

AI Client Data Reconciliation for Enhanced Financial Crime Detection

Introduction

As financial institutions (FIs) seek to enhance financial crime detection, many are exploring AI-backed solutions for transaction monitoring and customer due diligence. However, the effectiveness of these solutions hinges on the accuracy and consistency of client data. At most FIs, client data is scattered across multiple systems and stored in unstructured formats, which is a significant challenge to data quality.

To maximize the benefits of AI-driven tools, it is crucial to establish a high-quality data foundation by consolidating client

information into a single source of truth. Without this, even the most powerful models risk being compromised by incomplete or inconsistent data.

Traditionally, FIs have employed entity resolution software with human-in-the-loop workflows to clean and match data. This approach typically involves matching multiple keys, such as names, postcodes, or social security numbers, to reconcile disparate data sets. But these methods struggle with poor quality or unstructured data, such as information from negative media reports or blogs, which limits their effectiveness.

In this whitepaper we propose an alternative approach to client data reconciliation that leverages embeddings and graphs for more effective matching, especially under challenging data conditions. It begins by defining client data reconciliation and its importance for financial crime detection, then explores the current challenges and shortcomings of traditional methods. Finally, it then outlines new techniques for matching and demonstrates its real-life application through a use case. ➔

Client data reconciliation in financial crime detection

With global data volumes increasing by 90% in the past two years alone, the potential for leveraging data in financial crime detection grows every day. Organizations now strive to enrich their understanding of customers, products, and markets. Yet the sheer volume of data—often stored in various formats across disparate systems and frequently unstructured—has made it increasingly difficult to extract meaningful insights, turning many data lakes into data swamps where valuable information is lost in a sea of noise.

Data management is the foundation for insights

Master data management (MDM) is the umbrella term for many different disciplines relating to the management of master data, including its governance, accuracy, and stewardship.

A solid master data management practice systematically collects, matches and maintains data.

The goal is to avoid multiple mastered versions of the data concept and build a single client record across systems, improving the ability to automate compliance reporting and enhancing decision-making.

Most MDM solutions aim to create a unified view of key entities such as customers, suppliers, and products, often stored as a Golden Master Record or a System of Record. For organizations of all sizes, effective client data management is critical for extracting greater value from data assets, reducing operational costs, and identifying new business opportunities.

The role of data reconciliation

An essential component of robust data management is data reconciliation—the process of comparing and matching data from different sources to ensure consistency and accuracy. This process involves deduplication, resolution, and connection of data, ensuring it accurately represents a single real-world entity, be it an organization, individual, product, or more conceptual abstractions such as a

topic or event. Data reconciliation can also reveal the connections across different entities, such as parent/child relationships, shareholders, ultimate beneficial owners (UBOs), corporate hierarchies, and links to relevant news events.

A prerequisite for effective financial crime detection

For financial institutions, one of the most critical applications of data reconciliation is creating a single, accurate view of a client—a golden record.

This is especially important in the context of anti-financial crime (AFC) detection and prevention. Without reliable and comprehensive client data, even the most advanced AFC solutions will struggle to deliver meaningful results.

Current AFC systems often generate an overwhelming number of alerts, with over 90% of them being false positives as determined after thorough investigation. To mitigate this issue, FIs are increasingly adopting machine learning and generative AI-based solutions for faster and more accurate decision-making. But these advanced technologies rely heavily on high quality client data within a rich and complete client context. Incomplete or contradictory data sets, such as those with duplicate records or inconsistent attributes, can mislead models, leading to irrelevant or incorrect results.

The consequences of poor client data reconciliation

In the fight against financial crime, de-prioritizing client data reconciliation can have serious consequences. It is not uncommon, for instance, for an organization's activities to be split across multiple, independent entities. Such fragmentation can falsely flag a legitimate client as suspicious or, conversely, allow a criminal to evade detection by flying under the radar.

Fragmented data enabling criminal evasion

Consider the case of a criminal enterprise deliberately structuring its transactions to exploit gaps in data reconciliation. The organization could be involved in money

laundering, using multiple shell companies to move illicit funds across borders. It may run two companies which regularly transfer funds between them, but without accurate reconciliation, these companies can be treated as unrelated entities and the transactions dismissed as routine business operations.

