloed view of what is going on (**Challenge 1**). This is further complicated by data not being standardised or readily consumable, making collaboration across institutions more difficult (**Challenge 2**). Finally, the data required to undertake effective AML/CFT measures have to be balanced against the need to protect individuals' privacy and personal data, which further complicates collaboration and sharing (**Challenge 3**). These challenges are detailed below:

- **Challenge 1:** money launderers use a combination of methods to hide the source and destination of funds, such as accounts at different financial institutions and payment service providers (PSPs), cryptoasset payments, correspondent banking relationships, domestic instant payment systems and cash purchases. This creates a complex network of transactions across financial institutions and borders. This is illustrated in Graph 2. These complex networks of money laundering can remain undetected with siloed approaches to monitoring and analysis by individual financial institutions are used versus when capabilities to securely share data and conduct analysis using collaborative approaches are used (ie leveraging privacy enhancing technologies and advanced analytics). This is illustrated in Graph 3.

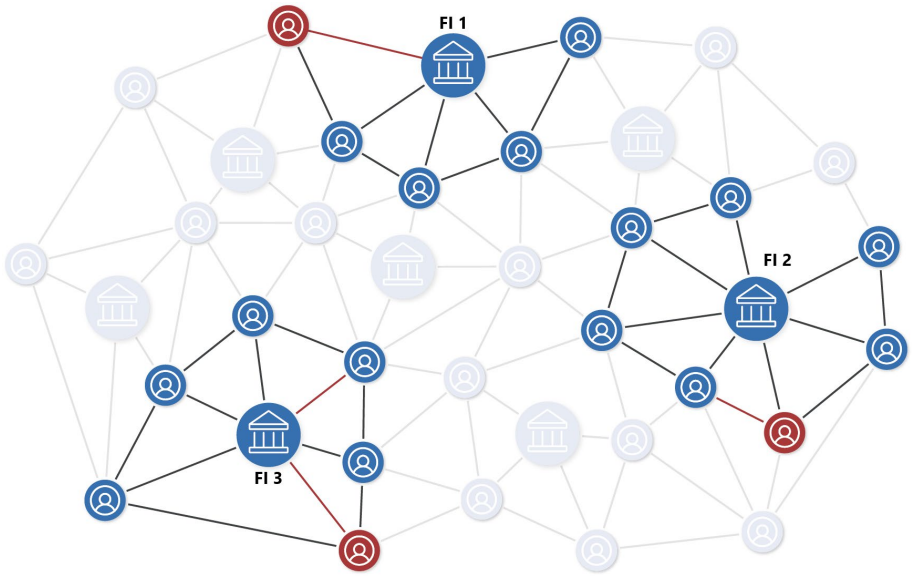
Graph 2: Simplified view of different payment ecosystems used by money launderers



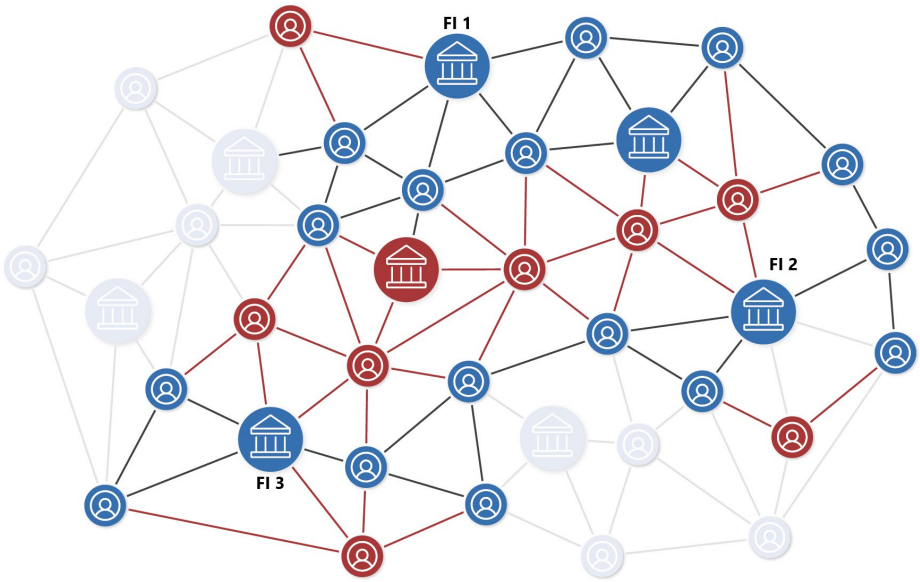
In reality, this process can be more complex than shown in this diagram.

Graph 3: A simplified view of the visibility of suspicious networks in isolation vs when payments data are analysed in CAL arrangements

Isolated view



With collaborative analysis and learning (CAL) arrangements



- **Challenge 2:** data fields, formats and processes are not standardised, hindering data quality, efficient analysis, comparability and systems integration. In addition, data definitions describing financial crime typologies and other related information are neither standardised nor available in standard machine-readable form for analytical systems to use.

- **Challenge 3:** the balance between objectives for data protection and privacy, and those for AML/CFT can also be a challenge. On the one hand, there is a need to protect individuals' privacy and personal data, ensuring compliance with data protection regulations. On the other hand, effective AML/CFT measures require access to relevant data and information to detect and prevent illicit activities. Balancing privacy protection and effective AML/CFT measures is complex, as it involves navigating legal, ethical and technical considerations to address the objectives of privacy and security.

3.5 Technology

The use of advanced technologies in AML transaction monitoring has the potential to change the way financial institutions detect and prevent financial crimes. The FATF has published several papers on digital transformation which encourage greater use of technology to improve AML/CFT efforts.[35]

The FATF has identified several technologies and approaches that could be used to improve AML/CFT efforts, including different approaches to data-sharing,[36] privacy-enhancing technologies,[37] advanced analytics,[38] data standardisation and data protection.[39] These technologies and approaches offer the benefits of data-sharing and advanced analytics while preserving data privacy.

3.5.1 Privacy-enhancing technologies (PETs)

One of the key challenges in AML efforts is balancing the need to detect suspicious activity with the need to protect privacy. PETs could offer a solution, as they are designed to protect sensitive information, even if data are distributed across multiple organisations, while enabling advanced analytical methods to be applied to protected data. The field of PETs is a fast-growing area of innovation and different PETs are available. Each has specific strengths, limitations, and suitability to given use cases. Table 1 provides an overview of the PETs used in this project. Further information on each PET can be found in Annex C.

---

35 See FATF (2021a).

36 Sharing information could also support customer due diligence measures, such as institutional risk assessment, onboarding customers, risk management of a business relationship, identification of the beneficial owner, and can help identify and share patterns and flows, such as typologies.

37 Privacy-enhancing technologies (also referred to as cryptography/encryption technologies) such as homomorphic encryption, secure multi-party computation, differential privacy and zero-knowledge proofs can facilitate secure and privacy-protected information-sharing and analysis.

38 Advanced analytics such as machine learning, federated learning, deep learning, network analysis and natural processing can be applied to analyse large amounts of structured and unstructured data more efficiently and identify patterns and trends more effectively.

39 FATF (2021b).

Table 1: PETs used in this project, applicability to AML and challenges

Type of PET	AML use case	Challenges and limitations
Synthetic data	Creating realistic data sets for AML testing	Difficulty in replicating real-world data accurately Trade-off between information security and realistic replication of the real data
Differential privacy	Protecting privacy of sensitive data while allowing for data analysis	Risk of re-identification Balancing privacy needs with data quality and model accuracy
Homomorphic encryption	Allowing for data processing without revealing underlying data	Higher computational costs and slower processing times Ensuring that information does not leak
Federated learning	Multiple parties can train a shared model without having to share their data	Risk of data leakage Higher computational costs, and a potentially less accurate model

3.5.2 Graph data structures

Graph data structures are useful in modelling various types of networks (eg transaction data). They can be used to represent and analyse a wide range of relationships between different entities (ie individuals and business), helping identify key actors and patterns of relationships between them. They can be used by machine learning models, data clustering graph algorithms and network analysis which makes them suitable for AML monitoring in combination with machine learning and network analysis.

3.5.3 Machine learning

Artificial intelligence and machine learning methods offer the potential to enhance AML suspicious transaction monitoring by identifying patterns and anomalies in transaction data that traditional methods cannot. One potential example is the use of graph neural networks, a type of machine learning model that can analyse graph-structured data (eg transaction data). By using models such as graph neural networks in the context of AML transaction monitoring and analysis, it may be possible to detect suspicious patterns and anomalies in transaction networks that can be difficult to identify with traditional methods.

However, their use also poses challenges. Machine learning models using network features could suffer from data bias and interpretability issues. Data ethics and the explainability of automated decisions, as well as the role of human review, would be important considerations.

3.5.4 Network analysis

Network analysis can help uncover hidden patterns of suspicious money laundering networks in financial transaction data. Network analysis can complement machine learning methods to improve AML monitoring by leveraging features of data (eg ratios, accumulations) in the network of transactions. Examples of such features are detailed in Annex B.

By examining the connections between entities (eg individuals and businesses) as well as information contained in transaction data, network analysis can enable the detection of money laundering networks and provide insights through data visualization, statistical analysis, and interpretation of results. This could allow AML experts to identify suspicious activities, understand the flow of illicit funds, and pinpoint the core money laundering networks that require investigation.

3.6 Summary of trends and opportunities

A number of trends, initiatives and developments in payments, data standards and transaction monitoring could address some of the challenges facing AML efforts. The main ones are summarised below with further details on each in Annex A.

Standardisation, transparency and and harmonisation

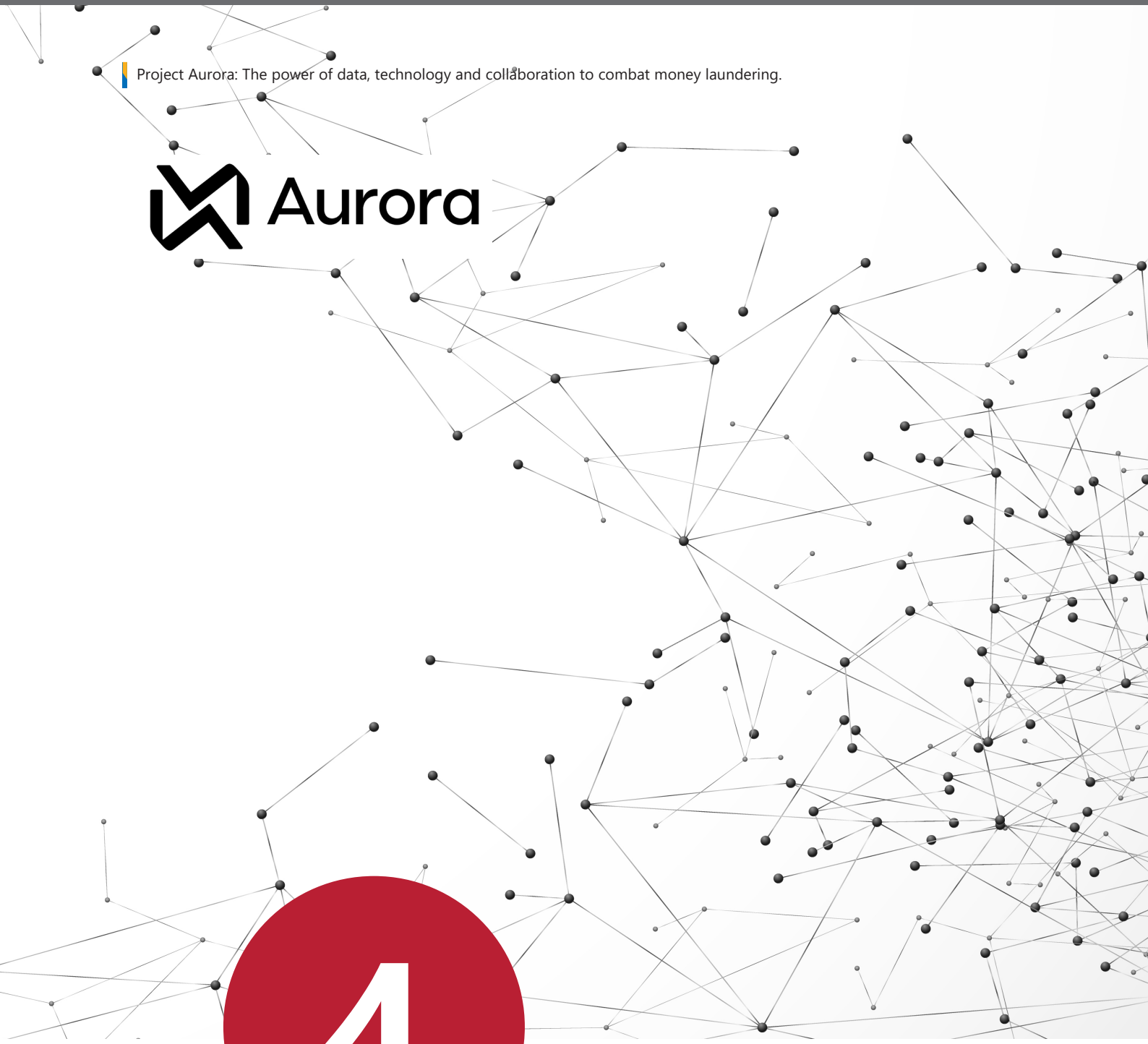
- The G20 Roadmap for enhancing cross-border payments - seeks to enhance cross-border payments via several building blocks addressing various issues including financial crime prevention.
- Harmonisation with the ISO 20022 standard which provides a common language and structure for financial messages that can be used by different payments, enabling greater interoperability and straight-through processing, A richer set of structured data with ISO 20022 messages would benefit AML analysis.
- Data standards such as the legal entity identifier (LEI) and beneficial ownership also provide a common language, structure and ability to link different data sets associated to legal entities and beneficial owners. For example the LEI maps to other identifiers used in payments or securities transactions.
- Payments transparency standards published by The Wolfsberg Group consist of ten principles financial institutions should adhere to when processing payments.

Transaction monitoring utilities

- In certain jurisdictions, transaction monitoring utilities (TMUs) (eg TMNL in the Netherlands) have been established in private-private collaborative analysis arrangements to detecting and preventing financial crime.

Instant payment systems and potential CBDC systems

Instant payment systems can provide a broad view of transaction data and could provide the ability to unveil suspicious networks across several financial institutions. This could also apply to CBDC systems too.



4 Project Aurora - proof of concept

4. Project Aurora – proof of concept

4.1 Objectives and scope

Project Aurora utilises the power of connected payments data to combat money laundering across financial institutions and borders. For the purposes of money laundering detection, the project tests privacy-enhancing technologies coupled with machine learning and network analysis in different monitoring scenarios and with different approaches to collaborative analysis and learning.

4.1.1 Objectives

The objectives of the project are to:

- Generate a synthetic data set containing a minimum set of common data points and representing real-world transactions and flows between many financial institutions across several countries, with money laundering activities embedded into the data.
- Based on the synthetic data, experiment with three different simulated monitoring scenarios (at a siloed financial institution level, a national level and a cross-border level) with machine learning models and network analysis to test and compare the performance and effectiveness of each scenario in detecting money laundering networks.
- Based on the synthetic data and the optimal machine learning models from the previous objective, experiment with different CAL approaches such as centralised, decentralised or hybrid at national and cross-border levels[40] when privacy-enhancing technologies are applied to the data in each CAL approach. Analyse the data using machine learning and network analysis, to test and compare the performance and effectiveness of each CAL approach in detecting money laundering networks.

[40] These are discussed in more detail in Section 4.4.

4.1.2 Scope

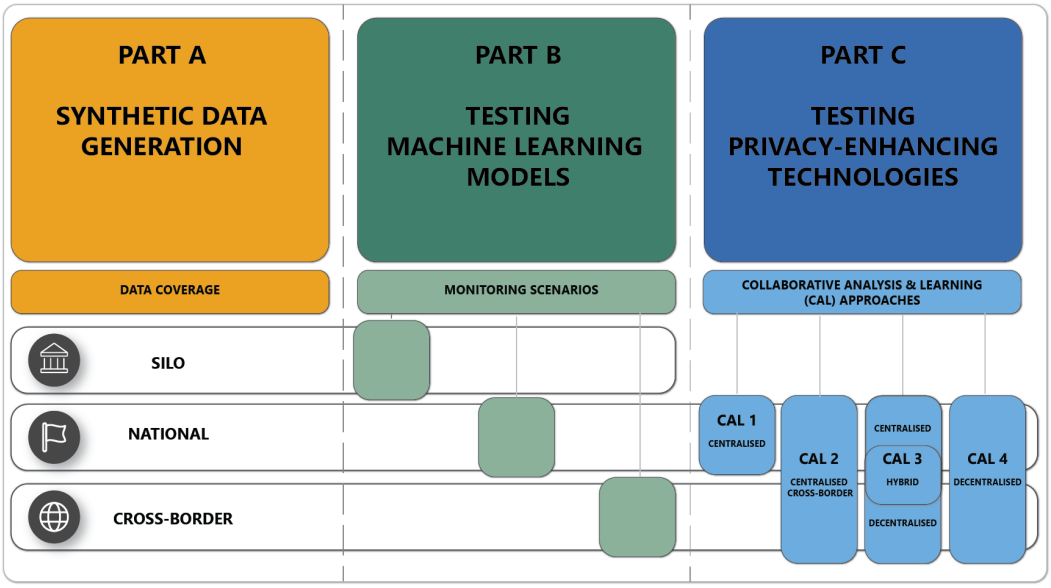
Several experiments were conducted in Project Aurora. The PoC is structured into **three main parts**, which address the challenges outlined earlier:

- **Part A – generation of a synthetic data set** representing domestic and cross-border payments between individuals and businesses across multiple financial institutions operating in one or more jurisdictions, and with several money laundering events embedded into the data set.

- **Part B – application of machine learning models** on the generated synthetic data set to detect networks and patterns of suspicious flows of funds against different views of the data – at the single financial institution level, at the national level and at the cross-border level.

- **Part C – application of PETs** on the generated synthetic data to test CAL arrangements (centralised, decentralised and hybrid) and the performance of machine learning models in each.

Graph 4: The three parts of Project Aurora



The following are **out of scope** of this phase of Project Aurora:

- providing a comparison of all available machine learning models that could be used for AML;
- finding the best performing algorithm for any class of machine learning models;
- presenting a complete comparison of privacy-enhancing technologies; and
- including a broader and comprehensive set of typologies and studying the detection results on real-world data.

4A Part A - Synthetic data generation

4.2 Part A: Synthetic data generation

Synthetic data generation has emerged as a useful technique for creating realistic and representative data that can be used to train, test and measure the effectiveness of different analytical tools. It is particularly useful for simulating data that would contain sensitive information and would typically be difficult to obtain, for example payments data.

Synthetic data can be generated through different approaches. The project utilises a step-by-step layered approach to generate synthetic data, mimicking realistic domestic and cross-border transactions, and money laundering activities across multiple financial institutions and several countries, to test the objectives and technologies used in this PoC.

4.2.1 Purpose of generating synthetic data

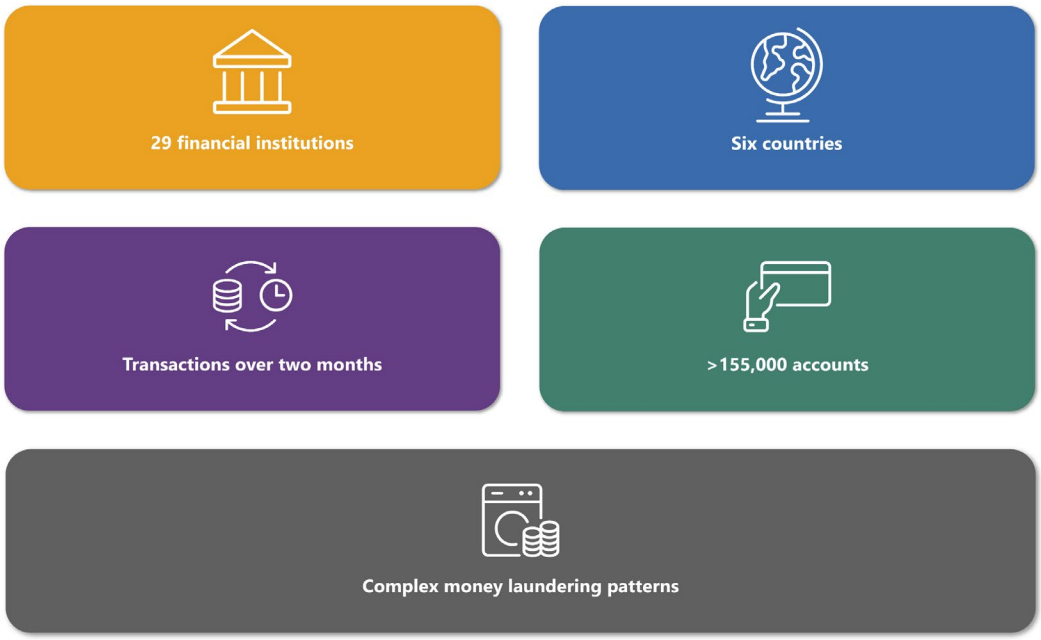
The generation of synthetic data for this project should enable the following:

- Training machine learning models on a representation of real transaction data.
- Running experiments against different views of the data: siloed financial institution, national and cross-border, using machine learning to compare any differences between them.
- Understanding the efficacy of detecting money laundering networks using a minimum set of common data points.
- Demonstrating a minimum set of common data points could be used by all participants in a CAL arrangement or for analysis at an instant payment or central bank digital currency (CBDC) system level.
- Applying different PETs to the data to better understand the strength and limitations of each.
- Simulating different CAL arrangements.
- Simulating transactions in an instant payment system (or another system, eg a CBDC).

4.2.2 Generating the synthetic data

The synthetic data generated for this PoC represents domestic and cross-border transactions over two months among 155,250 entities, including individuals and corporations, across six countries and between 29 financial institutions.[41]

Graph 5: High-level overview of the generated synthetic data



The synthetic data set generated for the project contains the minimum essential information, including:

- the sender (payer) and the recipient (payee) details;
- the transaction amount;
- the date and time of the transaction;
- the transaction method (eg card payment, bank transfer or cash deposit);[42]
- the country the payment was sent from;
- the country in which the payment was received; and
- the details of the financial institutions involved (eg banks and payment service providers).

