



ONE – Workshop

Business Glossary

Prepared for: v15.4

Prepared by: Ataccama

Dated: November 2024



Contents of the Document

Introduction	3
1. Creating a new Business Glossary term	3
2. Creating a Business hierarchy	7
3. Creating a Detection Rule (Based on data)	8
4. Creating a Detection Rule (Based on metadata)	14
5. Adding Detection rules to a Glossary Term	15
6. Reprofileing after term creation	16
Conclusion	19

Introduction

In this workshop, we will focus on the business glossary section to learn how terms can be manually created and related, as well as understand their impact on catalog items. Additionally, we will explore the creation of detection rules and their role in attribute term assignment.

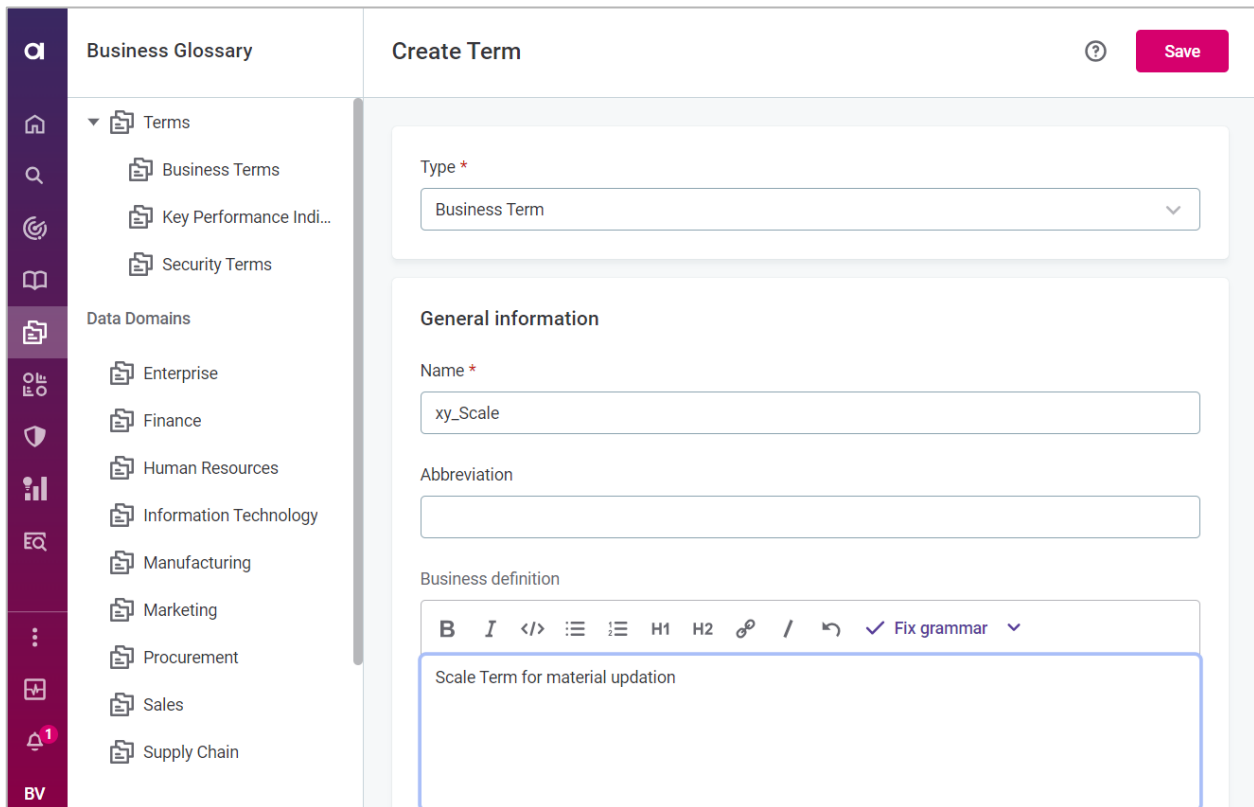
Tasks

In this workshop, we have six tasks. Two of these tasks focus on creating glossary terms and establishing a hierarchy between them. The next two tasks involve creating detection rules using data and metadata. The final two tasks are dedicated to assigning detection rules to terms and re-profiling a relevant catalog item.