But if properly reconciled, it will become clear that both companies share the same ultimate beneficial owner (UBO) who, when linked to media reports or blogs, should have been known to be a bad actor. Recognizing this connection is critical, since it could reveal a larger pattern of illicit activity and prompt a deeper investigation. Without reconciliation, the criminal strategy to fly under the radar by exploiting data fragmentation would likely succeed, allowing illicit funds to move undetected.

Reconciliation as a core strategy

To improve financial crime detection, FIs must prioritize client data reconciliation as a foundational step. Only with accurate, consolidated client data can advanced detection systems operate effectively, uncovering hidden networks and detecting financial crime. In the sections that follow, we will explore innovative approaches to client data reconciliation that deliver superior results even under challenging data conditions.

Where do traditional approaches to client data reconciliation work?

Traditional entity resolution systems, primarily built around the use of compound keys or key matching techniques, have long served as the backbone of data management in financial institutions (FIs). These systems excel in scenarios where data attributes are well-defined, stable, and structured, making them particularly effective in environments that deal with high volumes of consistently formatted data, such as customer due diligence (CDD), transaction monitoring, and fraud detection.

The legacy and strengths of traditional systems

The legacy of traditional entity resolution systems is deeply rooted in their ability to handle structured data reliably and consistently. Over decades, these systems have proven to be robust and efficient, especially when applied to use cases where entities can be uniquely identified by a combination of fixed attributes, including social security number, account number, and date of birth.

The expertise required in traditional systems

A significant strength of traditional entity resolution systems is their reliance on expertly crafted parsers and cleansers for common attributes such as phone numbers, addresses, and emails.

The development and maintenance of these components require deep domain knowledge because they must account for the myriad ways in which such data can be formatted, entered, and updated. The accuracy of these systems is heavily dependent on the quality of these parsers and cleansers, which must be continually refined to handle new data formats and inconsistencies.

The requirement for expertise extends to the maintenance of these systems, where the consistency and reliability of entity resolution hinge on the monitoring and maintenance of the resolution rulesets. This deep understanding of data patterns, along with the application of well-established algorithms, has made

traditional systems a trusted solution at financial institutions for many years.

Limitations and the need for evolution

Yet despite their strengths, traditional entity resolution systems face significant challenges when dealing with certain data complexities. The proliferation of unstructured data, the variety of data sources, and the increasing prevalence of data quality issues reveal the limitations of these systems.

When data does not fit into predefined formats, or when identifiers are missing or inconsistent, traditional systems struggle to maintain the same level of accuracy and efficiency. Moreover, the rigid nature of traditional systems, which rely heavily on hardcoded rules, makes them difficult to adapt to new data requirements or to leverage more nuanced contextual information.

As financial institutions increasingly encounter data that is unstructured, dynamic, or sourced from diverse channels, there is a growing need for more flexible and intelligent approaches to entity resolution.

Bridging the gap: The need for advanced resolution systems

Newer systems offer a more general approach that is not limited to a few common attributes and adapt to new and meaningful challenges such as extracting and matching information stored in documentation, matching events in the news, and providing a more holistic and accurate view of the data.

By integrating the strengths of traditional systems with the capabilities of advanced AI-driven frameworks, financial institutions can achieve a more comprehensive and effective approach to entity resolution, one that is better suited to the complexities of today's data landscape.



Where does AI make the difference?

The science and art of entity resolution: Learning from how we think

One of the biggest challenges in entity resolution is the scarcity of human-labeled data. Practitioners are eager for large data sets filled with real and tricky resolution problems. Yet when applying entity resolution to a new, unlabeled data set, the goal is to replicate how humans would think through the problem. This is where AI-driven solutions hold promise. While there is no silver bullet, AI provides an opportunity to enable more flexible, user-driven solutions.

These new frameworks should empower users to fine-tune their extraction, transformation, and comparison processes without forcing them to rigidly conform to predefined schemas. Instead of merely converting unstructured data into structured formats and trying to fit it into existing models, the new generation of entity resolution tools should allow business users to adjust models to their specific needs, disambiguating knowledge across various entity types, whether conceptual abstractions like events and diseases, or real-world entities like companies, products, and beyond.

Configurability, explainability, and reliability: Essential components

Despite the promise of AI, it is critical that users remain in control of the process. AI-driven models should be highly configurable, allowing users to fine-tune their approaches while ensuring explainability and transparency.

Hallucinations, or unsupported conclusions, are not acceptable in the AFC domain, especially when legal and regulatory scrutiny requires lineage, reproducibility, and human review. Tooling must enable human feedback, and monitor models must be used throughout the system.

