Quantexa Financial Investigation Solution for Microsoft Azure

How to manage analytics, prevention, and detection for financial crime.

By Jerome Bryssinck, Director of Government Solutions, Quantexa EMEA
Karl Heinz Krug, Industry Advisor Public Finance, Microsoft Western Europe

quantexa | Microsoft Azure

Index

Executive Summary

Anti-money laundering efforts have become a global priority in an attempt to thwart the two to five percent of GDP ($800 billion and $2 trillion) that the United Nations Office on Drugs and Crime (UNODC) claims is laundered each year[1].

In their battle against money laundering, authorities face the daily operational challenges of data sharing and data collaboration as well as utilizing data and artificial intelligence (AI) in analysis and investigation. They are not only limited by technological abilities but by existing requirements and restrictions from each country's legislation. Existing regulations on data protection and data privacy, for example, limit the possibility to openly share information between authorities at the national and international level.

There is a need for data-privacy-compliance, in-depth-analysis and cross-authority case management and data sharing. Trusted Execution Environment (TEE) is a technology for managing these boundaries. This technology has also been in the scope of the FATF[2] in the evaluation of possible solutions.

Quantexa's Decision Intelligence (DI) platform is a data-agnostic solution, which does not have a predefined data model. Any dataset and all entity attributes within a dataset can be ingested, cleansed, and combined for matching, analysis, and monitoring. Quantexa's platform provides a broad set of capabilities with a number of proprietary features for parsing, cleansing, and standardizing data. This world-class platform provides a single, authority-wide entity and network generation for many user profiles; allowing several investigation methods while avoiding data duplication. With the broad capabilities of Quantexa's platform, intelligence officers have a powerful tool that can run different analytical tasks, covering a risk-based approach and integrating third-party data and external sources for deeper investigation.

Providing the right solutions and technology is crucial for an effective and efficient way to combat financial crimes. These solutions must also comply with existing legislations concerning data protection and data privacy. Investigation results also need to be documented and evidence must be stored in a case management system for cross-authority collaboration and prosecution. By seamlessly integrating Microsoft Dynamics 365 case management, case and file handling can be managed end-to-end.

Microsoft provides Confidential Computing and TEE on which the Quantexa solution is built. This approach allows sensitive data to be handled in a highly protected environment of hardware-based encryption and technical enforcement rather than just implementing business policies. These technologies even render the workloads not visible to the cloud provider and limit the processing of secrets to only a specified part of the code modules in an enclave of the TEE.

The solution enables authorities' collaboration with the possibility of controlled data sharing and collaboration in confidential cloud services, with strong technical guarantees regarding the use of pooled data.
The confidential cloud services work across Quantexa instances ensures that none of the contributing participants can have access to the dataset, which is key-encrypted and is only accessible to the hardware protected TEE.

Therefore, our proposed solution offers a powerful toolbox for combatting money laundering and financial crime while complying with the highest standards of data privacy and data protection. It also enables authorities on a national and international level to work hand-in-glove to fight financial crime.

[1] Money Laundering | Europol (europa.eu)
[2] Stocktake on data pooling, collaborative analytics, and data protection, FATF, July 2021

"Quantexa's platform provides a broad set of capabilities with a number of proprietary features for parsing, cleansing, and standardizing data. This world-class platform provides a single, authority-wide entity and network generation for many user profiles; allowing several investigation methods while avoiding data duplication."

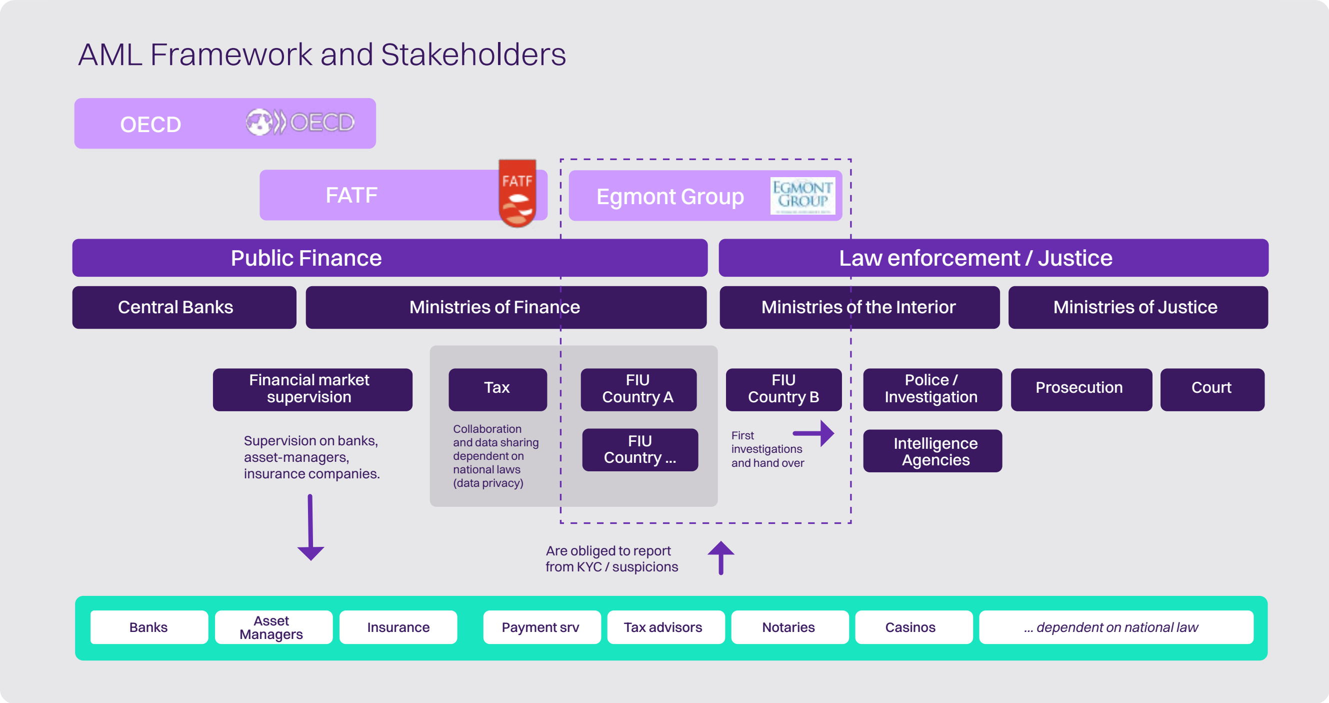
Challenges and requirements in AML

Combatting money laundering (AML) and the financing of terrorism (CFT) is a top priority for the Organization for Economic Cooperation and Development (OECD) and G7/G20, and it has been on the agenda of each summit since 2008.

Illegal money from drugs, human trafficking, wildlife trade, and weapons sales is laundered globally with proceeds being used to finance terrorism and destabilize democracies. And as the ongoing war in Ukraine has made clear, sanctions and the need to identify the beneficial owners of shell-companies cloaking Russian oligarchs, show the need to "follow the money" to identify the source of illegal funds as being a top priority. Yet estimations worldwide show that up to 90% of money-laundering crimes go undetected[3].

While there is consensus for closer and more aligned collaboration based on standardized AML procedures, in practice, Financial Intelligence Units (FIU) and law enforcement still struggle with operational challenges. These operational challenges often refer to existing limitations, which are not only technological, but often refer to existing requirements and restrictions from each country's national legislation. Data privacy and protection regulation obliges authorities to comply and limits their ability to counterattack. This is especially challenging in federated environments, where many authorities or AML bodies, working together with private financial institutions, make collaboration more difficult.

High level overview on AML framework and data sharing needs



AML Framework and Stakeholders
OECD
OECD
FATF
FATF
Egmont Group
EGMONT GROUP
Public Finance
Law enforcement / Justice
Central Banks
Ministries of Finance
Ministries of the Interior
Ministries of Justice
Financial market supervision
Tax
FIU Country A
FIU Country B
Police / Investigation
Prosecution
Court
Supervision on banks, asset-managers, insurance companies.
Collaboration and data sharing dependent on national laws (data privacy)
FIU Country ...
First investigations and hand over
Intelligence Agencies
Are obliged to report from KYC / suspicions
Banks
Asset Managers
Insurance
Payment srv
Tax advisors
Notaries
Casinos
... dependent on national law

[3] Zippia. "20 Money Laundering Statistics [2022] Facts About Money Laundering In The U.S." Zippia.com. Oct. 18, 2022

Therefore, it is not only about enhancing the capabilities to analyze data, but also about technologies for sharing data. Yet each has its challenges. Where analytics is concerned, more and more complex data from different sources, suspicious transaction reports (STRs), government data, third-party (KYC[4]) data, data from crypto- and brokerage platforms, and social media are but a few of the obstacles.

This vast amount of data needs to be ingested, structured, and analyzed, ideally in a real-time-manner. This requires extensive capabilities in big data and AI, and 360-degree views on objects by taking into consideration across all connected data points. For fast action, a streamlined case management system is needed that provides the right information to the right authority in a timely manner.

A Heterogenous And Complex Data Ingestion

Over the years, sources of data for AML have been evolving. The continuous digital transformation also impacts criminal operations and behavior, and requires law enforcement and AML bodies to keep up with the developments and to enhance their digital skills. The authorities need extensive data management capabilities to ingest complex and heterogenous data from different sources across all kinds of transactions and registers. Almost all criminal transactions can be traced, assuming that one has access to the data sources and the ability to ingest, analyze, and interpret the data. Handling the data requires a thorough process of data collection, data triage, data fusion, data analysis, and data communication and dissemination[6]. Data quality and data cleaning are some of the most important tasks of the data-collection process, apart from the selection of the right data sources.

Data used can consist of structured and unstructured data, and plain text documents, which could be analyzed by entity recognition and OCR. It could also consist of images and graphs as well as social media content and sentiment analysis. The types of data as well as the means of incorporating it into the analytic capabilities can have different characteristics, hence solutions being used need to be capable of handling and analyzing these types of data to track and trace illicit activities.

B Real-Time Monitoring And Decision-Making

Due to the speed of money flows and electronic transactions, real-time-monitoring and immediate actions are crucial. This requires continuous streaming analytics in real-time. Continuous data flows and transaction monitoring on real economy insights provide authorities with the ability to act.