[41] The size of the data set used in the PoC is smaller compared with the actual size of real transaction networks. This decision was made intentionally to ensure the feasibility and simplicity of testing various technologies within the PoC. While the data set may not accurately reflect the scale of a real transaction network, it still serves the purpose of evaluating different technologies in a controlled environment.

[42] It is important to include the form of transaction as a variable, as criminals often use multiple transaction methods to conceal the origin and flow of illicit funds.

This approach ensures that the results can be compared and used across various arrangements and systems while adhering to the principle of data minimisation.

4.2.3 Leveraging a three-step approach to generate synthetic data

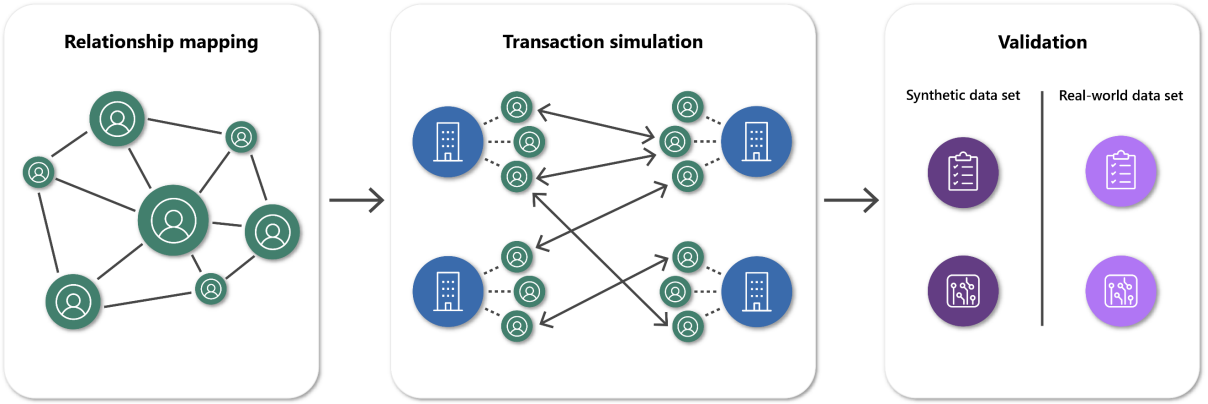
The synthetic data are generated in three steps:

1. **Relationship mapping** generates a network graph that represents relationships between individuals and businesses.

2. **Transaction simulation** of transaction flows based on the relationships in the relationship map.

3. **Validation** of the generated synthetic data.

Graph 6: Synthetic data generation process



Relationship mapping

The relationship map is a graph representing the relationships and behavioural patterns of individuals and businesses, which form the basis of the simulated transactions. The actions and interactions of individuals and businesses are created by using a statistical model tuned to understand behaviours and relationships observed in domestic payments data. This means that the synthetic data are mimicking real behaviour and relationships instead of being a direct one-to-one replication of original data.

The relationship map is created in the following way:

1. **Simulating the connections between individuals and businesses.** These data provide insights into the transactional behaviour of individuals and businesses, including the number of transactions they receive and send to others, as well as the corresponding transaction amount distribution.
2. **Establishing relationships between individuals, businesses and FIs within each country.** These relationships involve linking individuals and businesses to various financial institutions. The probability of an individual or business being a

customer of a particular institution is influenced by a number of factors, including the relative size of the institution. Additionally, relationships between individuals and businesses are established based on behaviours (eg employment, shopping, remittance), transaction types (card, cash or transfer) and transaction volumes. These relationships are built to approximate the observed aggregated payment pattern in real-world data.

3. **Establishing relationships between individuals, businesses and FIs across borders.** These relationships involve links between individuals and businesses across borders based on various behaviours (cross-border purchases, remittance payments etc), payment methods (card, cash or transfer) and transaction volumes.

The businesses in the data set encompass a variety of entities, including VASPs. This represents the growing trend towards using alternative payment methods such as cryptoassets and crypto exchanges to facilitate money laundering.[43] VASPs in the data set are characterised by a substantial volume of incoming transactions but limited visibility into their outgoing transactions. This mirrors the scenario in which outgoing transactions from VASPs (eg conversion into cryptoassets) are opaque, sometimes seen in connection with VASPs located or operating in jurisdictions with minimal AML regulations.

The output of this process is a set of interconnected relationships between individuals, businesses, financial institutions, and countries that are organised into clusters.

The transaction simulation

The transaction simulator is an additional layer on top of the relationship map and generates a set of simulated transactions between the individuals and businesses based on information from the relationship mapping. These simulated transactions are carried out across multiple financial institutions located in various jurisdictions.

Each day, the simulator will gather information about individuals and businesses such as the transaction date, the connections they have with others and how often they typically make a transaction. Using this information, the simulator decides when and with whom the individual or business should have its next transaction. It also determines the payment method and value of the transaction. This process is continuously repeated to simulate the information in the relationship mapping. Graph 7 visualises a network based on transactions from a single day.

[43] See FATF (2020).

Graph 7: Minimal example of a one-day simulated transaction network

Example of a one-day transaction graph representing transactions between individuals and businesses across six countries. Each country is represented as a different colour and each transaction is represented as a grey line. Transactions on the first day of the month are shown. The transactions have different purposes.

Validation

The validity of the synthetic data is demonstrated by combining an actor behaviour perspective with an aggregate pattern perspective:

- **Actor behaviour** – the behaviour of individuals and businesses is compared with real-world data[44] to ensure consistency.

- **Aggregated patterns** – overall patterns at the national and cross-border levels are validated by comparing them with statistics from central banks,

[44] Real-world data used for comparison were based on the proprietary data of a third party.

intergovernmental bodies, and others to ensure that the synthetic data accurately reflect real-world information.

4.2.4 Constructing money laundering activities in the synthetic data set

Transactions flows representing money laundering patterns are embedded in the synthetic data set. These are based on known complex money laundering techniques, which are ideal for demonstrating the objectives of the project.

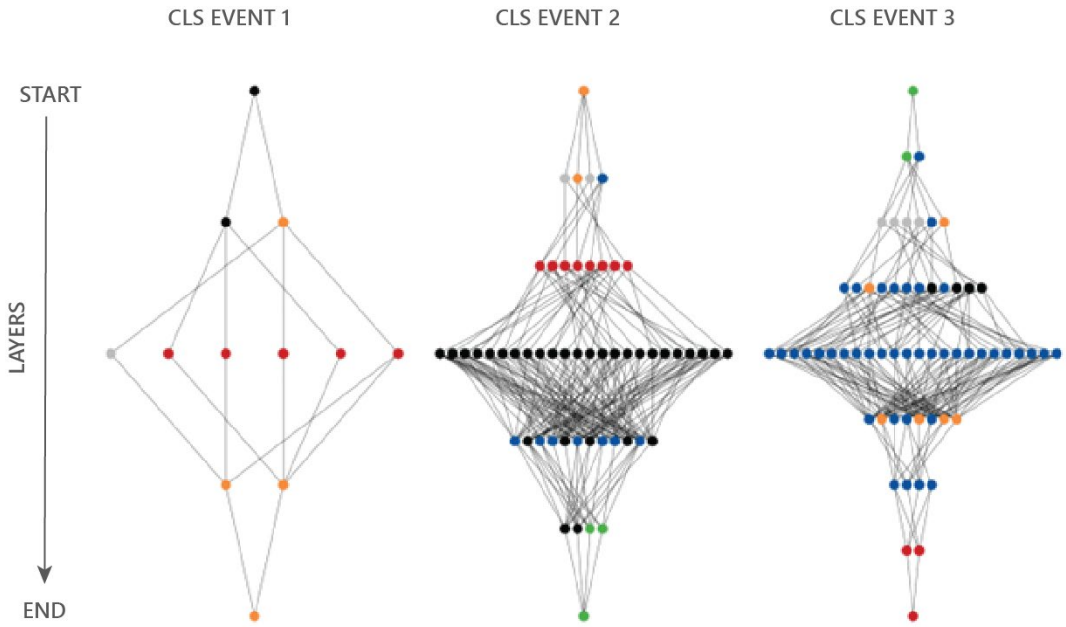
The money laundering techniques used in this PoC are:

1. **Complex layering schemes (CLSs).**[45]
2. **Smurfing.**[46]

Complex layering scheme typology

A CLS event (covering several typologies, sometimes used in combination) can be represented as a network structure, with a clear starting point (the source) and endpoint (the sink), and with layers of transactions and accounts in between. A few examples of network structures representing a CLS event are illustrated in Graph 8.

Graph 8: Simplified example patterns of complex layering schemes (CLS)



The plot illustrates three different network structures of the CLS events. The coloured nodes represent accounts located in various FIs from different countries involved in the money laundering network. The line between entities represents transaction flows.

[45] This includes techniques/typologies such as money muling, abnormal cross-border transactions, transactions from dormant accounts, sudden account emptying and more.

[46] The main purpose of smurfing is to avoid reporting thresholds, but it can also contain abnormal cross-border transactions, sudden account emptying and other behaviours.

The CLSs in Graph 8 are embedded into the synthetic data and are represented in the following way:[47]

1. **Start and end** – the CLS begins with a starting point (top node) and ends with an endpoint (bottom node). These two mark the beginning and end accounts of the money laundering process.
2. **Funds allocation** – at the starting point (top node), a certain amount of illicit funds are assigned. These funds are used as the initial funds for money laundering activities.
3. **Layers** – between the starting point and endpoint, there are several layers, each representing a step in the money laundering process. These layers consist of different accounts at financial institutions.
4. **Layer size** – the layers are then organised in a particular way. The first half of the layers are larger than the subsequent layer (branching-out), and the second half are smaller (branching-in) than the previous layers. This arrangement creates a complex layering structure throughout the money laundering process.
5. **Connection between layers** – each layer is connected to the next layer, ensuring a progression from one layer to the other. This connection simulates the flow of funds during the money laundering process.
6. **Flow of funds** – as the process moves from one layer to another, the illicit funds are split and transferred to the next layer, with a slight decrease in the amount of funds at each subsequent layer. This decrease represents the payment made to the people (eg mules) involved in the money laundering process.

The steps described above lead to the creation of graphs that represent complex layering schemes, as shown in Graph 8. These graphs are then incorporated into the relationship mapping to generate synthetic data that accurately capture these money laundering activities.

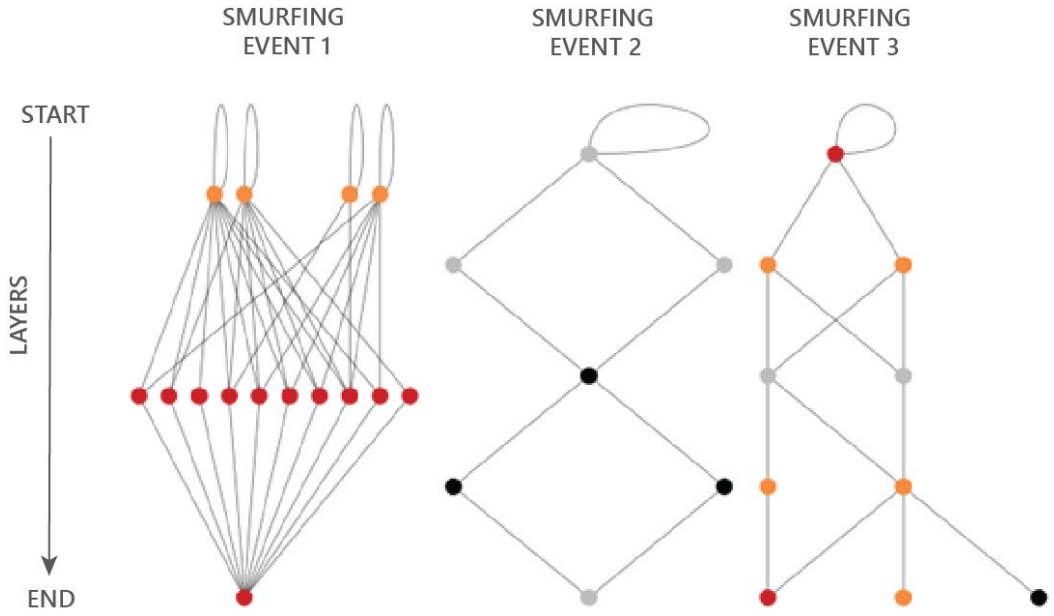
Smurfing typology

Many similarities exist between the smurfing typology and the complex layering scheme typology, and both are able to generate complex networks of money laundering entities. The main way in which smurfing differs is that transactions falling within this typology are all designed to evade reporting requirements.

Moreover, smurfing events can take place over multiple days and the number of accounts involved can vary from a small network to a large one, meaning the complexity of a smurfing event can also differ.

[47] This is a simplified explanation of the CLS event. The CLS event embedded in the synthetic data includes additional factors such as different types of accounts, varying transaction sizes, intermediaries involved and transactions taking place across multiple countries.

Graph 9: Simplified example patterns of smurfing events



The plot illustrates three different network structures of smurfing events. The coloured nodes represent actors involved in the money laundering network. The lines between entities represent transaction flows.

The simulation of the smurfing typology, as shown in Graph 9, is structured as follows:

1. **Start and end** – the smurfing scheme begins with one or more starting points (top nodes) and one or more endpoints (bottom nodes). For example, smurfing events 2 and 3 have one starting point and smurfing events 1 and 2 have one endpoint. The loop at the start point of each smurfing event represents the deposit of cash below the reporting threshold into one or more accounts that are under the control of a money launderer. In smurfing event 1, multiple accounts are involved, whereas in smurfing events 2 and 3, a single account is used.
2. **Funds allocation** – at the starting points, illicit funds are assigned and used as the initial funds for the money laundering activities.
3. **Layers** – between the starting point and endpoint, there can be one or several layers, each representing one step in the money laundering process. For example, smurfing event 1 consists of one layer between the starting and endpoint, while smurfing event 3 consists of multiple layers.
4. **Layer size** – for each layer between the start and endpoints, the number of accounts is either increased or decreased.
5. **Flow of funds** – the illicit funds are then transferred over a pre-defined period. If there are multiple layers (eg in smurfing events 2 and 3), the accounts at the starting point transfer the illicit funds to accounts in the next layer with a slight decrease in the amount of funds at each subsequent layer.
6. **Repeat** – the above step is repeated until the money has reached the end accounts.

The steps described above lead to the creation of graphs that represent smurfing events, as shown in Graph 9. These graphs are then incorporated into the relationship

mapping to generate synthetic data that accurately capture these money laundering activities.

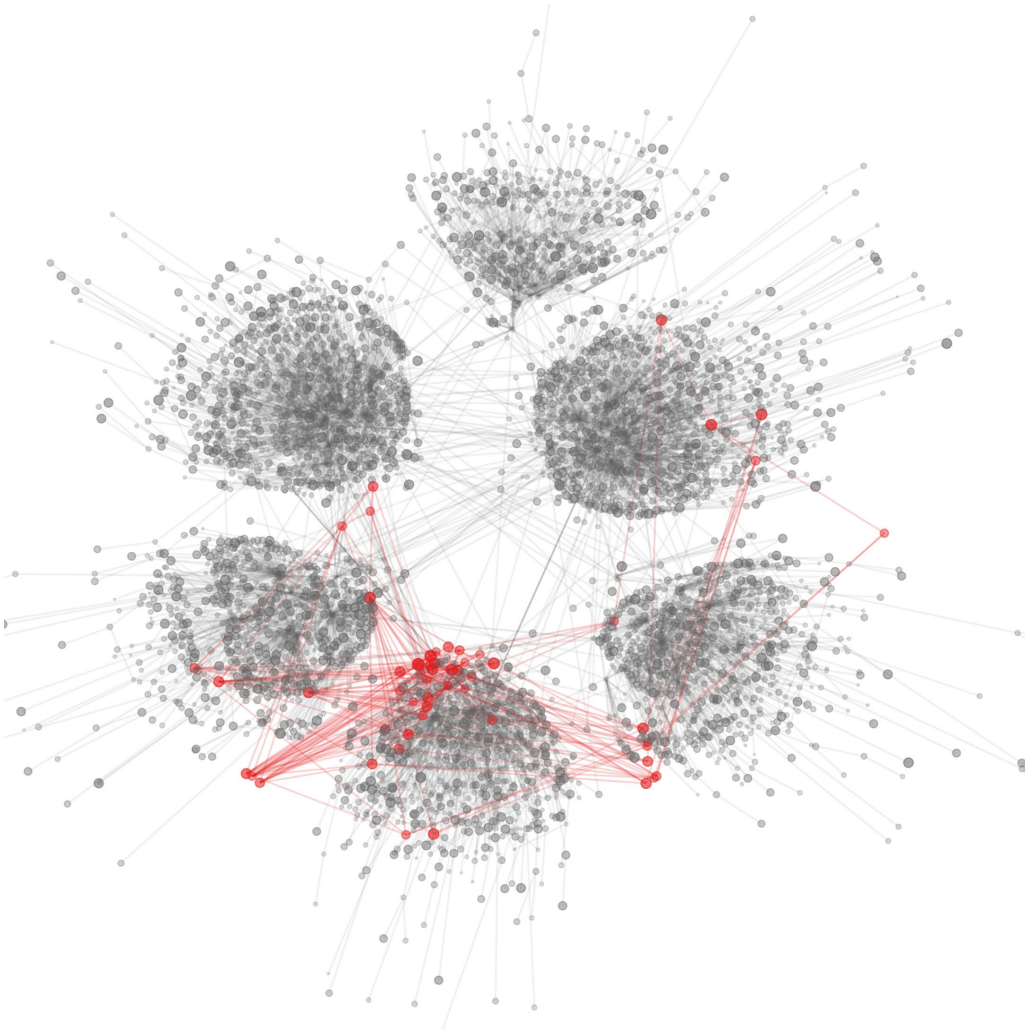
Key assumptions for the synthetic data generation

Duration of the money laundering event: money laundering events can happen over varying time periods. For the synthetic data, the maximum duration of a money laundering event was set at 14 days.

Standardised data and consistent data structure: the data fields and structure are standardised across FIs and different countries.

Complete data points: there are no missing data fields in the data set.

Graph 10: Example of a one-week transaction graph with integrated CLS events



The grey clusters indicate transaction networks in a given country. Money laundering events are indicated in red.

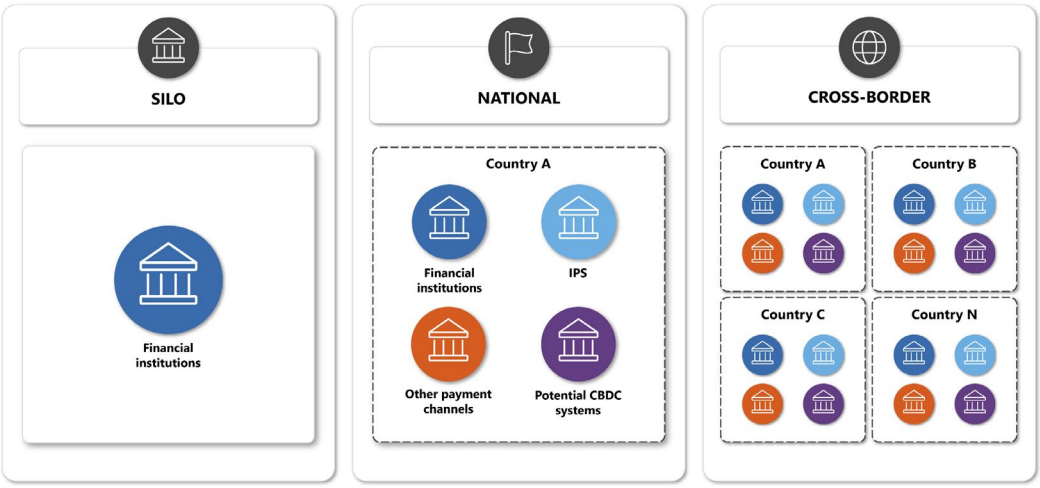
Aurora

4B Part B - Machine learning

4.3 Part B: Application of machine learning to the synthetic data set

This part of the project demonstrates the application of a few machine learning models on the synthetic data set generated in Part A. The performance of machine learning models is compared in different monitoring scenarios, shown in Graph 11, to explore the possibilities and limitations of suspicious transaction monitoring under different views of the data across FIs and borders.

Graph 11: Three monitoring scenarios



Silo: each financial institution conducts transaction monitoring independently, using only their own transaction data.
National: transaction data are visible on a country level. Each country can only see its own transaction data. Transaction monitoring and model training is done on a country level. This could be an instant payment system, national transaction monitoring utility, other collaborative arrangement or potential CBDC system.
Cross-border: transaction data are visible on a cross-border level. Transaction monitoring and model training are undertaken at a cross-border level.

4.3.1 Machine learning models

The machine learning models explored in the project can be categorised into two types: supervised and unsupervised.

- **Supervised models** are trained on data containing labels indicating whether a transaction is illicit or legitimate. These models can leverage prior knowledge to identify suspicious patterns and transactions within the data. An advantage of supervised learning is that the model can be trained to predict specific outcomes with higher accuracy, provided there are sufficient labelled representative data. Supervised learning models can be easily shared where the data structure is similar.

- **Unsupervised models** do not rely on labelled data; instead, they aim to identify patterns, outliers and relationships within the data without prior knowledge. This approach can be advantageous when labelled data are scarce, difficult to create or expensive to purchase, which can often be the case in the AML space. An advantage of unsupervised learning is that it can uncover hidden or unexpected

patterns in the data that might not be obvious to AML experts. However, assessing the quality of the model's output without the use of labelled data can be a challenge.

