

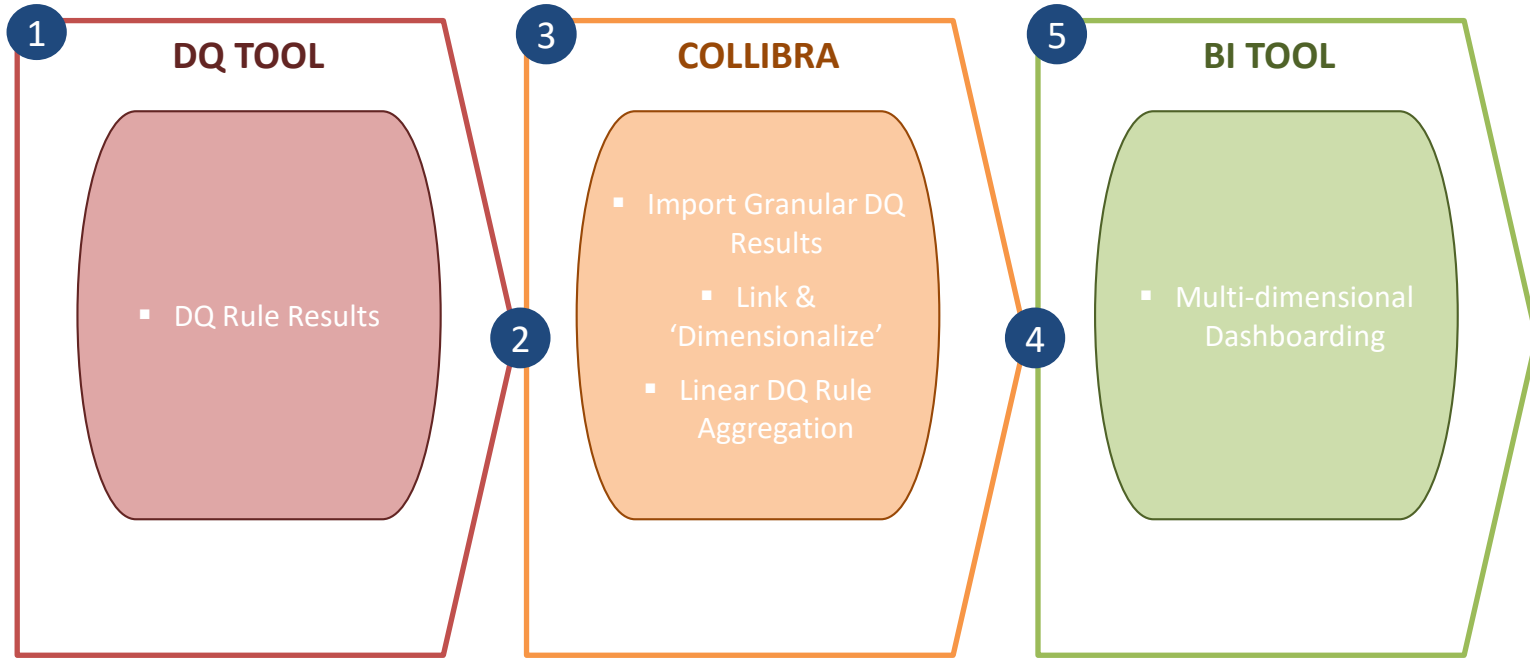


Data Quality in Collibra

Document Control

Date	Author	Version	Comments
18/04/2017	John Lenaghan	0.1	Internal Draft
21/04/2017	John Lenaghan, Courtney Small, Chris Pegler	1.0	Added BI Integration