Quantexa addresses these complex challenges with transformative technologies that provide context to data operations. With a real-time, AI-driven approach to Decision Intelligence, Quantexa uniquely integrates and connects internal and external data to provide a single view of risk necessary to safeguard public systems, cut through rapidly increasing data complexities, exceed the demands of complex regulatory requirements, and holistically understand risk.

The process of data ingestion and monitoring needs to be accelerated, which requires that repetitive, time-consuming activities be transformed into value-creating activities in analysis and decision-making. These tasks also need to be manageable within a reasonable time limit. It follows then that more automation and less manual interaction are required. Focusing on the right information, by prioritization STRs using a risk-based approach, helps to focus valuable human resources on the right cases. Tools can support this approach by helping to identify cases with evidence of money laundering while at the same time avoiding too many false positives.

Decision-making requires the management and handling of cases. This kind of solution integration is an essential part of a technology environment in AML/CFT. The time it takes to move from decisions to actions must be minimized for the process to be effective.

[4] KYC = Know Your Customer)
[5] Stocktake on data pooling, collaborative analytics, and data protection, FATF, July 2021
[6] Digital transformation of AML/CFT for operational agencies, FATF, October 2021

C Data Sharing and Collaboration In Compliance With Data Privacy

One of the most critical aspects within AML/CFT is data sharing and data collaboration. Existing legislative frameworks on data privacy and protection reduce the ability to openly share data and collaborate between national and international authorities, as well as in the private sector. If data pooling is not possible, data sharing needs to be followed. This is often allowed only in a very small bandwidth of activities and only on limited datasets of legitimate interest.

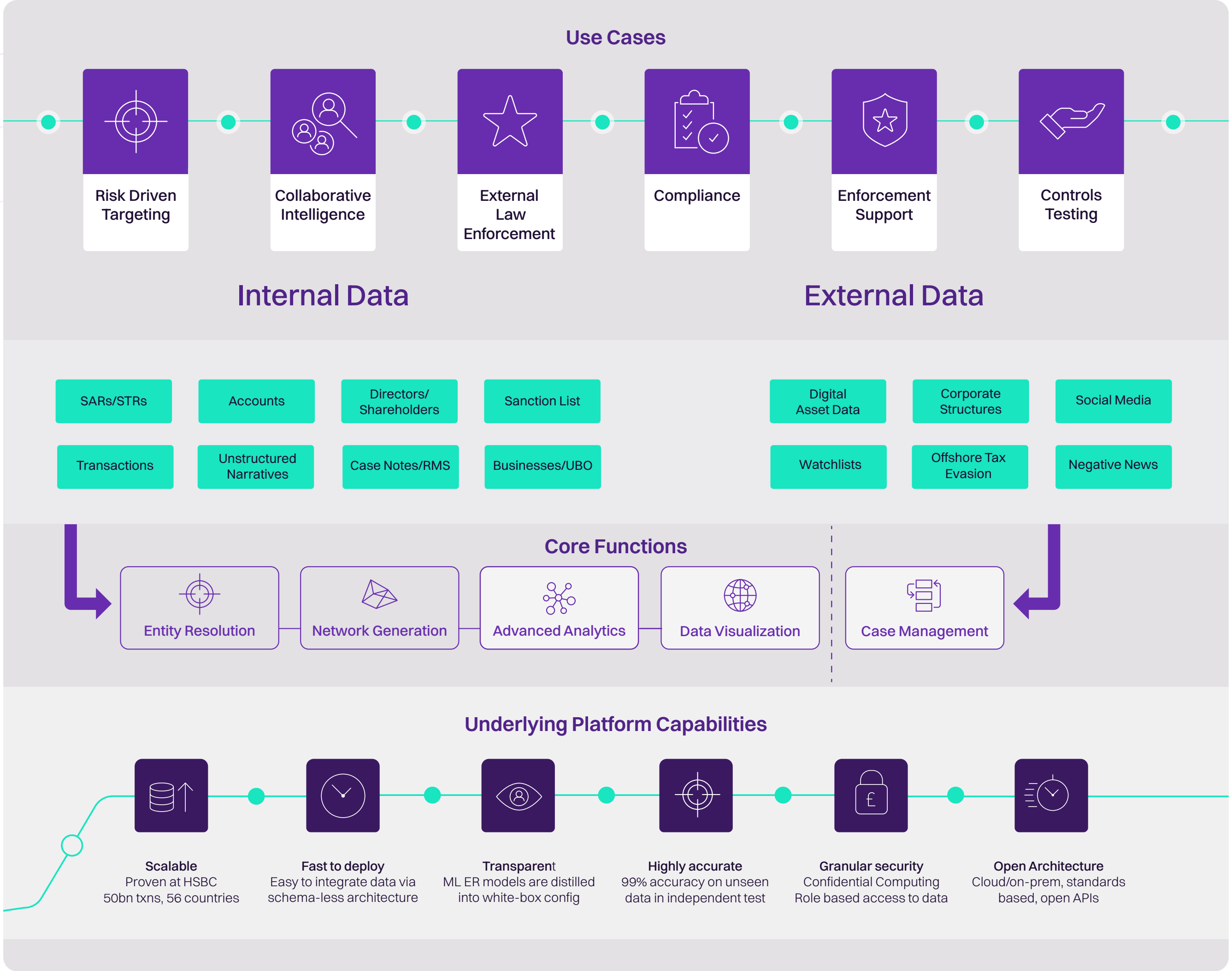
The rules and procedures for this can slow down processes, reducing the ability to act fast. In the trade-off between AML/CFT activities and data privacy protection, there is a need for solutions that enable AML bodies to share data on agreed procedures and frameworks in a timely manner. This requires technology-enabled data minimization of personal data for very specific legitimate interests. It is also a question of where to retain data and who should be provided with access. In federated and multi-national stakeholder environments, each party will be the owner of its own data estate but providing access for legitimate purposes is key for an effective AML approach. This access needs to be encrypted and secured by establishing privacy-enhancing technologies.

> **The challenge in data sharing** not only impacts cross-authority collaboration, but also the collaboration between the public and the private sector. In this instance, **additional requirements for data sharing and data collaboration must be considered.**

Solution Capabilities

To address the challenges in AML/CFT and to provide a holistic solution, Quantexa and Microsoft designed an AML/CFT solution combining strengths and capabilities from both parties. The broad Decision Intelligence capabilities from Quantexa build on Azure data and AI services, and the extended encryption for data sharing and data collaboration. This core of the solution is complemented by Microsoft Dynamics 365 case management.

Quantexa CDI: Core Functions for FIU



Use Cases

Risk Driven Targeting

Collaborative Intelligence

External Law Enforcement

Compliance

Enforcement Support

Controls Testing

Internal Data

SARs/STRs

Accounts

Directors/ Shareholders

Sanction List

Transactions

Unstructured Narratives

Case Notes/RMS

Businesses/UBO

External Data

Digital Asset Data

Corporate Structures

Social Media

Watchlists

Offshore Tax Evasion

Negative News

Core Functions

Entity Resolution

Network Generation

Advanced Analytics

Data Visualization

Case Management

Underlying Platform Capabilities

Scalable
Proven at HSBC
50bn txns, 56 countries

Fast to deploy
Easy to integrate data via
schema-less architecture

Transparent
ML ER models are distilled
into white-box config

Highly accurate
99% accuracy on unseen
data in independent test

Granular security
Confidential Computing
Role based access to data

Open Architecture
Cloud/on-prem, standards
based, open APIs

The following pages will describe the process from data ingestion, data modeling, and analytics to case management. An introduction into Confidential Computing and the capabilities of the Confidential Consortium Framework will provide deeper insights into encryption technologies for secure data sharing and collaboration between authorities and third parties.

A Data Ingestion and Data Models

Both batch and real-time ETL (extract, transform, load) is configured in Quantexa's Data Fusion ETL framework (where batch data is loaded from Apache Hadoop Distributed File System (HDFS), and real-time data is loaded via Kafka). Data Fusion provides high-performance functions, which parse, cleanse, and standardize key linking information for further batch processing or loading to Elasticsearch.

This linking information includes the names, dates, addresses, contact details, and identifiers of all entity types such as individuals, businesses, places, vehicles, devices, or any "real-world entity." Matching the linking information for these entities is the process of Entity Resolution described in the next section. The Quantexa platform is data-agnostic and does not have a predefined data model, meaning any dataset and all entity attributes within a dataset can be ingested and combined for matching, analytics, and monitoring.

Quantexa's out-of-the-box parsers, cleansers, and standardizers allow users to extract linking information from data with variable structures and quality.

Data Fusion includes individual name parsing, business name classification/standardization/cleansing, phone standardization, and email standardization. A set of name synonyms, produced from analyzing global name data, are also supplied. Multiple transliteration schemes can be used for key entity data, and support for double metaphone encoding allows for phonetic matching to aid with the fuzzy matching process. These natural language processing (NLP) functions can be used on structured and semi-structured data.

The standard functions within Data Fusion can also be customized to address the nuances of implementation-specific data sources to maximize resolution performance, so they are provided as a fully customizable source code. In addition, third-party/open source ETL functions (such as NLP libraries) can be deployed within Data Fusion to extend this functionality. For example, Quantexa has previously leveraged Tensorflow NLP and Stanford NLP libraries within our ETL pipeline for additional entity extraction on unstructured datasets. Custom ETL can also be configured and deployed.

Parsing, Cleansing and Standardisation Functions

Quantexa provides several out-of-the-box functions for performing parsing, cleansing, and standardization of raw data. The Quantexa process for **batch data ingestion** is as follows:

- Raw data is extracted from up-stream systems and copied to HDFS, without any prior transformation/cleansing needed
- Raw data is mapped to a hierarchical model
- Raw data is processed by Data Fusion to create new versions, which have been cleansed, parsed, and standardized. All cleansing is additive, so that new versions are created and raw values are kept
- All processed data elements (i.e., raw and new) are combined into "compound keys," which can be used to match original records together within Entity Resolution
- All elements and compounds are persisted to HDFS for subsequent use within batch processing (Entity Resolution, network generation, scoring, and analytics) and also loaded into Elasticsearch indexes for use by Quantexa's real-time processing engines (dynamic resolver and dynamic scoring)

A similar process is used for real-time data ingestion:

- Raw data is copied from relevant Kafka streams
- Raw data is processed by Data Fusion to create cleansed, parsed, and standardized elements and compound keys
- All elements and compounds are loaded into Elasticsearch indexes for use by Quantexa's real-time processing engines (dynamic resolver and dynamic scoring). They are also be copied to HDFS for use within subsequent batch processing.