Table 2 summarises the models used in this PoC and their capabilities in AML transaction monitoring.[48]

Table 2: Machine learning models explored

Model name	Type	Use in AML
Logistic regression	Supervised learning	Widely used linear statistical model. It estimates the probability of a money laundering event based on the input data. It is a simple and interpretable method.
Artificial neural network	Supervised learning	Neural network model that can discover complex, non-linear relationships in the data for detecting money laundering. It approximates the unknown function that links transaction data properties to the money laundering labels.
Graph neural network	Supervised learning	An advanced model that can identify complex relations in data represented as graphs, making them useful for detecting money laundering. It uses the structure of the transaction network to detect money laundering events.
Isolation forest	Unsupervised learning	Detects unusual patterns in data by randomly dividing the observations into groups, and isolating anomalies that appear early in the process, which makes it a good choice for detecting money laundering activities.

In the model comparison, a **rule-based approach** is introduced to replicate the common practice of using rule-based methods to detect suspicious transactions. This is introduced as a benchmark, so that both a rule-based approach and the machine learning approaches use the same data and can be compared.

[48] The project incorporates three supervised learning models with varying model complexities, along with one unsupervised learning model. These models are chosen because they are widely employed in AML research. The logistic regression model represents a simple linear classification model, while the artificial neural network has the ability to capture both linear and non-linear relationships. Additionally, the graph neural network can extract information from the network efficiently. While isolation forest is unsupervised, it is capable of adapting to the data to detect outliers and anomalies potentially linked to money laundering.

Assumptions

The assumptions underlying the models and the rationale for selecting specific features[49] used for monitoring are:

- **Selection and use of features.** The features used in the models have been derived from financial crime research and consultations with compliance experts. A range of features are included in the models, for example the ratios of unique counterparties, transaction values, counts, ratios, and accumulated values of different types of transactions, speed of movement of funds and the sum of squared distances to report thresholds or accumulated funds. The PoC assumes consistent use of these features. Please see Annex B for further information about the model features.

- **Stationary patterns.** It is assumed that the patterns of illicit transactions and the different data features associated with money laundering remain stable over time. Therefore there should be sufficient information to train machine learning models with historical data to make predictions.

4.3.2 Testing the models

After selecting and training the machine learning models to be used (see Annex B for a description of the training process), they were tested and evaluated for performance and efficiency on the three different monitoring scenarios (see Graph 11).[50]

This testing phase involved assessing the models' predictive abilities based on the true positive[51] and false negative,[52] in terms of the model's ability to identify suspicious transactions and networks.[53] This is important because certain transactions may not appear suspicious on their own but may exhibit patterns of abnormal behaviour when viewed as part of the overall network.

The primary metrics used to measure the performance and efficiency of the model are:

- **recall**[54] – measures the detection rate of money launderers, which shows how well the model can detect actual money laundering activities; and

---

49 These features are also known as attributes or input variables. Features are used to represent the underlying patterns and relationships in the data and are used by machine learning algorithms to make predictions or decisions.

50 The machine learning models evaluated were not fine-tuned to achieve the best possible performance on the synthetic data, but were used in their most general form. The comparison assumes equal time and effort spent on setting up each model.

51 The fraction of correctly identified money launderers.

52 The fraction of incorrectly identified money launderers.

53 In the synthetic data, money launderers are flagged when illicit payments are embedded in the transaction data.

54 It is the ratio of true positive predictions to the sum of true positives and false negatives (instances of money laundering activities that the model failed to identify). A high recall indicates that the model is effective in detecting most money laundering activities, minimising the number of undetected instances.

- **reduction of false positive cases** – is a metric used to capture the machine learning models' ability to reduce false positives compared with monitoring using the rule-based method.

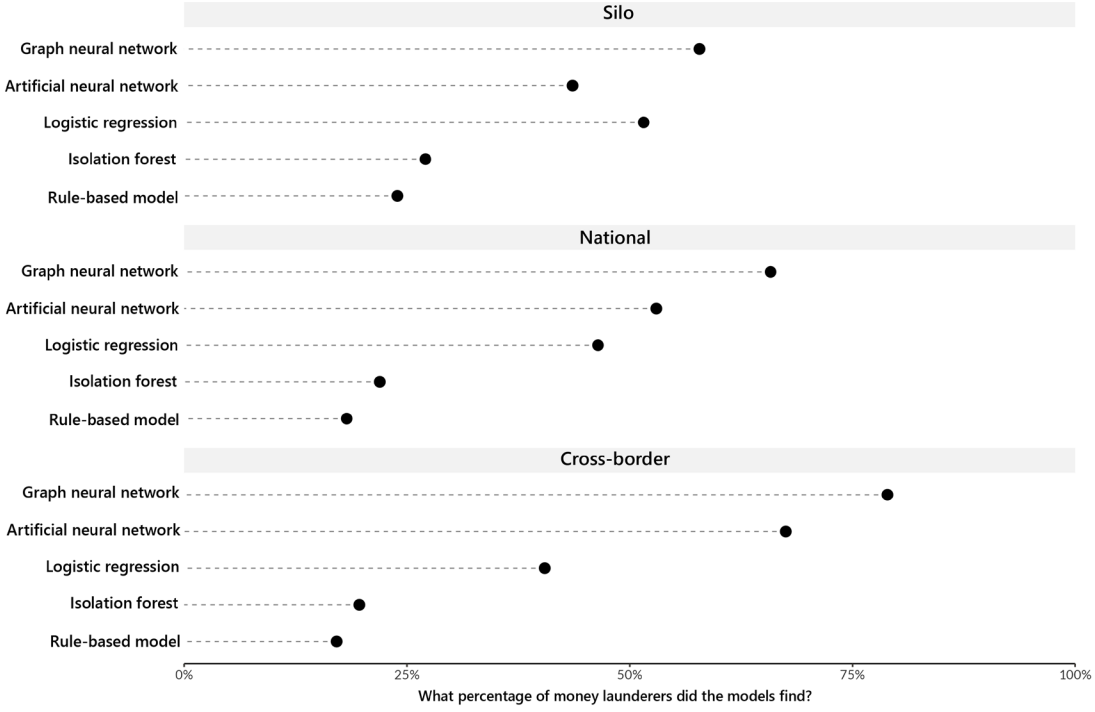
Other metrics included precision[55] and F1 score.[56, 57]

4.3.3 Results

Result 1: Machine learning models outperform rule-based monitoring.

Graph 12 illustrates the performance of four different machine learning models and one rule-based model applied to each of the three monitoring scenarios for detecting money laundering events.

Graph 12: Machine learning models' performance in different monitoring scenarios



Two key findings from the comparison are highlighted:[58]

[55] It is the ratio of correctly identified money launderers to the sum of all identified money laundering activities (correctly and incorrectly). A high precision indicates that the model successfully identifies money laundering activities while minimising false positives.

[56] The F1 score is useful when dealing with imbalanced data sets, as is often the case in anti-money laundering contexts. Money laundering instances are relatively rare compared with legitimate transactions.

[57] In this experiment, the precision and F1 score findings are similar to the recall ones.

[58] The improvements in the logistic regression and isolation forest are relatively small in the various monitoring scenarios. This is mainly due to the fact that the models have been treated and trained in the same way without further specification to each model in order to make them comparable. In a real case, more specification would be needed to train each model individually. In comparison, graph neural networks are, by design, able to leverage graph structures and capture complex relationships better, such that they show an overall better performance.

1. **Machine learning models are more effective than the rule-based model in detecting money launderers.**

   In this experiment, the rule-based monitoring approach detected only up to 25% of the money launderers in a siloed monitoring scenario, whereas the machine learning models were able to detect more than twice as many.

2. **Machine learning models that incorporate network features are optimal.**

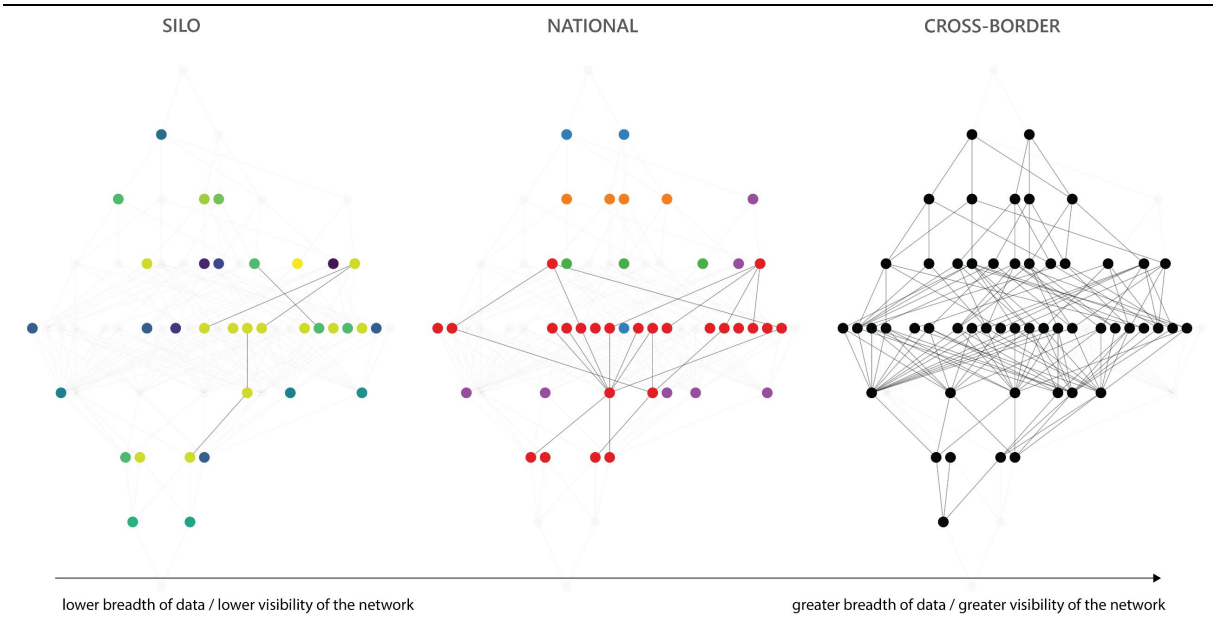
   The graph neural network (GNN) model performs best out of the five tested models when a broader set of data is available.

   With a cross-border view of transaction data, the model can detect approximately 80% of the money launderers in the synthetic data.

Result 2: A holistic view and monitoring of data enhances the detection of complex money laundering networks.

Machine learning models become more efficient when they are based on larger sets of transaction data, thereby providing a better view of transaction networks. Graph 13 illustrates a comparison of each model in each of the three monitoring scenarios.

Graph 13: Visualisation of the detection of money laundering networks with graph neural networks under different monitoring scenarios



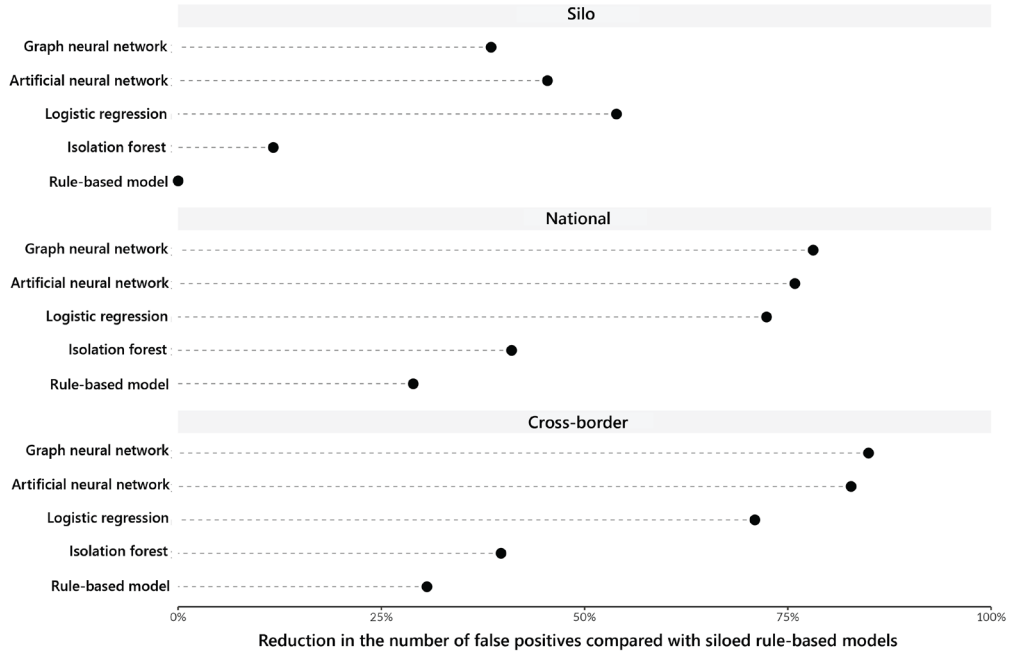
Silo: colour coding represents what a financial institution can detect.
National: colour coding represents what can be detected at a national level.
Cross-border: black dots represent what can be detected at a cross-border level.
The dark lines represent the detected suspicious accounts or transactions.

- A cross-border monitoring scenario is more effective than national and siloed ones in detecting actors in complex layering scheme.
- Siloed and national monitoring scenarios may enable the detection of a subset of the individual entities involved in a money laundering network, but they would lack broader context and would be unable to gain visibility on the broader network.
- However, a national monitoring scenario is still useful for detecting suspicious networks within a country's own jurisdictional boundaries.

Result 3: Supervised learning models flag more suspicious activities and create fewer false positives.

Graph 14 illustrates that when data are analysed in a national monitoring scenario, all models, including the rule-based monitoring model, become more efficient. Further improvements are also observed for some models in a cross-border monitoring scenario.

Graph 14: Potential reduction in false positives compared with rule-based model in siloed monitoring



The rule-based model in the siloed monitoring scenario is used as the benchmark to which the potential reductions are compared. For example, the GNN in the siloed monitoring scenario could reduce the number of false positives by 40% compared with the rule-based approach.

1. **National and cross-border monitoring of transaction data are more effective in reducing the number of false positives than the current rule-based model**:

- Most supervised machine learning models reduce the number of false positives, compared with the rule-based model. The reduction is approximately 40% when monitoring is done in silos, whereas it is approximately 75% when undertaken with a national view of transaction data. Further improvements are observed with the cross-border view.
- Machine learning models are more effective than rule-based ones due to their ability to learn complex data patterns and adjust to sophisticated scenarios.

2. **Supervised machine learning models can be better at reducing false positives:**

- The overall performance of supervised machine learning models is better than for unsupervised models.
- Labelling[59] helps the model better distinguish between suspicious and normal transactions.

[59] Model trained on data that contains labels about whether the transaction is illicit or legitimate.

- A graph neural network model works best when monitoring is done at the national or cross-border levels, because the model can learn more from a larger observed network structure.

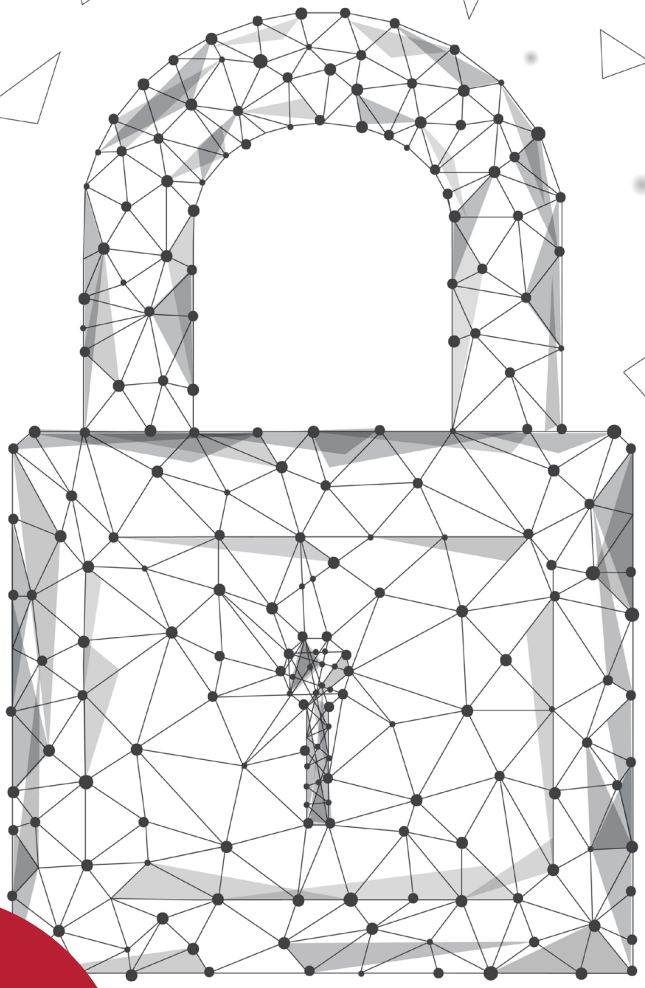
4.3.4 Summary

In conclusion, Project Aurora successfully tested and compared different machine learning methods and a rule-based monitoring tool on different views of synthetic transaction data. These included siloed, national, and cross-border views representing monitoring scenarios to detect money laundering networks. The results showed that machine learning models can be more effective than the traditional rule-based approach, particularly when those models incorporate network features.

Moreover, holistic transaction monitoring, which includes monitoring across multiple payment systems and borders, enhances the detection of sophisticated money laundering networks while also reducing the number of false positives.

Finally, supervised machine learning models are more effective in reducing false positives. These findings suggest that machine learning models, when combined with national or cross-border transaction monitoring arrangements, could be valuable in AML efforts.

Aurora



4C Part C: Privacy-enhancing technologies

4.4 Part C: Testing privacy-enhancing technologies for AML

Part B demonstrated the potential benefits of using machine learning models on transaction data to detect complex money laundering networks at the national and cross-border levels. However, implementing such CAL arrangements could be challenging due to concerns relating to data protection, privacy, security, competition and legal compliance.

To address some of these challenges, this part explores the use of privacy-enhancing technologies (PETs) to facilitate secure and privacy-preserving CAL arrangements, and assesses the effectiveness of machine learning models in combination with different PETs. This part of the PoC is structured into three stages:

1. Testing different PETs in different simulated CAL arrangements (centralised, decentralised and hybrid).

2. Applying the best performing machine learning models from Part B to each arrangement.

3. Evaluating the detection capabilities of these models on privacy-enhanced data and assessing the feasibility of using each PET in each CAL arrangement.

4.4.1 Privacy-enhancing technologies explored

Three PETs are explored in this PoC:

- **Homomorphic encryption (HE)** is an emerging and evolving technology that enables computations to be performed on encrypted data. This can be useful for privacy-preserving deep learning and cloud computing, for example when analysing transaction data in AML efforts while protecting personally identifiable information (PII).[60]

- **Local differential privacy** obscures individual records in a data set (ie locally at the financial institution level) while allowing for accurate analysis, making it a useful technique for ensuring some level of privacy-protection of non-PII data in money laundering detection. However, the process of adding noise to raw data can lead to reduced model performance.[61]

---

[60] To date, the computational overhead associated with HE can lead to increased processing times and resource requirements, however this is rapidly changing. Some hardware manufacturers and technology companies believe in the potential of the technology and are investing in further research and development. An ISO standard (ISO/IEC 18033-6:2019) for homomorphic encryption has also been developed.

[61] If the model is sensitive to the noise introduced by local differential privacy, the model performance can be worse after applying the PET. In the project, the machine learning methods are tested under local differential privacy to show whether the performance is affected by the noise.

- **Federated learning** is a decentralised machine learning technique that enables multiple entities to train a shared model collaboratively without the need to share raw data. This technique can help address issues such as data privacy, data security, data access rights, data localisation and access to heterogeneous data. However, it faces challenges such as increased processing and communication overhead, data quality, trustworthiness of the data and participants and a reduced ability to detect money laundering networks across different financial institutions versus centralised approaches.

The choice of PETs depends on the specific requirements, resources and constraints of the institutions involved in anti-money laundering efforts. A hybrid approach that makes use of multiple PETs may achieve better results by leveraging the unique strengths of each PET.

While other emerging PETs such as multi-party computation (MPC), zero-knowledge proofs (ZKPs) and private set intersection (PSI) could offer innovative privacy-preserving solutions, they were not tested at this stage.[62] Nevertheless, these PETs are worth considering for future investigations. (Please see Annex C for more details on PETs.)

Limitations in using homomorphic encryption

In the first part of the experiment, homomorphic encryption was used as the primary PET.[63] This method appeared to offer promise because it could allow all data to be encrypted and shared into a centralised data pool that could then analyse it whilst protecting the security and privacy of the shared data.

However, implementing this PET fully in the PoC proved unfeasible due to the size of the encrypted financial transaction data.[64] Therefore alternative approaches involving a combination of HE and other PETs, such as local differential privacy were explored.

4.4.2 Testing a combination of privacy-enhancing technologies in four different collaborative analytics and learning arrangements

Project Aurora experiments with privacy-enhancing technologies in four different collaborative analytics and learning (CAL) arrangements.