1. Creating a new Business Glossary term

In the first task we want to practice adding a business glossary term to the ONE. As the **products** catalog item doesn't have any terms mapped to it, we decided to create a glossary term (**Scale**) for one of its attributes (**productscale**).

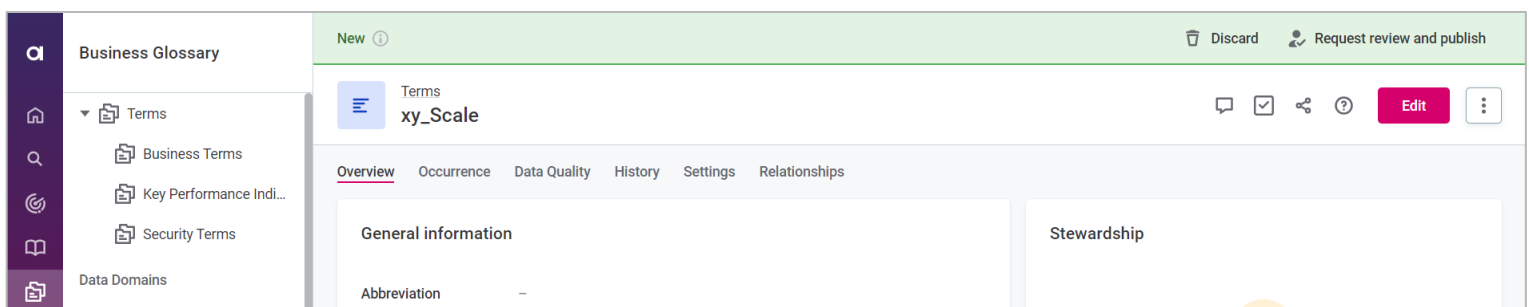
- › Navigate to the **Business Glossary** in the left panel and click on the **Create** button in the upper right corner:
- › As a **Type**, choose the **Business term** option and fill in the name "**prefix_Scale**". Optionally, you can provide Business definition for the term and create an **Abbreviation** for it.
- › Click on the **Save** button in the upper right corner.




The newly created term is in the **NEW** state of the **approval workflow**. Any change made to an asset needs to be approved by the assignee of the workflow, which is by default the stewardship group members of the asset.

 **NOTE** New assets have **green** color, changes are displayed with **orange** color and deletions are displayed as **red**.


➤ To approve the change, click on the **Request review and publish**



- › If you are satisfied with the changes, write a short comment describing your change (the description is mandatory at this step) and click the **'Submit'** button and choose **View publishing task**.


Request review
×


Workflow

 Review request

Task name *


Priority *

=
Medium
▼

Description * 


Creating a new term - Scale




Assets


 Use Show different only filter on the left to list assets for review and publish.





Submit

Cancel

New | Pending review 

 View comments
 Review and publish



Metadata model > 36ed325c-0000-7000-0000-00000155093f

View publishing task
Cancel review request

Overview
Occurrence
Data Quality
History
Settings
Relationships

General information

Stewardship



- › Click on **Approve or Reject** and Confirm changes by clicking on the **Approve** button.

Tasks and Workflows

☒ Tasks

☐ Launched Workflows

☐ Workflows

☐ Task Types

Review Request

Overview

Comments (0)

Review Request

Type

Review Request

Workflow step

Waiting for review

Author

bliss.varghese@ataccama.c...

