IFC-Bank of Italy Workshop on "Data science in central banking: enhancing the access to and sharing of data"

17-19 October 2023

Project Aurora: the power of data, technology and collaboration to combat money laundering across institutions and borders[1]

Beju Shah,
Bank for International Settlements

[1] This contribution was prepared for the workshop. The views expressed are those of the authors and do not necessarily reflect the views of the Bank of Italy, the BIS, the IFC or the other central banks and institutions represented at the event.

BIS Innovation Hub

Project Aurora

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

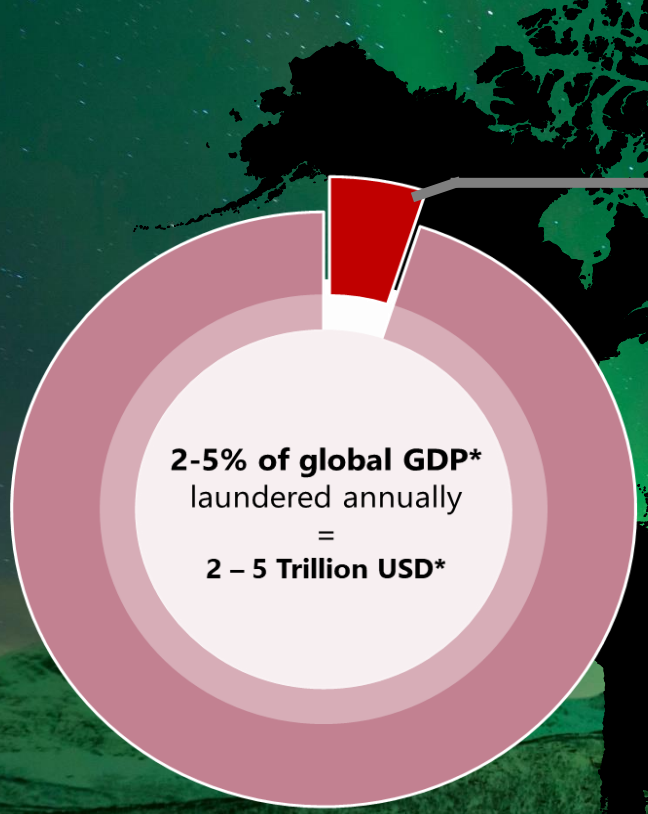
Irving Fisher Committee Data Science Workshop – Rome – October 2023

Beju Shah, Head of BIS Innovation Hub Nordic Centre

BIS Innovation Hub

Aurora

Money laundering is a global and costly problem



Increasing costs of financial crime compliance

2022:
$274 billion

28% increase
from 2020

80% increase
From 2019

Source: 1) UN Office of Drugs & Crime - https://www.unodc.org/unodc/en/money-laundering/overview.html and 2) https://www.unodc.org/documents/data-and-analysis/Studies/Illicit_financial_flows_2011_web.pdf

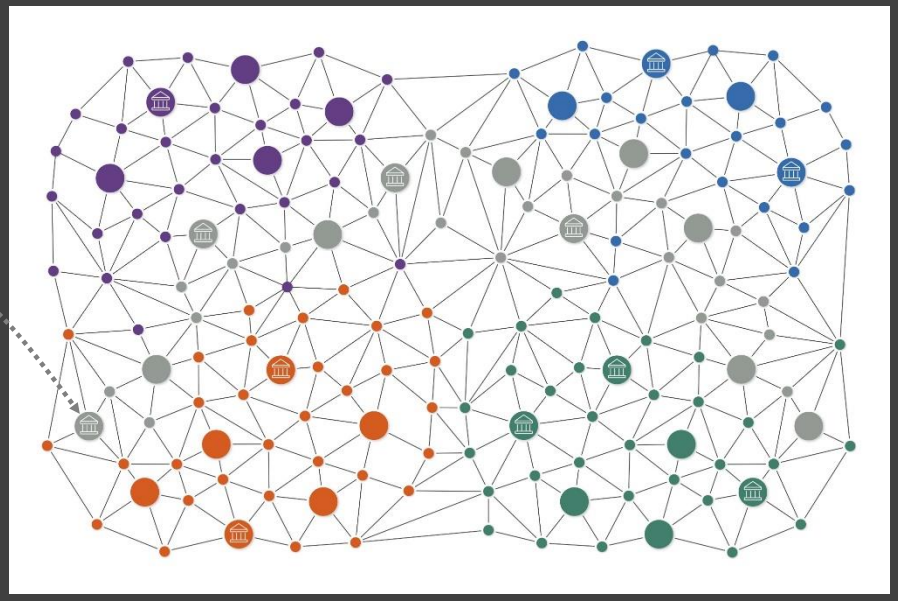
* Estimates are calculated relative to recent GDP

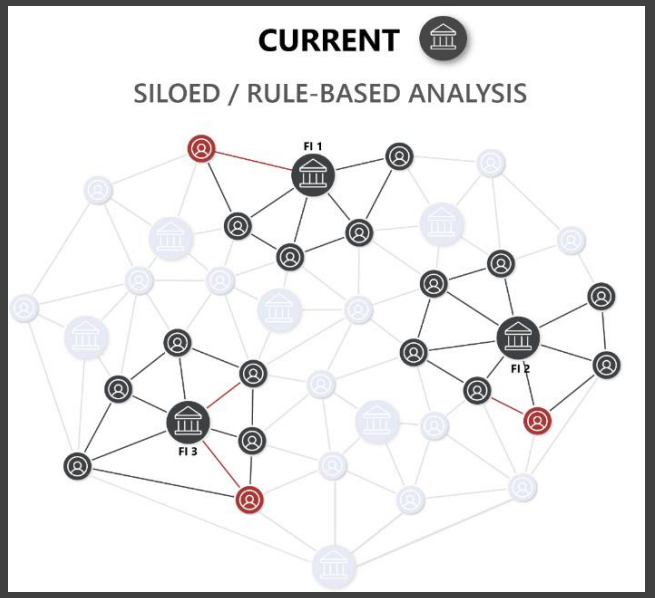
** True Cost Financial Crime Compliance 2022 | LexisNexis Risk Solutions. **Note: an increase of 28% from 2020 to 2022 in compliance costs, 23%-80% increase between 2019 and 2022.**

Challenges with today's approach(es)