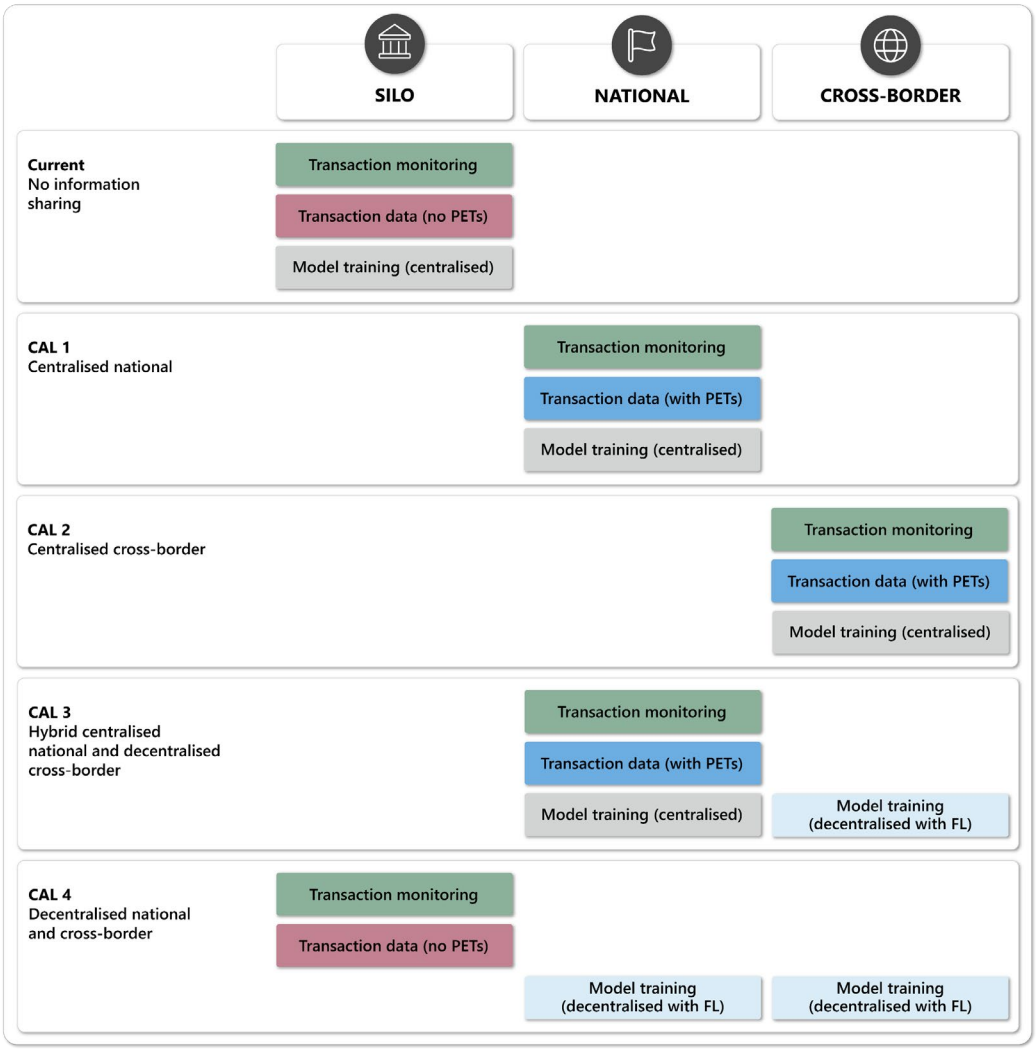
---

62 For example, The Alliance for Privacy Preserving Detection of Financial Crime (APP DFC) is researching the use of MPC and ZKPs, and synthetic data generation for application in KYC, AML and fraud detection.

63 See Cheon et al (2017). A fully HE approach was used in this project.

64 Even though this was tested using open source FHE libraries on distributed infrastructure with up to 70GB of RAM, it was not possible to overcome the challenges posed by the size of the encrypted data.

Graph 15: Overview of the different PET-enabled CAL arrangements explored



The PETs used in the project where homomorphic encryption (HE) and local differential privacy (LDP). The decentralised model training utilises federated learning (FL) with collaboration at the FI or national level.

Graph 15 illustrates the different CAL approaches in this experiment:

- **CAL 1 (centralised national)** – data are encrypted using HE, obfuscated using LDP and the privacy-enhanced data are shared in a centralised national system. The machine learning models are trained and applied on national-level data.

- **CAL 2 (centralised cross-border)** – data are encrypted using HE, obfuscated using LDP and the privacy-enhanced data are shared in a centralised cross-border system in which the data set represents all shared transactions. The machine learning models are trained and applied on data at the cross-border level.

- **CAL 3 (hybrid national and cross-border)**[65] – data are encrypted using HE and obfuscated using LDP. The privacy-enhanced data are shared into a centralised national system. Each country then collaborates on training a machine learning model using federated learning without the need to share transaction data across borders.

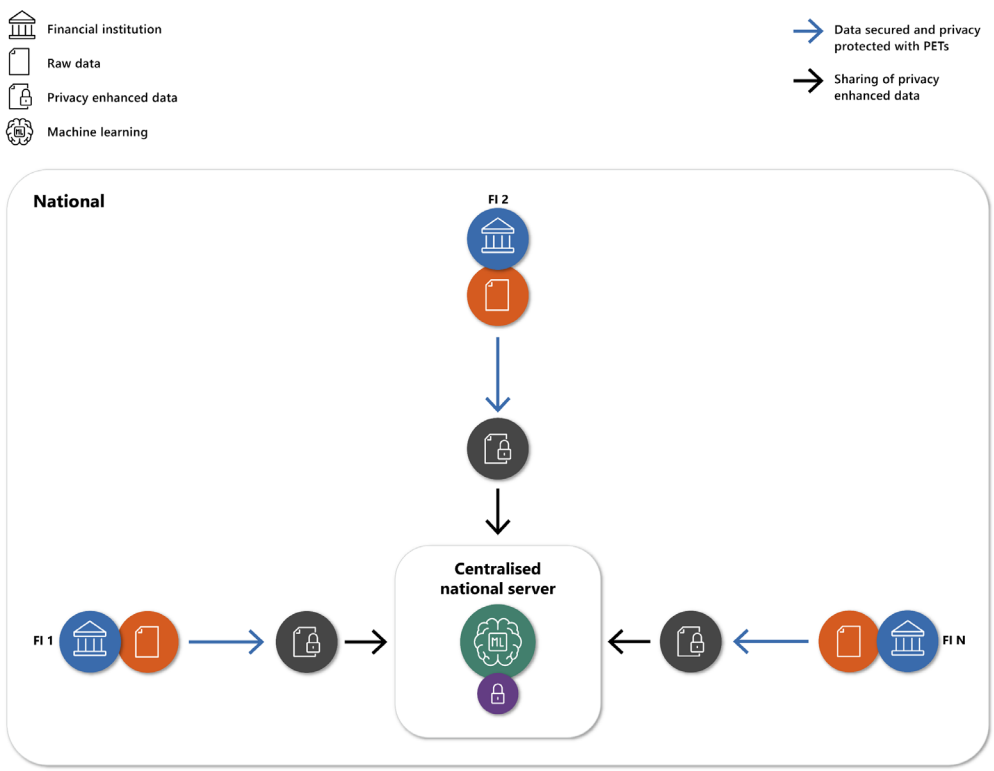
- **CAL 4 (decentralised national and cross-border)** – participants collaboratively train a machine learning model using federated learning locally on their own data and share the model updates to a common global model. This aggregates the local model updates into a global model, which is then, in turn, shared.

Combining homomorphic encryption and local differential privacy – applying PETs according to the sensitivity of the data

This approach was tested to address the limitations of using HE as the primary PET. In this approach, personal identifiable information (PII) data, including sensitive account identifiers, are protected using HE, while local differential privacy (LDP) is used as the PET on transaction flow data (see Annex C for further details). As illustrated in Graphs 16 and 17, this approach shows that different PETs could be used in combination and appropriate to the sensitivity of the data fields.

[65] Transaction data can be processed on a central server (eg in each country) while still allowing each country to collaboratively train models and improve the results of the national analysis without sharing the transaction data directly.

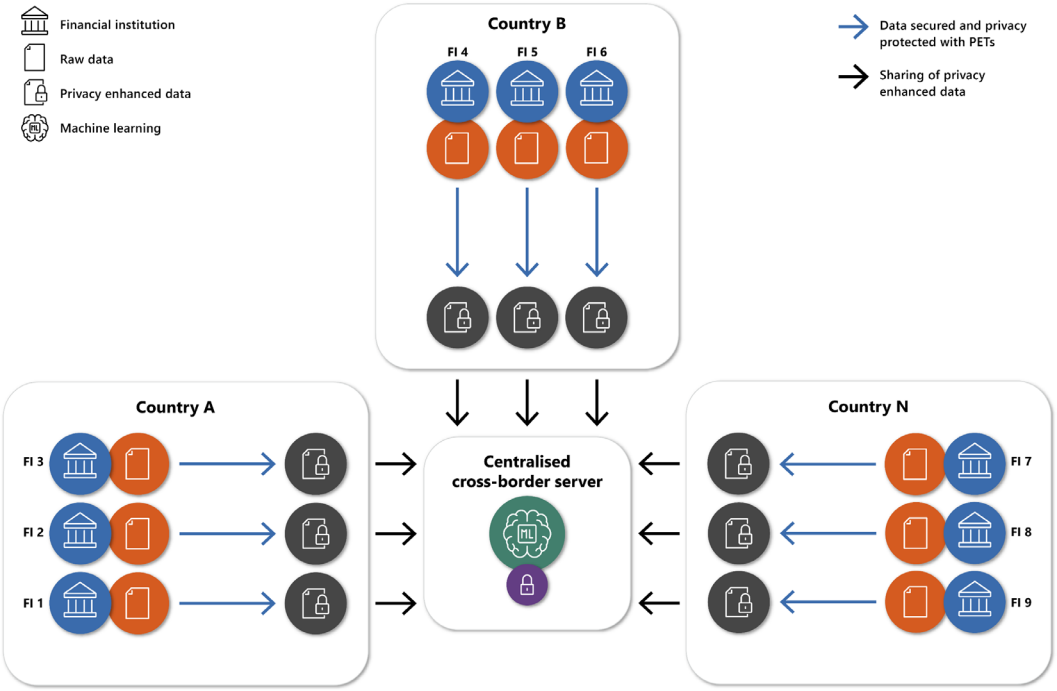
Graph 16: CAL 1 – a centralised national approach using PETs



Each financial institution encrypts and obfuscates its own transaction data using HE and LDP, respectively. Financial institutions then share the encrypted data into a central server located within the financial institution's home country for transaction monitoring purposes.

The analysis of findings (flags) can be approached in various ways. For instance, they can be analyzed centrally at a server, or they can be sent back to each financial institution. However, this experimentation did not make any assumptions about where the results and findings would be shared.

Graph 17: CAL 2 – a centralised cross-border approach using PETs



Each financial institution encrypts and obfuscates its own transaction data using HE and LDP, respectively. Each financial institution shares transaction data on a cross-border central system.

The analysis of findings (flags) can be approached in various ways. For instance, they can be analyzed centrally at a server, or they can be sent back to each country or each financial institution. However, this experimentation did not make any assumptions about where the results and findings would be shared.

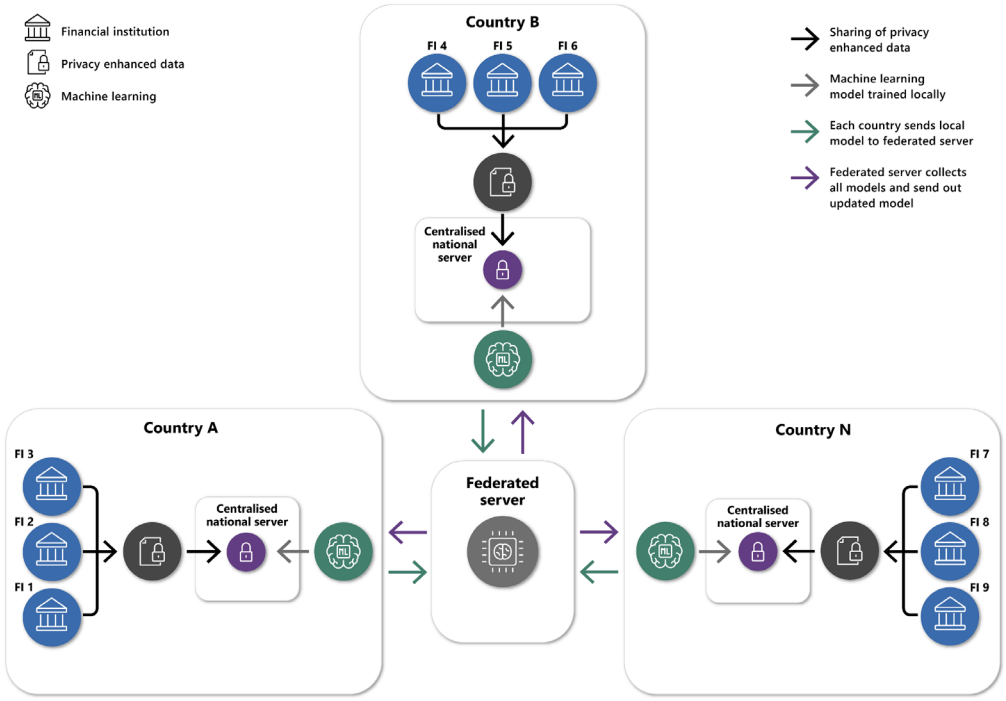
Using federated learning (FL) for decentralised and hybrid CAL arrangements

For the decentralised CAL arrangement (CAL 4), and partially for the hybrid CAL arrangement (CAL 3), federated learning is utilised.

The main objective of using FL is not to share transaction data but to collaborate with other parties in training a machine learning model capable of detecting money laundering activities. Financial institutions train the model on their own transaction data and share only the model updates (learnings) with other financial institutions through a trusted central server.

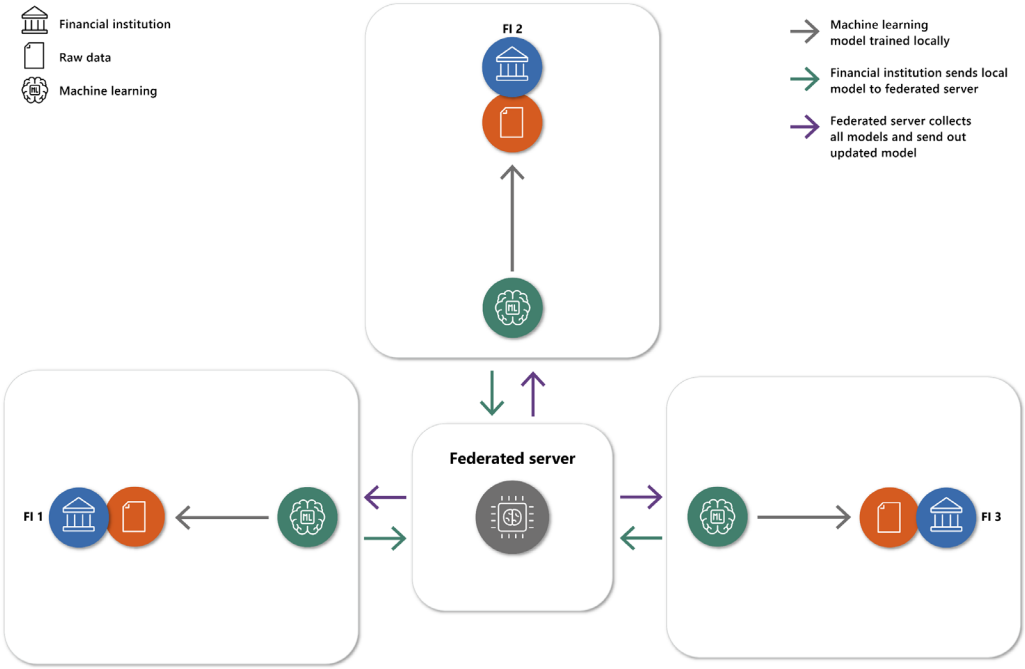
By training the model together, its accuracy and robustness can be improved compared with models trained by financial institutions in isolation (see Annex C for a technical description). Graphs 18 and 19 illustrate the setup for CALs 3 and 4.

Graph 18: CAL 3 – a hybrid national and cross-border approach using PETs



As with CAL 1, each financial institution encrypts its own transaction data and shares them with a national central server. Additionally, each country trains the FL machine learning model on their own data and shares the model updates with other countries using federated learning.

Graph 19: CAL 4 – a fully decentralised national and cross-border approach using FL



Each financial institution trains the FL model on their own transaction data and shares the model updates with other financial institutions located within and across borders. No transaction data are shared with other countries.

4.4.3 Results: applying machine learning models in combination with privacy-enhancing technologies

The second part of the experiment consists of applying the best performing machine learning models from Part B on the privacy-enhanced data in each simulated CAL arrangement, to test and compare the performance of each in detecting money laundering networks. This section presents the results obtained from the experiment, comparing the performance of these CAL arrangements.

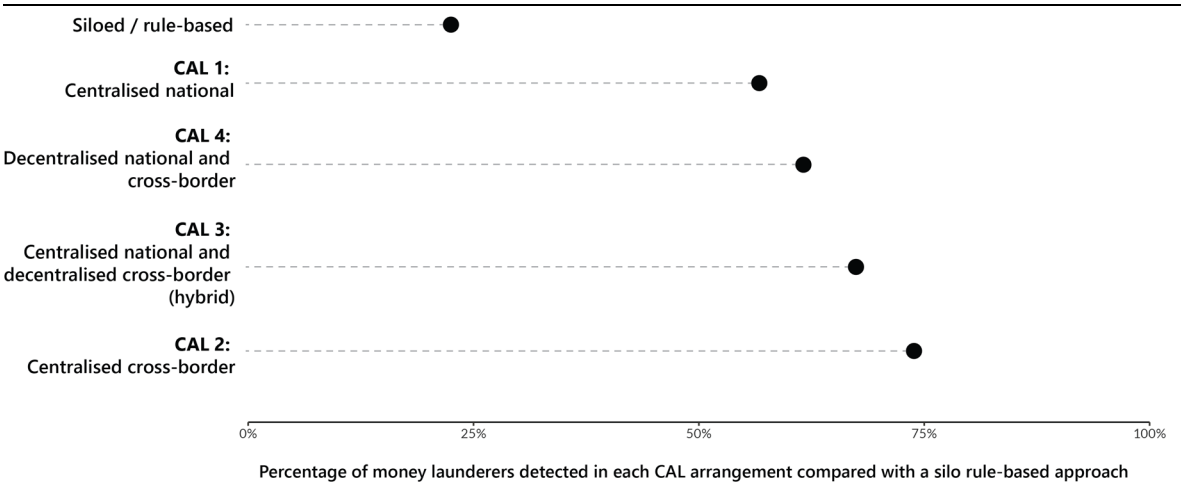
Three machine learning models – the graph neural network, artificial neural network and logistic regression – were found to have superior performance in Part B. These models are used in each CAL arrangement to evaluate their performance on the privacy-enhanced data, the potential loss of information when applying PETs and the optimal CAL arrangement for detecting complex money laundering networks.

Additionally, a rule-based model is tested against a CAL 2 (centralised cross-border) arrangement versus a siloed view to compare its performance against machine learning models and evaluate potential efficiency gains from each CAL arrangement.[66]

Result 1: PET-enabled CAL approaches can improve performance by up to three times in comparison with rule-based monitoring without CAL.

Graph 20 shows the results of the effectiveness of siloed rule-based monitoring compared with each CAL arrangement in detecting money launderers. using the best performing machine learning model in each CAL arrangement. The evaluation metric is the percentage of money launderers found by the model (known as "recall").

Graph 20: Machine learning models under PET enabled CAL arrangements could detect a larger proportion of money launderers



The results correspond to the best performing machine learning model under each CAL arrangement.

[66] The models will be compared using the same metrics as in Part B, namely recall and reduction in incorrectly identified money launderers. The metrics will be presented as the percentage of money launderers each model finds, and time savings from a reduced number of false positives.

1. **PET enabled CAL arrangements lead to more effective identification of money launderers when using machine learning models**.

- On average, the machine learning model under national or cross-border CALs with PETs can detect two or three times more money launderers than with rule-based monitoring with no CAL arrangement.

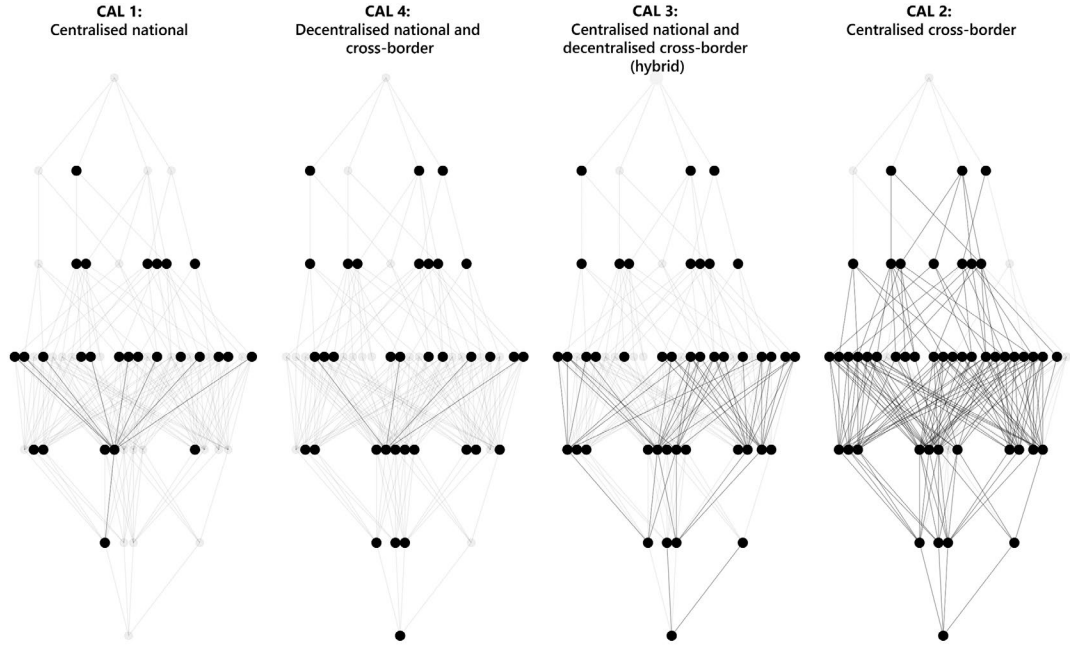
2. **PETs can facilitate secure CAL arrangements and improve cross-border detection capabilities**.

- CAL 1 (centralised national model) yields the lowest recall score amongst the four CAL arrangements. Approximately 60% of all money launderers are detected by CAL 1.
- CAL 2 (centralised cross-border model) yields the highest recall score, indicating that it can detect money launderers most effectively. Approximately 75% of all money launderers are detected by CAL 2.
- CAL 3 (hybrid model) can detect 70% of money launderers, a slightly lower recall compared with CAL 2.
- CAL 4 (fully decentralised model) exhibits slightly lower performance and results than CAL 3. This implies that a fully decentralised approach that preserves privacy and protects sensitive information whilst enabling increased global collaboration, may involve a trade-off in terms of its detection capabilities compared with the centralised and hybrid CAL arrangements.

Result 2: a centralised cross-border CAL arrangement performs better in detecting cross-border money laundering networks.