Vectors in entity resolution

A new technique with good potential is vector-based entity resolution. It allows users to perform similarity searches for entities (organizations, individuals, concepts, topics, events, etc.) by comparing the distinguishing features that show semantic and/or contextual similarity to a particular entity. This is particularly helpful when data is incomplete, inconsistent or unstructured.

A simplified explanation of vector-based entity resolutions:

- 1. Representation of clients as vectors:** Firstly, clients are broken down into dimensions based on their features: age, gender, occupation, etc. The dimensions are then represented in (high-dimensional) mathematical vectors for each customer. These vectors capture the semantic meaning of each attribute and not just its plain value (it understands that a consultant is very similar to an advisor).
- 2. Similarity calculation of two or more client data sets:** The two vector-values can be used to compare two entities by calculating how (mathematically) close the two entities are. Cosine similarities or Euclidean distance, for instance, can break down the similarity (somewhere between 0 and 1).
- 3. Threshold setting to define a match:** Users can then define a minimum similarity score for it to be considered a match. For example, only scores above 0.9 might be considered similar enough to be grouped into one consolidated client file.
- 4. Validation of true matches and edge cases:** Until enough trust has been established to let the machine decide independently, a human being should review matches before grouping them, and sample-based validations should be performed over time. This helps refine the model and ensure accuracy, especially for outliers.

Compared to traditional entity resolution software, vector-based entity resolution software shows improvements in:

- **Accuracy and reliability:** Superior matching performance when dealing with unstructured and imperfect data to avoid incorrect matches and extensive human review.
- **Connectivity:** Showing the entity in its context helps visualize how everything is related to everything.
- **Lineage:** Data lineage is essential to demonstrate how the data reconciliation solution comes to a decision and how the data is linked to the raw information.

Vendors following this approach should aim to solve some technical challenges as well:

- **Scalability:** Due to the high number of potential connections, the solution must be able to cope with very high volumes of data.
- **Ease of use:** Product vs. project. Vector-based entity resolution software must overcome some of the challenges of key-based solutions, which are often highly customised.

Use case: The impact of entity resolution on the monitoring of transaction flows

There are many ways in which large FIs can leverage advanced entity resolution to make better use of their data. Below we provide an example of how AI-enabled entity resolution can be used within a project setup. For reasons of confidentiality, certain specifics cannot be revealed.

Introduction

In a recent project, a large bank used entity resolution to support its transaction monitoring efforts. The client had identified a need to track how money moves across different accounts and entities.

The Knights Analytics solution was selected so the client could significantly improve the identification and association of financial transactions across both internal customer profiles (referred to as "Profiles") and external transaction systems like Swift.

The ambition was to provide a clearer, more comprehensive understanding of transaction flows and resolve accounts belonging to the same real-world entity.

Understanding the baseline: Before advanced entity resolution

- **Profiles and Swift account matching:**

The data set initially contained roughly 350 thousand Profiles with accounts that explicitly matched Swift accounts (IBANs). Some entities had multiple associated accounts linking them to roughly 400 thousand Swift accounts.

- **Total number of transactions:** The data contained approximately 1.2 billion transactions associated with 34.5m Swift accounts.

The Knights Analytics impact: Enhanced entity resolution

Knights Analytics introduced an advanced entity resolution technique that built on this baseline by identifying additional relationships between Profiles and Swift accounts.

- **New Profiles linked to Swift:** The client identified previously unknown links between Profiles and Swift accounts, marking a 5.3% increase in entity coverage.
- **Expanded account mapping:** The total number of Swift accounts linked to Profiles grew by 23.1%.

Impact on transaction monitoring

The improvements in entity resolution directly impact the ability to understand transaction flows:

- **Resolved entities in transactions:** After resolution, 64.1% of all Swift transactions were associated with an entity which had been resolved to at least one other account, significantly reducing fragmentation in the flow of transactions.
- **Transaction value impact:** 72.3% of the total euro value of transactions involved resolved entities, indicating that the enhanced resolution disproportionately affects entities involved in larger transaction volumes.

Duplicate detection in Profiles data

Entity resolution also identified potential duplicates within the Profiles data:

- **Duplicate resolution:** Ten percent of the 350 thousand Profiles were marked as highly similar or matched to another Profile, suggesting that there might have been more duplicates in the internal Profiles than was expected and client data integrity could be improved.