- 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 the network**

money laundering pattern



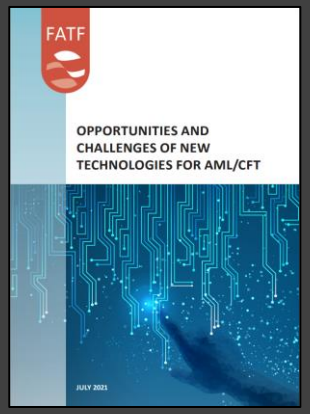


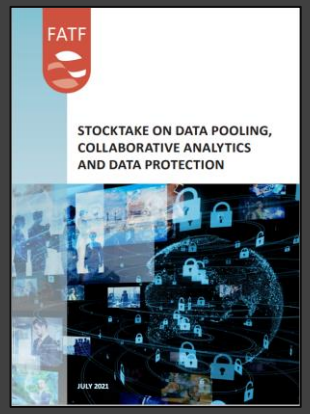
- **Non-standardised and / or readily and easily consumable data**, making collaboration and analysis across institutions (and borders) **more difficult**
- Rule base approaches are not suitable for analysing behavioural pattern
- Data required (based on the typology) for effective **AML/CFT measures have to be balanced with the need to protect individuals' privacy, personal data and human rights**.

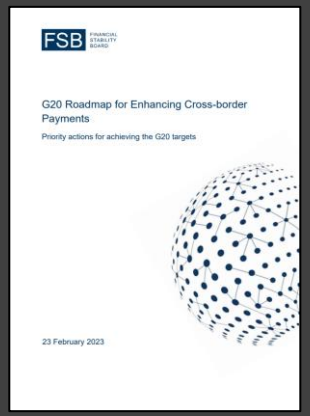
BIS Innovation Hub

Aurora

Leveraging data and technology





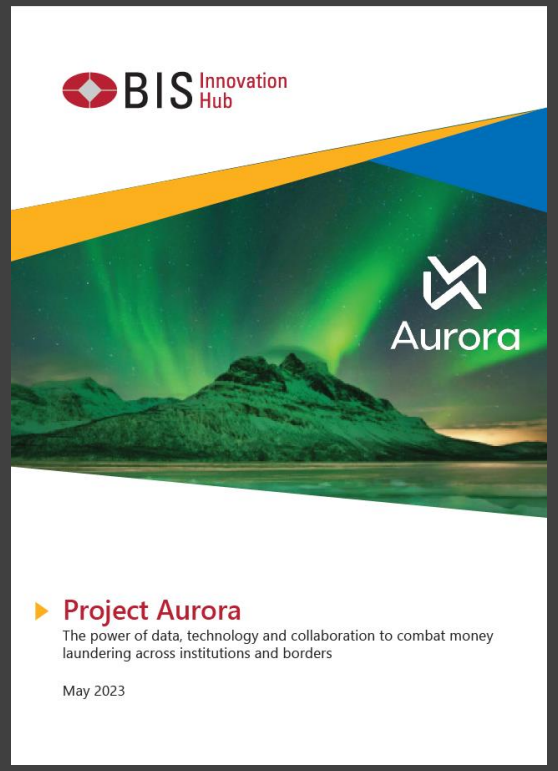


- FATF stocktake highlights several approaches / technologies to improve AML efforts:
  - **Data sharing / data pooling**
  - **Network analytics**
  - **Machine learning / Federated learning**
  - **Privacy enhancing technologies (PETs)**
- **Payments data could be utilised more.** Analysis of this data where **privacy is protected, sensitive information is encrypted** and, in some cases, where such **data need not leave a jurisdiction or institution** could be **game changing**.
- Cross border payments modernisation, ISO20022, legal entity identifier (LEI), FATF Recommendation 16, public private partnerships, beneficial ownership standards and further data standardisation initiatives could help.
- Network nature of payments data is well suited to graph data structures.
- When combined with the latest technologies could unlock the power of this data in AML efforts.

BIS Innovation Hub

Aurora

Project overview



PROBLEMS

Money laundering is a **cross institutional, payment system and borders problem.** It is a complex data challenge

Siloed and rule-based monitoring is **ineffective when a network view of behavioural patterns is required**

Compliance costs > **$274 billion***

< 1% laundered money seized (**$20-$50 billion****)

Holistic network view needed but FIs **cannot share data**

OBJECTIVES

Show how to **improve the performance and effectiveness** of AML analysis while **protecting privacy and sensitive information**

Showcase the potential of AI and PETs with connected payments data to detect money laundering

Test and compare the performance of advanced technologies and different analysis approaches

Underscore **the power of public private partnerships, and collaboration** when enabled by technology

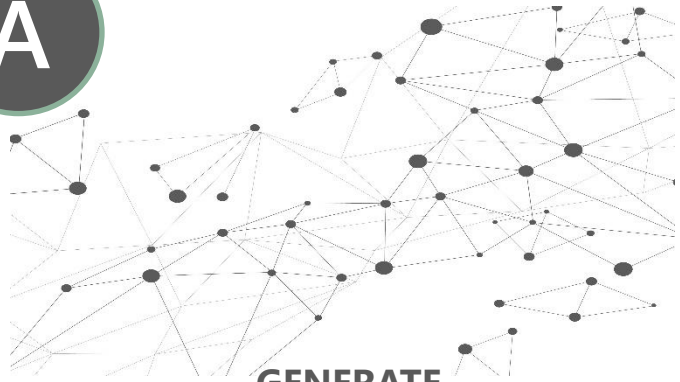
* True Cost Financial Crime Compliance 2022 | LexisNexis Risk Solutions. **Note: an increase of 28% from 2020 to 2022 in compliance costs, 23%-80% increase between 2019 and 2022.**

**UN Office of Drugs & Crime - https://www.unodc.org/unodc/en/money-laundering/overview.html & https://www.unodc.org/documents/data-and-analysis/Studies/Illicit_financial_flows_2011_web.pdf

BIS Innovation Hub

Aurora

A



GENERATE SYNTHETIC DATA

DATA SET

6 COUNTRIES

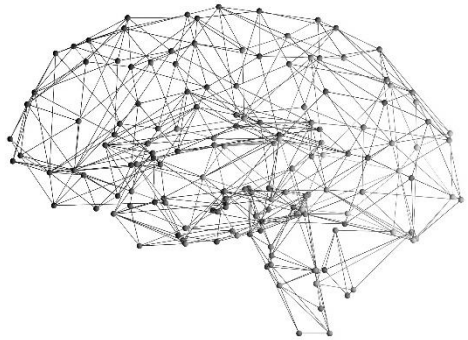
29 FINANCIAL INSTITUTIONS

>155,000 ACCOUNTS

MONEY LAUNDERING PATTERNS EMBEDDED INCLUDE:

