ONE - Workshop

Business Glossary

Prepared for: v15.4

Prepared by: Ataccama

Dated: November 2024

Contents of the Document

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.

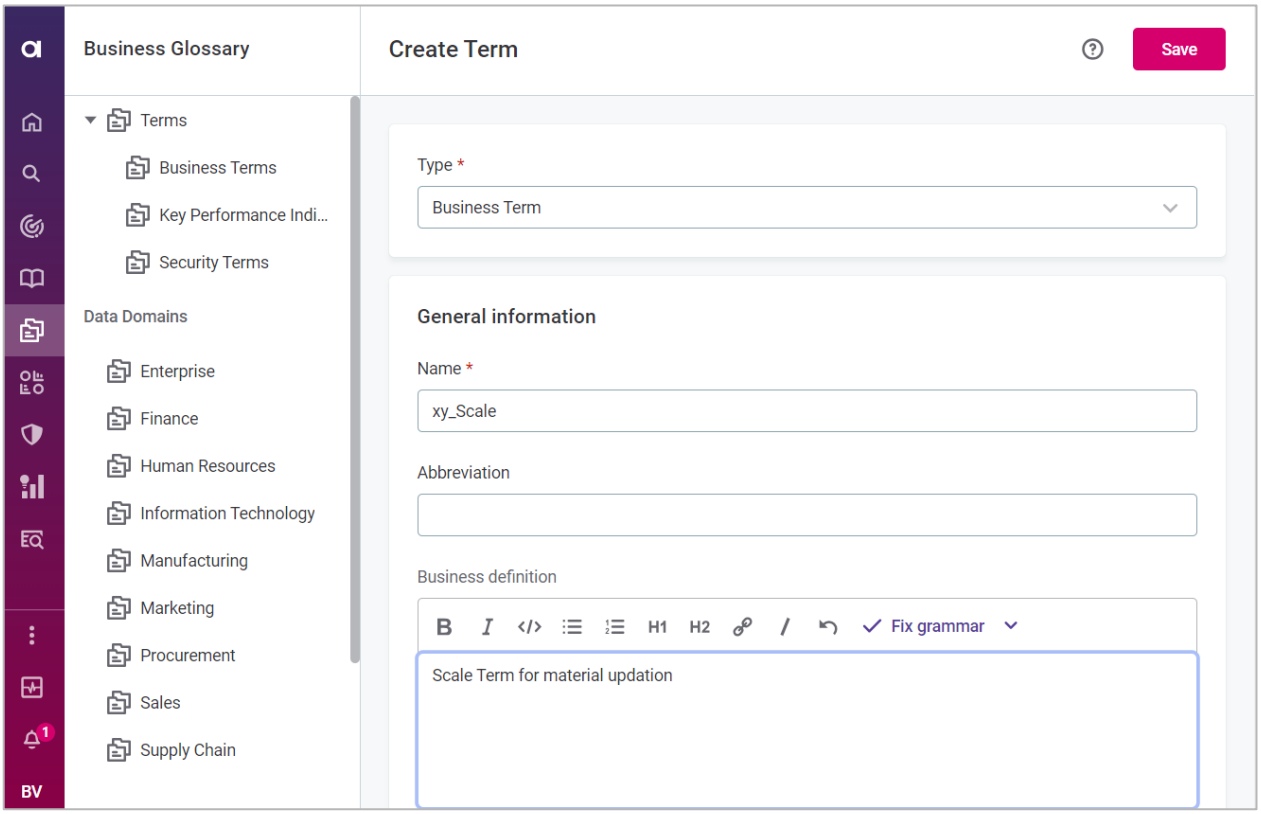
Tasks

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)**.