While various CALs may appear to perform equally in terms of the number of money launderers detected, it is essential to differentiate between the quantity of money launderers detected and the identification of money laundering networks. A broad set of data is needed to identify complex criminal networks. Graph 21 compares results from different CAL arrangementss in detecting "CLS event 3" from Part CAL 2 (centralised cross-border), can detect more money launderers and discover the underlying network structure across countries.

Graph 21: Comparison of network detection capabilities in different CAL arrangements



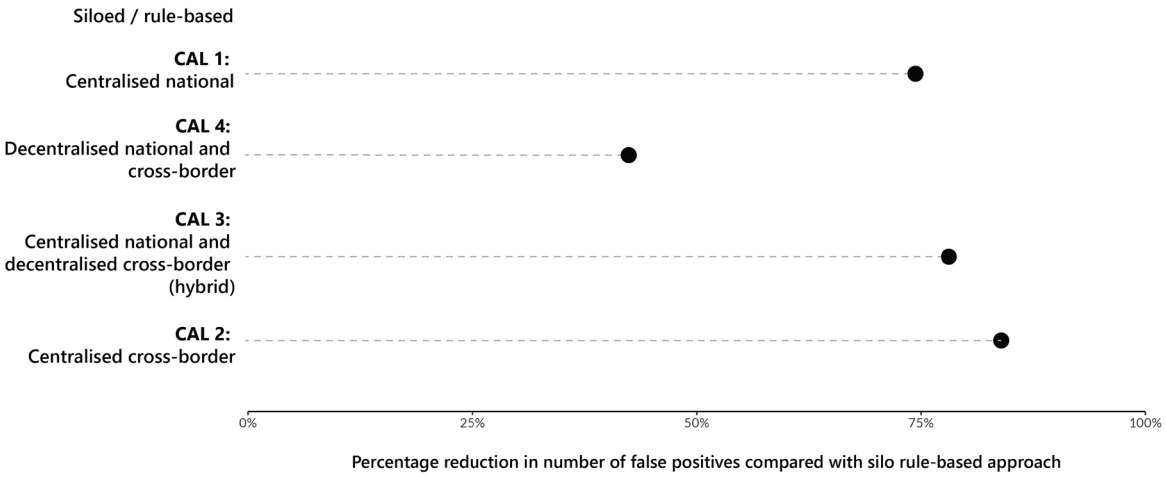
The grey and black nodes represent actors involved in a money laundering network. The dark lines between actors represent the suspicious transaction flows that are detected.

The improvement of the machine learning models' detection capabilities under different CALs is due to increased data availability and an improved view of the transaction network. Results 3 to 5 below were discovered from the comparison.

Result 3: PET-enabled CAL together with machine learning-based network analysis appears to reduce the number of false positives by up to 80% compared with the siloed rule-based method.

Graph 22 demonstrates that CAL arrangements and machine learning models could reduce the number of false positives compared with siloed rule-based models (which serve as a benchmark).

Graph 22: Machine learning models and PET-enabled CAL arrangements could reduce the number of false positives



The results correspond to the best performing model under each CAL.

A. PET-enabled CAL arrangements lead to more efficient transaction monitoring with a reduced number of false positives.

- Graph 22 suggests that the four PET-enabled CAL arrangements can reduce false positives by between 40 and 80% compared with the siloed rule-based method, with CAL 2 (centralised cross-border) having the highest reduction in false positives.

B. National and cross-border CAL arrangements have the highest reduction in false positives.

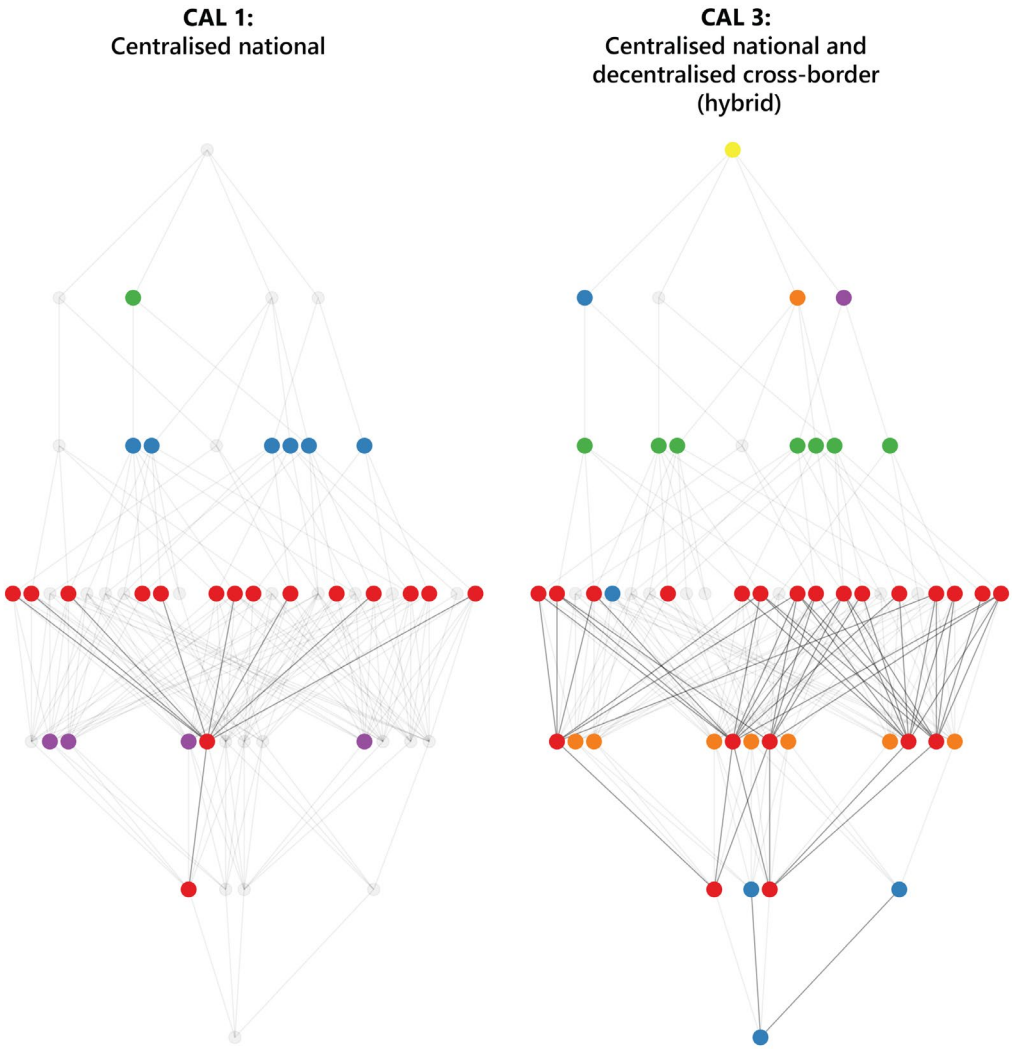
- The number of false positives differs in each CAL arrangement.
- Notably, among the four CALs, CAL 2 (centralised cross-border) demonstrates the greatest reduction in false positives compared with the siloed rule-based method.
- CAL 1 (centralised national) and 3 (hybrid), show reductions in false positives that are comparable to CAL 2.
- CAL 4 (decentralised national and cross-border) was only able to reduce false positive cases by 40%.

Result 4: National CALs are limited to detecting national networks only.

The detection capabilities of CALs 1 and 3 to identify money laundering networks are shown in Graph 23:

- CAL 1 enables the discovery of a part of the money laundering network within a country's own borders but it cannot detect a network outside its own country.
- Similar to CAL 1, CAL 3 also enables the detection of money laundering networks within a country's borders. However, the implementation of a federated learning model enhanced each country's capability to detect money laundering networks within its own jurisdiction.

Graph 23: Comparison of network findings between national CAL arrangements

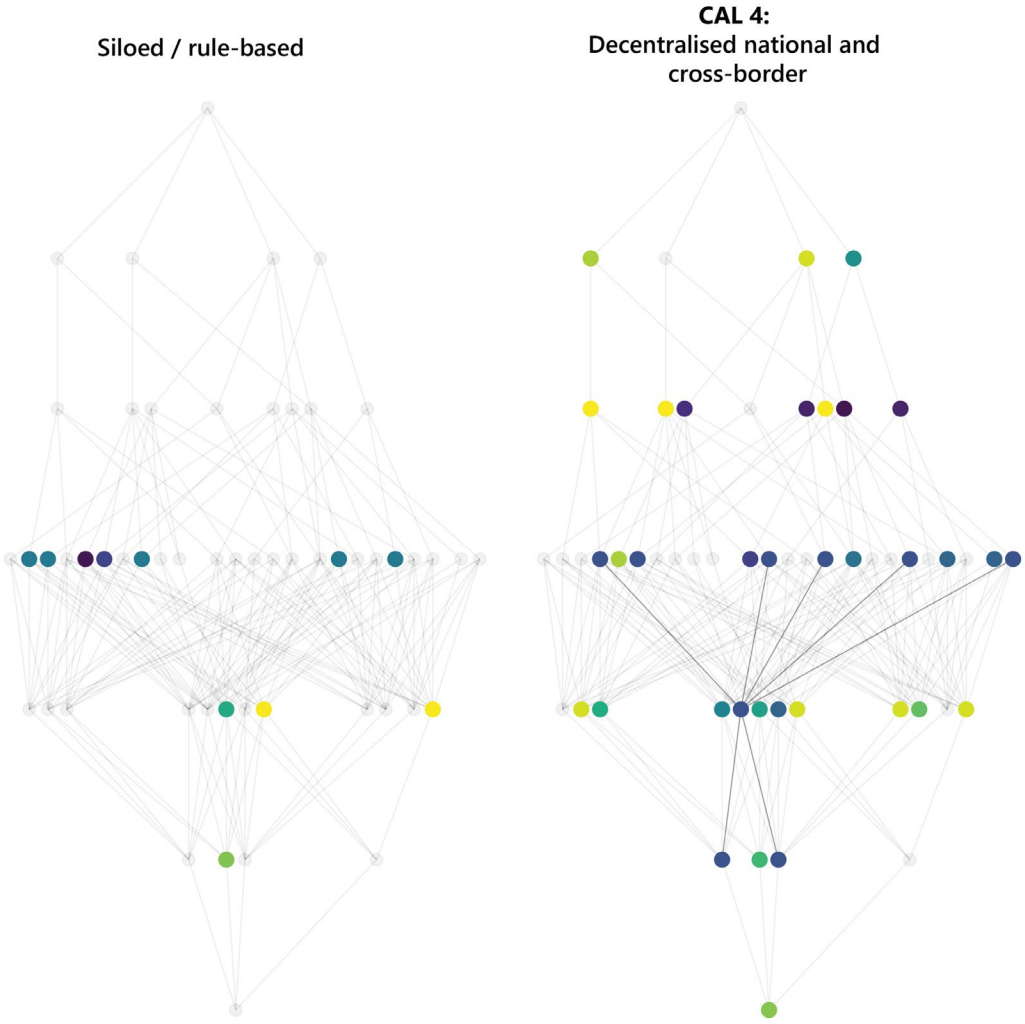


The coloured nodes represent actors in different countries (one colour per country) involved in the money laundering network. The line between entities represents transaction flows.

Result 5: Fully decentralised CALs can improve local detection.

As illustrated in Graph 25, CAL 4 improves each financial institution's local detection of money launderers, but may miss the network linkages between launderers.

Graph 25: Comparison between CAL 4 and the current siloed approach with no CAL arrangement



The coloured nodes represent actors in different financial institutions (one colour per financial institution) involved in the money laundering network. The lines between entities represents transaction flows.

4.4.4 Privacy evaluation of PETs when encrypting transaction data.

The experiment showed that homomorphic encryption (HE) and Local Differential Privacy (LDP) are techniques that can be used to protect personally identifiable information data and transaction data.

HE allowed computations to be performed on encrypted data without the need to decrypt it, while LDP obfuscated the data before sharing, thereby making it difficult to link data to specific individuals. Additionally, the implementation of federated learning showed that FL is a promising PET that could provide stronger privacy guarantees, as the data remain in each party's systems and only model updates are shared.

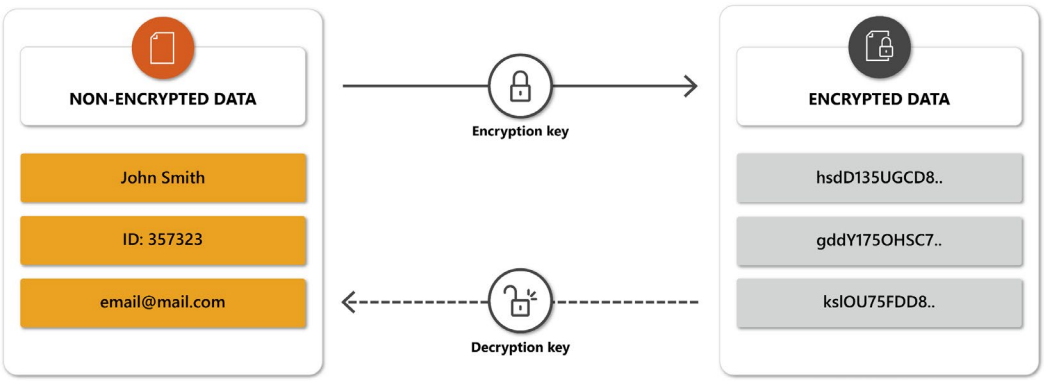
Privacy risks and attacks

Although these PETs provide privacy advantages, they, like any other technology, have limitations and potential risks.

It is crucial to obtain legal clarity regarding the classification of privacy-enhanced data, especially personal data. This legal clarity is important for individuals and businesses to understand the extent of their rights regarding their personal data, as well as how those data are being used and protected by financial institutions. This promotes transparency and trust between financial institutions and their clients, and ensures that everyone is operating within the boundaries of the law.

This section does not aim to provide a legal assessment, but instead it will discuss various perspectives and risks on the use of PETs for data protection.

Graph 26: Example of pseudonymisation, encryption and decryption



It is important to note that encrypting data does not necessarily mean that the data become anonymous. While anonymisation ensures that data cannot be traced back to individuals, pseudonymisation often involves replacing identifiable information with a unique identifier. Therefore, depending on the PET used, PII data are pseudonymised, rather than anonymised, when encrypted. It may still be possible for

the original data to be traced back to an individual if an identifier is linked with external information (eg by the use of a decryption key) as illustrated in Graph 26.[67]

- With **HE**, the encrypted data are still linked to an individual or business, and the individual can be re-identified with the necessary decryption keys, so the encryption of the personal data could possibly be considered pseudonymous rather than anonymous. Furthermore, there is a risk that an attacker could use side-channel attacks[68] or other methods to decrypt data.

- Similarly, **LDP** provides some level of protection for sensitive information by adding random noise to obfuscate the data before sharing, but it also has limitations. By adding noise to the data, it becomes more difficult to link data to specific entities, however it may still be possible for a threat actor to use external information to identify individuals.

- Lastly, **FL** could be vulnerable to data leakage attacks, where the original data could be reconstructed from the model updates. This vulnerability could pose a risk for both CAL 3 and 4 which use federated learning. The potential risk of model leakage could be mitigated with secure multiparty computation and secure aggregation. Further experimentation would be beneficial. It is important to evaluate the strengths and weaknesses of each technique in the context of a use case and consider combining different privacy-enhancing solutions.

4.4.5 Summary

In conclusion, the experiment shows that all CAL arrangements are superior to scenarios in which there is no such arrangement in place for detecting financial crimes. Each arrangement has its strengths and weaknesses, but the overall outcome is improved detection capabilities and fewer false positives. The key takeaway from this experiment is that sharing information and network analysis can provide innovative solutions in the fight against financial crime.

Centralised cross-border (CAL 2) is the most effective method for detecting money laundering when using machine learning models. It performs three times better in detecting money launderers and reduces the number of false positives by 80% compared with rule-based monitoring with no CAL arrangement. This arrangement is also the best at detecting a much larger network of money launderers, making it the most efficient at detecting cross-border money laundering.

The centralised national (CAL 1) and hybrid centralised national plus decentralised cross-border (CAL 3) methods also showed significant improvements, compared with rule-based monitoring with no CAL arrangement. The hybrid model, CAL 3, detected almost twice as many money launderers, reduced false positives by

---

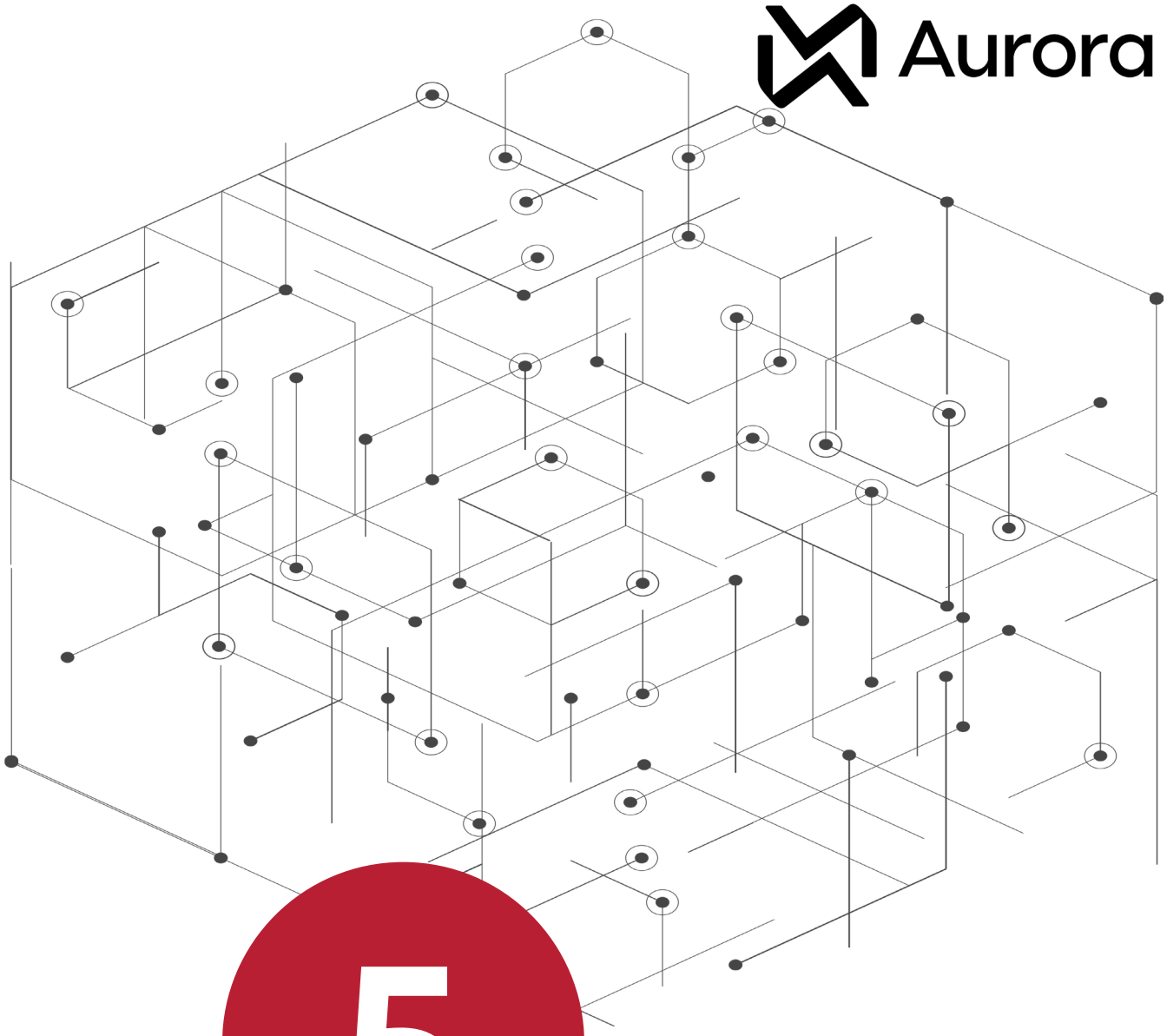
[67] The concept of link-ability could be relevant if illicit activity is identified in the pseudonymised data set and it needs to be traced back to an identified person, subject to appropriate controls.

[68] A side-channel attack is a method of hacking that exploits weaknesses in a system's physical implementation, such as power consumption, electromagnetic leaks, or sound, to extract sensitive data.

over 75% and detected a larger part of the money laundering network, compared with rule-based monitoring with no CAL arrangement. However, it is important to note that the national models cannot detect networks outside their borders, even if they perform better than rule-based monitoring.

The fully decentralised (CAL 4) method improved the performance of each simulated financial institution's local detection, by detecting twice as many money launderers compared with the rule-based monitoring with no CAL arrangement. It was able to reduce the number of false positives by 40% but was not as effective as other CAL arrangements in detecting the full money laundering network since the underlying transaction data are not shared and each institution could only analyse its own data.

Furthermore, the PoC successfully protected personal and sensitive data by encrypting and obfuscating specific data fields using a combination of PETs. As PETs have limitations and potential risks, it is important that there is legal clarity regarding the classification of privacy levels of the data when encrypted. Encryption does not necessarily mean data are anonymous, and some PETs may only provide pseudonymisation, making data potentially traceable to an individual. Combining different privacy-enhancing solutions could mitigate information security and privacy risks.



5 Further considerations

5. Further considerations

Project Aurora combined machine learning and network analysis, privacy-enhancing technologies on the synthetic transaction data, which demonstrated that CAL approaches leveraging these technologies and the value of payments data could be a more effective way to detect suspicious networks and illicit flows across institutions and borders.

The PoC is based on several simplifying assumptions listed earlier, however in reality such an undertaking would face several challenges that would need to be addressed.

This chapter covers a few discussion points on possible data, technology and policy considerations that could be useful in future work that leverages the findings of this project to improve AML efforts.

5.1 Data

5.1.1 Additional data and money laundering typologies

Not all money laundering typologies can be detected solely through transaction data analysis and using a minimum set of data points. Depending on the typology, additional data sources and data points may be required to aid detection, for example:

- **Personal use of business accounts.** This typology can be detected by monitoring know-your-customer (KYC) data such as ultimate beneficial owner (UBO) labels, tax reports, business expense receipts or through the observation of large transactions from legal entities to individuals.