- **COMPLEX LAYERING SCHEMES**
- **SMURFING**

B

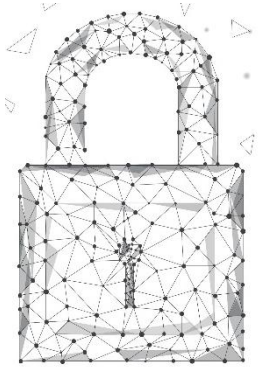


TESTING MACHINE LEARNING MODELS

MONITORING SCENARIOS

- **SILOED**
- **NATIONAL**
- **CROSS-BORDER**

C



TESTING PRIVACY-ENHANCING TECHNOLOGIES

COLLABORATIVE ANALYSIS & LEARNING

- **CENTRALISED**
- **CENTRALISED XB**
- **DECENTRALISED**
- **HYBRID**

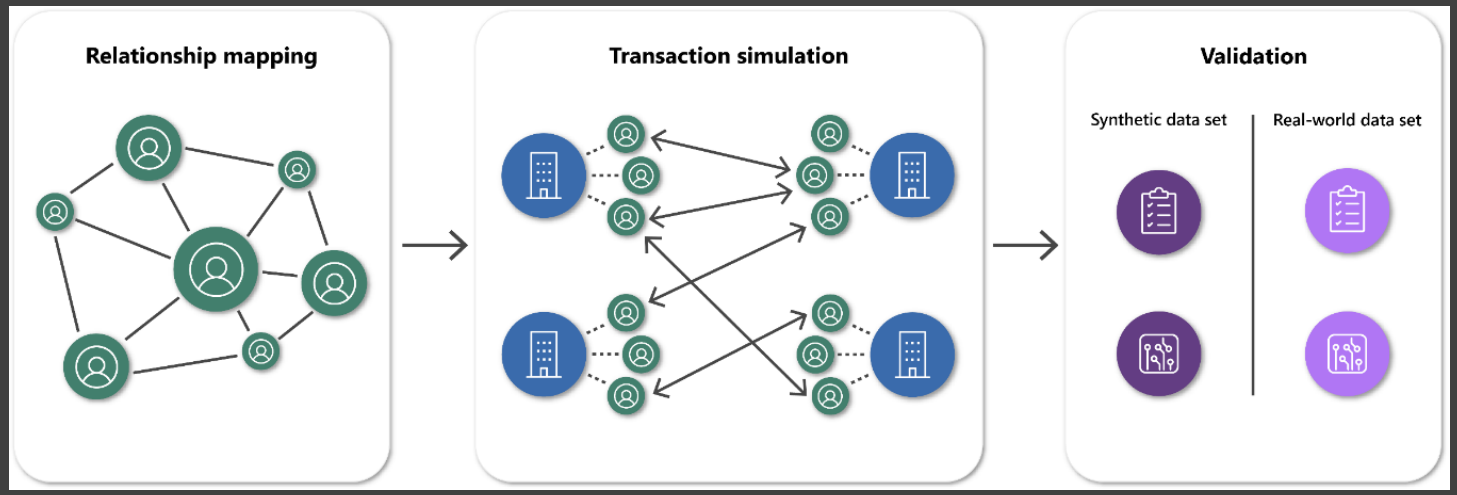