Description

Creating a new term - Scale

View all changes

Entity reference

xy_Scale

Created by system, September 3, 2024

Review changes

Review changed entities. You can navigate to them through the attached link.

xy_Scale

Approve or reject

Assignee

Administrators

Reassign

Status

New

Priority

Medium

Back

View changes

View publishing task

Reject

Approve

Show different only

xy_Scale Terms

Detection Rules

Validation Rules

Standardization R...

xy_Scale	Published version	Changed version
General information		
Name	-	xy_Scale
Business definition	-	Scale Term for material updation
Pushdown enabled	-	✗

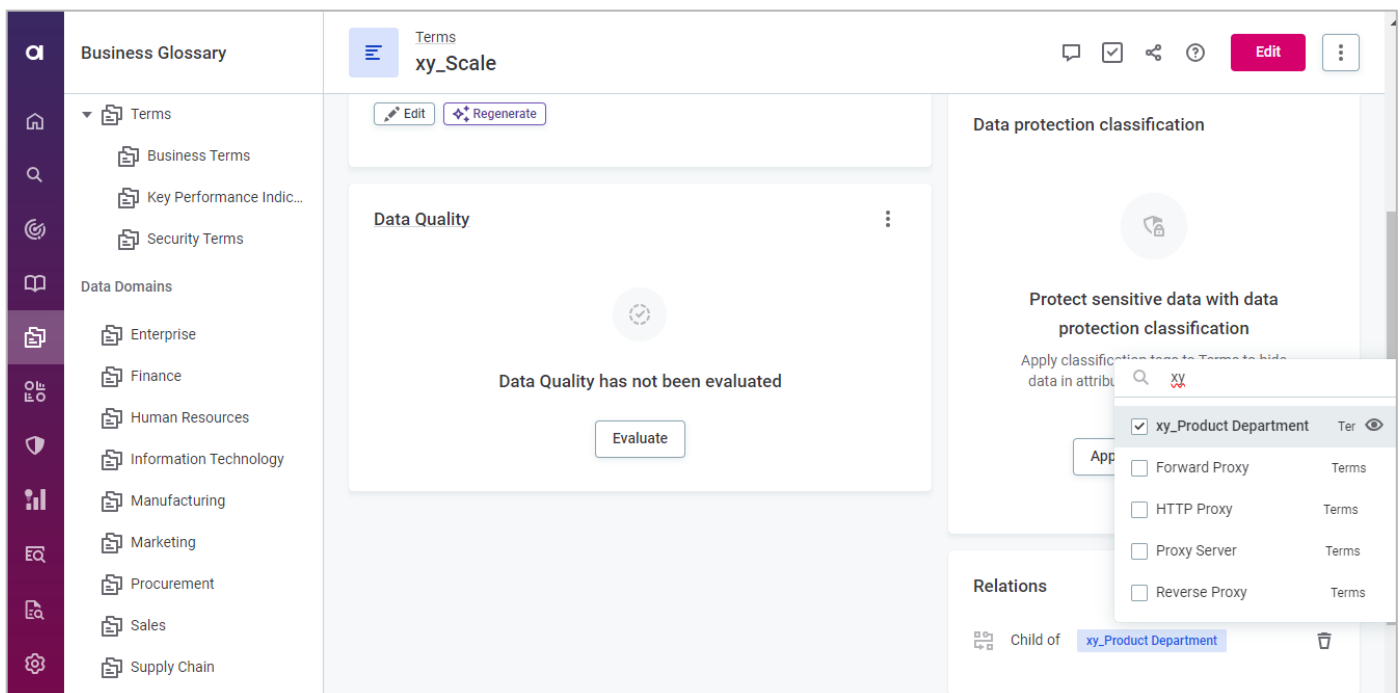


Your turn now! Create another business term called **Product department** and publish it using the same procedure. Don't forget a distinguishing factor when naming it!

2. Creating a Business hierarchy

If we consider the **Product department** term belongs to the product teams, the **Scale** term should belong to this group. This will allow us to practice creating a new business hierarchy in the **Business Glossary**.

- › Open the term **<prefix>_Scale** you have just created.
- › On the right side of the Overview tab in the **Relations** section, select **Child of** as the type and add the term **<prefix>_Product department** as a parent.