- **Transacting with politically exposed persons (PEPs).** This typology can be detected by flagging entities and counterparties of transactions as PEPs. In a collaborative data aggregation setting, this would require the sharing of this flag among the relevant parties.

- **Off-ramping into crypto assets.** With the growing popularity of cryptoassets and VASPs, it is important to note that money launderers can use these channels to hide their trail of money. These typologies may begin in the traditional financial system but can quickly move into public blockchains in which it can be difficult to link specific transactions to individuals or entities due to the pseudonymity of blockchain addresses. Additional information from the VASPs (eg KYC data) and analysis of the relevant blockchain are necessary to detect these typologies.

- **Trade-based money laundering (TBML).** TBML involves the exploitation of international trade to move value around the world, often using complex financial transactions and multiple intermediaries to obscure the origin and movement of funds, and artificially inflating or deflating the prices of legitimate products. While Project Aurora shows that collaboration through a CAL arrangement can yield better results in detecting complex schemes like TBMLs, there can still be a lack

of access to vital data that would help with these kinds of investigations, such as invoices, shipping records, product price indices and more.

5.1.2 Real-world data are crucial for understanding the feasibility and impact

Although CAL arrangements could offer many potential benefits, there is a need to undertake real-world pilots at scale, ideally conducted over longer periods of time, running in parallel with current approaches. This would assist in understanding their actual effectiveness and uncover further issues or questions that need to be addressed. Real-world pilots at scale would provide important learnings and results that support policy, data and legal discussions.

There is currently no clear consensus on how CAL approaches, particularly in a cross-border context, can be arranged. The CAL arrangements explored in Project Aurora could be fed into these discussions.

Designing CAL arrangements and collaborative analytical capabilities would require consideration of a number of key issues. These include funding, membership, governance, objectives and performance monitoring. Further, consideration of the nature of analytical capabilities and the scope of the crimes covered by the analysis would be necessary. Issues relating to visibility, control, liability for data and analysis, audit and assurance, and the role of public agencies also require examination.

ISO 20022 messaging standards include additional standardised data points which could potentially improve CAL efforts to combat financial crime more effectively. Further practical work to better understand how ISO 20022-based data could be leveraged for a variety of use cases, including for financial crime, could be beneficial. This includes identifying further enhancements to the standard. Objectives for achieving a standardised data framework for national and cross-border financial crime detection could be considered.

Additional issues relevant to the development of CAL arrangements, highlighted by Project Aurora, include defining the types of data to be utilised, the use of privacy-enhancing technologies and desired privacy protection levels, ensuring cyber and information security, defining operating procedures and professional standards, and engaging law enforcement agencies. Questions to support the design of real-world pilots can be found in Annex D.

5.1.3 Limitations of payments data and the need for other data

Network-wide transaction data provide a large analytical advantage in discovering previously unknown accounts linked to high-risk or suspicious activity. However, for approaches with less visibility, or limits and thresholds on the input data, there is a benefit in achieving data minimisation and a decreased level of intrusion into privacy. Nevertheless, there is a trade-off, in terms of reduced efficacy, in being able to identify network-wide risks that are not visible to individual members.

The principles of data minimisation and purpose limitation are important. With this in mind, it should be noted that payments-level data alone may not provide a complete view of financial crime risk and additional data could be required depending on the

financial crime being targeted. For example, KYC, tax authorities or corporate registries data.

The principal challenge in the design of current AML systems is that the visibility on payments of individual financial institutions is limited to those made by customers and relevant counterparties. Their awareness of KYC is also limited to their direct interactions with customers and does not necessarily extend to the wider range of financial relationships and accounts that the customer may have with other financial institutions. Conversely, national and international payments infrastructure could have much greater visibility on payment flows between financial institutions, but no visibility on KYC data.

5.1.4 Data protection

To ensure responsible and ethical sharing and processing of data, it is important to establish a transparent process and open dialogue. This can facilitate the creation of clear legislation and guidelines, which can further promote innovation and consistency in national and cross-border collaboration frameworks.

Protecting data and safeguarding privacy, as well as combatting financial crime, are significant public interests that need to be balanced. These objectives are not mutually exclusive or conflicting, but complementary. Finding the right balance between these objectives is essential to ensuring the responsible and ethical use of data while also maintaining the integrity of the financial system. It is crucial to uphold the rights of those whose data are being shared, while also strictly limiting the processing of data to its intended purpose to prevent potential misuse of personal information.

5.2 Technology

5.2.1 Technical challenges with CAL arrangements

Any CAL arrangement could encounter several technical challenges. These include:

- **Data quality and consistency**. Differences in data collection, storage and processing methods among financial institutions may lead to inconsistencies, discrepancies or biases in the shared data – affecting the performance of detection models. To address this challenge, FIs must agree on standard data collection and processing methods and implement data quality control mechanisms to ensure consistency and accuracy in the shared data.

- **Feature selection and engineering.** Ensuring consistent identification of the most important factors or characteristics for detecting money laundering becomes challenging in collaborative settings involving multiple institutions. To solve this problem, financial institutions can work together to establish a shared framework for selecting and analysing these important factors.

- **Computational cost and information overload.** Many machine learning models and PETs currently face the challenge of high computational costs, which may increase in a non-linear manner when handling a large amount of data. Data

volumes would be very large depending on the jurisdiction and the institutions involved. The private and public sectors should cooperate to find an efficient solution for the monitoring system.

- **Model divergence due to underfitting and overfitting.** When using the federated learning approach, there is a risk that the model might not work well, either underfitting or overfitting, if the different data sets used for training contain limited information or differ a lot in data structure. To prevent this, financial institutions should collaborate on reviewing the model, and consolidate the data structure and standards.

- **Lack of real-world labels globally.** In order to identify money laundering activities globally, it can be difficult to obtain accurate and trustworthy information. Even if individual institutions or countries have some information, it may not apply to other places. To solve this problem, financial institutions can create a way for experts to provide feedback on the system, which can help ensure that the system is receiving accurate information and improving over time.

- **Fairness and impartiality.** When machine learning models are utilised for AML, it is essential to evaluate the fairness and impartiality of the monitoring model and the policy recommendations. In Project Aurora, the potential risk of machine learning bias is mitigated because the synthetic data generation process is simple and transparent, and the statistical simulation does not introduce bias in the training sample. The variables usually associated with bias and discrimination, eg gender and other demographic information, are not part of the synthetic data. Moreover, the machine learning models in the project are explainable, which helps to open up the "black box" and examine any potential model biases.

- **Governance, risk management and information security.** Depending on the CAL arrangement, the capability may have access to a very large amount of data. As a potential "honeypot" of data and given its relevance to disrupting organised crime, CAL arrangements could be at risk of cyber security attacks, infiltration and data corruption attempts from threats such as organised crime, insiders, kleptocrats and malign state actors. Governance, risk management, conduct and assurance processes will need to mitigate cyber security and information security risks.

- **Utility.** Any CAL arrangement would only be as good as the number of participants and data sources available.

5.2.2 Machine readable typologies that facilitate information sharing

Information-sharing is essential for combatting a wide range of threats, from financial crimes to cyber attacks. For instance, in the context of cyber security, real-time

indicators are being shared,[69] and standards to enable the automated exchange of cyber threat intelligence have been implemented,[70] to combat immediate threats.

In the context of AML, data that may need to be shared, such as those relating to financial crime typologies, are not standardised in their definitions or descriptions, and are unstructured and analogue in format. This makes it challenging to use this valuable information for data analysis.

Representing typologies as knowledge graphs, that may be suitable for information-sharing and analysis by advanced analytical systems, could provide a solution. This could be something that the public and private sectors collaborate on to enhance cross-border collaboration and knowledge-sharing. Indeed, it could build on existing efforts, such as those of the European Financial Intelligence Public Private Partnership (EFIPPP).[71]

What is a knowledge graph and how could it apply to AML efforts?

A knowledge graph is a type of dynamic data structure that encodes information as a network of different data points (nodes) and the relationships between them (edges). They provide the ability to include semantics such as synonyms, taxonomies or ontologies (which help provide context and meaning to the data). In particular, they can support the determination of whether one or more data fields with different names or descriptions are in fact the same thing (entity resolution). This is useful in circumstances in which this needs to be carried out at scale and data consistency is important, such as in AML efforts.

Knowledge graphs provide a comprehensive view of information, including relationships to other information, context and meaning, by connecting many different data on a subject. This can drive new insights. Knowledge graphs can be updated dynamically as new information becomes available, for example missing data items, and missing or previously hidden relationships.

Knowledge graphs can provide insights by focusing on exploration, deduction and inferences, making them useful for AML. For example, a graph query can look for specific features that warrant further investigation and graph algorithms can identify new patterns. Graph embeddings can learn from data in the knowledge graph and find new connections in the data that may not have been discovered, based on shared characteristics. These include customer behaviour features and other data points in the knowledge graph that can help to detect suspicious patterns.

Knowledge graphs could be particularly useful for AML by representing knowledge about typologies in digital form. This could be used in public-private partnerships to

---

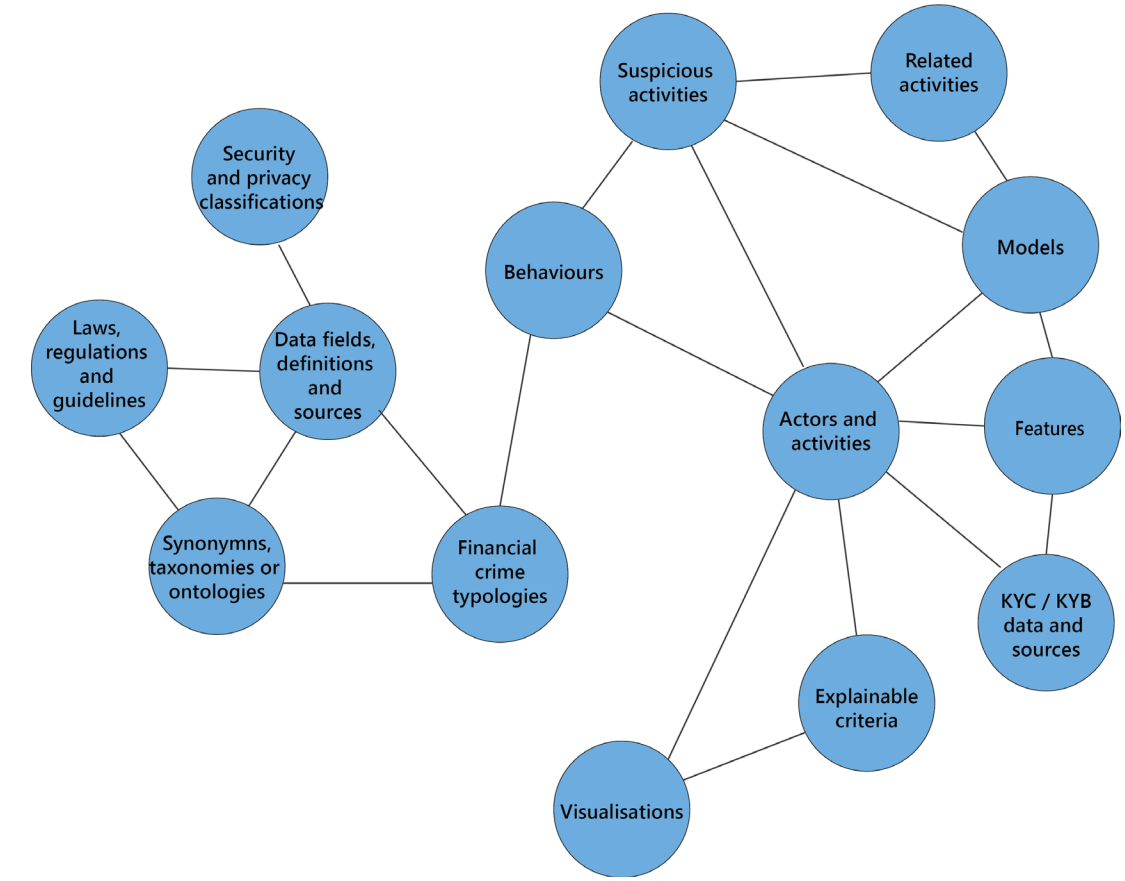
69 See the Center for Internet Security's website at: www.cisecurity.org/ms-isac/services/real-time-indicator-feeds.

70 See OASIS Open's website at: www.oasis-open.org/2021/07/14/new-versions-of-stix-and-taxii-approved-as-oasis-standards-to-enable-automated-exchange-of-cyber-threat-intelligence/.

71 EFIPPP consists of large financial institutions, FIUs and law enforcement agencies from multiple jurisdictions. EFIPPP shares information about different typologies such as investment frauds, trade-based money laundering, virtual assets, narcotics and more.

encode knowledge on typologies and make such code available to financial institutions and authorities for use in advanced analytical systems. This could support greater collaboration, increased information-sharing, better insights and detection capabilities, as well as timely and effective responses to potential threats and risks. Graph 27 shows a high-level illustration of a typology knowledge graph.

Graph 27: High-level illustration of a typology knowledge graph



Conceptually, a knowledge graph could include the following:

- data fields, data types, data sources, security and privacy classifications;
- synonyms, taxonomies or ontologies;
- transaction features;
- behaviours;
- actors and activities;
- explainable criteria;
- suspicious activities and related activities; and
- references to laws, regulations or guidelines.

Further analysis and experimentation would be required.

5.2.3 Explainability

Understanding how machine learning models arrive at their final results is critical, this is also known as model-explainability. Different machine learning models require different approaches to explainability, with some models being more transparent than others. Explainability is crucial for promoting transparency, fairness and accountability in machine learning models.

Being flagged as a suspicious customer can have a significant impact on individuals and businesses, ranging from inconvenience to financial exclusion. It is therefore important for financial institutions to understand and explain how the models work and how findings are calculated. This is to avoid "black box" models that can lead to unfair, biased or discriminatory practices in AML efforts.

It is important to note that the features used in the PoC are inherently explainable. The features are created by aggregating various data points based on pre-defined criteria. By investigating the data points without aggregation, one can understand the information that went into each feature.

While explainability is critical, it is also crucial to consider other factors such as model governance, legal risks and regulatory compliance at an early stage.

For example, model governance involves establishing a framework to ensure that machine learning models are developed, deployed and maintained in a responsible and accountable manner. This could include creating and maintaining an overview of the machine learning models that are being used, allocating appropriate resources for ongoing evaluation of these models, weighing the trade-off between improved performance and high complexity/low explainability, and fostering the sharing of best practices.[72]

5.3 Looking ahead

5.3.1 Instant payment systems, CBDC systems and financial crime

The rising adoption of instant payment systems (IPS) and the potential introduction of CBDC systems offer novel possibilities to strengthen domestic and cross-border payments. The implementation of CBDC systems may result in new participants and may enable different ways of facilitating payments, creating new vectors for financial crime.

It is crucial to assess the potential misuse of these systems by criminals and determine if and how they could create new avenues for combating financial crime. It is therefore important that operators (eg private sector or central banks) work together with financial institutions participating in these systems to ensure that appropriate financial crime controls are in place.

---

[72] See Danmarks Nationalbank (2022).

Operators could conduct preparatory work with public and private organisations to develop a clear understanding of the domestic and cross-border financial crime typologies that may emerge in instant payment and CBDC systems, including money laundering, terrorist financing and other illicit activities. This could be undertaken by collaborating with government agencies and financial institutions to share information and best practices for detecting and preventing financial crime.

The approaches explored in this project could be used by operators considering IPS or CBDC systems that include AML monitoring and analysis capabilities to enable greater detection of networks of suspicious activities.[73]

It is important to ensure that financial crime measures do not interfere with the efficiency and speed of payments, while still providing guidance and regulation to system participants to ensure that adequate measures are taken. A legal framework would need to be in place to govern the use of these systems for AML and other financial crime prevention measures. By addressing these issues, IPS and CBDC systems can be implemented in a way that ensures the safety and efficiency of payment systems.

5.3.2 Legal and regulatory considerations

To realise the potential benefits of CAL arrangements that leverage payments data, a national and cross-border strategy for public-private partnerships to combat financial crime would be beneficial. Such a public-private partnership would require a clear legal basis to enable information-sharing and collaborative analytics. Depending on the different CAL arrangements and technologies involved, there may be different legal, privacy, ethical, policy and regulatory implications to consider.

Legislation that allows for one type of collaboration may or may not support another. Policymakers could reflect on the holistic capabilities that may be needed. This could include the development of financial crime threat typologies, network detection, transaction monitoring, incident reporting, evidence recording and reporting packs.

Collaboration between public authorities and the private sector in creating an appropriate legal and regulatory framework is essential for supporting CAL arrangements. Without such a legal basis, data privacy risks, civil damages, defamation, competition law risk, bank secrecy and AML framework prohibitions against tipping off[74] could hinder any initiatives.

---

[73] Operators could encounter limitations on the types of money launderers that could be detected depending on the data available.

[74] A legal prohibition, for example, on disclosing the fact that a suspicious transaction report or related information is being filed with an FIU.

Aurora

6 Conclusion

6. Conclusion

Project Aurora has demonstrated that analysis of payments data is highly valuable for AML. It offers greater visibility and improved detection of suspicious networks and illicit payment flows across financial institutions and borders. The project simulated a synthetic data set of domestic and international transactions (adopting the principle of data minimisation). It then tested several privacy-enhancing technologies and advanced machine learning methods on siloed data compared with data analysis or learnings made available through four different CAL arrangements.

The main findings of Project Aurora suggest that behavioural-based transaction monitoring and analysis at national or international levels is more effective in detecting money launderers and suspicious networks than current siloed and rule-based monitoring. These approaches could be more effective at, potentially uncovering a larger proportion of money laundering networks by using advanced analytical tools. Graph neural networks are a tool that appears to be optimal for this use case.

Privacy-enhancing technologies can be leveraged to allow secure and privacy-preserving CAL arrangements, including machine training (using federated learning) across financial institutions and borders. This could support privacy-preserving monitoring and AML efforts at an international level.

The use of these technologies and approaches could play a role in catalysing approaches to public-private partnerships, and collaborative analytics and learning. There are several barriers to overcome, as well as some potential limitations associated with the technologies and approaches. Future advances in technology may remove some of the limitations and blockers. Strengthened cooperation between the private and public sectors, between different disciplines and across national borders are key drivers for innovation to improve AML efforts. Policymakers have already identified the need for a strategic approach to encouraging and co-designing the capabilities needed for private-private information-sharing to detect financial crime, however further strategic work on public-private initiatives would be valuable.

Various initiatives, such as the G20 roadmap to enhance cross-border payments, the establishment of national transaction monitoring utilities, the adoption of advanced analytics that focus on behaviour instead of rule-based monitoring, the migration of payments systems to the ISO 20022 messaging standard and the ongoing review of FATF recommendation 16, could have a positive impact on data-driven efforts to combat financial crime.

Greater cross-functional dialogue between experts on payments, financial crime and data protection could support the development of data standards. This could include, for example, leveraging or enhancing ISO 20022 messages, greater use of the legal entity identifier (LEI) in data sets associated to legal entities, the use beneficial ownership standards the design of systems and processes, or the standardisation and representation of typology information or behaviour model development.

Further work

Project Aurora contributes to both national and international discussions on leveraging connected payments data to combat money laundering across institutions and borders. It does so by highlighting the possible opportunities and limitations of different collaborative analytics and learning approaches that are enabled by applying privacy-enhancing technologies together with machine learning and network analysis.

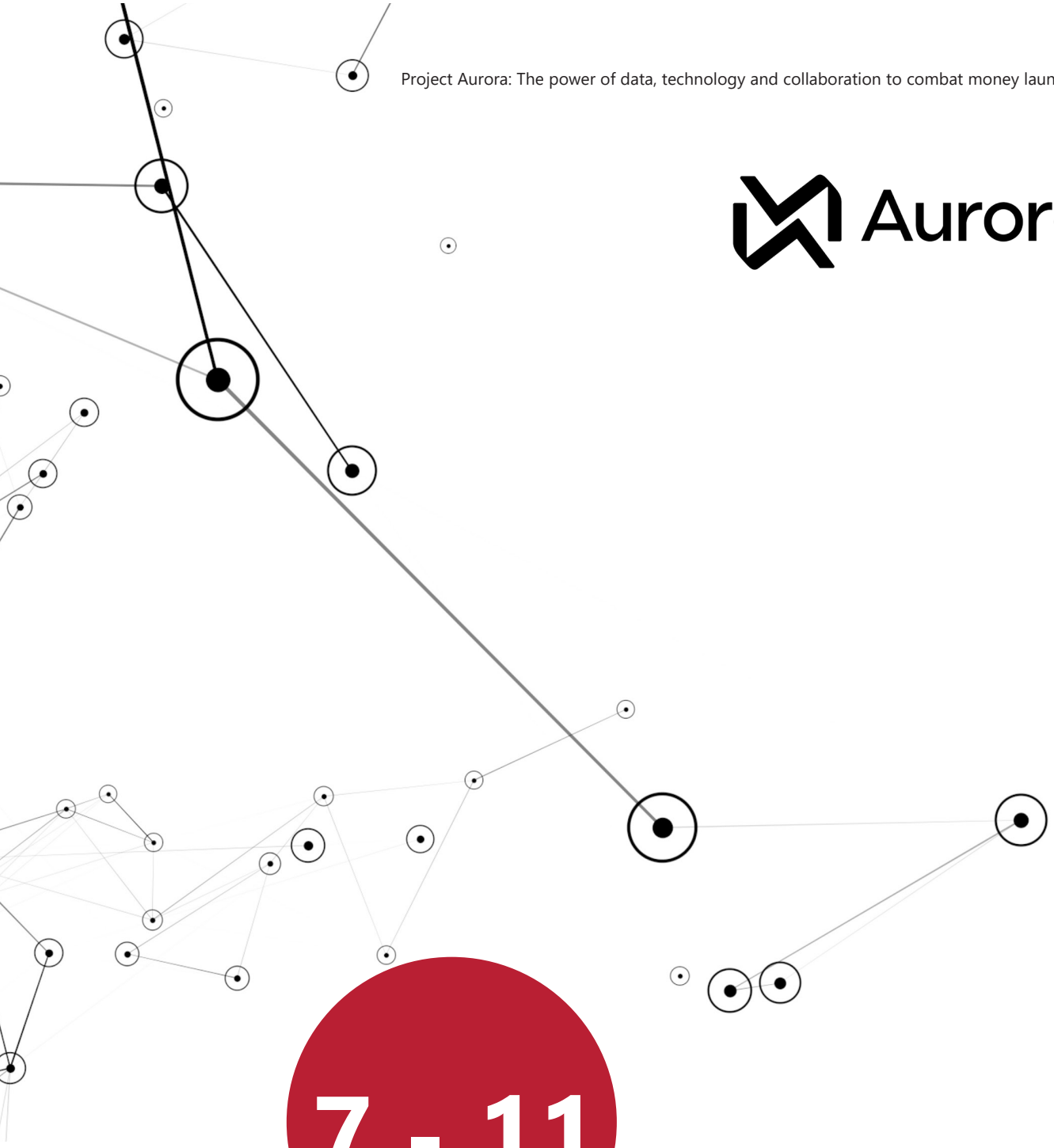
In order to test the feasibility and effectiveness of the technologies and CAL approaches at scale, real-world proofs of concept (or pilots) with a range of different actors across the payments landscape would be needed. Such real-world initiatives could also be useful in surfacing legal, regulatory, data protection, and technical issues and questions that would need to be addressed.