The Data Quality Result Flow



Step 1 - DQ Tool

ID	First Name	Last Name	DOB	Product	Division	Legal Entity	DQ2 Result	DQ1 Result
1	Jon	Snow	12-Dec-12	Current	Retail Banking	Iron Bank	PASS	FAIL
2	Arya	Stark	NULL	Current	Retail Banking	Iron Bank	FAIL	N/A
3	Ned	Stark	NULL	Current	Retail Banking	Iron Bank	FAIL	N/A
4	Robb	Stark	03-May-68	Current	Retail Banking	Iron Bank	PASS	PASS
5	Jamie	Lannister	01-Oct-07	Current	Retail Banking	Iron Bank	PASS	FAIL
6	Sansa	Stark	17-Jan-83	Current	Retail Banking	Iron Bank	PASS	PASS
7	Daenerys	Targaryen	NULL	Current	Retail Banking	Gringotts	FAIL	N/A
8	Cersei	Lannister	07-Aug-91	Current	Retail Banking	Gringotts	PASS	PASS
9	Tyrion	Lannister	04-Apr-88	Joint	Retail Banking	Iron Bank	PASS	PASS
10	Gregor	Clegane	01-Jan-83	Joint	Retail Banking	Iron Bank	PASS	PASS
11	Sandor	Clegane	06-Nov-93	Joint	Retail Banking	Iron Bank	PASS	PASS
12	Ransay	Bolton	27-Dec-89	Joint	Retail Banking	Gringotts	PASS	PASS
13	Margaery	Tyrell	26-Dec-16	Student	Retail Banking	Iron Bank	PASS	FAIL
14	Petyr	Baelish	05-Nov-88	Student	Retail Banking	Iron Bank	PASS	
15	Bran	Stark	03-Jun-73	Student	Retail Banking	Iron Bank	PASS	PASS
16	Theon	Greyjoy	17-Jul-95	Student	Retail Banking	Gringotts	PASS	PASS
17	Stannis	Baratheon	01-May-05	Junior	Retail Banking	Iron Bank	PASS	N/A
18	Robert	Baratheon	09-Sep-05	Junior	Retail Banking	Iron Bank	PASS	N/A
19	Renly	Baratheon	19-Jun-05	Junior	Retail Banking	Iron Bank	PASS	N/A
20	Jorah	Mormont	03-Feb-05	Junior	Retail Banking	Iron Bank	PASS	N/A

DQ Rule 1 - Customers age cannot be under 16 to open a **Current Account** in **Iron Banks Retail** division

Total Rows: 4
 Passed Count: 2
 Failed Count: 2
 Passing Percent: 50%
 Threshold: 100%
 Result: Fail

DQ Rule 2 - Customers age cannot be under 16 to open a **Current Account** in **Gringotts Retail** division

Total Rows: 1
 Passed Count: 1
 Failed Count: 0
 Passing Percent: 100%
 Threshold: 100%
 Result: Pass

DQ Rule 3 - Customers date of birth cannot be null

Total Rows: 20
 Passed Count: 17
 Failed Count: 3
 Passing Percent: 85%
 Threshold: 100%
 Result: Fail

Step 2 – DGC Integration

Profiling Engine (PE)

1

This granularity of the rule being linked to assets at the lowest level to achieve rolling up and consolidation of figures across the model

Rule:IsUserEligibleToOpenAccount	Type of Rule
Condition: date_of_birth < Now – 18 years: evaluate Boolean true,false	Conformity
Table, Column : User, date_of_birth	
Total Rows, Total Evaluated Rows	

2

Rule	Total Rows	Rows Evaluated
IsUserEligibleToOpenAccount	10	1



Tabular Row Level Metrics

> CSV/XLS

Integration/Transformation Engine

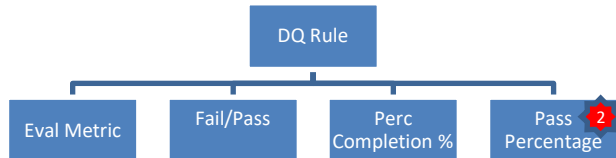
3

PE: IsUserEligibleToOpenAccount	DGC: IsUserEligibleToOpenAccount Total Rows: 10 Rows Evaluated: 1 Status: Failure ; Metric: 90% Pass;10% Fail
---------------------------------	--

This information is calculated using values persisted and exposed from Collibra to the Integration Engine