BIS Innovation Hub

Aurora

Part A – Generating the 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

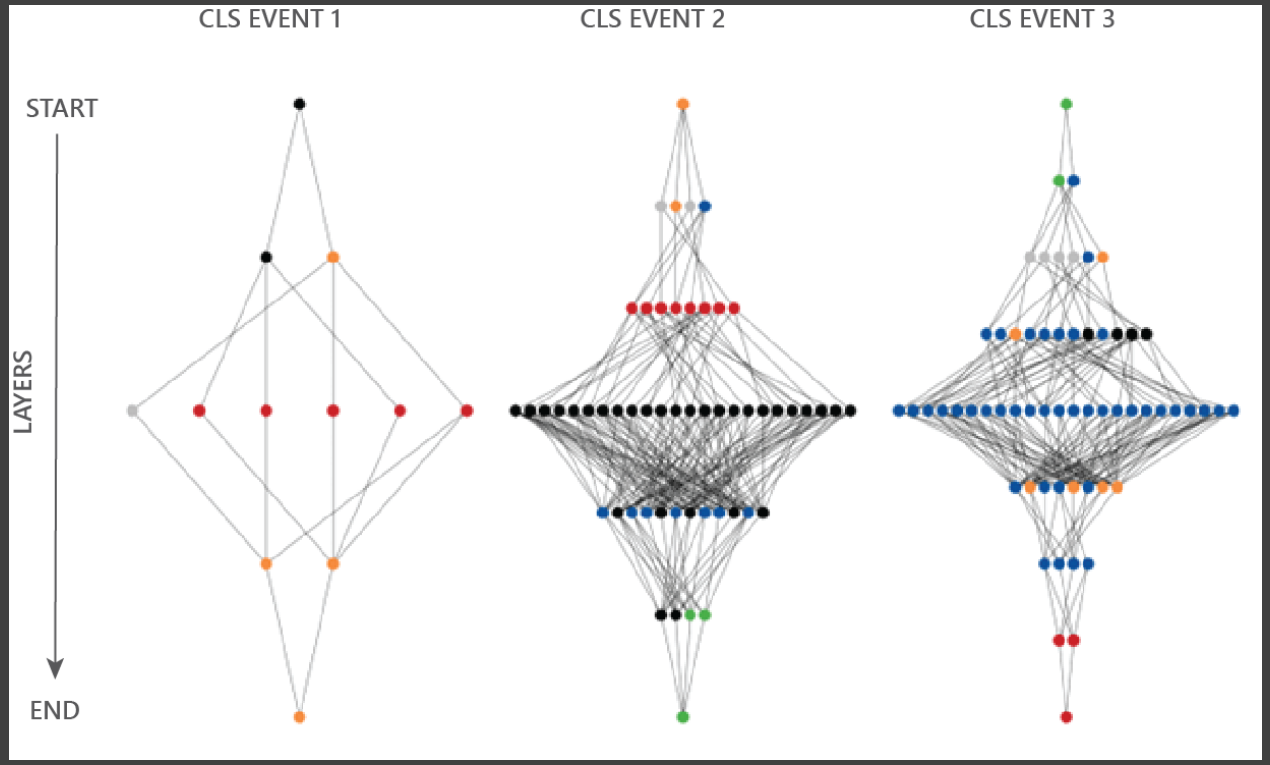


BIS Innovation Hub

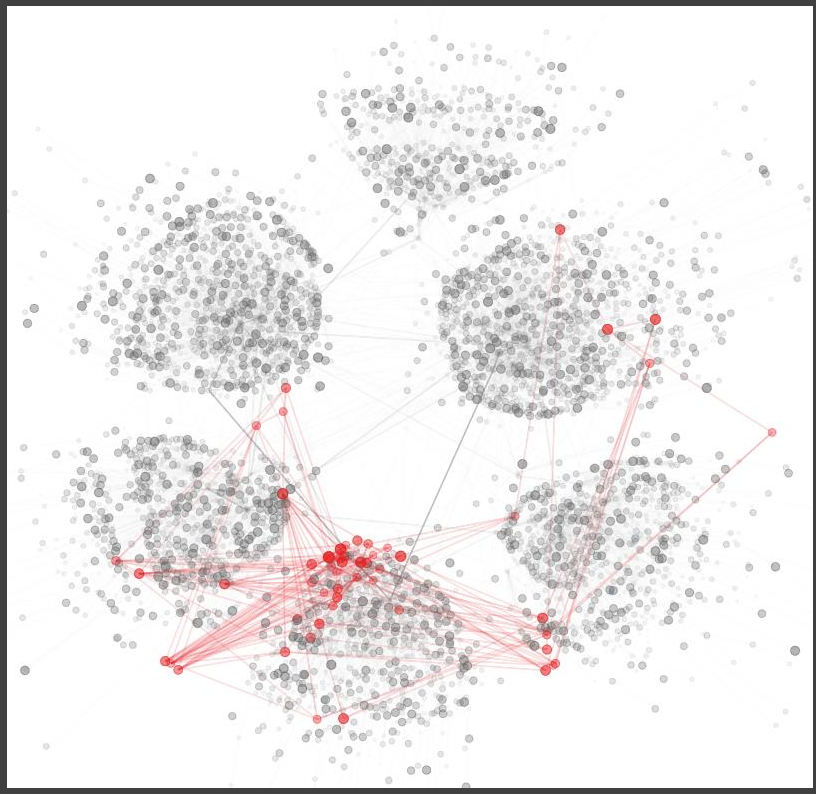
Aurora

Part A - Synthetic data

Examples of complex layering scheme (CLS) patterns in the synthetic data



Example of transactions over 1 day with money laundering activities highlighted in red



