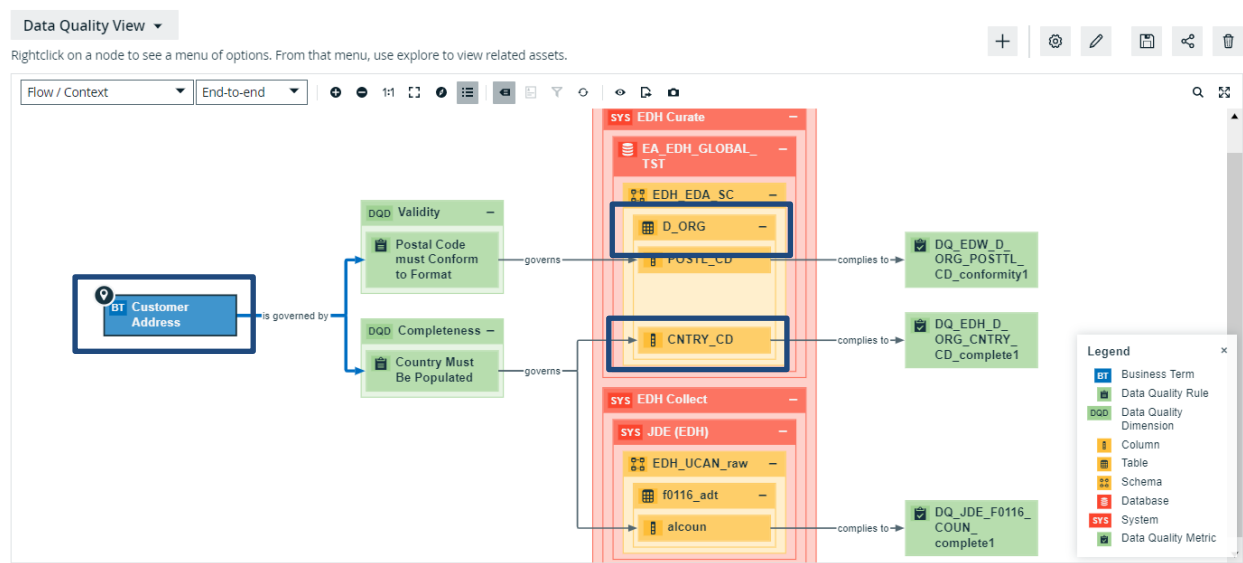


Collibra Data Quality Configuration

The following steps describe how to configure and enable data quality results to be viewed in Collibra for various asset types. When configuring data quality rules in the Settings of Collibra you are defining how data quality information will be presented in the system. This is not the creation of actual data quality rule metadata (which is built out by the data steward) or data quality metric values (created by a technical architect and fed by a data quality system).

The first step in configuring data quality in Collibra is to determine which assets you would like data quality metrics to show and how the metric information should roll up. In this example, we want to show quality metrics at the Column, Table and Business Term level.



Creating a view diagram with the labels on will help you when you configure the data quality rules in the settings.

Settings > Data Quality Rules

	Name	Description	Resource Id
General	Total Data Quality Results for Parent Table	Shows the full data quality results on a column asset for the full table	7730a97e-844c-4e47-8177-58c...
Asset types	Table Level Data Quality	Shows the table level data quality for each column with associated DQ metrics	23029e8f-fbe6-423e-9222-da3...
Domain types	Column Level Data Quality	Displays all the data quality metrics associated with a particular column.	e7737f74-cacd-4477-a971-45f...
Statuses	Business Term Data Quality	Shows the combined data quality results for all physical data columns associated with a business term	10c201e8-8982-4c15-927a-10e...
Characteristics			
Data quality rules			
Scopes			
Workflows			

In this example, we set up four different data quality rule configurations, reporting the metrics at different levels for different asset types.

Asset	Data Quality Rule	Description
Column	Column Level Data Quality	Reports quality at the column level
Column	Total Data Quality Results for Parent Table	Reports total quality for the parent table of the column, allowing consumer to compare column quality to overall table quality
Table	Table Level Data Quality	Reports quality at the table level for all columns within that table
Business Term	Business Term Data Quality	Reports quality at the table level for all columns associated to the Business Term

Column

The simplest data quality rule configuration is for the asset type that the data quality metric measures. In this example, we are running our data quality metrics for specific columns in a table.

Name*

Column Level Data Quality

Description

Displays all the data quality metrics associated with a particular column.

Aggregation Path

Path*

Asset complies to Governance Asset

Construct the path through which you want to calculate your aggregated scores.