For parsing, cleansing, and standardizing data, Quantexa has developed a number of proprietary features including several machine-learning classifiers for individual/business names, as well as global address parsers built on natural language processing. These enable Quantexa to standardize data of variable structures and quality, and create 'alternate' versions of fields, which are optimized for Entity Resolution.

Entity Resolution allows a unified view of a unique real-world entity (such as a person, a business, or an address) to be created from multiple disparate data sources, as well as within a single data source where there are duplications. Quantexa's Entity Resolution engine can be operated dynamically giving the system the flexibility to manage numerous use cases, as well as varied user requirements and security considerations.

The challenge with batch-only Entity Resolution systems is that a single deployment of the system can only cater for a single configuration of Entity Resolution. Different user requirements, investigation types, security protocols, and use cases will have different requirements for both the confidence levels required in the resolution and the input data to be resolved into the entity. As a result, we have seen many organizations have to implement batch-based systems multiple times to service different needs. With Quantexa, a single instance of the technology and a single underlying data layer can service an unlimited number of resolution templates for different requirements. The dynamic engine even allows users to alter the Entity Resolution parameters in real-time during an investigation. This is crucial for discovering potential connections, especially when there are challenges with data quality and availability.

To overcome the challenges with batch-only systems, Quantexa has developed a world class engine for dynamic Entity Resolution, which allows a single enterprise-wide entity and network generation engine servicing many user profiles, investigation methods and use cases while avoiding data replication. The Entity Resolution process was designed to work in the domain of fraud and financial crime, which is characterised by fraudsters who will try to provide incomplete and incorrect information to avoid detection. As a result, Quantexa's Entity Resolution capability performs well in the face of poor quality or manipulated data. This means that data quality issues do not need to be fixed prior to deploying Quantexa – it is designed with data challenges in mind.

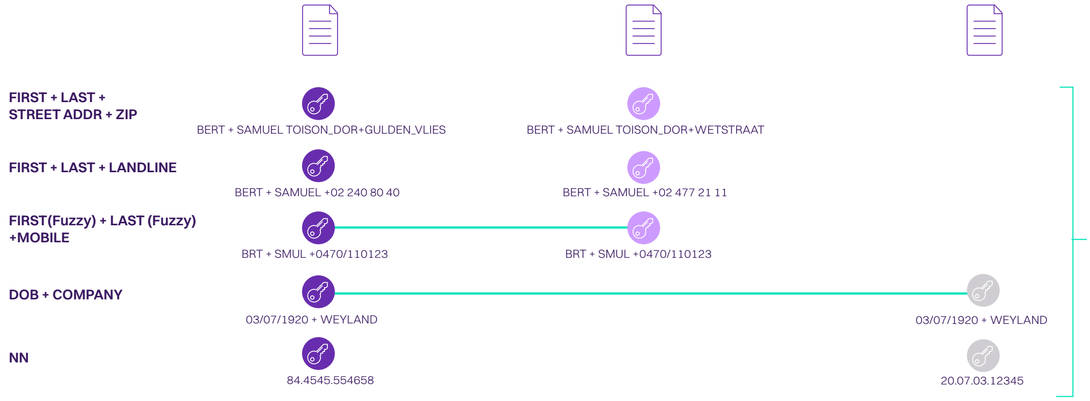
B Entity Resolution in Quantexa's Decision Intelligence platform

To achieve maximum accuracy, Quantexa creates every combination of attributes which can be used for matching to help to identify a unique entity, based on what is available in the input data. Such combinations are called "compounds" and form the basis of the Entity Resolution process.
In addition, Quantexa can generate enriched compounds including phonetic or "sounds-like" matching and can also apply uniqueness measures to each compound to assess and control the strength of the resolution in the context of the available data. For example, common names combined with common DOBs (e.g. Wei Tan + 01/01/1960) create weaker compounds than rare names combined with normally distributed DOBs (e.g. Felix Hoddinott + 08/06/1980). Recipes (i.e. different combinations) of compounds can then be selected to service different Entity Resolution requirements. Users can also dynamically control the resolution process from the User Interface (UI). This is crucial for discovering potential connections between entities when data quality is a challenge. Quantexa has leveraged a host of machine-learning techniques using extremely high volumes of real-world data to identify the most performant compounds for resolving the common entity types for many use cases. This has made Quantexa's Entity Resolution capability market-leading and proven to achieve up to 99% accuracy.

The Entity Resolution capability will also automatically calculate "counter-evidence" for each compound.
This caters for compounds which look abnormal/weak based on other data elements. For example, the system can opt to use "Name + Street Address + Zip" unless there are more than X apartment numbers at that street address or "Name + Date of Birth" unless this connects more than 1 National Number.

Documents as input for Entity Resolution

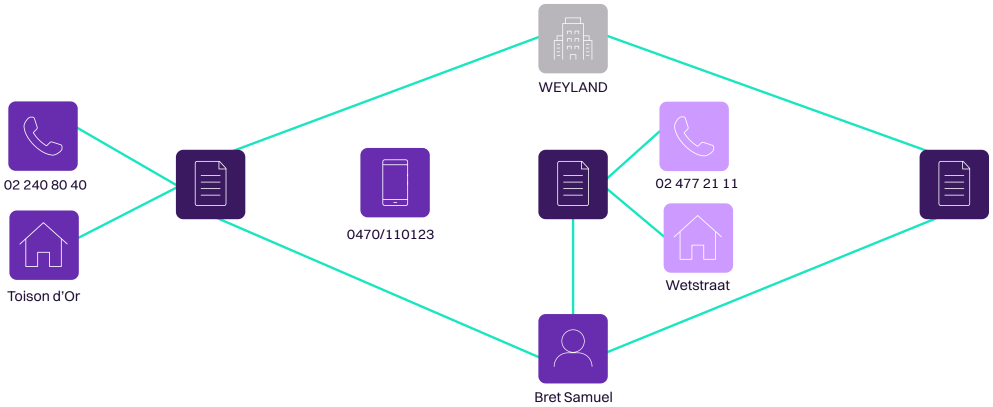
Document	Company	Cust Name (Common Name)	NN	First Name	Last Name	Fuzzy First	Fuzzy Last	Street Address	Zip	DoB	Cell Phone	Phone
SAR_1	WEYLAND	Bret Samuel	84.4545.554658	BERT	SAMUEL	BRT	SMUL	Toison d'Or	1050	03/07/1920	0470/110123	02 240 80 40
SAR_2		Samuel Bret		BERT	SAMUEL	BRT	SMUL	Wetstraat	1000		0470/110123	02 477 21 11
SAR_3	WEYLAND	B.Samuel	20.07.03.12345	B	SAMUEL	B	SMUL			03/07/1920		



There are compounds that connect across all the rows so this forms one entity

Quantexa has developed an off-the-shelf (but still open and configurable) entity model for the most frequently encountered entities like person, company, address, telephone, bank account, etc.

Network based on the resolved entities



Document	Company	Cust Name (Common Name)	NN	First Name	Last Name	Fuzzy First	Fuzzy Last	Street Address	Zip	DoB	Cell Phone	Phone
SAR_1	WEYLAND	Bret Samuel	84.4545.554658	BERT	SAMUEL	BRT	SMUL	Toison d'Or	1050	03/07/1920	0470/110123	02 240 80 40
SAR_2		Samuel Bret		BERT	SAMUEL	BRT	SMUL	Wetstraat	1000		0470/110123	02 477 21 11
SAR_3	WEYLAND	B.Samuel	20.07.03.12345	B	SAMUEL	B	SMUL			03/07/1920		

Network generation is the process of connecting entities through events or other relationships in the data such as shared addresses, telephone numbers, email addresses, corporate structures, watchlists entries, and transactional flows. It is common within intelligence agencies for investigators to construct knowledge graphs from the data and intelligence that they gather from diverse sources and to analyse them for risks. Working in combination with Entity Resolution & scoring, Quantexa's network generation engine is automating this process on a large scale using the full volume of available data.

Using the full volume of data across all available sources, Quantexa's network generation process goes beyond simply identifying connections within the data, and also determines which connections are useful or relevant for the use case, avoiding confusing overlinking and the visualization of the numerous meaningless connections which exist in the data (such as commonly used ATMs connecting thousands of nodes in the network together). Like Entity Resolution, networks can be generated in batch and dynamically. Technical users can configure "graph scripts," which tell Quantexa how the networks should be built, and which connections are relevant for a particular use case. Quantexa provides numerous graph scripts out of the box and will tune these to specific requirements during the implementation.

Graph theory algorithms are also supported out-of-the-box for use within the network generation logic. Quantexa uses GraphX's extensive library of graph algorithms (including PageRank, Connected Component, Triangle Count, etc.) to identify clusters and communities within the entity connections. These communities can then be visualized within the networks loaded in the Quantexa UI. GraphX runs natively on Apache Spark, which is Quantexa's core batch processing technology.

As well as being critical for visual investigation, networks are used within the automated risk detection process, where the risk detection models can identify patterns of behavior and concentrations of risk within the network. These network scores form a key part of the scorecards, which are used to identify and prioritize risk across numerous different typologies. Network generation can either be triggered by scheduled, predefined batch configuration, or on an on-demand basis via Quantexa's dynamic microservices.

Quantexa's Scoring Framework enables a revolutionary approach to look across the full context of data and provide an aggregate view of risks and insights. Compared to traditional approaches to detection, Quantexa's contextual approach delivers significantly improved efficiency and effectiveness. Quantexa builds this contextual view by applying analytics across multiple levels:

1. **The event or document of information** – a transaction, a piece of intelligence, a citizen profile, etc.

2. **The entity** – the individual, company, address, vehicle, telephone number, and its entire history and behavior over time.

3. **The network** – the group of entities and how they link and interact.