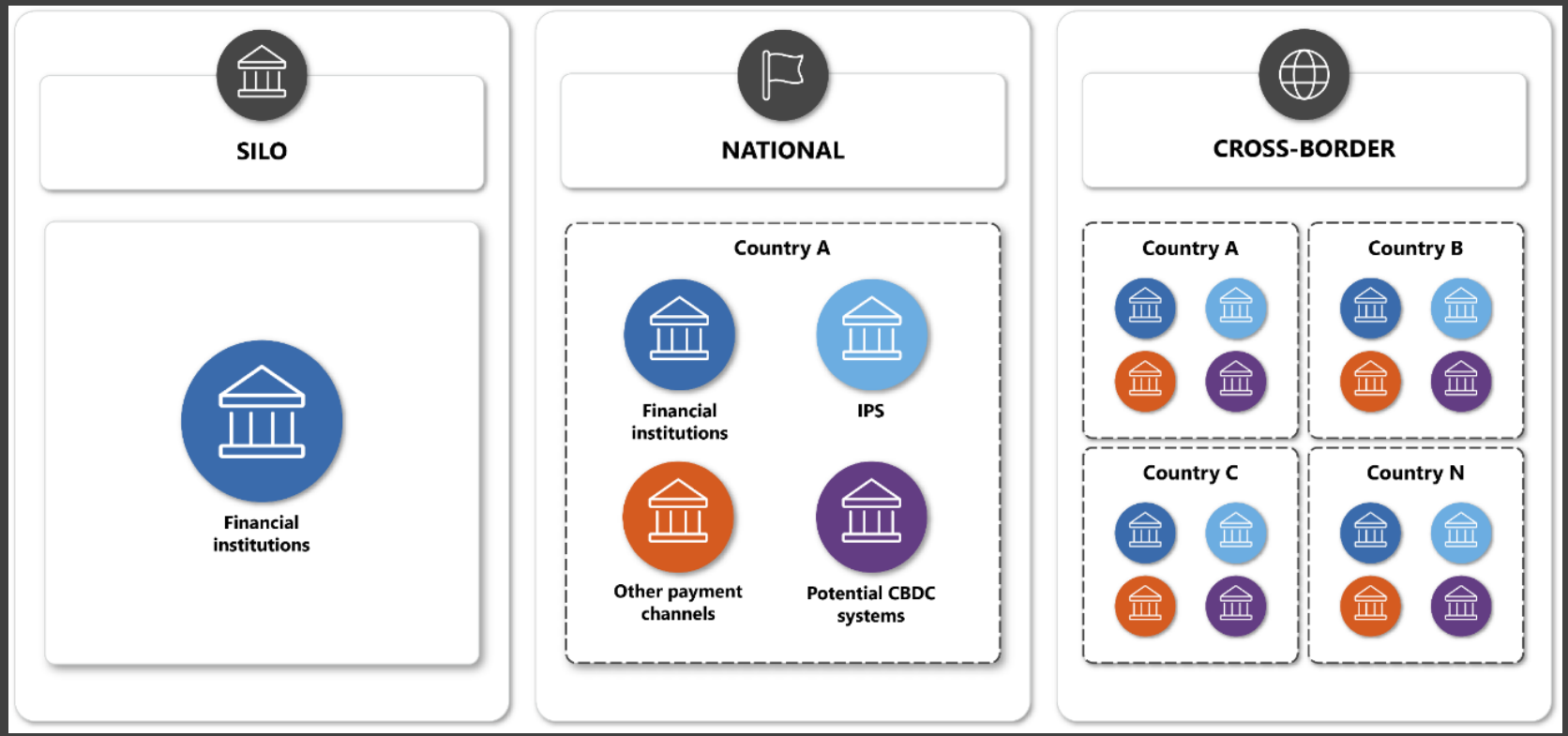
BIS Innovation Hub

Aurora

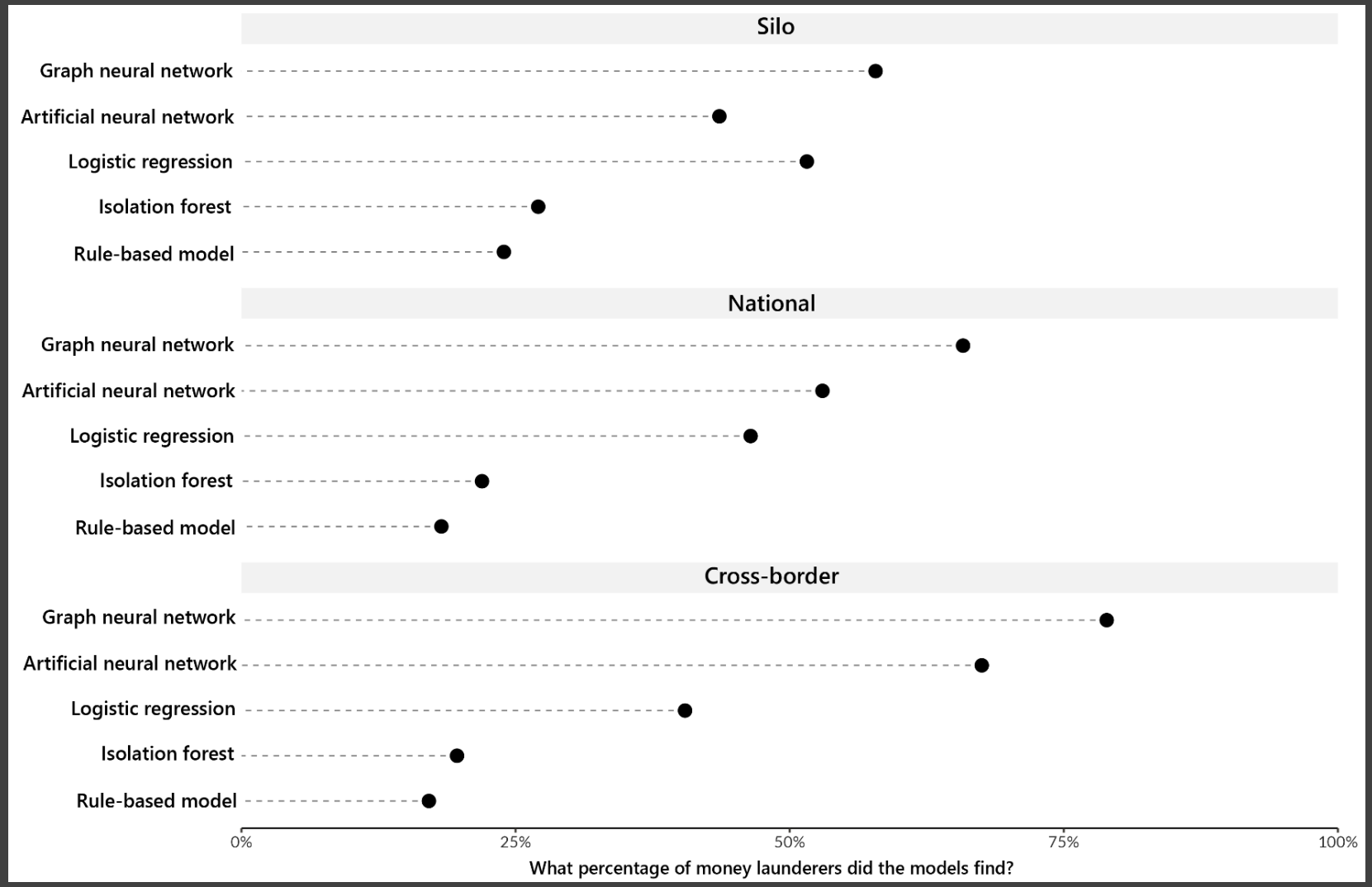
Part B – Applying machine learning models

Several machine learning models tested against different views of data for performance /effectiveness

- **Graph neural-networks**
- **Artificial neural networks**
- **Logistic regression**
- **Isolation forest**
- **Rule based as baseline**