Categorization*

grouped by Data Quality Dimension

We can categorize all Rule elements based on a relation type. You will have the ability to

CNTRY_CD

complies to

DQ_EDH_D_ORG_CNTRY_CD_complete1

Legend

Column

Data Quality Metric

Enter the Name of the Data Quality Rule Configuration and a brief description. NOTE: The name and description will appear on the data quality page for the asset so make sure the name describes what you are presenting

Column Level Data Quality

95.00

Overview

Displays all the data quality metrics associated with a particular column.

1 Data Quality Metric

100 Rows

95

The **Aggregation Path** tells the system how the results stored in the Data Quality Metric should be aggregated. In this example, because data quality metrics are associated with Columns and this is a column configuration, it is a simple one step path: Asset Complied to Governance Asset.

Column Asset complies to the Governance Asset (Data Quality Metric)

The Categorization is how the metrics will be categorized.
NOTE: This will always be grouped by **Quality Dimension**

Table Level Compare at Column

For this example document, we wanted to also show at the column level, what the total aggregated quality was for the table to which the column belongs. To do this we created a second data quality rule configuration, **Total Data Quality Results for Parent Table**

Name*

Total Data Quality Results for Parent Table

Description

Shows the full data quality results on a column asset for the full table

Aggregation Path

Path*

Column is part of Table • Table contains Column • Asset complies to Governance Asset •

Construct the path through which you want to calculate your aggregated scores.

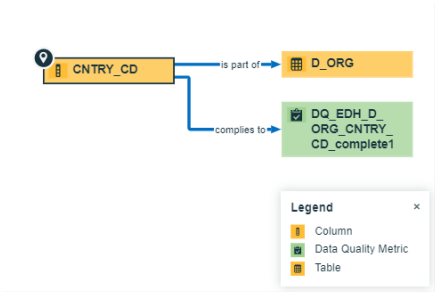
Categorization*

grouped by Data Quality Dimension

We can categorize all Rule elements based on a relation type. You will have the ability to see subscores on the asset types based on t

For this data quality rule configuration, the aggregation path needs more steps.

From the Column, we need to indicate that the **Column is part of a table** in order to pick up all metrics for a table. To complete that circle and to be able to show those Table level metrics on the column asset, we need to also indicate the reverse relationship, **Table contains Column**.

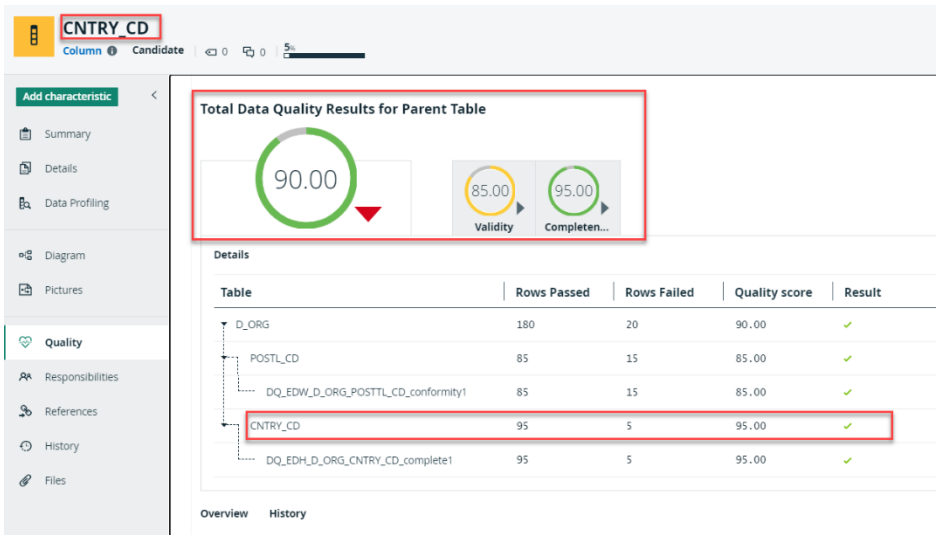


Finally, with the aggregation path back at column, we can link to the metric which contains the results – **Asset complies to Governance Asset**.

The result of this second aggregation path is to expose all of the data quality metrics associated with the parent table of the column for comparison.

In this example, when viewing the Column asset CNTRY_CD we can see the Total Data Quality Results for the Parent table D_ORG, exposing not just the CNTRY_CD quality, but also the other column in the table with a quality metric assigned, POSTL_CD.

While the quality score for the CNTRY_CD column is 95.00, total table quality is only 90.00 because of the other column's quality.

