

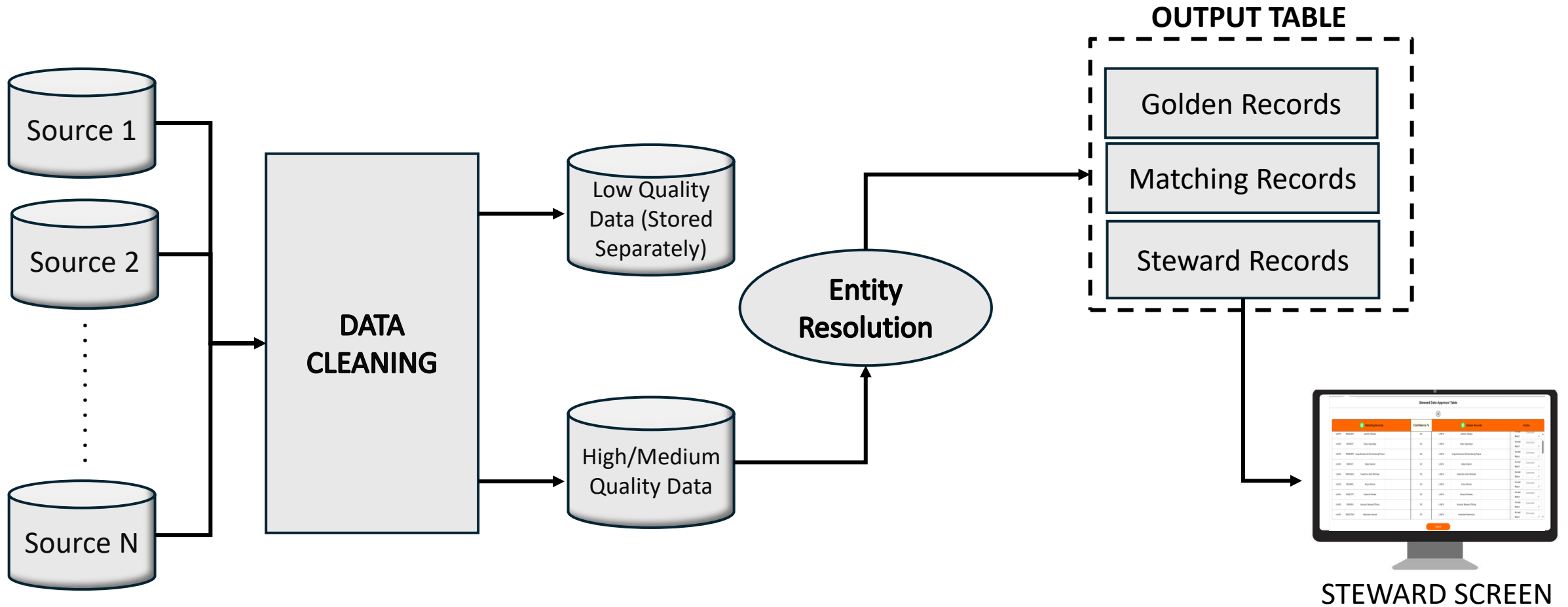
Master Data Management



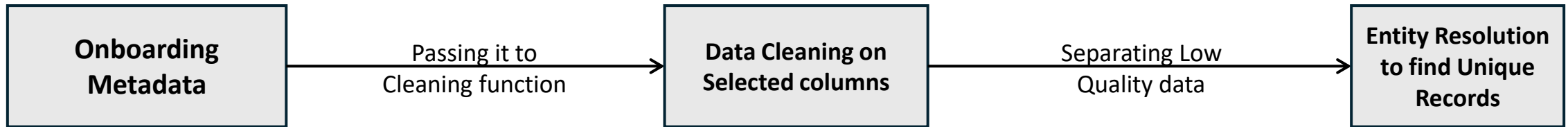
Objective



The solution aims to create a combined single-source table with unique customer information from all the data sources.



Methodology



Columns to be Extracted from each data source

- Unique ID
- Name
- Government IDs
- Phone Number
- Email Address
- Gender
- Postal Address
- Date of Birth

Creating clean columns from extracted columns

- Name
E.g., Mr. Robert Mitchel!!
-> Robert Mitchel
- Number
E.g., +234 98-903-39310
-> 09890339310
- Email
- Gender
- Address
- DOB

Based on Rulebook divide records in 3 categories.