Part B – Results

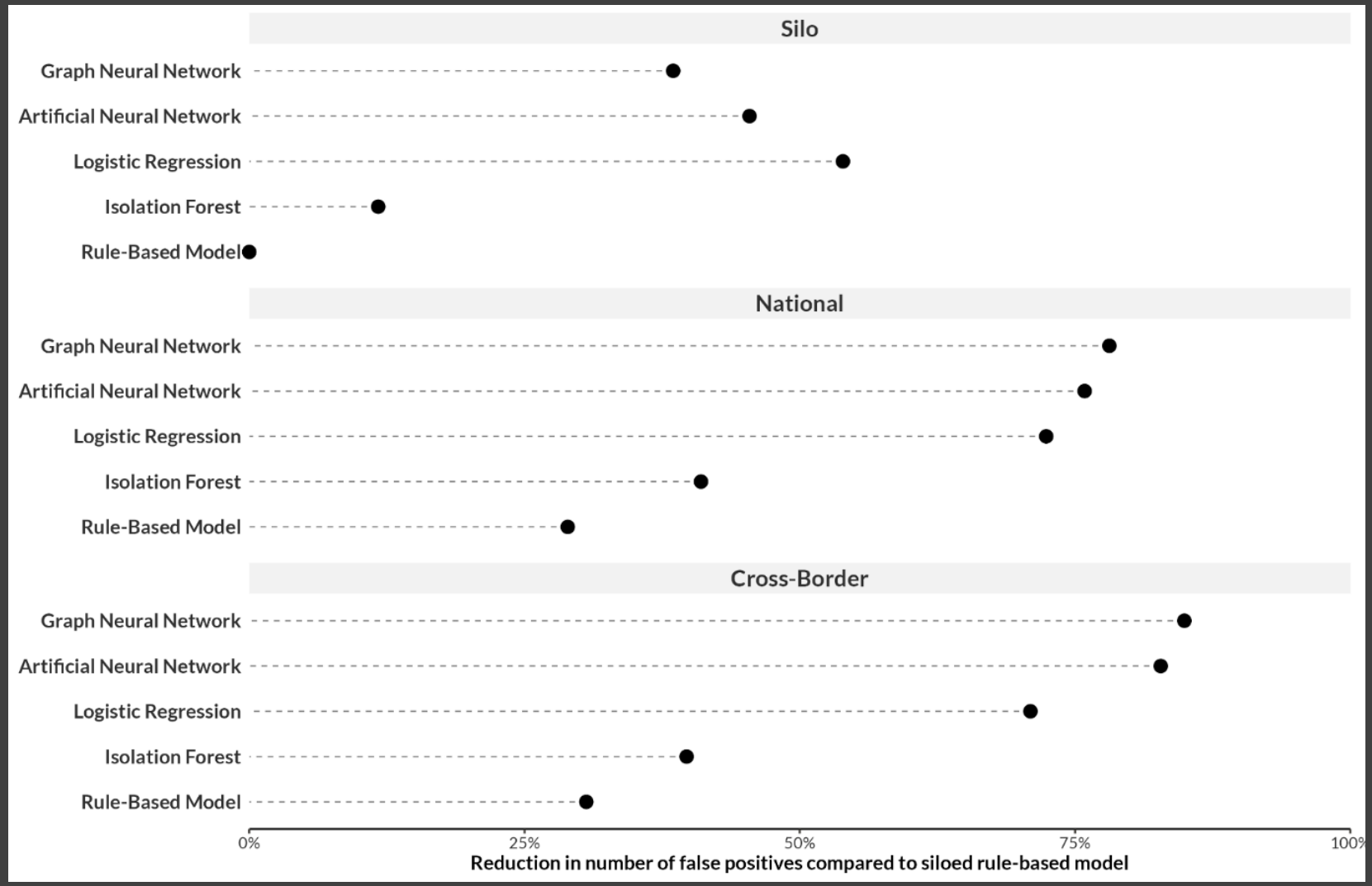


2 – 3x increase in detection of money launderers

BIS Innovation Hub

Aurora

Part B – Results



80% reduction in false positives

BIS Innovation Hub

Aurora

Part C – Applying PETs

Part B demonstrated the potential **benefits of using machine learning models on shared transaction data to detect complex money laundering**.

Implementing such arrangements could be challenging due to **concerns relating to data protection, privacy, security, competition and legal compliance**.

Part C explores the use of **privacy-enhancing technologies (PETs) to facilitate secure and privacy-preserving CAL (Collaborative Analysis and Learning) arrangements**, and assessed the effectiveness of machine learning models in combination with different PETs.

BIS Innovation Hub

Aurora

Part C – CAL arrangements explored

COLLABORATIVE ANALYSIS & LEARNING (CAL)

CAL 1 (centralised national)

- Data are encrypted and obfuscated using PETs
- Data are shared in a centralised national system.
- Machine learning models are trained and applied at a national-level data.

CAL 2 (centralised cross-border)

- Data are encrypted and obfuscated using PETs
- Data are shared in a centralised cross-border system.
- Machine learning models are trained and applied on data at a cross-border level.

CAL 3 (hybrid national and cross-border)

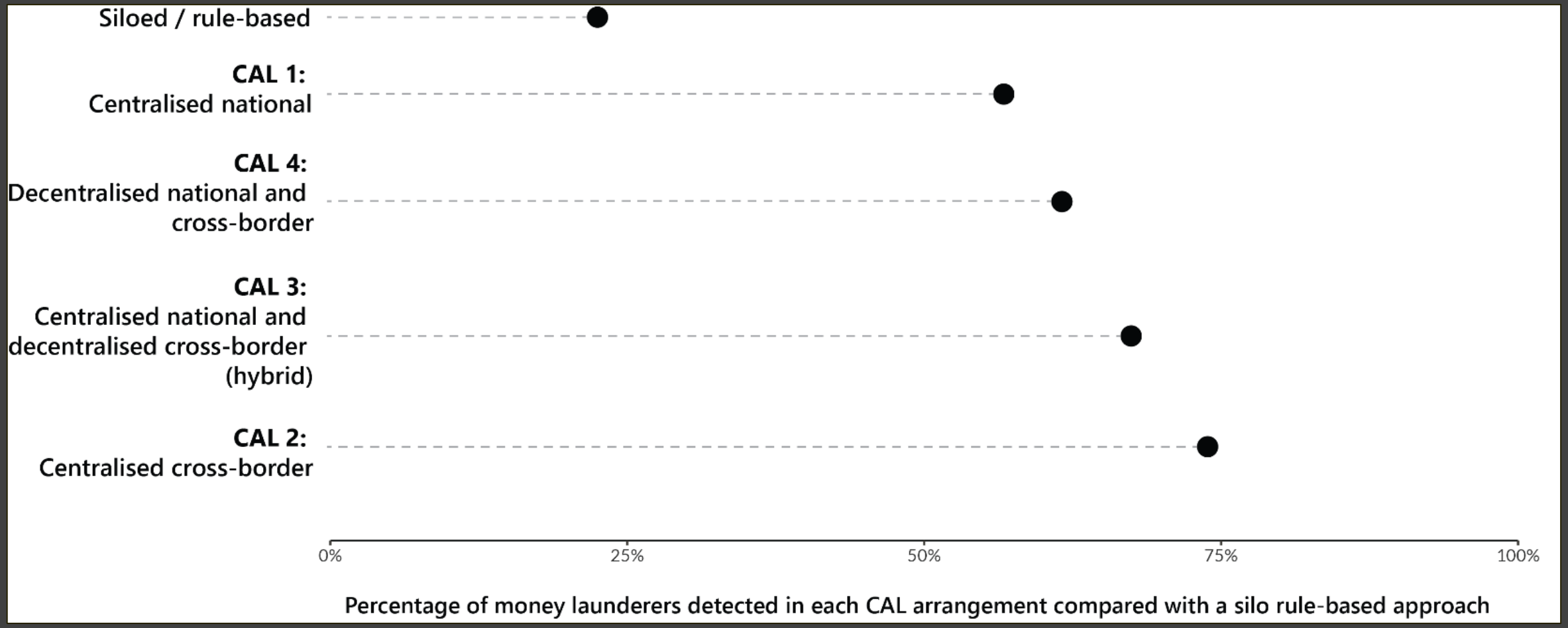
- Data are encrypted and obfuscated using PETs.
- Data are shared into a centralised national system.
- Countries collaborate on training a machine learning model using **federated learning** without the need to share transaction data across borders.