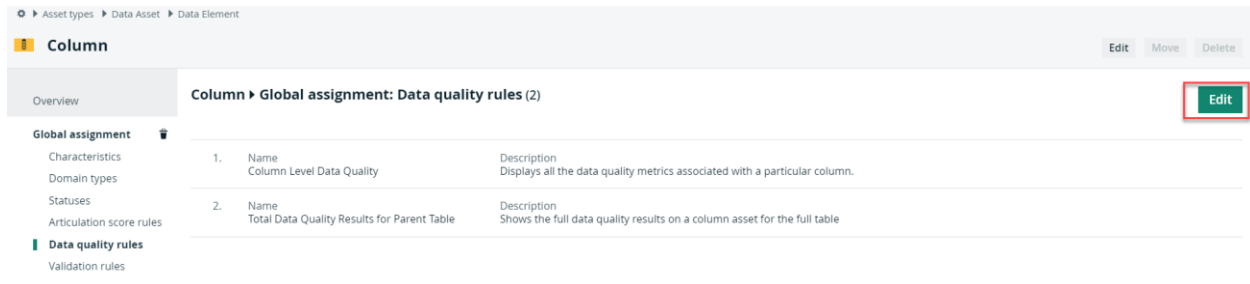


Assigning Data Quality Rule Configuration to Assets

The next step in configuring how data quality is exposed in Collibra is to assign the data quality rule configurations you created to the assets you want them to appear. In this example, the two data quality rules configurations we set up were for the column asset.

From Settings, go to Assets Types and find the Asset Type: Column

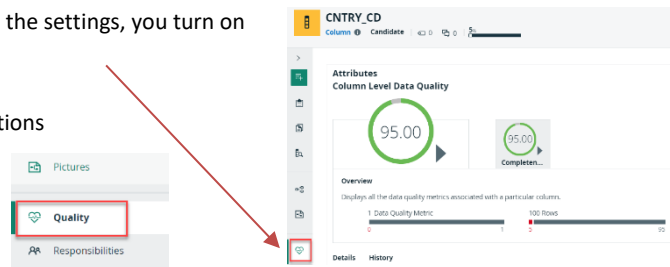
In the Global Assignments, Select Data quality rules and click the Edit button



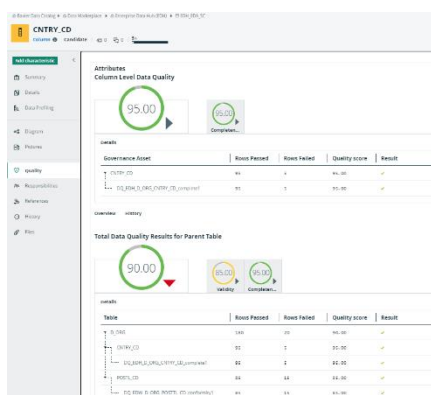
Click the Add button to select the data quality rule configurations you would like to add to that asset and click save.

By adding a data quality rule configuration to an asset in the settings, you turn on the menu option for that asset to see data quality rules.

If an asset does not have any data quality rule configurations associated with it, this menu option does not appear.



At this point, the configuration for how data quality information will be displayed for an asset (column) is complete. For an end user to actually see quality results for a particular column, a data quality rule and data quality metric must be set up. These items are metadata – and will be created by data stewards and their technical architect counterpart. Data Quality Metric (result) information can also be imported from a data quality system like IBM's Information Analyzer. More information on Data Quality Rules and Data Quality Metrics are found in this document (link to future document).



This completes the configuration of Column data quality to show both the specific column level quality and a full table view for comparison.

Additional Data Quality Rule Configurations

As stated at the beginning of this document, we also wanted to show data quality metadata for the assets Table and Business Term. To configure this, you would follow the steps outlined above. Below are the aggregation paths that were defined for each of these asset types to provide you with more examples of aggregation paths.

Table – to enable aggregation of data quality metrics at the Table level you will need an additional step to the aggregation path to move from table to column to metric. Screen shots below show the results.

Create Table Level data quality rule configuration and aggregation path:

Name*
Table Level Data Quality

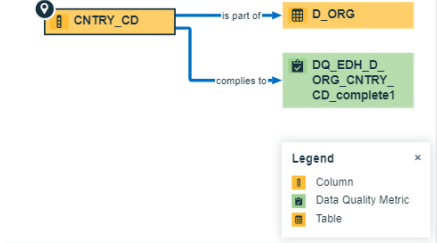
Description
Shows the table level data quality for each column with associated DQ metrics

Aggregation Path
Path*
Table contains Column → Asset complies to Governance Asset

Construct the path through which you want to calculate your aggregated scores.