The machine readable typology concept using knowledge graphs proposed in this report could be explored further with public and private sector experts as a technical and process design exercise.

Instant payment systems and potential CBDC systems could also play a key role in detecting money launderers and suspicious networks, depending on the design. Privacy-enhancing technologies and the principle of data minimisation could be important in achieving this.

CBDC systems could provide a new vector for money launderers. Therefore, undertaking some preparatory work at an early stage to simulate potential financial crime typologies[75] that may affect CBDC systems could be beneficial.

[75] Whilst some known typologies may extend to CBDC systems, new typologies emerge all the time, although some characteristics may be common.



7 - 11 Annexes A - E

- A: Trends and opportunities.
- B: Machine learning in this PoC.
- C: Privacy-enhancing technologies.
- D: Questions to support real-world pilots.
- E: Additional acronyms and definitions

7. Annex A: Trends and opportunities

This annex explores the trends, initiatives and developments in payments, data standards and transaction monitoring initiatives. It also explores the potential impact of addressing the challenges facing AML efforts that were discussed earlier in this report.

7.1 Standardisation, transparency and harmonisation in payments

There are various standardisation, transparency and harmonisation initiatives underway. These could help better connect data within and across institutions and borders, and improve data quality, compliance and AML efforts.

7.1.1 G20 roadmap for enhancing cross-border payments

In 2020, the G20 leaders endorsed a *Roadmap for enhancing cross-border payments*.[76] The roadmap aims to enhance cross-border payments and consists of 19 building blocks (BBs) that cover various issues, technologies and arrangements, including financial crime prevention.

7.1.2 ISO 20022 harmonisation

The ISO 20022 standard provides a common language and structure for financial messages that can be used within and across different payment systems and jurisdictions, enabling greater interoperability and straight through processing. ISO 20022 allows for richer and more structured data to be shared via messages, with the potential to enhance the efficiency of AML transaction monitoring systems.

The Bank for International Settlements' Committee on Payments and Market Infrastructures (CPMI) has been tasked with working with the industry to draft a report on the ISO 20022 harmonisation requirements for enhancing cross-border payments.[77] For example, the proposed requirements include:

- Stating the purpose of a transaction (eg cross-border transaction).
- Identification of all FIs involved in cross-border payments using a bank identifier code (BIC).
- Identification of all persons and businesses involved in a cross-border payment in a structured and standardised way by including an identifier such as the legal entity identifier (LEI) for businesses or other legal entities, or some form of standard personal identifier (passport number or national identity number) for individuals, together with other information such as name and postal address. The

[76] See FSB (2022).

[77] See BIS (2023).

use of global identifiers such as the LEI, enables the use of associated reference data[78] for further verification.[79]

The FATF recommendation 16 requires basic information about the sender and recipient of a payment transfer (both domestic and cross-border) such as names, account numbers and addresses, to be included in the transaction message.[80] As part of the planned review of recommendation 16, the FATF may consider updating this recommendation to account for developments such as the adoption of the ISO 20022 messaging standard.

7.1.3 Data standards for legal entity identification and beneficial ownership

Legal entity identifier

The legal entity identifier (LEI) is the only global standard for corporate identification. It is a 20 character, alphanumeric code based on the ISO 17442 standard. Each LEI contains information about the entity and the entity's ownership structure, answering questions on "who is who" and "who owns whom". The LEI data are publicly available and can enhance the transparency of corporate information.

The LEI can link corporate information in different sets of data that also use the LEI, but it can also map to other identifiers. For example, the Global LEI Foundation (GLEIF) provides mappings between:

- LEI and business identifier code (BIC);
- LEI and international securities identification number (ISIN);
- LEI and market identifier code (MIC);
- LEI and OpenCorporates ID; and
- LEI and S&P global company ID.

Work is also beginning to map entity ownership data associated to an LEI to the beneficial ownership data standard (BDOS).

The GLEIF has also introduced a digital version of the LEI known as a "verifiable LEI" (vLEI) based on the concept of self-sovereign identify (SSI). It can support automated and decentralised verification of corporate identity information and could be useful in some cross-border scenarios.

Beneficial ownership data standard

Information on beneficial ownership can increase visibility about how companies and other legal entities are owned and controlled. Such information can include the

---

78 The Global LEI Foundation (GLEIF) make LEI data available.

79 Verification can happen through lookups of the LEI data or could leverage vLEIs, which embed LEI information into a digital verified credential, which is a W3C standard.

80 The travel rule was updated in 2019 to include virtual asset service providers (VASPs) and expanded in 2021 to include private wallets, non-fungible tokens (NFTs) and decentralised finance (DeFi).

identification of individual owners, and other legal entities or intermediate entities in the ownership chain, as well as breakdowns of shareholdings and voting rights. Data are often spread across different sources including annual reports, articles and regular filings. It is increasingly the case that data on beneficial ownership must be disclosed and reported.

The Beneficial Ownership Data Standard (BDOS) provides a structured format and standard[81] for how such data should be collected, shared and used. The standard captures information on identifiers and details for both individuals and corporates, types of levels of involvement in companies, provenance, jurisdiction, and historical and current information. The standard could help beneficial ownership information to be clearly identified and tracked over time.

7.1.4 The Wolfsberg Group Payment Transparency Standards

The Wolfsberg Group was established in 2000, and its members consist of a number of global financial institutions. The group aims to develop industry standards and promote transparency in the global financial system. In 2017, the group published Payment Transparency Standards[82] that consist of ten principles that financial institutions should adhere to when conducting payments.

Furthermore, the Wolfsberg Group has been supportive of discussions around incorporating the LEI into payment messages and encourages further exploration of its potential benefits, particularly in terms of reducing false positive alerts generated by sanction screening systems and transaction monitoring systems.

7.2 Transaction monitoring utilities

Different information-sharing approaches are being explored and established by various stakeholders. These include sharing information directly between financial institutions, such as shared know-your-customer (KYC) utilities – to collect and share KYC information between financial institutions – or suspicious transaction monitoring utilities (TMUs).

In certain jurisdictions, transaction monitoring utilities have already been established and are operating, providing financial institutions with greater means by which to detect and prevent financial crime. For example, Transaction Monitoring Netherlands (TMNL) is a TMU that currently only monitors transactions by businesses.

TMUs are collaborative arrangements between participating financial institutions and, potentially, public authorities (eg law enforcement), mostly at a national level. Transaction data can be shared or pooled, and subsequently analysed. Accordingly, transaction monitoring utilities enable financial institutions and public authorities to

---

81 See Open Ownership for further details and specifications on the Beneficial Ownership Data Standard.

82 See The Wolfsberg Group (2017).

gain greater insights into criminal activities and identify patterns of illicit behaviour that are difficult to detect in isolation.

Opportunities with TMUs

- **Increased efficiency:** TMUs can increase the efficiency of transaction monitoring by consolidating data and providing a centralised platform.

- **Improved accuracy:** TMUs allow for the pooling of data from multiple sources that can increase accuracy in the detection of patterns and anomalies that may not be apparent when analysing data relating to individual financial institutions.

- **Enhanced risk management:** TMUs enable financial institutions to identify and mitigate risks more effectively. By sharing data and insights, financial institutions can better understand emerging risks and take proactive measures to mitigate them.

- **Cost savings:** TMUs can be more cost-effective than individual transaction monitoring solutions, by sharing the costs of developing and maintaining TMUs.

Challenges with TMUs

- **Data protection and privacy:** the sharing of transaction data among financial institutions can raise concerns about data privacy. Financial institutions must safeguard sensitive customer data and make sure that the sharing of data complies with data protection regulations.

- **Operational risks:** TMUs can be vulnerable to operational risks, including system failures, cyber attacks and unauthorised access. Financial institutions must ensure that appropriate security measures are in place to prevent such risks.

- **Legal and regulatory risks:** the use of TMUs can raise legal and regulatory issues, particularly around data-sharing, proportionality and responsibility for monitoring. Financial institutions must ensure that they have the appropriate legal and regulatory certainty before developing and participating in a TMU.

- **Trust and collaboration:** the success of TMUs depends on trust and collaboration among participating financial institutions. Financial institutions must be willing to share data and insights with their peers, which can be challenging in a competitive environment.

- **Standardisation:** TMUs require a high degree of standardisation to ensure that data are consistent and can be analysed effectively. Financial institutions must agree on common data formats and definitions to enable effective analysis.

- **Interoperability:** TMUs must be able to integrate with existing transaction monitoring solutions to ensure that they are effective. Financial institutions must ensure that TMUs can work seamlessly with their existing systems and processes.

- **Governance and supervision:** TMUs require effective governance and supervision to ensure that they are used appropriately, and that data shared are accurate and reliable. Financial institutions must establish appropriate governance structures to manage the use of TMUs effectively.

By addressing these challenges, financial institutions could realise the full potential of TMUs and enhance their ability to detect and prevent illicit activities, and manage financial crime risks.

7.2.1 TMU example: Transaction Monitoring Netherlands

Transaction Monitoring Netherlands (TMNL) is a joint project of five banks in the Netherlands with the objective of improving transaction monitoring and strengthening the role of banks as gatekeepers of the financial system. The ultimate goal is to establish an industry-wide utility for transaction monitoring, enabling financial institutions to detect money laundering and terrorist financing more effectively and providing law enforcement with high-quality information.

Type of data used by TMNL

- Currently, only transaction data from businesses are monitored.
- Data such as the time, destination and value of transactions are used to detect unusual transaction patterns.
- Only receives data that are necessary for monitoring purposes from the banks (eg Chamber of Commerce numbers, IBAN and company names) and they are made pseudonymous by the banks before they are provided to TMNL so they cannot be linked to individual customers and are meaningless without the bank's encryption key.

7.3 Instant payment systems and potential CBDC systems

The shift of certain types of payments from batched transaction processing (settled at specific intervals during a working day) to individual and instant settlement, presents both opportunities and challenges for financial institutions, operators and authorities in relation to monitoring transactions for suspicious activities. Payment systems, such as instant payment systems (or future CBDC systems),[83] make financial transactions faster, more efficient and more convenient for users. However, these same features can also make it easier for criminals to move illicit funds quickly across multiple financial institutions and countries.

[83] CBDCs are a form of digital money, denominated in the national unit of account, which is a direct liability of the central bank. CBDCs can be designed for use either among financial intermediaries only (ie wholesale CBDCs), or by the wider economy on a more individual level (ie retail CBDCs).

Opportunities

- Instant payment systems offer the possibility of following the money across several financial institutions by analysing transaction data centralised in real time.
- This holistic overview of payment system-level data enables the detection of suspicious activities that may not be visible at the individual institutional level.
- Both instant payment systems and future CBDC systems could be designed with built in privacy protections for sensitive user information, described by the principle of data minimisation and functionality for AML regulations by default.

Obstacles

- Criminals use multiple payment accounts and channels across financial institutions and borders. That makes them quickly disappear from one system and reappear in another. The identification of illicit flows with this pattern can be challenging for network analysis when data that is spread across these different ecosystems cannot be connected.
- Some data may reside outside IPS or CBDC systems, reducing visibility.
- CBDC payment systems, depending on their design, could be new vectors for money laundering and may produce new typologies.

7.4 Public blockchains used for payments

Public blockchains are transparent and could offer additional insights as all transactions are publicly recorded and visible to anyone.

Opportunities

- This transparency provides an opportunity for system-wide transaction monitoring, which can be used to identify suspicious patterns of activity that may indicate money laundering or other illicit activity.
- Transaction monitoring on public blockchains can be automated using blockchain analytics tools, which can analyse large volumes of transaction data in real time and flag any potentially suspicious activity.

Obstacles

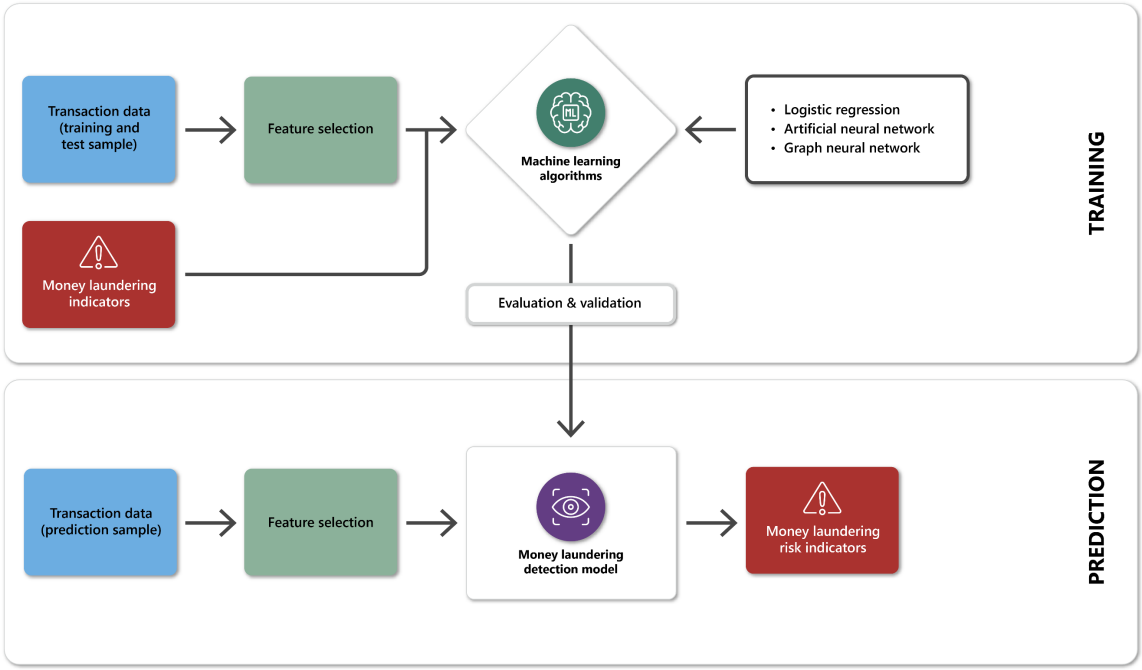
- Transaction monitoring on public blockchains is not a silver bullet, as it can be difficult to link transactions to individuals or entities due to the pseudonymity of blockchain addresses. Some blockchains provide a higher level of privacy by obscuring transaction details, making it harder to track the movement of funds.
- Data storage, retention, localisation and the right to be forgotten are also challenges.

8. Annex B: Machine learning in this PoC

8.1 Machine learning training, validation and evaluation

In this annex, the process of machine learning training, validation and evaluation is presented. Graph B1 illustrates the workflow of supervised machine learning models. Unsupervised models do not make use of the money laundering indicators.

Graph B1: Training, validation and evaluation of machine learning models



1. **Feature engineering**: experts choose relevant data features that can provide information for AML, for instance the regular transaction inflow and outflow for the actor.
2. **Model training**: machine learning models are trained using transaction data. The input features represent the characteristics of the transactions and the model finds the mapping from features to the outcome – the labels of known money laundering events.
3. **Model evaluation and validation**: the trained model is used in a test sample for predicting money laundering event labels. The known money laundering labels are used to evaluate and validate model performance.
4. **Model prediction**: after training, each model predicts the probability of a transaction being a money laundering event for the new prediction sample. The same set of input features are chosen as in the training step. It is a risk indicator for the transaction being part of the money laundering network.

5. **Money laundering events indicators**: it is feasible to map each model's risk scores to a classification (money laundering events or non-events) based on whether the risk score is larger than the chosen threshold.

8.2 Machine learning model feature engineering

Determining the list of features for the machine learning models is important to capture various aspects of transactional activities, for instance transaction counts, transaction values, parties involved in the payment flow and more. In the machine learning model exercises, the following features are used as key variables:

- accumulated sums of various transaction counts (eg inflowing/outflowing cash transactions, total transactions, international transactions);
- accumulated values of various transaction types;
- accumulated sums of unique counterparties for various transaction types;
- ratios of different types of transactions (eg incoming, outgoing, total) over historical data;
- ratios of transaction values for different types of transactions over historical accumulated sums of transaction values;
- ratios of unique counterparties for various transaction types over historical accumulated sums of counterparties;
- sum of squared distances to reporting thresholds for different types of transactions; and
- speed of fund movements and how frequently transactions happen between counterparties.

These aggregated features provide a comprehensive representation of transactional patterns that are input for both supervised and unsupervised machine learning models. It is necessary to note that the list of features is different for graph neural networks (GNN) compared with the other machine learning methods. The GNN model aims at learning the properties of transaction networks, enhancing its ability to detect graph-based money laundering events. The features employed in the GNN model are:

- in-degree and out-degree – these features represent the number of incoming and outgoing transactions for each entity in the relationship mapping;
- total amount sent and total amount received – individuals' and businesses' total amounts sent and received provide an overview of their transactional behaviour regarding transaction values;
- number of transactions sent and received by each entity, individual or business; and
- entity type – the entity type feature distinguishes between individuals and businesses in the transaction network.

These features enable the GNN model to learn the graph representation of transactions and to detect network-based money laundering events more effectively. This approach complements the features employed in the other supervised and unsupervised models, which use summarised information from transaction data.

9. Annex C: Privacy-enhancing technologies

9.1 Overview of PETs

9.1.1 Homomorphic encryption

Homomorphic encryption (HE) allows computations to be performed on ciphertexts which are data transformed by data encryption algorithm generating an encrypted result that, when decrypted, corresponds to the result of the operation performed on real data. This technique has the potential to provide strong privacy guarantees while still enabling useful analysis of encrypted data.[84] To facilitate financial institutions' collaboration, homomorphic encryption could play a role in preserving the privacy of sensitive information, such as personally identifiable information (PII).

Homomorphic encryption can be particularly useful for analysing transactional networks, as it preserves the network-based measures from the real data. This allows an AML monitoring system to verify transactions and identify potential typologies, such as muling and smurfing, without compromising the privacy of the individuals or corporations involved. Additionally, encrypted know-your-customer (KYC) and risk data can be utilised to improve detection accuracy.

However, the main challenge of applying HE is the computational overhead associated with performing operations on encrypted data, which can lead to increased processing times and resource requirements.[85]

9.1.2 Local differential privacy

Local differential privacy (LDP) is a mathematical framework that ensures the privacy of individual records in a data set while allowing for accurate analysis. It works by introducing a carefully calibrated amount of noise into the data queries, making attribution to a specific individual or corporation more difficult.[86] A mechanism, such as the Laplace mechanism, is applied to the data locally (ie at the financial institution) to achieve local differential privacy. The Laplace mechanism, for example, adds Laplace-distributed noise to the true result of a query, with the noise scale determined by the query's sensitivity and a privacy parameter. The privacy parameter controls the trade-off between the privacy level and the query results' accuracy. Local differential privacy can provide strong privacy guarantees while enabling meaningful statistical analysis.

For money laundering detection, local differential privacy can be applied to non-PII numerical data to protect the privacy of individuals and sensitive corporate information without compromising the detection of suspicious activities. However, the process of adding noise to the raw data could lead to reduced model

---

84 See Gentry (2009) and Cheon et al (2017).

85 See Aono et al (2016).

8686 See Dwork et al (2006).

performance, especially if the participating members do not follow the same practice rules.

9.1.3 Federated learning

Federated learning (FL) is a decentralised machine learning framework enabling multiple entities to train a shared model collaboratively without exchanging raw data.[87,88] This technique allows for advanced privacy protection, as sensitive information remains within each participating entity's boundaries while benefiting from the knowledge and shared insights. Federated learning can be used to train a shared machine learning model that detects money laundering activities without disclosing the data or adding noise to the data. Instead, only the model updates, such as model parameter changes, are shared among the participants.

Federated learning could enable the shared model to learn from a broader collection of data sets without the actual pooling of data from different institutions. This approach can lead to more robust and accurate models, as the learnings generated from them can be generalised to unexpected data features while maintaining privacy. Moreover, federated learning enables the use of a centralised model, which facilitates the comparability of risk scores and model results.

However, federated learning also faces challenges, such as increased communication overhead and the need for a compatible data structure to ensure the proper sharing of model updates. Furthermore, the shared model may have limited access to cross-institutional transaction networks, making it more challenging to detect certain types of money laundering activities.

9.1.4 Other privacy-enhancing technologies

Methods such as secure multi-party computation (SMPC), zero-knowledge proofs (ZKPs) and private set intersection (PSI) could also provide valuable privacy-preserving techniques.

SMPC enables multiple parties to jointly compute a function over their inputs while keeping those inputs private. This allows financial institutions to collaborate on data analysis without revealing sensitive information to each other, offering an additional layer of privacy protection. SMPC could address some of the limitations encountered in both centralised and decentralised approaches while maintaining robust detection capabilities for cross-institutional graph typologies.

ZKPs allow one party to prove the validity of a statement without revealing any additional information and PSI enables parties to find the common elements in their data sets without sharing the actual data.