The screenshot displays the 'Business Glossary' interface. On the left, a sidebar lists various categories like 'Terms', 'Data Domains', and 'Enterprise'. The main area shows the details for the term 'xy_Scale', including a 'Data Quality' section with an 'Evaluate' button. On the right, there's a 'Data protection classification' section and a 'Relations' section. A dropdown menu is open over the 'Relations' section, showing a list of terms to be added as parents, with 'xy_Product Department' selected.



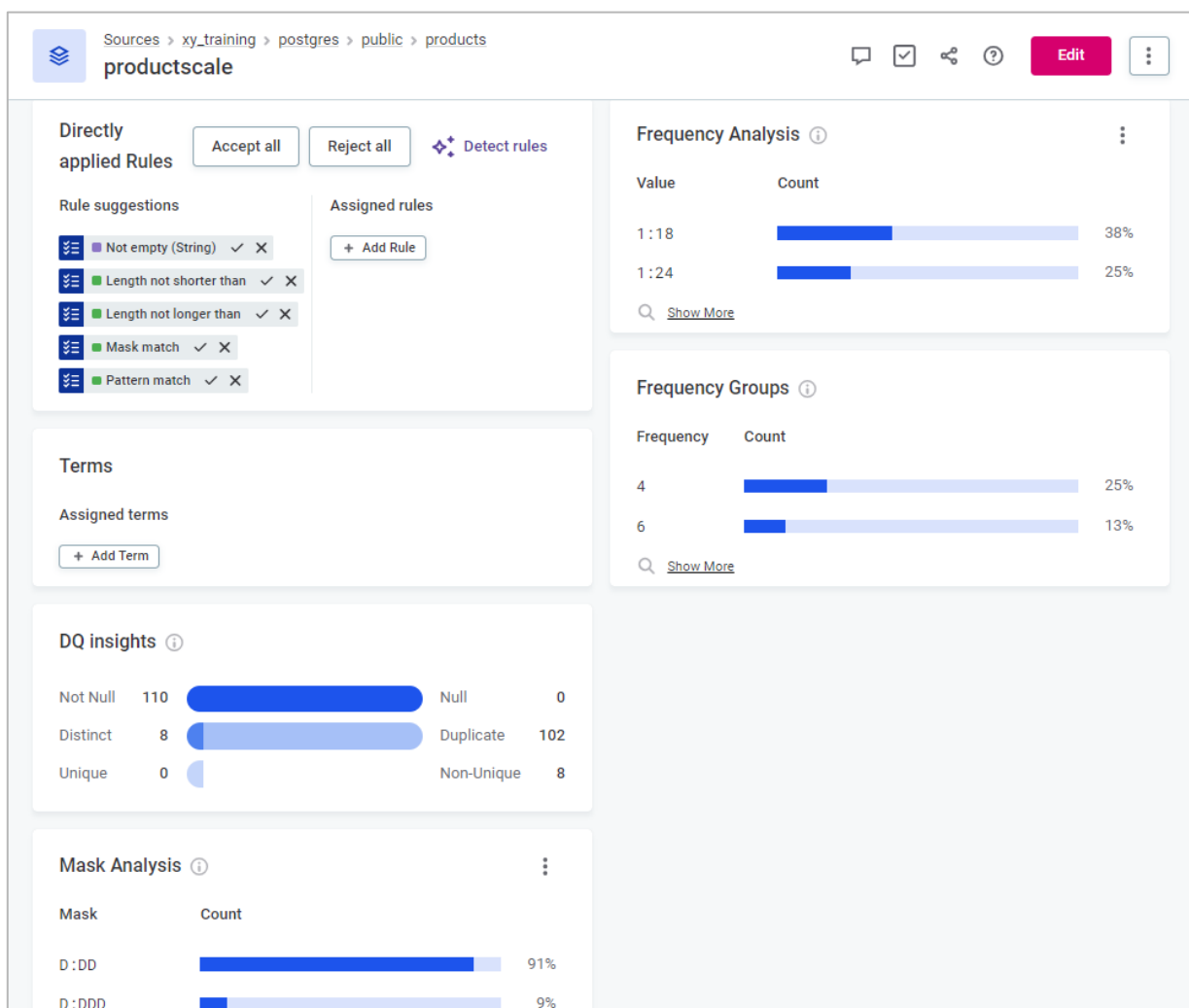
NOTE

Each **Term** can have multiple **Relations** defined. In each relation, multiple terms can be chosen.

3. Creating a Detection Rule (Based on data)

In order to make sure the **<prefix>_Scale** term gets assigned to the **productscale** attribute during the profiling, we need to create a Detection Rule. In this regard, we should have a closer look into the attribute.

- › Navigate to the **Catalog Item > Products > productscale** and select the **Profiling & DQ insights** tab.
- › Notice the patterns and stats in the **Pattern Analysis, Frequency Analysis,** and **Mask Analysis** sections.



As you can see, most of the values follow the **D:DD** pattern (e.g. 1:10) which can be a good foundation for our detection rule. Having this in mind, we can proceed to the rule creation.

- › Navigate to the **Data Quality** section on the left panel.
- › In the **Rules** section, click on the **Create** button.

- Now, we should select the **type** of the rule:

- **Detection rules** are used to detect terms in Catalog items,
- **Data Quality rules** are used to evaluate data quality in Catalog Items.

- Create Rule

?

Save

General information

Name *

xy_Sclae

Rule type *

Based on data

Description

B

I

</>

:

:

H1

H2

/

✓

Fix grammar

Rule definition source

- Page 9 / 19

Now that we have the rule defined, it is time to create its logic:

- › Click on the Rule's **Implementation** tab to define the rule logic.