Categorization*
grouped by Data Quality Dimension

We can categorize all Rule elements based on a relation type. You will have the ability to see subscores on the a



Attach Rule Configuration to Table Asset:

Asset types ▶ Data Asset ▶ Data Structure

Table

Overview

Global assignment

Characteristics

Domain types

Statuses

Articulation score rules

Data quality rules

Validation rules

Table ▶ Global assignment: Data quality rules (1)

1.	Name	Description
	Table Level Data Quality	Shows the table level data quality for each column with associated DQ metrics

Results in Table Asset Data Quality Page:

Baxter Data Catalog ▶ Data Marketplace ▶ Enterprise Data Hub (EDH) ▶ EDH_EDA_SC

D_ORG

Table Under Review

50%

Add characteristic

Summary

Details

Columns

Sample data

Diagram

Pictures

Quality

Responsibilities

References

History

Attributes

Table Level Data Quality

90.00

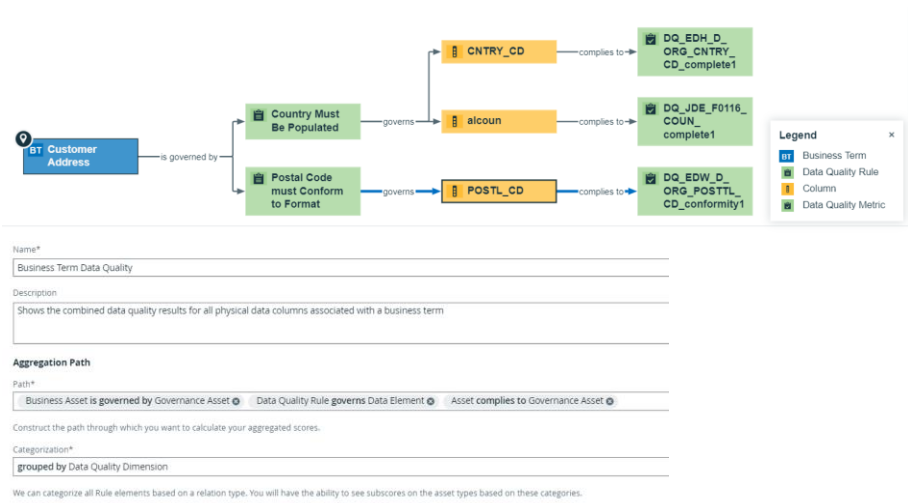
Validity 85.00

Completeness 95.00

Column	Rows Passed	Rows Failed	Quality score	Result
D_ORG	180	20	90.00	✓
CNTRY_CD	95	5	95.00	✓
DQ_EDH_D_ORG_CNTRY_CD_complete1	95	5	95.00	✓
POSTL_CD	85	15	85.00	✓
DQ_EDW_D_ORG_POSTL_CD_conformity1	85	15	85.00	✓

Business Term – to enable aggregation of data quality metrics for a business term, we took a slightly different approach. Because we wanted to include the data quality rules (which are business friendly definitions of the data quality metric tests) in our result set, we took a slightly longer path to aggregation. All other steps are the same. Screen shots below show the results.

Create Business Term Level data quality rule configuration and aggregation path:



Attach Rule Configuration to Table Asset:

Asset types Business Asset

BT Business Term

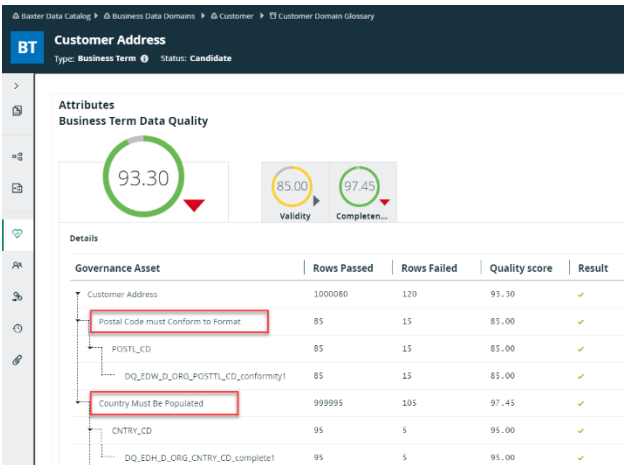
Overview

Business Term Global assignment: Data quality rules (1)

Name	Description
Business Term Data Quality	Shows the combined data quality results for all physical data columns associated with a business term

Add assignment

Results in Table Asset Data Quality Page:



Notice the inclusion of the business-friendly definition of what the quality rule is. This is the Data Quality Rule defined in business terms by the data steward.