Internal Reference – Unified CDD Schema Challenges & BA Action Plan

# Background:

As part of the Quantexa integration project, we initially began mapping CDD source systems individually into the landing layer. However, Quantexa's latest requirement involves designing a unified schema that consolidates all CDD data into a single structure. This introduces several complexities, especially due to fragmented data and inconsistent field representations across systems.

# Challenges and Corresponding BA Action Items:

## 1. Fragmented Data Across Tables

🔹 Challenge: Information related to a single customer is spread across multiple tables in the same system (T1 level).

* ✅ Action Items:
* - Identify primary and supporting tables in each source system.
* - Map relationships and dependencies between tables.
* - Request ERDs or source documentation from data owners.

## 2. Column Name Misalignment Across Systems

🔹 Challenge: Same data element is named differently in different systems.

* ✅ Action Items:
* - Create a synonym dictionary or crosswalk sheet.
* - Align with Quantexa team on naming convention to be followed in unified schema.

## 3. Data Semantics Confusion

🔹 Challenge: Similar fields might have different business logic or scales (e.g., different risk score ranges).

* ✅ Action Items:
* - Capture field definitions and logic from each source.
* - Confirm expected standard format with Quantexa (e.g., unified risk score scale).

## 4. Duplicate or Conflicting Records

🔹 Challenge: Merging sources may introduce duplicate customer records or conflicting details.

* ✅ Action Items:
* - Check Quantexa's preferred deduplication logic.
* - Ensure traceability by storing source and timestamp.

## 5. Missing or Optional Fields

🔹 Challenge: Some systems may lack certain fields expected in the unified schema.

* ✅ Action Items:
* - Use MoSCoW prioritization per field.
* - Seek clarification from Quantexa whether null/defaults are allowed.

## 6. Transformation Overhead

🔹 Challenge: Fields may require standardization or data enrichment.

* ✅ Action Items:
* - Prepare transformation logic for each required field.
* - Validate with Quantexa and internal dev teams.

## 7. Scalability of the Unified Schema

🔹 Challenge: Schema needs to handle future additions without full redesign.

* ✅ Action Items:
* - Check if schema should be versioned or extensible.
* - Recommend semi-flexible formats (e.g., JSON with optional fields).

## 8. Tight Timelines vs Requirement Clarity

🔹 Challenge: Risk of misdevelopment due to partial clarity.

* ✅ Action Items:
* - Split requirement delivery in phases (Must-have vs. Nice-to-have).
* - Regularly update field-level closure tracker with decisions from Quantexa.

# Suggested Talking Points for Quantexa Meeting:

Hi team, as part of our current scope, we are now focusing on consolidating multiple CDD sources into a unified schema for your contextual scoring engine. We’ve identified a few challenges that we believe need collaborative resolution.

1. Many systems store information across multiple tables. We’ll need help validating which tables are primary vs. supporting.  
2. Field names and logic vary across systems. We'd like to propose a field crosswalk and confirm preferred standards.  
3. In case of duplicate values across systems, can you help us finalize source priority rules?  
4. What transformation rules does Quantexa expect for inputs like risk score, customer type, and watchlist flags?  
5. Lastly, how flexible should the schema be in terms of future additions—strict schema or schema-on-read model?

Once these questions are aligned, we’ll prepare a clean requirement document and pass it to developers for implementation.

Extended Meeting Speech – Business Analyst View (Updated)

Hi team, thank you for joining. I just wanted to take a few minutes from the business side to walk through where we stand right now with the CDD integration activity for Quantexa.  
  
As you all know, initially we started planning to onboard each CDD source system separately into the Quantexa landing layer – iCDD, D-Cube, Client Central, Veritas Watchlist, BvD and so on. But now, as per the latest requirement from the Quantexa core team, the approach has shifted to building a unified schema where all this data should come in a single, consolidated structure.  
  
So our current focus is to gather clear requirements from Quantexa on what exactly you’re expecting in the unified schema. The goal from our side is to avoid any assumptions and ensure the developers get a clear, actionable set of mappings.  
  
Let me walk you through some key pain-points we’re seeing from the business/data perspective:  
  
1️⃣ Multiple Tables Per System   
In many of these systems like iCDD or D-Cube, there’s not just one customer table — the information is actually spread across 3–4 related tables. So we need to know which ones are critical and which are supporting, and how you expect the data to be joined before loading into the unified schema.  
  
2️⃣ Column Name Differences   
Even for basic fields like customer name or risk score, each system calls it differently. One might say cust\_name, another says client\_fullname. And sometimes, both exist with slightly different data. So we need help finalizing preferred field names and if any transformation is needed.  
  
3️⃣ Conflicting Data Across Sources   
Multiple systems may have similar data – for example, both iCDD and Client Central have customer names or types. If there’s a conflict, we want to confirm which system should be treated as golden source for each field.  
  
4️⃣ MoSCoW Prioritization   
We understand that Quantexa will be the one to define the MoSCoW classification of fields — what is Must-have, Should-have, and Could-have — along with the list of unified columns they expect. We’ll wait for your guidance on this and help align source systems accordingly.  
  
5️⃣ Transformation Requirements   
Many fields are not directly usable — like country names, risk scores, watchlist flags. We want to confirm if Quantexa is expecting us to send these in a transformed format (e.g., ISO codes, Boolean flags, numeric risk scale) or if that logic will be handled post-ingestion.  
  
6️⃣ Scalability   
Is the unified schema expected to be strict and final, or should we design it with some flexibility in case more fields come in later? That will affect how we build the schema – whether it’s tight table-based or semi-flexible like JSON.  
  
Action Plan:   
What we are planning to do is consolidate all your expectations field-by-field, source-by-source — including transformations, priority, and any conflict logic. Once we get that clarity, we will pass it as a final requirement set to the developers.  
  
Timelines are tight, so instead of trying to complete the entire schema in one shot, we’re thinking of closing it in phases – for example, first get alignment on must-have fields, and later extend to should-have and derived fields.  
  
From today’s call, we would like to get clarity on:  
1. List of all source systems you expect to include in scoring  
2. Unified CDD fields that should be designed in the schema  
3. MoSCoW priority classification per field (from Quantexa)  
4. For duplicate fields – your preference for which system to use  
5. Expected format/transformation for key fields  
6. Any future extensibility you expect in the schema  
  
Once we get this clarity, we’ll clean up the unified schema spec and move it to development right away. This will help reduce rework and give you a reliable data set to start scoring logic.  
  
We’re happy to share the Excel trackers and schema drafts we’ve prepared so far, and we’re open to having follow-up working sessions with your technical team if needed.