4

Collibra DGC



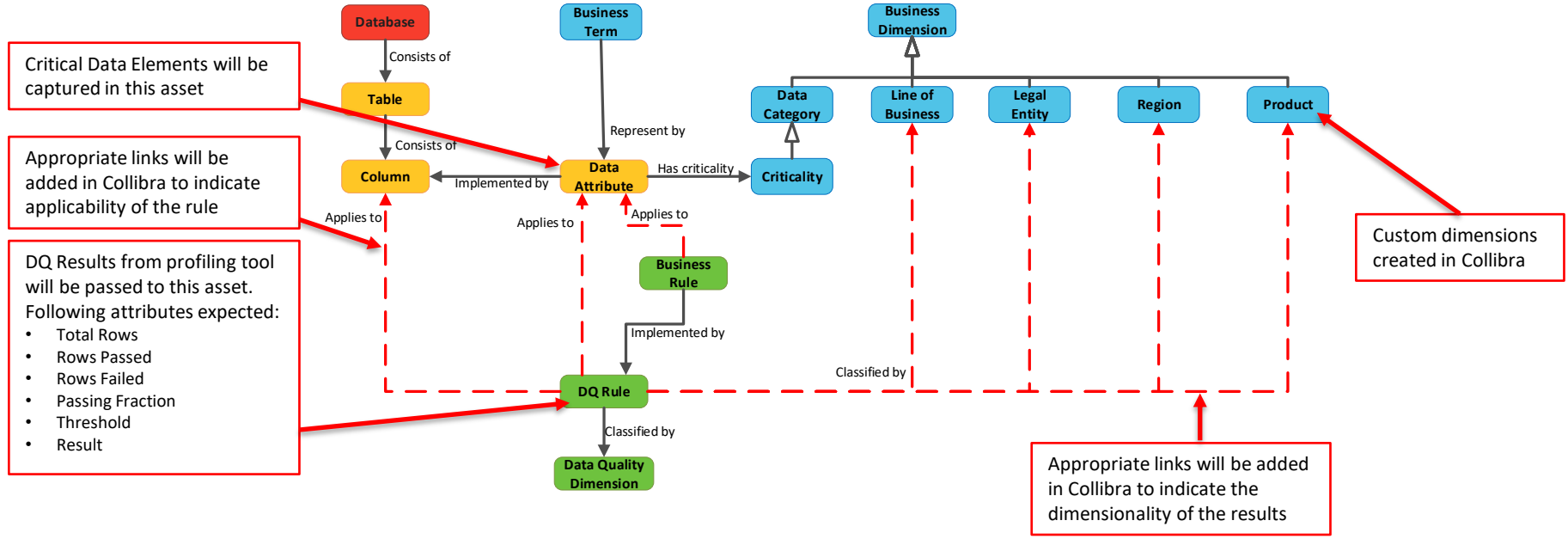
This Asset of type 'Data Quality Rule' in Collibra has attributes to persist the metrics, this could be reported via a View

5

DQ Rule	Status	Completion Quota
IsUserEligibleToOpenAccount	Fail	90%

The linkage/traceability of this solution is only as good as the depth of detail applied to the operating model and participants. But the basic principle is based on a

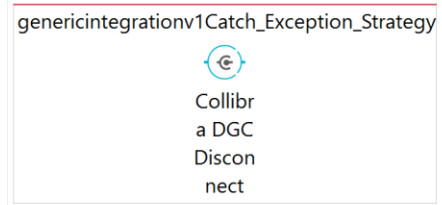
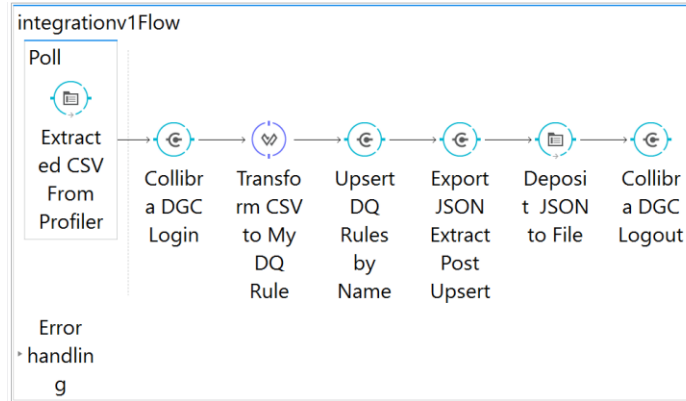
Step 3 - Collibra