Across each of these three levels, a series of detection methods can be used to identify both known and unknown threats:

Rules and scenarios

Patterns of behavior that have been observed by investigators and domain experts can be detected via targeted rules and scenarios. Quantexa enables these to be applied to the event, the individual or business at the center of that event, as well as relationships with other entities. Through the course of investigations and strategic analysis, as new scenarios are uncovered these findings can be integrated in the underlying threat assessment model.

High-risk connections

Any connections to high-risk data can be detected, (e.g. high-risk cases, known criminals, etc.) This is important for identifying entities that do not immediately appear to be risky, but in fact have hidden risks through their connections or relationships. Moreover, Quantexa's ability to fuzzy match records means that even where targets have attempted to manipulate or obfuscate their details (or there is poor data quality), these high-risk connections can still be detected. The strength of connection to high-risk data is also assessed, (e.g. a mobile/cell phone connection to a known criminal is deemed riskier than a historic address connection.)

Network analysis

Quantexa uses network analysis to automatically identify the key actors in a network or detect the significant events across the full volume of data. Network analysis is also used continually to identify suspicious patterns of interest. These suspicious network patterns can be used in the risk scoring together with other analytical models and expert rules.

Machine-Learning

Quantexa's Scoring Framework can utilize various machine-learning (ML) methods in combination with other rules, scenarios, and networks analytics:

Supervised Learning (reliant on sufficient volumes of labeled examples)

Predict Risk: Models using algorithms such as Decision Trees, Logistic Regression, and K-Nearest Neighbour identify the features indicative of historical known cases. Quantexa can use the widest set of open source and proprietary ML algorithms thanks to the platform's open API approach, but open, transparent models that are explainable to a governance process are preferred/recommended for financial crime use cases. Features are our model scores and can themselves be derived from ML (typically anomaly methods). Typically, the features used combine the network context, the network risk indicators and mitigators with the core characteristics of the customer/party (e.g. type).

Residual Analysis: Models can use the full context of the customer to predict how they would normally behave, and identify parties who do not behave as expected, e.g. the model predicts transactions of $10k and we observe transactions of $100k. Methods include Regressions, Gradient Boost, Decision Trees, and Nearest Neighbor. One strength of this approach is that the model strength is reported, unlike unsupervised methods. So, when looking for anomalies in a characteristic, which isn't predicted by network context, then the model informs us.

Temporal Analysis: This is a special case of Residual Analysis, where the model predicts the usual behavior as a trend and finds anomalies. Quantexa can extend normal approaches by not just using a customer's own behavior to define expected behavior but also to consider the behavior of similar customers. This avoids social/market events (e.g. Covid-19) being triggered as anomalies.

Predict Segment Membership: Models can use the full context of the network to predict which segment the customer should belong to, (e.g. predict the industry segment.) This can then be used to detect anomalous behavior for the segment.

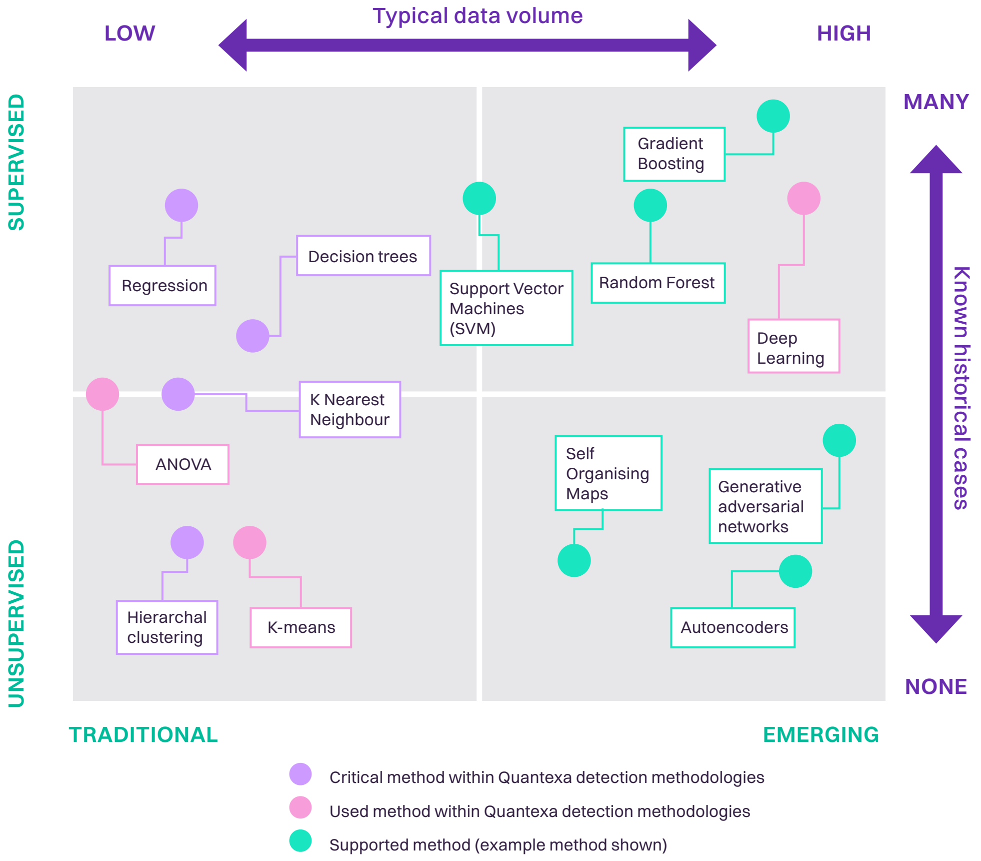
Unsupervised Learning

Segmentation : Models can group parties into a set of segments (a.k.a. clusters) based on a measure of similarity. Segmentation including behavior are described as Dynamic Segmentation, because the customer can move a segment dynamically without details such as KYC being updated. Typical methods include K-Means, K-Nearest Neighbor, and Persistent Homology.

Hierarchal Segmentation: We frequently find we have a classification in the input data with a large number of segments, (e.g. a client/party type or SIC code.) Models can identify segments as similar if they have historically behaved in similar ways or contain similar customers and can then hierarchically group segments to obtain a useable number of clusters. Specific methods include Ward Linkage or Average Linkage.

Peer Analysis: Given a segmentation (either an existing segmentation from the raw data or from the segmentation methods) models can look for records which don't appear similar to their peers. For example, clients/parties can be clustered based on their observed income levels (i.e. salary etc.) but the model can look to identify clients/parties making a large single payment which is abnormal compared to their peers. Key methods include Analysis of Variance (ANOVA). Models can understand whether the segmentation is indicative of the behavior being analyzed. If so, the segmentation is a good choice to use for peer analysis. This allows the model to keep multiple segmentations and use the best for a particular purpose.

Multivariate Anomaly Analysis: This is an extension of the above segmentation anomaly methods. Here the model learns the shape of the full data (deep learning) and identifies anomalous records. Usually, it is better to start with transparent-segmentation-based methods, then use these as a parallel method to identify more complex outliers.



In summary, the Quantexa Scoring Framework is:

- Data-scientist friendly, enabling customers/partners to build/maintain their own models
- Provides in-built Graph Scripting to rapidly configure new network-pattern logic for detection
- Structured framework with templates for model configuration
- Open API to Microsoft Synapse, with extended AI and ML capabilities
- Open API approach also means that the widest set of open source & proprietary ML algorithms can be leveraged (Python/R/Tensorflow etc.)
- Our architecture supports the building, deployment, and reporting on models using third-party tools such as MLFlow
- All triggered scores are stored/recorded for audit and model-retraining purposes
- The same scores can be used within both batch and real-time scoring pipelines where technically feasible (i.e. some batch scores may be too compute-heavy to run in real-time. These are run in batch and results made in real-time using lookups)

The Scoring Framework supports all score types across all levels:

	Business Rules / Scenarios / Known Typologies / Engineered Feautures (automatic & manual)	Anomalies / Peer Grouping / Unsupervised Learning	Overall Models: Weighted Scorecards / Expert models, Machine Learning
Transaction Level Scoring	e.g. Transaction made to High Risk Geography	e.g. Excessive txns in one day (volume / value)	
Entity Level Scoring	e.g. Counterparty connection to an STR	e.g. atypical STR scenario for individual based on behavioural profile	
Network Level Scoring	e.g. U-turn transaction activity	e.g. high connection of STR's on network	

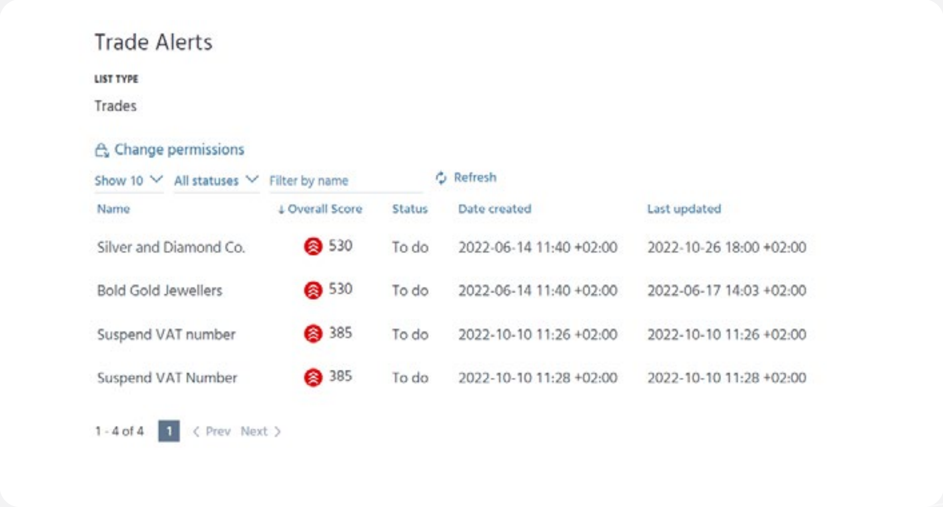
Aggregated risk across scoring levels has been proven to find Complex and Hidden AML risk

The Scoring Framework fully supports ensemble model deployment, leveraging a mixture of internally deployed scores and scenarios, scores derived from externally hosted models and rules engines (e.g. Python models, Tensorflow, AutoML, etc.), as well as lookups against batch-based scoring profiles.