- Golden Records
- Matching Records
- Steward Records

DATA QUALITY REPORT



COMPANY

Source

XXXXXXX

Column

All

Total Rows

12,286,989

Attribute

All

Date

All

Source Error Rate

41%

Data Completeness

Before Cleaning

43%

After Cleaning

38%

Data Validity

Before Cleaning

98.7%

After Cleaning

100%

Unclean Data

Field	Source Column	Total Count	Populated Values	Unique Values	Inaccuracy Count	Inaccuracy Rate	Data Type
Address	ADDRESS	12,286,989	659,250	204,678	1,272,807	10.36%	StringType()
Client ID	CLIENT_NO	12,286,989	12,286,962	12,286,294			IntegerType()
DOB	BIRTH_DT	12,286,989	11,767,210	31,666	0	0.00%	TimestampType()
Email	EMAIL_ADDR	12,286,989	304,554	209,165	504	0.00%	StringType()
Gender	SEX	12,286,989	6,967,511	72	595	0.00%	StringType()
Name	FIRST_NAME	12,286,989	12,269,247	1,696,211	837	0.01%	StringType()
Name	NAME	12,286,989	12,286,956	2,127,491	44,677	0.36%	StringType()
NIN	verified_nin	12,286,989	140,580	123,322			StringType()
Phone Number	MOBILE_NO	12,286,989	432,843	410,501	215,696	1.76%	StringType()
Phone Number	PHONE_NO	12,286,989	415,325	392,711	196,552	1.60%	StringType()
Phone Number	WORK_PHONE_NO	12,286,989	55,927	54,126	50,140	0.41%	StringType()
Total		135,156,879	57,586,365	17,536,237	1,781,808	1.32%	

Clean Data

Field	Total Rows	Populated Values	Unique Count	Inaccuracy Count	Data Type
DOB	12,478,864	11,951,344	31,666	0	DateType
NIN	12,478,864	198,756	134,506		IntegerType
Phone Number	12,478,864	713,908	475,363	0	IntegerType
Address	12,478,864	834,626	187,293	0	string
Email	12,478,864	326,022	197,561	0	string
Gender	12,478,864	7,151,695	3	0	string
Name	12,478,864	12,428,458	3,237,661	0	string
Total	87,352,048	33,604,809	4,264,053	0	

- Data Completeness :** The percentage of non-null values in the source compared to the total expected values.
- Data Validity:** The percentage of valid values in the source to the total values.
- Error Rate:** The proportion of records with invalid or missing values relative to the total number of records.

Entity Resolution Report

Source 1

Source 2

Source 3

Total Records Processed

7.09M

Golden Records
(Confidence = 0)

2.6M

Matching Records
(Confidence > 80)

1.9M

Steward Records
(Confidence 55 - 80)

2.6M

Reference

All

Golden ID

All

Confidence

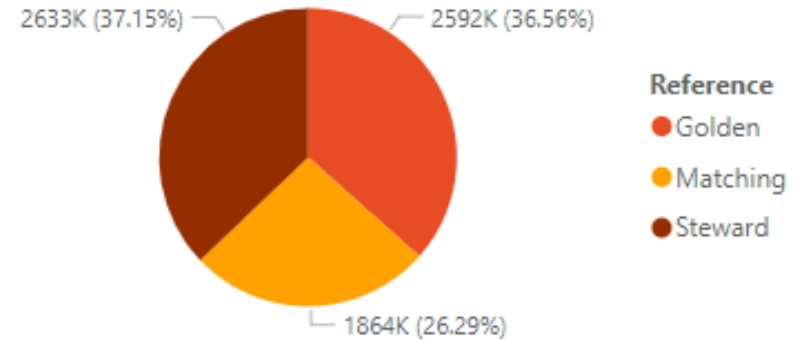
All

Golden ID Count

All

Golden_Record_Id	Data_Source	Reference	Confidence	Source_Client_Id	UIN
3040050	XXXXXXXX	Golden	0	10407130	
3040049	XXXXXXXX	Golden	0	6536424	
3040049	XXXXXXXX	Matching	100	6556859	
3040048	XXXXXXXX	Golden	0	2094430	
3040047	XXXXXXXX	Golden	0	3703346	
3040046	XXXXXXXX	Golden	0	3033884	
3040045	XXXXXXXX	Golden	0	5558708	
3040044	XXXXXXXX	Golden	0	9412355	
3040043	XXXXXXXX	Golden	0	4713847	
3040042	XXXXXXXX	Golden	0	10713047	
3040041	XXXXXXXX	Golden	0	1984831	
3040041	XXXXXXXX	Matching	100	2006623	
3040041	XXXXXXXX	Matching	100	2021637	
3040041	XXXXXXXX	Matching	100	2378297	
3040040	XXXXXXXX	Golden	0	5057362	

Total Record Distribution



Matching & Steward Records from Source 1