CAL 4 (decentralised national and cross-border)

- participants collaboratively train a machine learning model using **federated learning** locally on their own data
- share the model updates to a common global model

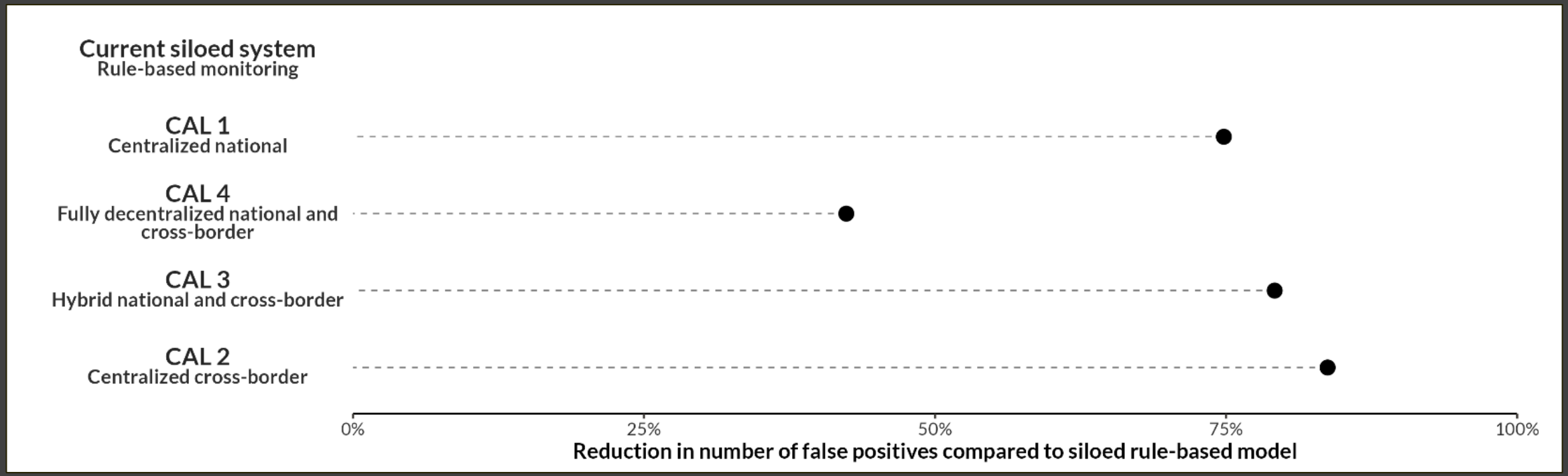
Part C – Results

When **machine learning is combined with PETs** that **performance and effectiveness is maintained**

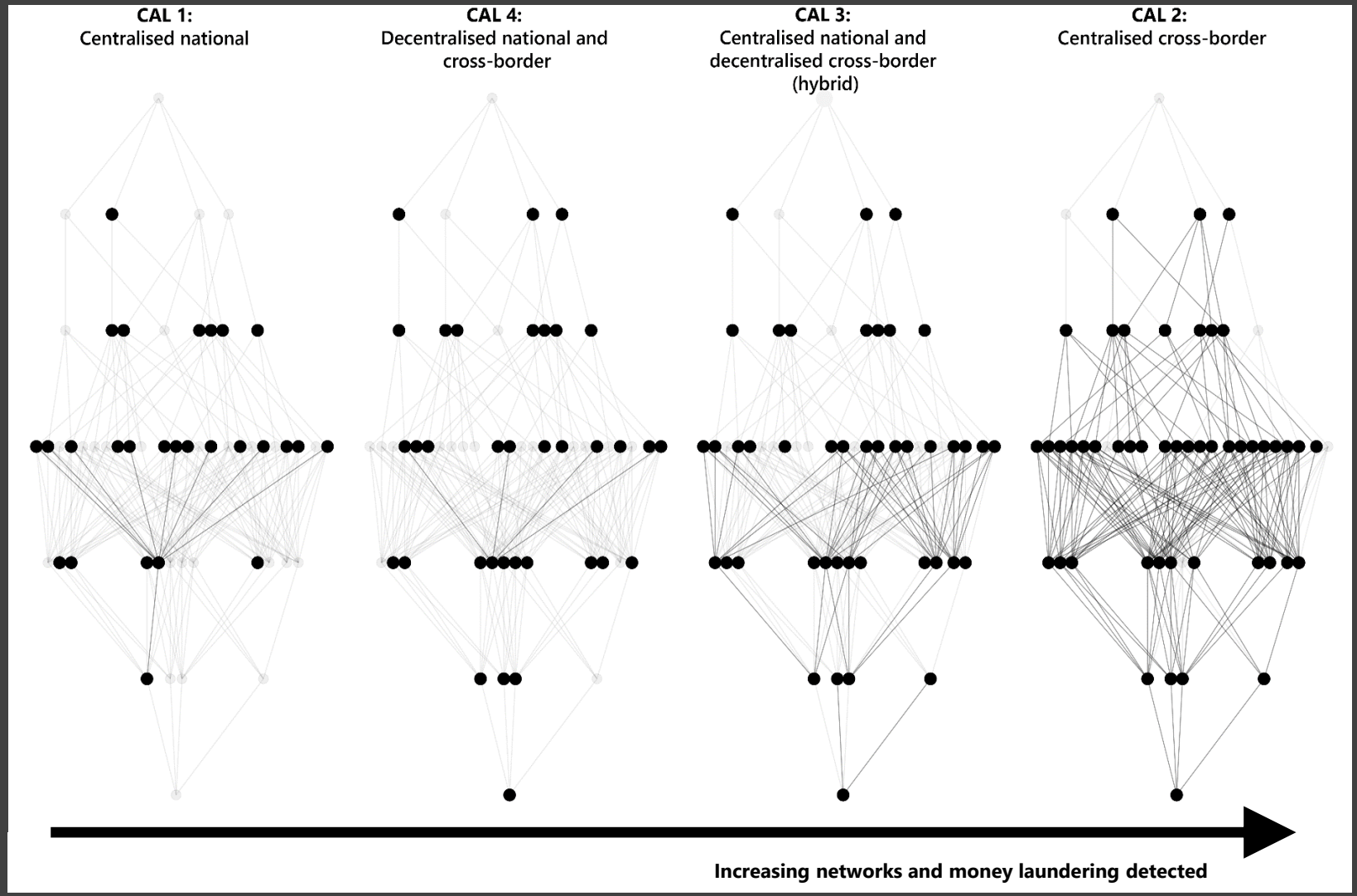


Part C – Results

When **machine learning is combined with PETs** that **performance and effectiveness is maintained**



Part C – Results



The broadest view of data is most performant and effective

CALS vary in performance and effectiveness but all are better than today's silo / rule based