C Contextual Monitoring transforms the view of risk

Whether responding to a request, running a search, preparing for an operation, or investigating an alert, whenever the situation is complex it will often require manual investigator effort in producing "intelligence product," which normally takes the form of network diagrams. The Quantexa Decision Intelligence Platform can create all the networks on demand in seconds, having resolved entities, matched all the data and links, and highlighted the risks. The intelligence officer has an immediate starting point, they can explore the data as far as it links, and they can add human intelligence to the background data. They may also be alerted to emerging risks based on inbound data or to changes in the cases. For example, consent Suspicious Activity Reports (SARs) may identify a list of high-risk transactions and the system can instantly score these and present them to the user for further investigation.

List of transaction requests of potential interest



Trade Alerts

LIST TYPE

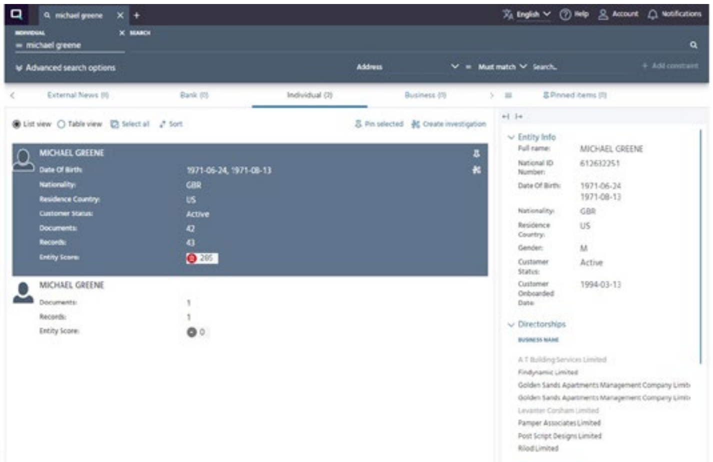
Trades

Name	↓ Overall Score	Status	Date created	Last updated
Silver and Diamond Co.	530	To do	2022-06-14 11:40 +02:00	2022-10-26 18:00 +02:00
Bold Gold Jewellers	530	To do	2022-06-14 11:40 +02:00	2022-06-17 14:03 +02:00
Suspend VAT number	385	To do	2022-10-10 11:26 +02:00	2022-10-10 11:26 +02:00
Suspend VAT Number	385	To do	2022-10-10 11:28 +02:00	2022-10-10 11:28 +02:00

1 - 4 of 4

Users can also search all the data in the system using any combination of criteria. The example shows the company director, shareholder, and ultimate beneficial ownership records.

Manual search and batch search



MICHAEL GREENE

Date Of Birth:	1971-06-24, 1971-08-13
Nationality:	GBR
Residence Country:	US
Customer Status:	Active
Documents:	42
Records:	43
Entity Score:	285

MICHAEL GREENE

Documents:	1
Records:	1
Entity Score:	0

Entity Info

Full name:	MICHAEL GREENE
National ID Number:	612632251
Date Of Birth:	1971-06-24 1971-08-13
Nationality:	GBR
Residence Country:	US
Gender:	M
Customer Status:	Active
Customer Onboarded Date:	1994-03-13

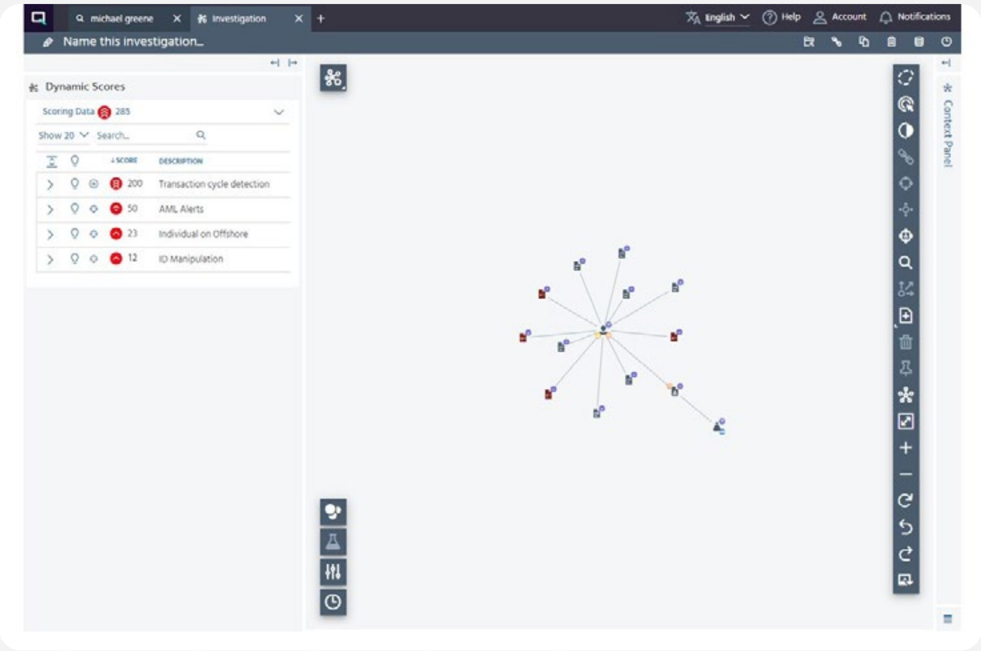
Directorships

BUSINESS NAME

A T Building Services Limited
Findynamic Limited
Golden Sands Apartments Management Company Limi
Golden Sands Apartments Management Company Limi
Levanter Corsham Limited
Pamper Associates Limited
Post Script Designs Limited
Rilod Limited

The real value, however, is having the ability to then immediately drill into the network of all the different data sources within the system. The Quantexa Decision Intelligence Platform takes care of the data-quality issues and the matching required to link data, and presents a result in seconds, with as many different risk-scoring highlights as required.

Investigating a risk network



Dynamic Scores

Scoring Data 285

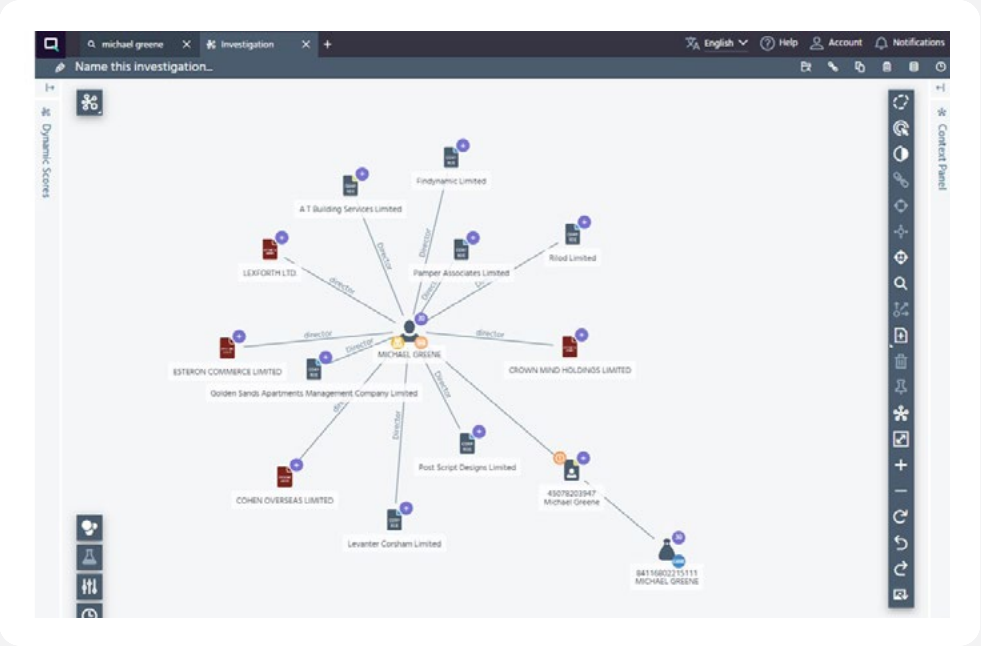
Show 20 Search…

SCORE	DESCRIPTION
200	Transaction cycle detection
50	AML Alerts
23	Individual on Offshore
12	ID Manipulation

The investigator clearly sees the connections to the individual, business, or other entity that the investigator has searched. This shows all the resolved entities, links to other records, and highlighted risk scores.

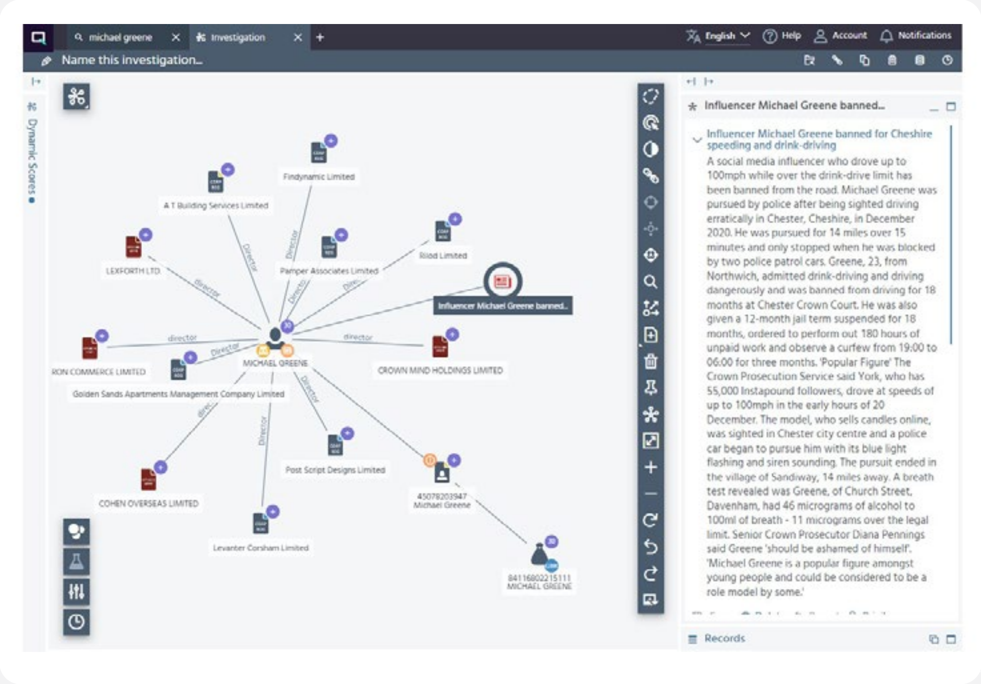
An investigator can also expand (or step out) from an entity without limit, to see what other connections exist. Also, the definitions of how entities should be built (i.e. the combination of attributes that will define an entity) are consistently applied in the initial network. However, it is possible for an investigator to loosen or strengthen a match dynamically for a particular entity, and to visualize the resultant network(s).