In the **Input Attributes** section, you can rename the Attribute from INPUT to **SCALE**. Leave the data type set as **String** type.

Based on the pattern we identified in our attribute, using **Regular Expression** is the best choice to construct our logic.

- › Make sure the condition builder is selected.

- › Select the **"matches mask"** from the drop down list and provide both of the recognized patterns using the OR operator.

D:DD or D:DDD

This is how your rule implementation should look like.



Rules xy_Scale

Overview Implementation Occurrence History

Term Detection Rule: Based on data

Rules based on data use values to determine terms, e.g. if the data follows mask or isn't empty.

Inputs ⓘ

ATTRIBUTES ⓘ

Abc String

SCALE

+ Add Term

VARIABLES ⓘ

+ Add variable

Rule Logic

Test Rule

Condition

Ask AI

</>

WHEN

Abc SCALE

value

matches mask

D:DD

or

D:DDD

+

+ Add expression

THEN

Detect Term

ATTRIBUTES 1

Abc String

SCALE

+ Add Term

PARAMETERS 1

+ Add parameter

VARIABLES 1

+ Add variable

Condition

Ask AI

</>

WHEN

Abc SCALE

value

matches mask

D:DD

or

D:DDD

+ Add expression

THEN

Detect Term

Test

SCALE		Message
1:10		Detected
1:100		Detected
1:10		Undetected
10.1		Undetected
STRING		Undetected
+ New Row		

- › Close the Test Rule window and **publish** the changes if you are happy with the expression.
- › If you are happy with the test results, publish the rule.

New ⓘ

Discard Publish

Rules
xy_Scale

Edit



NOTE

In this case, there is no **approval workflow** configured; therefore, you can publish the change right away.



Alternatively, you can switch to the advanced expressions and provide the same logic using the matches function and global expressions syntax.

matches(@"[0-9]:[0-9]+",SCALE)

Rules
xy_Scale

Edit

Overview
Implementation
Occurrence
History

Term Detection Rule: Based on data

Rules based on data use values to determine terms, e.g. if the data follows mask or isn't empty.