---

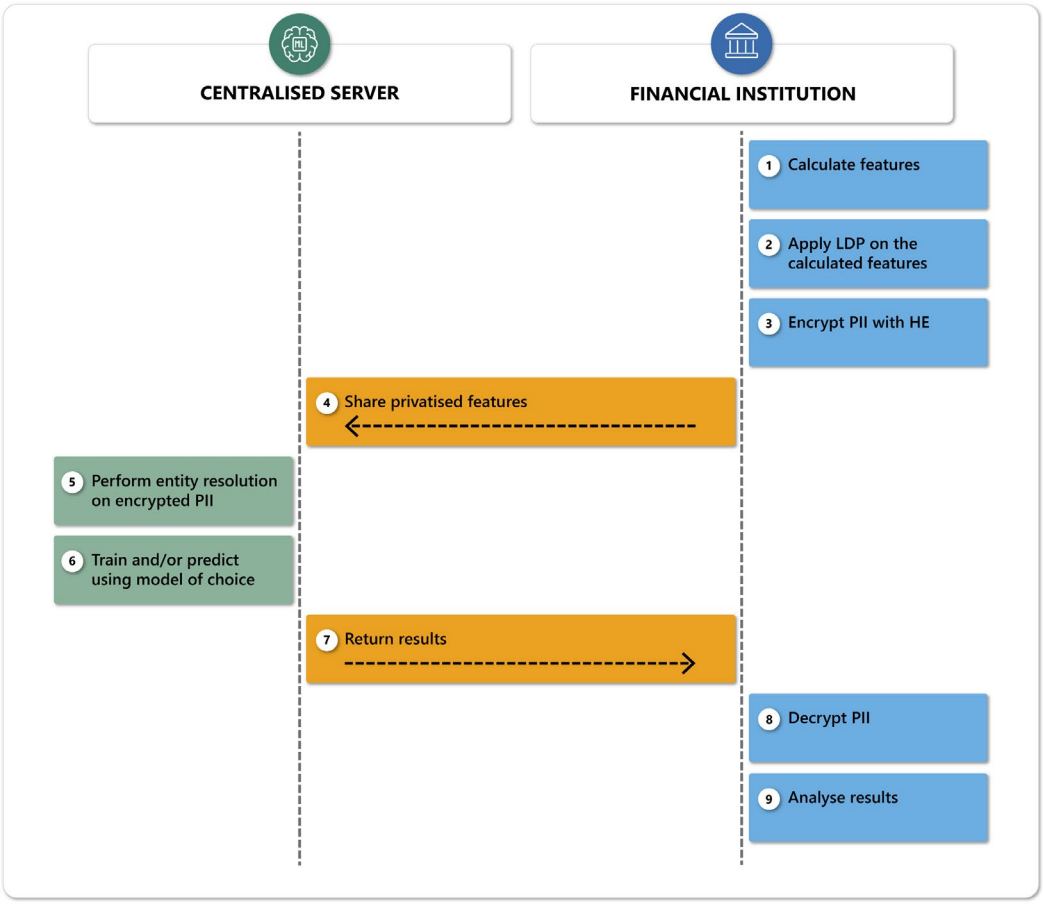
87 See Beutel et al (2022).

88 See McMahan et al (2017).

9.2 Application of PETs

Applying HE and local differential privacy

Graph C1: Implementation of HE and LDP in a centralised CAL arrangement



In the method for this approach, HE and LDP are employed. The process involves the following steps, as depicted in the figure above:

1. Each financial institution calculates a pre-defined set of actor-level features (characteristics for individuals and businesses). These features will help with understanding their behaviours and patterns.

2. To protect privacy, each institution adds some random noise to the calculated features by using LDP. This makes it difficult to identify specific individuals or businesses, while still preserving the overall pattern.

3. All financial institutions encrypt their data into a special identifier by using HE (possibly other PETs) that allows for comparisons without revealing personal information.

4. The encrypted features and identifiers are shared with a central server, which acts as a secure hub for analysis.

5. The central server compares the encrypted identifiers and determines if they belong to the same person or organisation. It does this by checking if the differences between the identifiers are below a certain threshold. If they are, it means they likely represent the same entity.

6. Using this combined information, a machine learning model makes predictions or analyses based on the data. The central server can also use pre-defined rules or involve a team of experts who have access to the protected data.

7. The predictions or analyses are then sent back to the respective financial institutions, along with any additional information that helps them to better understand the results.

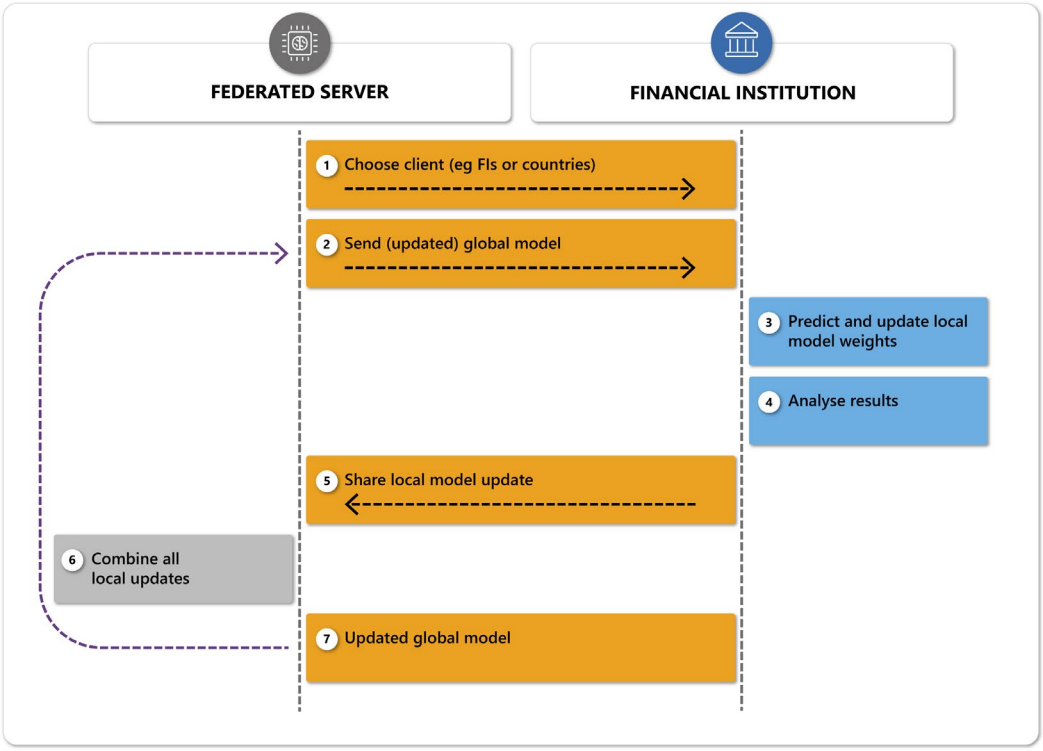
8. Each institution decrypts the identifier and maps it back to their own internal identification system.

9. Each institution can then analyse the results based on their own data.

This process can be repeated for new batches of data, allowing continuous analysis while still protecting privacy.

Decentralised approach using federated learning

Graph C2: Federated learning setup



The federated learning procedure in a decentralised setting:

1. The process starts with a federated server that chooses a group of clients (eg FIs or countries) from a larger network.

2. These selected clients receive a global model from the federated server.

3. Each client then updates the model's weights using their own local data, meaning that they each contribute their own knowledge to improve the model.

4. Each client can then analyse the results based on their own findings.

5. After updating the weights, the clients send their updates back to the federated server.

6. The federated server combines the updates from all the clients to improve the global model. It takes into account the contribution from each client to make the model more accurate and effective.

7. Steps 1–6 are repeated in an iterative manner. This means that the process keeps happening over and over again for continuous collaborative learning and improvement.

The application of federated learning in the decentralised setting enables financial institutions and countries to collaborate on detecting money laundering activities while preserving the privacy of personally identifiable information. By sharing model updates rather than raw data, institutions and countries can benefit from the collective knowledge and insights of the entire network, resulting in more robust and effective AML systems.

10. Annex D: Questions to support real-world pilots

Project Aurora demonstrates the potential benefits of different approaches to analysing payments data when combined with privacy-enhancing technologies and advanced analytics. However, real-world pilots would be necessary to derive real-world data on the impact and to address any practical issues. Such data and performance information are essential to support policy discussions.

Constant innovation is important as there is no clear "ideal" model for how collaborative analytics and learning (CAL) initiatives should be arranged. More practitioner and policy dialogue and consideration would be beneficial in the design of models to make use of connected payments or transaction data. As highlighted by the FATF, public authorities have a key role to support such initiatives, either as partners/stakeholders in a CAL platform or as sources of feedback on the design or output of CAL initiatives.

Some questions to consider when designing models that may be answered by a real-world pilot are listed below.

10.1 Objectives, performance monitoring and scope

- What is the ultimate objective of a CAL-based approach?
- What are the success criteria?
- What performance metrics are required to measure the impact of a CAL-based approach?
- How will these metrics be collected, stored and reported?
- What is the scope of financial crime activities that could lawfully be analysed in any pilot?
- Is the scope national, international or both national and international combined?
- What should the role of data protection authorities, regulators, central banks, law enforcement agencies and financial intelligence agencies be in such a pilot? How should they interact with each other?

10.2 Data and analysis

The type and volume of data required for financial crime analysis will have a key bearing on legal, security, data protection and technical issues relevant for a real-world pilot. It will be important to consider the type of data that will be shared, how they will be shared and with whom.

- What sources of payments data will be used or shared?
- What are the daily transaction volumes from each source?
    - How much would be required for a pilot?
- What data standards are used?
- What data formats are used?
- Is the usage of synthetic data generated on the basis of real-world transactions feasible for a pilot?
- How much historical data would need to be used and made available for analysis?
    - Would these data include previously identified money launderers?
- What are the minimum required data fields for a transaction?
- What features of the data will be measured?
- What other data could be required and for what purposes?
    - How will these other data be provided?
    - What is the volume, frequency and format?
- Who is responsible for ensuring that the information is accurate, timely, reliable and proportionate?
- Are centralised, decentralised or hybrid CAL approaches being explored and tested?
    - Which approach is a preferred starting point?
    - What type and level of views or insights on the data will be shared and with whom?
    - Would certain public authorities receive additional information (eg a full overview of a confirmed suspicious network)?
- How does the current legal framework for processing data support this pilot?
    - Who are the data controllers and the data processors?
- How will privacy-enhancing technologies be applied?
- How will any record keeping operate to explain actions taken in respect of privacy-enhanced data?
- How will data quality issues, particularly with privacy-preserved data, be observed?
    - How would data be corrected?
- How will any findings and detections from a CAL be shared with the participating parties or with agencies that are not members/participants of the CAL?

10.3 Post-pilot questions

- What data quality issues were observed with data subject to privacy preservation?
    - How can these be addressed?
- Who would be the owner(s) of a CAL capability?
    - What would the funding model need to be for such a capability?
    - Is it for profit, a cooperative non-profit utility or a national public asset?
- Who are the members and beneficiaries of the analysis?
- Who is the capability accountable to and on what basis?
- Who is accountable for the management of the operational risks and service levels of the platform?
- What duties, incentives and liabilities will the different stakeholders in a CAL capability have with regard to the appropriate use of the information for the intended purpose?
- Who is accountable for any damage caused by the use of information held by the platform?
- Who would manage liability issues relating to regulatory or legal risk?
- How will the future direction and investment roadmap for the platform be managed?
- What will the criteria be for participation and/or exits from the platform?
- What is the role of centralised or decentralised analysis at the payments level to identify risk compared with the responsibility of individual regulated entities to identify risk?
- Will the payments-level analysis include manual investigations or only automated alerts?
- What is the balance of security, privacy and utility issues relating to the use of analysis of privacy-enhanced data?
- What are the audit and inspection limitations on the analysis of privacy-enhanced data?
- What governance and security controls would be required to mitigate physical, cyber and information security risks?
    - What obligations would be placed on participants in such an initiative?
    - What validation, audit and assurance actions would be required?

11. Annex E: Additional acronyms and definitions

BIC	Bank identifier code.
CPU	Central processing unit.
Data pooling	A process in which data sets from different sources are combined and pooled in a centralised repository.
ECID	Encrypted comparable identifier can be a code or other means of identification to allow individuals to be tracked across data and systems without revealing their identity.
EFIPPP	European Financial Intelligence Public Private Partnership.
F1 score	The F1 score is used in statistical analysis to measure the accuracy and overall performance of a model. It is represented by a combination of precision and recall. These are two elements used for statistical testing.
False negatives	A metric used for statistical analysis that shows the number of illicit transactions that have not been flagged as suspicious by the model.
False positives	A metric used for statistical analysis that shows the number of persons or transactions that have been falsely classified as positive.
Payments data	Payments or transaction data refers to any personal or financial information that can be collected from credit cards, debit cards or other payment methods, including, but not limited to, a personal account.
PEP	Politically exposed persons.
PoC	Proof of concept.
Precision	True positive/(true positive + false positive). It measures how accurate the model is at identifying money laundering activities.
PSI	Private set intersection is a secure multi-party computation cryptographic technique that allows two parties to compare encrypted versions and to compute the intersection of those sets without revealing anything except the intersection.
PSP	Payment service providers.
RDF	A resource description framework is a framework for representing interconnected data in a structured and standardised way.
SMPC	Secure multi-party computation is a cryptographic protocol that distributes computational power across multiple parties, and in which no party can see the other parties' data.
Wallet	An app or online service used to make payments electronically or to store electronic representations of tickets, documents or other credentials.

12 References

12. References

Aono, Y, T Hayashi, L Trieu and S Yamada (2016): "Scalable and secure logistic regression via homomorphic encryption", in *Proceedings of the sixth ACM conference on data and application security and privacy*, Association for Computing Machinery, pp 142–44.

Bank for International Settlements (BIS) (2019): "FSI Insights: Suptech applications for anti-money laundering", August, "https://www.bis.org/fsi/publ/insights18.pdf".

——— (2023): "Committee on Payments and Market Infrastructures (CPMI) consultative report: ISO 20022 harmonisation requirements for enhancing cross-border payments", March, "https://www.bis.org/cpmi/publ/d215.pdf".

Beutel, D J, T Topal, A Mathur, X Qiu, J Fernandez-Marques, Y Gao, L Sani, K H Li, T Parcollet, P P B de Gusmão and N D Lane (2022): "Flower: A friendly federated learning framework".

Cheon, J H, A Kim, M Kim and Y Song (2017): "Homomorphic encryption for arithmetic of approximate numbers", in T Takagi and T Peyrin (eds), *Advances in cryptology–ASIACRYPT 2017: 23rd international conference on the theory and applications of cryptology and information security, Hong Kong, China, December 3-7, 2017, proceedings, part I*, Springer International Publishing, pp 409–37.

Danmarks Nationalbank (2022): "AI and machine learning in the financial sector: five focus points", April, "https://www.nationalbanken.dk/en/publications/Documents/2022/04/EM%202022%203.pdf".

Dwork, C, F McSherry, K Nissim and A Smith (2006): "Calibrating noise to sensitivity in private data analysis", in S Halevi and T Rabin (eds), *Theory of cryptography: third theory of cryptography conference, TCC 2006, New York, NY, USA, March 4-7, 2006, proceedings*, Springer, pp 265–84.

Financial Action Task Force (FATF) (2018): "Professional money laundering", July, "https://www.fatf-gafi.org/en/publications/Methodsandtrends/Professional-money-laundering.html".

——— (2020): "Virtual assets red flag indications of money laundering and terrorist financing", September, "https://www.fatf-gafi.org/en/publications/Methodsandtrends/Virtual-assets-red-flag-indicators.html".

——— (2021a): "Opportunities and challenges of new technologies for AML/CFT", July, www.fatf-gafi.org/en/publications/Digitaltransformation/Opportunities-challenges-new-technologies-for-aml-cft.html.

——— (2021b): "Stocktake on data pooling, collaborative analytics and data protection", July, www.fatf-gafi.org/en/publications/Digitaltransformation/Data-pooling-collaborative-analytics-data-protection.html.

——— (2021c): "Suggested actions to support the use of new technologies for AML/CFT".

——— (2022): "Conference on digital transformation".

——— (2023): Frequently asked questions – money laundering, https://www.fatf-gafi.org/en/pages/frequently-asked-questions.html#tabs-36503a8663-item-6ff811783c-tab.

Financial Stability Board (FSB) (2022): "G20 roadmap for enhancing cross-border payments", October, "https://www.fsb.org/wp-content/uploads/P101022-1.pdf".

Finanstilsynet (2021): "Project AML/TEK", May, "https://www.dfsa.dk/-/media/Nyhedscenter/2021/Consultation_Project_AML_TEK.pdf".

Gentry, C (2009): "A fully homomorphic encryption scheme", Stanford University, mimeo.

International Federation of Accountants (IFAC) (2022): "Anti-money laundering: The basics", March, "https://www.ifac.org/_flysystem/azure-private/publications/files/AML-Installment-8-Crime-Trends.pdf".

International Monetary Fund (IMF) (2021): "Factsheet: IMF and the fight against money laundering and the financing of terrorism", July, "https://iw-imf.org/en/About/Factsheets/Sheets/2016/08/01/16/31/Fight-Against-Money-Laundering-the-Financing-of-Terrorism.html".

Lexis Nexis (2021): "True cost of financial crime global summary".

——— (2022) : "True cost of financial crime global summary", "https://risk.lexisnexis.com/global/-/media/files/financial%20services/infographics/lnrs-global-tcoc-infographic2022_v2-nxr15749-00-1122-en-us.pdf".

McMahan B, E Moore, D Ramage, S Hampson, and B A Arcas (2017): "Communication-efficient learning of deep networks from decentralized data", in *Proceedings of machine learning research: Volume 54 Artificial intelligence and statistics 20 – 22 April 17, Fort Launderdale, Florida, USA,* ,pp. 1273-1282.

Open Ownership: "Beneficial Ownership Data Standard", "https://www.openownership.org/en/topics/beneficial-ownership-data-standard/".

Oracle (2019): "Disrupting status quo in AML compliance", March, "https://www.oracle.com/a/ocom/docs/industries/financial-services/fs-disrupting-status-quo-aml-complaince-wp.pdf".

Statista (2023): "Global gross domestic product (GDP) at current prices from 1985 to 2028", May," https://www.statista.com/statistics/268750/global-gross-domestic-product-gdp/".

United Nations Office on Drugs and Crime (UNODC): "Money laundering", "https://www.unodc.org/unodc/en/money-laundering/overview.html".

The White House (2021): "US and UK to partner on prize challenges to advance privacy enhancing technologies", December, "https://www.whitehouse.gov/ostp/news-updates/2021/12/08/us-and-uk-to-partner-on-a-prize-challenges-to-advance-privacy-enhancing-technologies".

The Wolfsberg Group (2017): "Payment transparency standards", October, "https://wb-db.basel.institute/assets/116e9ff8-e4af-4481-bb34-3421a93b22e0/1.%20Wolfsberg-Payment-Transparency-Standards-October-2017.pdf".

13 Acknowledgements

13. Acknowledgements

Bank for International Settlements

Beju Shah (Head of Nordic Centre, BIS Innovation Hub)

Hachem Hassan (Adviser)

Xin Zhang (Adviser)

William Zhang (Adviser)

Sidney Lampart (Adviser)

Björn Segendorff (Adviser)

Caroline Leung (Adviser)

Grimur Sigurdarson (Adviser)

Susanne Bohman (Adviser)

Ben Dovey (Adviser)

Karen Martin (Executive Assistant)

Lucinity

Óli Páll Geirsson

Guðmundur Kristjánsson

Jón Kristinn Þórðarson

Brynjólfur Gauti Guðrúnar Jónsson

Kristín Björg Bergþórsdóttir

Daníel Pálmason

Francisco Mainez

Consultants

Nick Maxwell (Head of Future of Financial Intelligence Sharing Programme)

Special acknowledgements

The BIS Innovation Hub Nordic Centre team would like to thank the following people for their support, input or feedback:

Cecilia Skingsley (Head of BIS Innovation Hub)

Ross Leckow (Head of Strategy and Operations, and Deputy Head of the BIS Innovation Hub)

Miguel Diaz (Head of Toronto Centre, BIS Innovation Hub)

Cristina Picillo (Adviser, BIS Innovation Hub)

Esther Rey Losada (Senior Operations Manager, BIS innovation Hub)

Lucy Wong (Adviser, BIS Innovation Hub)

Andreas Adriano (Communications Advisor, BIS Innovation Hub)

Vivienne Artz OBE FSCI (Hon)

Pavle Avramović (Financial Conduct Authority, UK)

Jo Ann Barefoot (Alliance for Innovative Regulation)

Theresa Bercich (Lucinity)

Ingvar Bjarki Einarsson (Lucinity)

Michael Dewar (Mastercard)

Helga Rut Eysteinsdóttir (Seðlabanki Íslands)

Baptiste Forestier (Hero)

Gabriela Guiborg (Sveriges Riksbank)

Leo Gosland (Financial Conduct Authority, UK)

Kristoffer Hansson (Finansinspektionen)

Maciej Janas (Klarna)

Thais Lærkholm Jensen (Danmarks Nationalbank)

Friðrik Laxdal Kárason (Lucinity)

Magnus Karlsson (Finanspolisen)

Matt Lowe (Financial Conduct Authority, UK)

Klas Malmen (Finansinspektionen)

Justo Manrique (Klarna)

Valerie Marshall (Financial Conduct Authority, UK)

Lasse Meholm (Norges Bank)

Viktor Möllborg (Sveriges Riksbank)

Henrike Muelle (Financial Conduct Authority, UK)

Dilan Ölcer (Sveriges Riksbank)

Mathias Lien Oskarsson (Finansinspektionen)

Helena Salomonsson (Klarna)

Che Sidanius (Global coalition to fight financial crime)

Hjörtur Stefánsson (Lucinity)

Luisa Stock (Klarna)

D. Edward Wilson Jr.

Daniël Worm (TNO)

Bank for International Settlements (BIS)

ISBN 978-92-9259-657-6