Geographic and time profile



The geographic profile enables investigators to understand where individuals or companies are located. Time is a critical aspect in an investigation including the ability to see whether two people lived at the same address at the same time.

Open source Intelligence



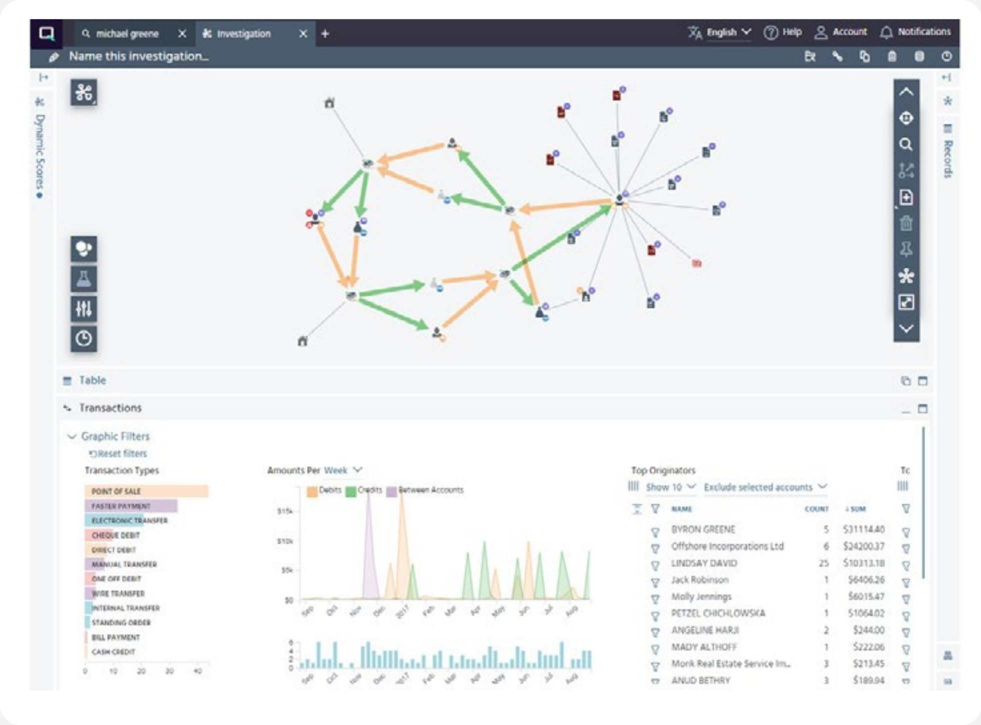
Influencer Michael Greene banned...

Influencer Michael Greene banned for Cheshire speeding and drink-driving

A social media influencer who drove up to 100mph while over the drink-drive limit has been banned from the road. Michael Greene was pursued by police after being sighted driving erratically in Chester, Cheshire, in December 2020. He was pursued for 14 miles over 15 minutes and only stopped when he was blocked by two police patrol cars. Greene, 23, from Northwich, admitted drink-driving and driving dangerously and was banned from driving for 18 months at Chester Crown Court. He was also given a 12-month jail term suspended for 18 months, ordered to perform out 180 hours of unpaid work and observe a curfew from 19:00 to 06:00 for three months. 'Popular Figure' The Crown Prosecution Service said York, who has 55,000 Instapound followers, drove at speeds of up to 100mph in the early hours of 20 December. The model, who sells candles online, was sighted in Chester city centre and a police car began to pursue him with its blue light flashing and siren sounding. The pursuit ended in the village of Sandiway, 14 miles away. A breath test revealed was Greene, of Church Street, Davenham, had 46 micrograms of alcohol to 100ml of breath - 11 micrograms over the legal limit. Senior Crown Prosecutor Diana Pennings said Greene 'should be ashamed of himself'. 'Michael Greene is a popular figure amongst young people and could be considered to be a role model by some.'

The system is also capable of reaching out to the Internet to drag in open-source intelligence (OSINT), or it can ingest OSINT provided elsewhere within the organization or through external providers.

Transaction viewer and visualization



NAME	COUNT	↓ SUM
BYRON GREENE	5	$31114.40
Offshore Incorporations Ltd	6	$24200.37
LINDSAY DAVID	25	$10313.18
Jack Robinson	1	$6406.26
Molly Jennings	1	$6015.47
PETZEL CHICHLOWSKA	1	$1064.02
ANGELINE HARJI	2	$244.00
MADY ALTHOFF	1	$222.06
Monk Real Estate Service Im...	3	$213.45
ANUD BETHRY	3	$189.94

There are a range of different visualizations that are possible to help intelligence officers understand transactional events such as financial flows, call records, or travel events.

The proposed solution also has a “bulk-search” capability that allows users to search for a list of entities (e.g.: companies, persons, addresses, etc.). The investigator can upload a list of entities, and thereafter the entities will automatically be resolved against all the data in the intelligence repository. Based on the resulting network, a risk score is calculated.

D Case Management and File Handling with Microsoft Dynamics 365

According to Gartner "Case management solutions are applications designed to support a complex process that requires a combination of human tasks and electronic workflow". As such, case management is applicable across sectors and processes including financial crime investigations and AML/CFT specifically.

Case management activities arrive across different channels of engagement; most obviously from external sources, but also from internal sources such as across departments or different agencies in a government organization and in international collaboration. The ability to reconcile, manage, and perform to agreed service levels in these circumstances is essential to quality of service and containing operating costs, compliance issues, and efficiencies. Dislocated and distributed information on activities with a certain case can lead to problems of higher costs, increased risk, and compliance issues.

As mentioned, case management is a process used inside of organizations, enabling the same platform of Dynamics 365 to be used for multiple reasons and case management functions or processes. This helps increase the value from the platform and aids re-use of skills and configured features. Internal processes relating to case management include dealing with criminal objects issues passing between departments, managing activities as part of an investigation or audit activity, or even in enabling colleagues to collaborate on activities in accordance with a consistent process and required response times, while providing traceability and analysis of outcomes.

Case management is a fundamental component of law enforcement and is the core record that tracks processes across channels and agents/authorities over time.
The motivations for case management in the public sector are often more focused on, compliance, risk, efficiency, and quality.

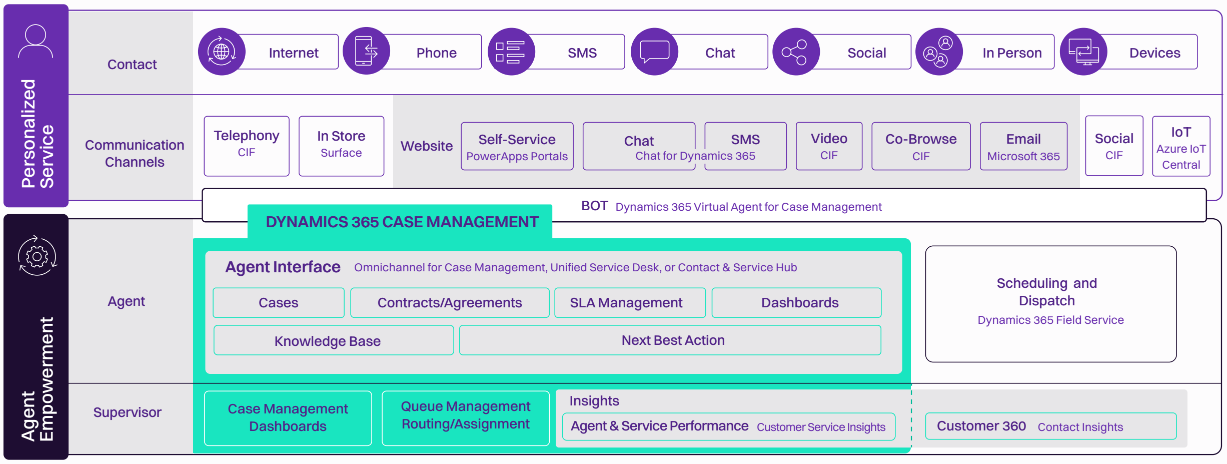
A case typically represents a SAR/STR that's reported and requires a resolution. Cases are designed to track the process from the initial intake of an incident/finding, through the remediation process, to the final resolution. Dynamics 365 has several components that work together to provide an end-to-end case management solution that not only helps identify cases but also routes each case to the most appropriate agent/authority who can provide guidance and resolve the case.

Listed below are some of the commonly used components:

- **Cases**: A case represents a single incident of service. It represents anything, in the context of an interaction, that requires some type of resolution or answer. Multiple cases can be associated with a single object at any time.
- **Activities**: An activity typically represents an interaction with an object, such as research. Multiple activities can be associated with a single case.
- **Entitlements**: Entitlements specify the rights that a case is entitled to.
- **Knowledge articles**: The knowledge base is a repository of informational background information that helps intelligence officers resolve cases.
- **Queues**: A queue is a place to organize and store activities and cases that are waiting to be processed.
- **Service-level agreements (SLAs)**: SLAs are a way to track and define what should happen when a case is opened, like how long it should take to take any action.
- **Record creation and update rules**: Record creation and update rules can be applied to different activity types to automatically create Dynamics 365 records.
- **Routing rules**: Routing rules are applied to cases to automatically route them to a specific queue or user internally or externally.
- **Business process flows**: A business process flow represents a guided process that has different stages and steps that are used to resolve a specific item, like a case.

The core capabilities above are accessible to staff through specific interfaces and applications of Dynamics 365, and on a variety of devices, so they can work with focus and full fidelity.