Overall resolution outcomes

Across the total of all Profiles and Swift data, the new entity resolution process worked on 34.5 million unresolved entities (13.6 million of which from Profiles). Of these, 3.4 million entities were resolved down to 1.3 million, while the rest remained single entities. It would have been an enormous task to work this out using traditional techniques.

Key takeaways

1. Enhanced account connectivity:

The advanced entity resolution process significantly increased the number of Swift accounts associated with Profiles.

2. Significant transaction impact:

A large share of euro value transactions was found to be tied to resolved entities, underscoring that the entities identified by this process are disproportionately responsible for moving higher transaction volumes.

3. Potential Profiles data duplication:

While a small percentage of Profiles were identified as duplicates, this finding proved to be valuable for further verification since the Profiles data had generally been expected to be duplication-free.

Solutioning

After the initial proposal and POC phase, Knights Analytics supports the project bank and provides their software, documentation and best practices support. The bank's technology and cloud teams are responsible for deploying the software within their private cloud environment, while the business is responsible for data sourcing and creating the business rules.

Deloitte supports the project, working with the bank on mapping data sets and working with the business to create the business logic and data scenarios. Next, Deloitte works with the client project team on iterating merging logic configurations and supporting build comparison reviews. Feedback collected through an active learning process is used to iteratively improve the matching logic. Lastly, Deloitte helps with overall architecture and upstream and downstream system

integration. During the project Deloitte supports the client with workshops, planning and logistics, staffing of resources, deliverables tracking, and stakeholder management.

This is done by a small team of experts from Deloitte, ensuring the support cost is kept to a minimum and the wider project can be run by the bank itself. It takes only a few weeks to produce the first results and the solution goes into UAT within six months from contract signature.

entity resolution can improve the efficiency and accuracy of transaction monitoring processes. There are many other use cases for the software. At other banks, the Knights Analytics software is used for client data reconciliation, cleaning up CRM data, data migration, matching data from documents to internal client data, etc.

Contact us to find out if Knights Analytics and Deloitte can help you with your data challenge.



Learn more



Martin Hirtreiter

Deloitte | Partner
Anti-Financial Crime Advisory
mhirtreiter@deloitte.de



Dr. Robert Schmuck

Deloitte | Director
Anti-Financial Crime Advisory
rschmuck@deloitte.de



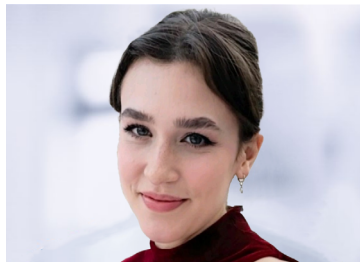
Özlem Tan-Tayik

Deloitte | Senior Consultant
Anti-Financial Crime Advisory
oetan@deloitte.de



Mazen Al-Yarimi

Deloitte | Senior Consultant
Anti-Financial Crime Advisory
mal-yarimi@deloitte.de



Alex Ridden

Knights Analytics | CEO

alex@knightsanalytics.com



Edo de Vries Robbé

Knights Analytics | CCO

edo@knightsanalytics.com

Deloitte.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited (DTTL), its global network of member firms, and their related entities (collectively, the "Deloitte organization"). DTTL (also referred to as "Deloitte Global") and each of its member firms and related entities are legally separate and independent entities, which cannot obligate or bind each other in respect of third parties. DTTL and each DTTL member firm and related entity is liable only for its own acts and omissions, and not those of each other. DTTL does not provide services to clients. Please see www.deloitte.com/de/UeberUns to learn more.

Deloitte provides industry-leading audit and assurance, tax and legal, consulting, financial advisory, and risk advisory services to nearly 90% of the Fortune Global 500® and thousands of private companies. Legal advisory services in Germany are provided by Deloitte Legal. Our people deliver measurable and lasting results that help reinforce public trust in capital markets, enable clients to transform and thrive, and lead the way toward a stronger economy, a more equitable society and a sustainable world. Building on its 175-plus year history, Deloitte spans more than 150 countries and territories. Learn how Deloitte's approximately 457,000 people worldwide make an impact that matters at www.deloitte.com/de.

This communication contains general information only, and none of Deloitte GmbH Wirtschaftsprüfungsgesellschaft or Deloitte Touche Tohmatsu Limited (DTTL), its global network of member firms or their related entities (collectively, the "Deloitte organization") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.