Note: It may also be necessary to specify the dimensionality of a data attribute (CDE). This will depend on the community / domain structure organisation

Step 4 - BI Integration

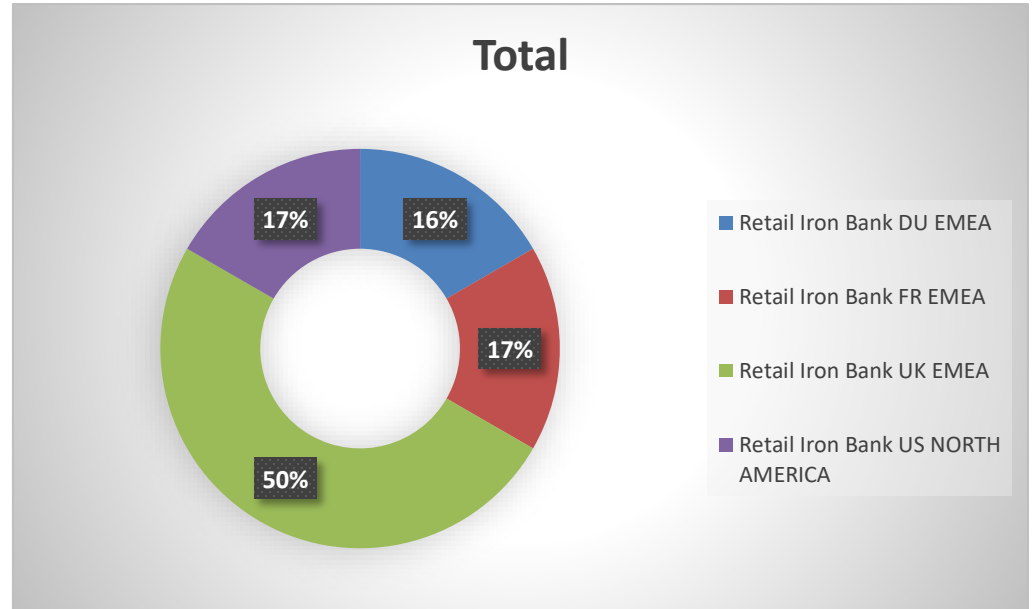
This is an example of the Flow required to integrate the Output from DQ Tool into Collibra Data Quality repository.



Message Flow Global Elements Configuration XML

Step 5 - BI Reporting

Row Labels	Faliure Rate
Retail	
Iron Bank	
DU	
EMEA	2
DU Total	2
FR	
EMEA	2
FR Total	2
UK	
EMEA	6
UK Total	6
US	
NORTH AMERICA	2
US Total	2
Iron Bank Total	12
Retail Total	12
Grand Total	12



Design Conditions

- DQ Results must be compiled in profiling tool with consideration to reporting requirements
- DQ Rule must be represented in Collibra before DQ Results can be passed between tools
- DQ Results must be loaded “as is” from profiling tool. DQ Rule in Collibra must reflect same grain as results being passed
 - Aggregated high level results passed to Collibra cannot be decomposed during load to establish dimensional composition
 - E.g. If there is a need to report on overall DQ of Current Accounts in Iron Banks Retail division then results passed to Collibra must be reflective of these dimensions
 - Granular DQ results in profiling tool will not be aggregated on load to remove dimensional composition
 - E.g. If rules have been created per Account Type but reporting requirements are irrespective of Account Type, aggregation will not occur during integration process. Results must be passed “as is” prior to aggregation
- A unique id must be added to the DQ Rule in Collibra which is cross referenceable as a unique id in the profiling tool
- DQ results generated from multiple different profiling tools will not be accommodated. Single result set from single tool only
- Collibra will be used to establish links that enable DQ result aggregation
- Provisions for basic DQ result aggregation can be made within Collibra
 - Condition: The aggregation path must follow a linear route, grouped by a single dimension only
- Provisions for multi-dimensional DQ result aggregation must be made in a BI tool leveraging Collibra content