Case Management Components


Personalized Service
Contact
Internet
Phone
SMS
Chat
Social
In Person
Devices
Communication Channels
Telephony CIF
In Store Surface
Website
Self-Service PowerApps Portals
Chat Chat for Dynamics 365
SMS
Video CIF
Co-Browse CIF
Email Microsoft 365
Social CIF
IoT Azure IoT Central
BOT Dynamics 365 Virtual Agent for Case Management
DYNAMICS 365 CASE MANAGEMENT
Agent Empowerment
Agent
Agent Interface Omnichannel for Case Management, Unified Service Desk, or Contact & Service Hub
Cases
Contracts/Agreements
SLA Management
Dashboards
Knowledge Base
Next Best Action
Scheduling and Dispatch Dynamics 365 Field Service
Supervisor
Case Management Dashboards
Queue Management Routing/Assignment
Insights
Agent & Service Performance Customer Service Insights
Customer 360 Contact Insights

For others in an organization who have partial involvement and responsibility in case management activities, Microsoft Dynamics 365 case management can be viewed, interacted with, and analyzed in different ways that are more familiar such as having case management tasks and information presented to users via Microsoft Teams, or embedded in other applications. This flexibility can increase staff engagement with case management, improve quality of process and data accuracy, and adherence to factors such as data controls and collaboration.

Not forcing a one-size-fits-all approach to software means that people and processes can be better applied so that colleagues and teams are able to:

- Track files through cases
- Record all interactions related to a case
- Share information in the knowledge base
- Use unified routing to efficiently route work items
- Manage conversations across channels, including voice
- Use AI-driven embedded insights and analytics to improve operations
- Collaborate with experts in Microsoft Teams across the organization
- Create and track service levels through service-level agreements (SLAs)
- Define service terms through entitlements
- Manage performance and productivity through reports and dashboards
- Create and schedule services associated with a case
- Participate in chats relevant to the case/matter and for these to be captured
- Get rich visualization of case management performance and activity and, through combining with AI-driven functions, be able to get increased levels of automation and actionable insight

As your organization tracks metrics on the types of cases that come in, you can create queues, routing rules, and flows to triage cases in the most efficient way.

Microsoft's case management capabilities allow you to unify cross-functional support for the entire journey in a comprehensive and scalable solution with our services portfolio in Dynamics 365. Powered by the security of Azure, organizations can create personalized engagements, elevate employee productivity, and optimize operations resulting in a transformed arc of law enforcement that strengthens intelligence officers and justice.

E Collaboration and Data Sharing with Azure Confidential Consortium Framework

In our solution, each Quantexa customer (FIU/law enforcement agency) has its own instance of the platform in their Azure subscription, where all the available data of the customer is ingested through HDFS (for batch processing) and Kafka (for real-time processing). FIUs only have a partial view of entities, accounts, and transactions based on the data available in their local instance. Having the ability to connect with instances from other countries would greatly improve the effectiveness of international investigations, providing an effective way to stop the financing of organized crime. However, there are many obstacles to this goal.

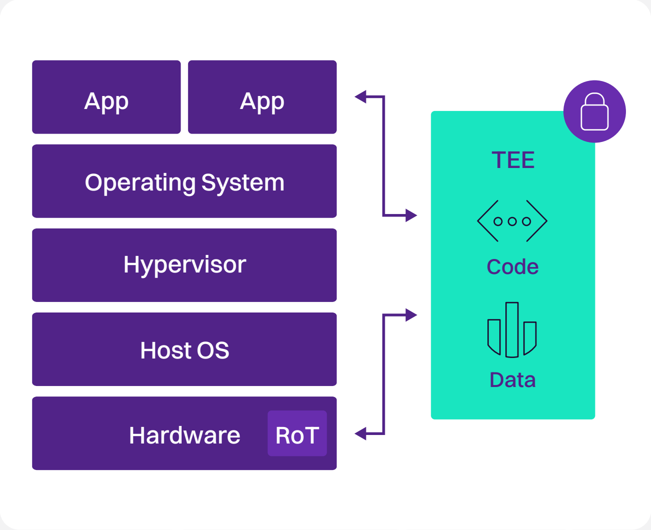
Even between countries with established cooperation agreements, such as within the European Union, national data privacy regulations (and other national policies) make any direct access to datasets from other countries highly unlikely without an established official request and review process. In this section, we will discuss the use of Confidential Computing technology to enable data augmentation (such as Entity Resolution, network visualization, etc.) across Quantexa instances, without having to directly expose national datasets to international investigators.

Confidential Computing protects sensitive data during processing by providing:

i. Minimal hardware, software, and operational Trusted Computing bases (TCBs) for their sensitive workloads.
ii. Technical enforcement (verifiable not just by the tenant, but also independently by their own regulators and auditors), rather than just business policies.
iii. Transparency about the guarantees, residual risks, and mitigations that they get.

Confidential Computing allows tenants to exercise full control over the Trusted Computing base (TCB) used to run their cloud workloads:

Application Enclave as an Example of TEE

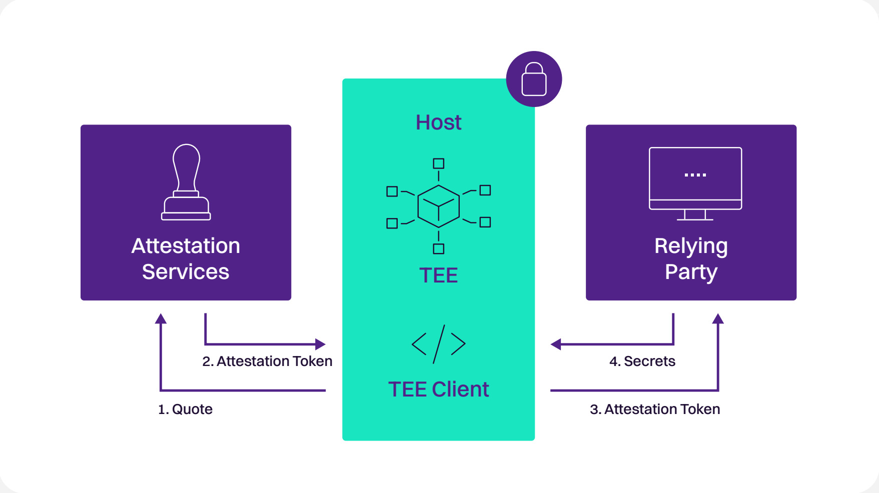


Confidential Computing allows tenants to precisely define all the hardware and software that has access to their workloads (data and code), and it provides the technical mechanisms to verifiably enforce this guarantee. In short, tenants retain full control over their secrets.

For instance, the Intel SGX implementation of TEEs isolates a single process on the system from all other software components (including other applications, the guest OS, the hypervisor, and host OS).

Confidential Computing can render workloads opaque to the cloud provider because tenants can use this precise level of control to prevent access to their secrets by the hypervisor and other cloud hosting infrastructure. This prevents attacks from the cloud fabric and its operators and complements the more traditional security goal of protecting the cloud fabric from potentially malicious tenants.

This level of control goes beyond preventing accesses by the cloud hosting infrastructure: it allows a tenant to specify that a particular set of secrets can only be processed by a specific code module. This capability is powerful, because it can be used to design resilient systems with reduced attack surfaces. Precise control over the trust placed in confidential cloud services enables useful scenarios between multiple parties that do not fully trust one another. For example, a tenant may in turn deploy a service with strong privacy assurances for its own customers; and competing parties may jointly configure and run a multi-party cloud computation (such as data analytics or machine-learning) with strong technical guarantees about the use of their pooled data.



Remote attestation is the main technical mechanism to allow a remote entity to verify the code running in a TEE, by validating the signed measurements of the code computed by the root of trust on the CPU, using a key that is unique to the hardware.

With Confidential Computing, customers can move the data to Azure knowing that it is safe from the following threats not only at rest, but also when in use:

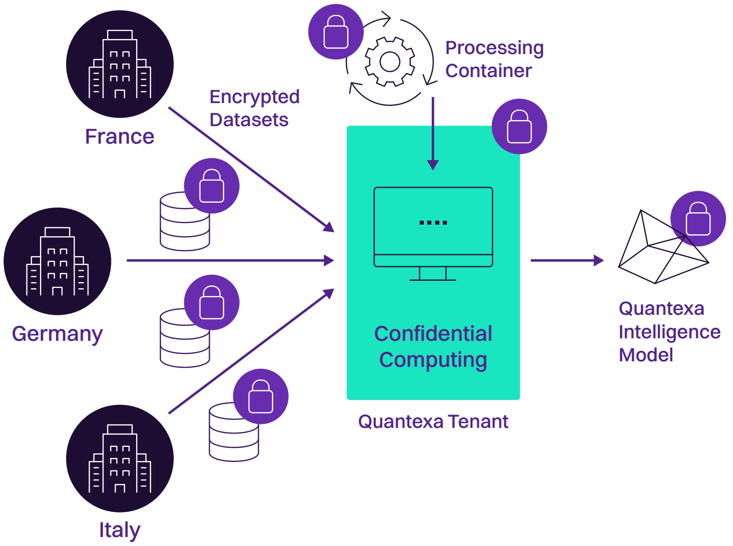
- Malicious insiders with admin privileges on guest or host virtual machines
- Hackers and malware that exploit bugs in OS, application, or hypervisor
- Third parties accessing it without their consent

Confidential Computing ensures that when data is “in the clear,” which is required for efficient processing, the data is protected inside a Trusted Execution Environment (TEE - also known as an “enclave”). TEEs ensure there is no way to view data or the operations inside from the outside. They even ensure that only authorized, unaltered code is permitted to access data.

The TEE enforces these protections throughout the execution of code within it. With Microsoft Azure Confidential Computing (ACC), we can minimize customers’ operational trust in Azure and bring the risk and liability profile closer to on-premises, but with even better standardized security controls.