Inputs

ATTRIBUTES 1

Abc String

SCALE

+ Add Term

VARIABLES

+ Add variable

Rule Logic

Test Rule

Condition

Ask AI

WHEN

1
matches(@"[0-9]:[0-9]+",SCALE)

THEN

Detect Term

Description

B I </> H1 H2

Generate

Detecting Terms

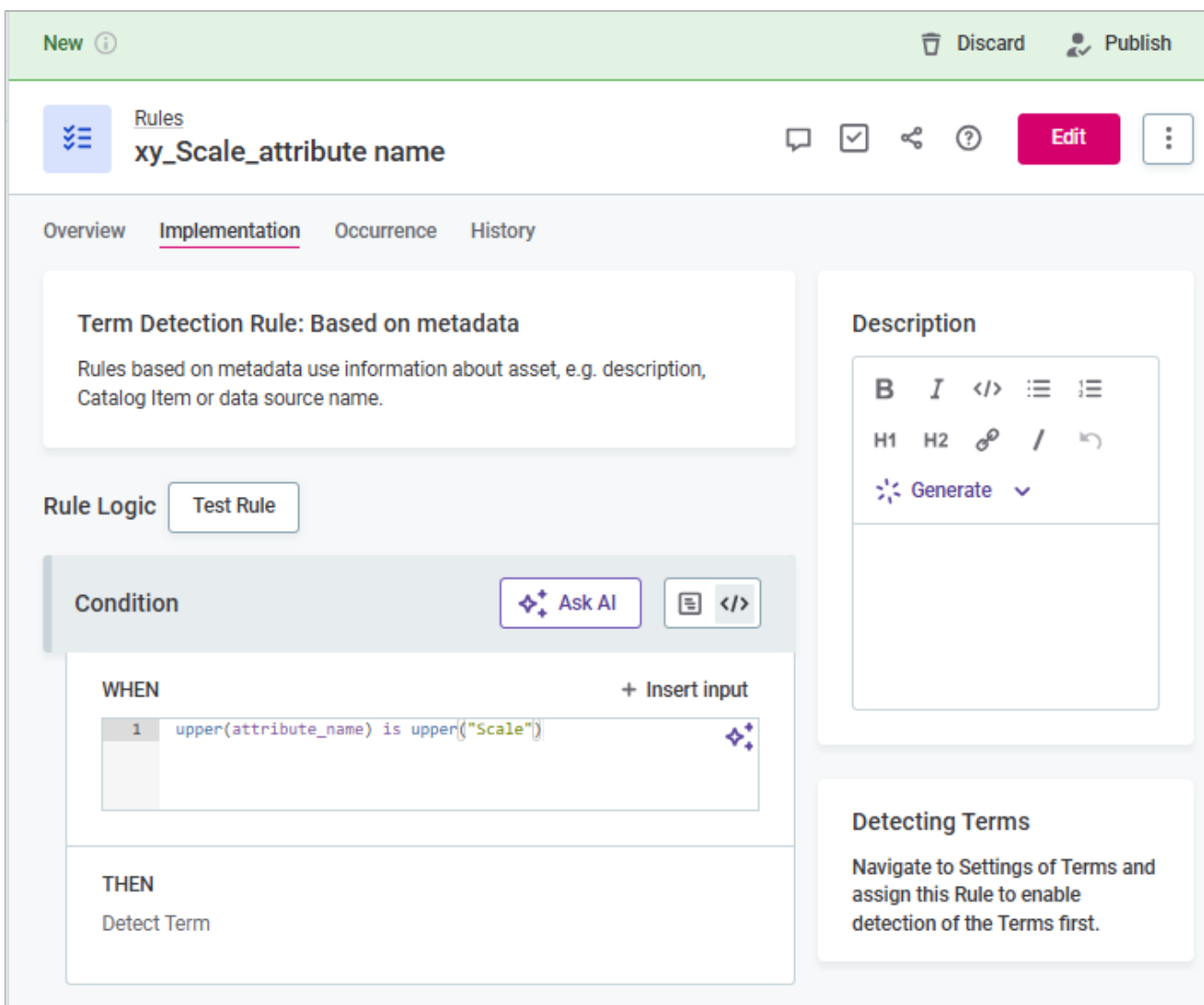
Navigate to Settings of Terms and assign this Rule to enable detection of the Terms first.