Different data points and sources = different typologies

- Not all money laundering typologies can be detected solely through transaction data analysis using a minimum set of data points.
- Depending on the typology, additional data sources and data points may be required to aid detection
- A small number of examples of this are:

N.B. there are > 11,000 typologies

- **Personal use of business accounts:** 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):** 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:** money launderers can use these channels to hide their trail of money. 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

Data standards example – Legal Entity Identifiers (LEI)

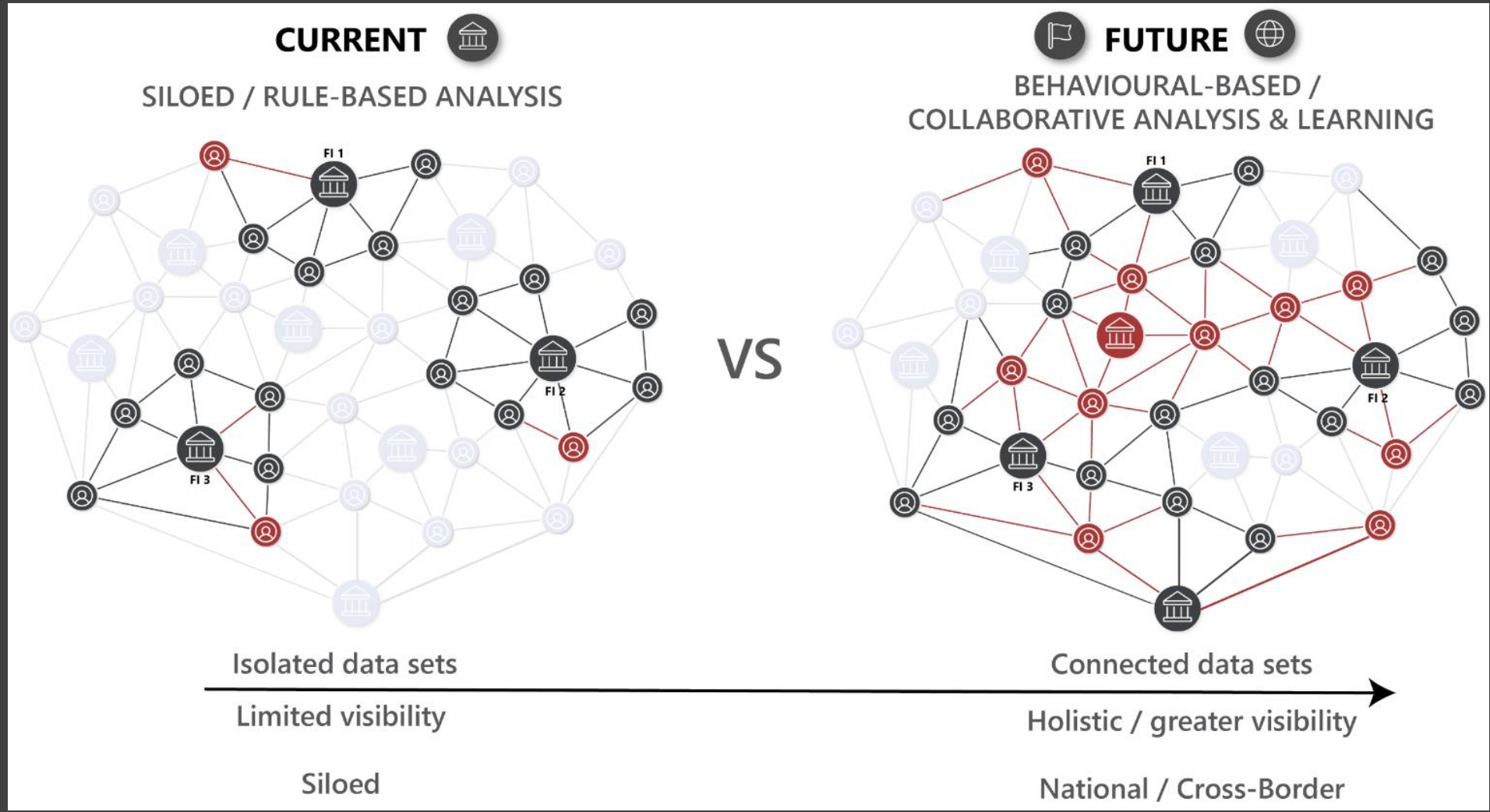
- Legal entities were represented by **name, entity number in the country, country, and address**
- Data constructed to allow **identification and linkage of legal entities**, while **LEIs can easily be utilized as unique identifiers for representing legal entities** as opposed to a generic entity number
- Approach showed the potential benefits of **using unique identifiers such as LEIs to enable effective analysis and detection of money laundering typologies and networks** involving legal entities

- **Enhanced Entity Resolution:** By utilizing unique identifiers like LEIs, network analysis could potentially accurately link and track entities across multiple transactions and financial networks. This could enable a comprehensive view of the connections between different entities involved in money laundering activities.
- **Uncovering Complex Relationships:** Graph structures built using identifiers could allow for the visualization and exploration of complex relationships within a network. These structures reveal hidden patterns, connections, and dependencies that can aid in identifying intricate money laundering networks that may be difficult to detect through traditional analysis methods.
- **Real-Time Monitoring and Predictive Analytics:** Identifiers could enable proactive detection of emerging money laundering networks and predictive insights.

BIS Innovation Hub

Aurora

What the future could be



BIS Innovation Hub

Aurora

Conclusion

CAL approaches using AI / machine-learning models and PETs showed greater effectiveness in detecting money launderers and suspicious networks compared to siloed and rule-based monitoring. Could be **game changing**.

PETs could enable secure and privacy-preserving CAL arrangements. Investment and development of PETs and PPC is advancing exponentially and in time commoditized and scalable services will be available. Many applications.

Collaboration between the private and public sectors is crucial for innovation in AML efforts. **Cross-functional dialogue is needed** to support the development and / or adoption of data standards (e.g. LEI, ISO20022, UBO)

Real-world proofs of concept (or pilots) would be needed to test the feasibility and effectiveness of the technologies and CAL approaches at scale. Necessary to surface legal, regulatory, data protection, and technical issues and questions.

New technologies, such as **Knowledge Graphs,** could be explored further with public and private sector experts as a public good exercise, which could enable more better standardised information sharing between the private the public sectors.

A data driven approach is essential. Though some data would be common, **different typologies require different data fields / data sets** – some of which may come from other public or private sources. **Collaboration is key.**


BIS Innovation Hub
Aurora
Questions