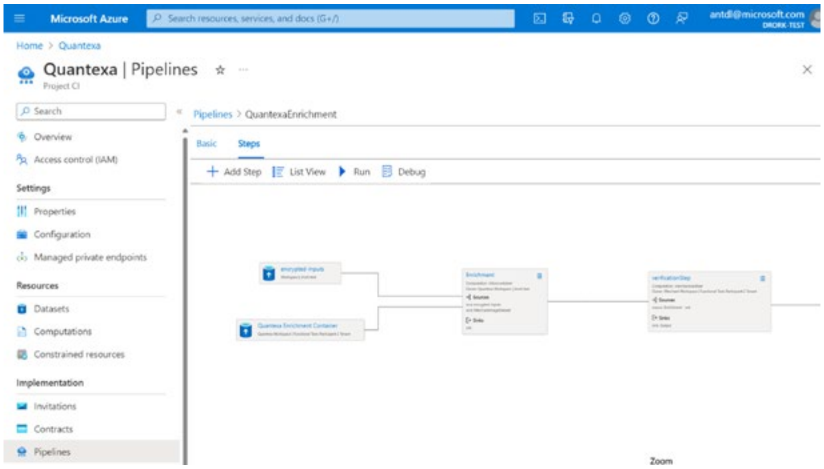
Enabling collaboration across Quantexa instances: In order to enable this collaboration scenario, we break down the collaboration process into two steps. First, we need to run the offline/batched analysis of Quantexa over the aggregation of the national datasets. However, the goal is to run this task in a Confidential Computing environment that guarantees that neither Quantexa, nor Microsoft, nor any of the contributing participants can have access to the datasets.

Each dataset is encrypted with a key that is only accessible to the hardware-protected/Trusted Execution Environment that will execute the processing code. The output of this process includes the scoring models, extracted features, and graphs that are consumed by the Decision Intelligence platform previously described.

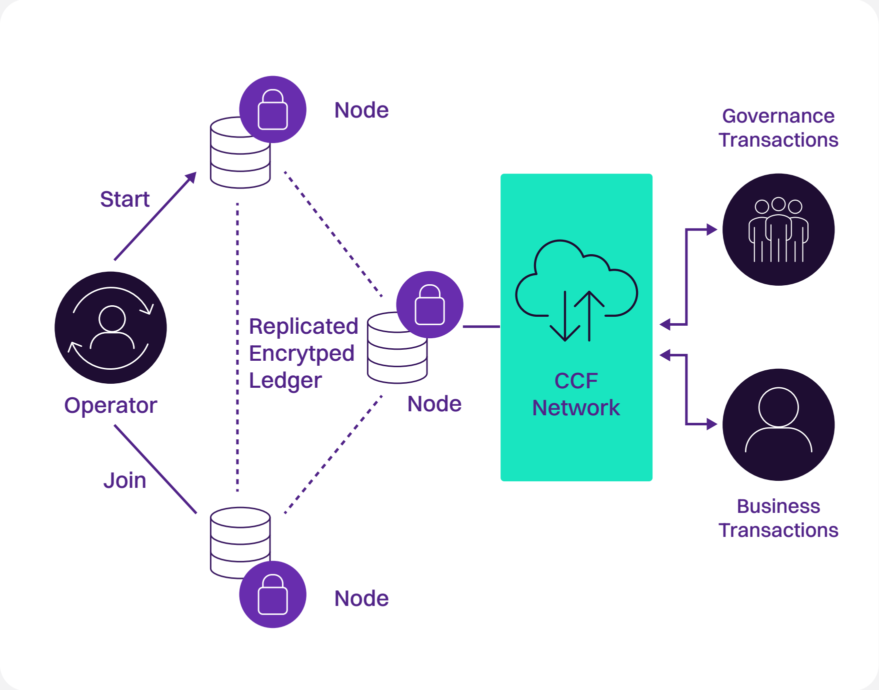


This output of the analysis over the joint dataset is likely to contain Personally Identifiable Information (PII) data from each participant's dataset. Therefore, it must also be encrypted using hardware-protected keys. In addition, each participant may have to sanitize the data before the aggregated processing begins to comply with their local regulations. For instance, the datasets may be anonymized or pseudonymized, though this process may still link international pseudonymous entities with identified national entities during investigations, which can be followed up on with an official request for information on the person of interest, account, or transaction.

The experience for defining the augmentation pipeline and data-sharing agreements between Quantexa and national FIUs can use the Azure Oakes platform, which depends on the Azure Key Vault Managed HSM offering to guarantee that the encryption keys can only be exported to the authorized augmentation container.

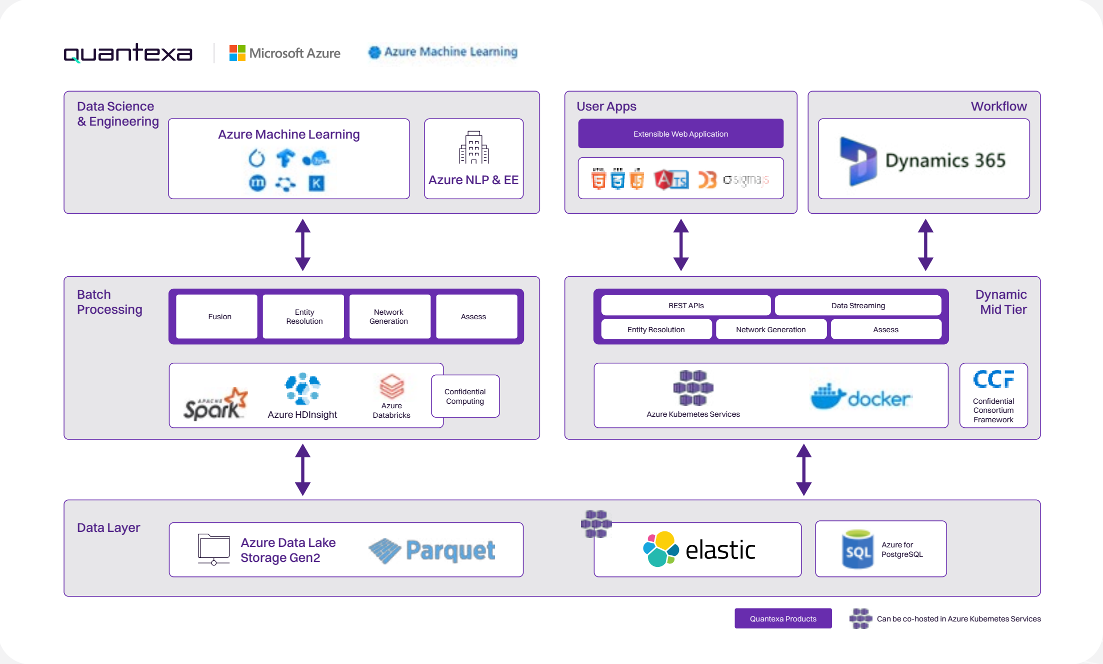


Real-time use of aggregated intelligence: we propose to expose query access to the aggregate intelligence model through a confidential consortium framework (CCF). CCF provides a distributed governance model where the FIUs that participate in the data collaboration form a consortium. They define a constitution, which sets the condition for participation into the system. The business transaction logic of this application enables query access to the aggregate model. CCF runs on Confidential Computing nodes, which guarantee that the confidentiality of the model, as well as the confidentiality of every query, can remain protected with respect to other participants.

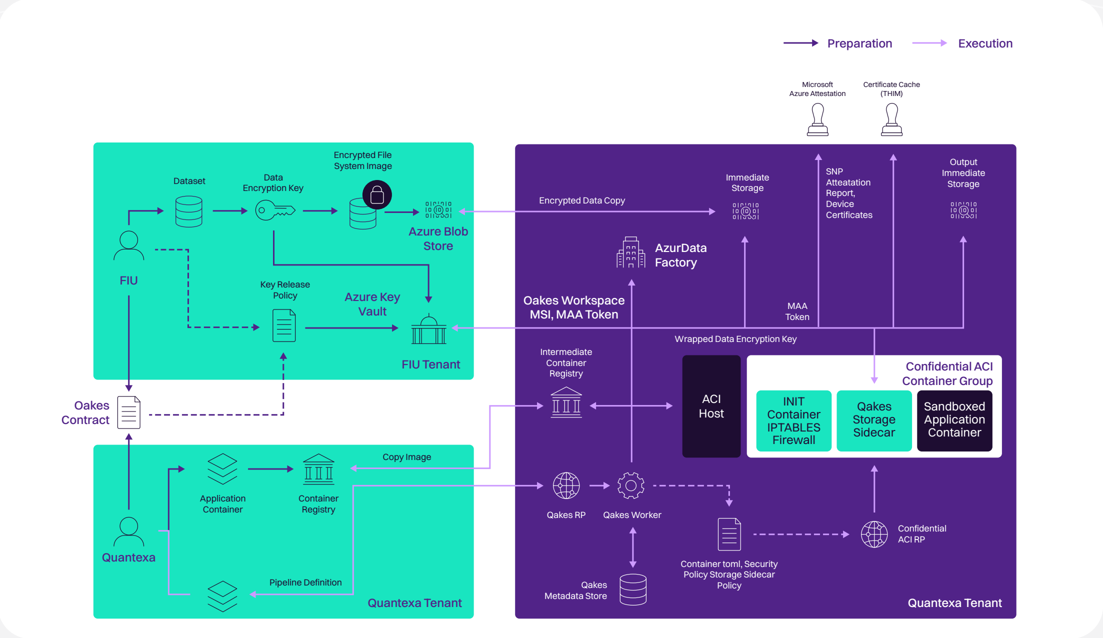


Additionally, CCF produces a tamper-proof ledger of all the queries and evidence processed by the system, which can be used to audit the use of the cross-border information by national investigators and establishes a verifiable "chain of custody" for processing the evidence extracted from cross-border data.

Functional and technical architecture



The full capabilities of the solution combine technology stacks of Quantexa and Microsoft. A reference architecture provides a visual overview.



The solution can be realized on a technical architecture, which ensures safe and secure data sharing and collaboration in a confidential compute environment with encrypted and protected execution.

quantexa | Microsoft Azure

Quantexa built a solution with Microsoft technology to effectively run AML/CFT investigations, collaborate, and manage cross-authority communication.

As the money laundering methods and technologies used by criminals become more sophisticated, AML authorities must keep up with their digital capabilities. Based on services from data, AI, and the Cloud, authorities can combat AML/CFT. One of the big hurdles to overcome is to have the possibility of data sharing in a data-privacy-compliant manner. Confidential Computing provides the technology to manage this requirement.

To request more information and schedule a demonstration, click here.

Jerome Bryssinck, Director of Government Solutions, Quantexa EMEA
Karl Heinz Krug, Industry Advisor Public Finance, Microsoft Western Europe

Quantexa.com | azure.microsoft.com