If the expression is not valid, a symbol  will appear next to the expression.

4. Creating a Detection Rule (Based on metadata)

- › Navigate to the Data Quality section on the left panel.
- › In the Rules section, click on the **Create** button.
- › Fill in the Name of the rule "**prefix_Scale_attribute name**".
- › Select **Detection – Based on metadata** as the **Rule Logic**.
- › Save the rule and switch to the implementation tab.
- › In the condition section, switch to the advanced expression.
- › Fill in the WHEN clause with the following the expression to check the attribute name :
upper(attribute_name) is upper("Scale")
- › **Test** and **Publish** the new rule.



The screenshot displays the Open Data Kit interface for creating a new rule. At the top, there's a green header bar with 'New' and 'Publish' buttons. Below this, the 'Rules' section is active, showing the rule name 'xy_Scale_attribute name'. The 'Implementation' tab is selected, showing the rule logic 'Term Detection Rule: Based on metadata'. The 'Condition' section is expanded, showing the 'WHEN' clause with the expression 'upper(attribute_name) is upper("Scale")'. The 'THEN' clause is set to 'Detect Term'. The right sidebar contains a 'Description' section with a rich text editor and a 'Detecting Terms' section with instructions to navigate to the settings of terms and assign this rule to enable detection of the terms first.

New **Discard** **Publish**

Rules
xy_Scale_attribute name **Edit**

Overview **Implementation** **Occurrence** **History**

Term Detection Rule: Based on metadata
Rules based on metadata use information about asset, e.g. description, Catalog Item or data source name.

Rule Logic **Test Rule**

Condition **Ask AI** **</>**

WHEN **+ Insert input**

1 upper(attribute_name) is upper("Scale")

THEN
Detect Term

Description

B **I** **</>** **≡** **≡**
H1 **H2** **🔗** **/** **↶**
Generate **▼**

Detecting Terms
Navigate to Settings of Terms and assign this Rule to enable detection of the Terms first.

5. Adding Detection rules to a Glossary Term

Now that we have the detection rules defined, it is time to link them to the glossary term **<prefix>_Scale**.

- › Navigate to the **Business Glossary** and find your term **<prefix>_Scale**.
- › Go to the **Settings** tab, the **Detection on attributes** section.
- › Add the new rule **Scale** and set the threshold to 80%.



When you set the threshold too low, the term can be detected for *wrong attributes*. On the other hand, too high threshold may cause that the *term not detected* at all. Therefore, setting the threshold properly requires some *practice*.

- › Add the new rule **Scale_attribute name** and let the operator between the rules be “or”.

Now that we are set to assign the detection rule to the term, we can add a data quality evaluation rule too to enable DQ evaluation after reprofiling the catalog item.

- › In the **DQ Evaluation** section add an existing rule, for example, **String Completeness** (to check if the attribute, where the term was assigned, is not empty).

The screenshot shows the 'Business Glossary' interface. On the left, a sidebar lists 'Terms' (Business Terms, Key Performance Indicators, Security Terms) and 'Data Domains' (Enterprise, Finance, Human Resources, Information Technology, Manufacturing, Marketing, Procurement, Sales, Supply Chain). The main area displays the settings for the term 'xy_Scale'. The 'Settings' tab is active, showing 'Detection on attributes' and 'Data Quality Evaluation' sections. In 'Detection on attributes', there is a checkbox for 'AI detection' and a list of rules: 'xy_Scale' (Based on data, 80% threshold) and 'xy_Scale_attribute_name' (Based on metadata). The 'Data Quality Evaluation' section has a checkbox for 'Apply the following rules automatically on all attributes with this term' and several categories: Validity, Completeness (with 'String Completeness' selected), Uniqueness, Accuracy, and Timeliness, each with an 'Add Rule' button. The 'Detection on tables' section shows 'No conditions'.

- › Press the **Send for Approval** button and follow all the required steps until the changes are published.

6. Reprofileing after term creation

Now, it's time to reprofile your Catalog Item to see if the new term gets properly assigned.

- › Navigate to the **catalog items > products** and run **Full profiling**.

NOTE

As you can see, there is no data quality available even though the Catalog Item has already been profiled; the data quality is only available for catalog items that have attributes with associated data quality rules.

Knowledge Catalog

Data Catalog
Catalog Items
Reports
Anomaly Over...
Data Export P...
Master Data
Reference Data
Not Monitored
Sources
Term Suggestions
Lineage Assets
Lineage Import

Sources > xy_training > postgres > public

products

Overview
History
Data
Data Structure
Lineage
Data Quality
Profile & DQ insights
Relationships
Data Export & Transformations

Attributes

productcode 68% [USA] Zip code ✓ X

productname 70% Stock Symbol ✓ X 70% Country ✓ X

productline

productscale

productvendor

productdescription

quantityinstock 79% Customer ID ✓ X

buyprice 100% by_buyprice ✓ X 68% Credit limit ✓ X

msrp 85% Average salary ✓ X

Add Attribute

Glossary terms

+ Add Term

Summary

Description

Source xy_training > postgres > public

Origin Postgres

Table type TABLE

Number of records ⓘ 110

Sample profiling

Sample profiling

Profile a sample of the dataset

Sample profile & DQ Evaluation

Profile a sample of the dataset and DQ evaluation

Full profiling

Profile the full dataset

Full profile & DQ Evaluation

Profile the full dataset and DQ evaluation

- › Check the results of the profiling (you may need to refresh the page); the **Scale** term should now be assigned to the **productscale** attribute.
- › Re-check the **Data Quality** tab on the catalog item as well as on the **productscale** attribute; notice that the **'Evaluate'** button is now available on the catalog item as now one of its attributes has a DQ rule associated with it.

Sources > xy_training > postgres > public

products

Overview History Data Data Structure Lineage Data Quality Profile & DQ insights More ▾

Data Quality Load failed records to ONE Data

Data Quality has not been evaluated

Full Evaluate

Filter Attributes and Rules Show only attributes with Rules Check for Rule suggestions

Standard view ▾ Hidden columns ▾ Clear sort

Name	Terms	Last evaluated	Overall Quality	Applied Rules
productscale	xy_Scale	-	-	+ Add String Completeness

- Now, hit the evaluate button to see the data quality results, which in this case will only be based on the **String Completeness** rule which was added to the **<prefix>_Scale** term.

Sources > xy_training > postgres > public
products

Attributes
Add Attribute

productcode	
productname	
productline	
productscale	xy_Scale
productvendor	
productdescription	
quantityinstock	
buyprice	
msrp	

Glossary terms
Add Term

Summary

Description
Add
Generate

Source
xy_training > postgres > public

Origin
postgres

Table type
TABLE

Number of records
110

Data Slices
0 View

Data Quality
September 4, 2024, 2:09:48 AM
Full Evaluate

Overall Quality
100%

Completeness
100%

Stewardship
Edit



NOTE

When we profiled the **customers** table in the previous workshop, the data quality evaluation was available at the table level as some of its attributes already had terms mapped to them.



TASK

Choose one or two attributes in the **customers** catalog item with no terms attached to them. Create terms and detection rules to map the terms to these attributes.

Add appropriate data quality rules and re-run the profiling to see your results and review the differences.

Conclusion

We have come to the end of this workshop!

We have gone through the definition of a new glossary term as well as creating a business hierarchy. We also learned how to create detection rules and how to apply them in a term's settings.