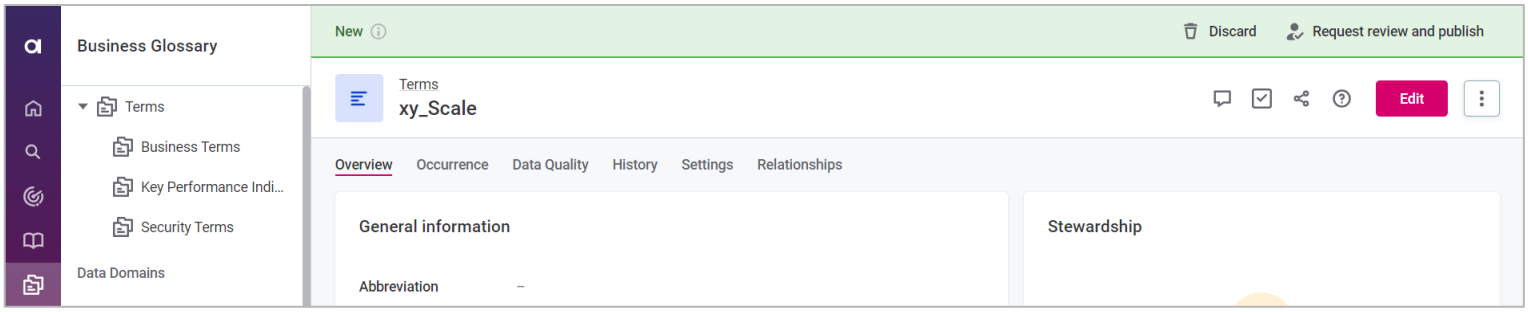
- Navigate to the **Business Glossary** in the left panel and click on the **Create** button in the upper right corner:
- As a **Type**, choose the **Business term** option and fill in the name **"prefix_Scale"**. Optionally, you can provide Business definition for the term and create an **Abbreviation**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



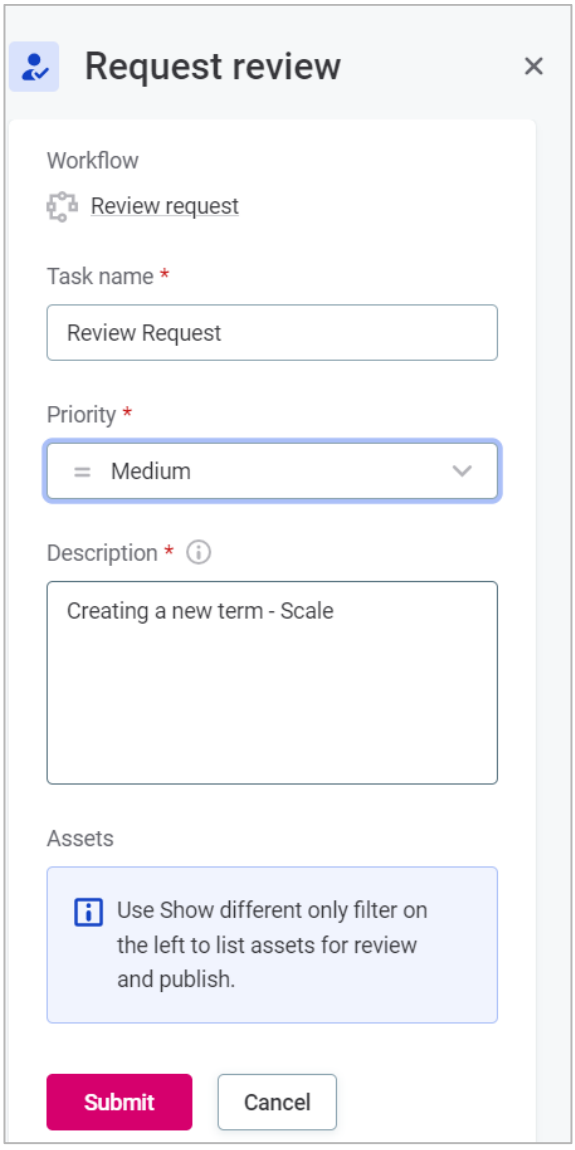
The newly created term is in the **NEW** state of the **approval workflow**. Any change made to an asset needs to be approved by the assignee of the workflow, which is by default the stewardship group members of the asset.

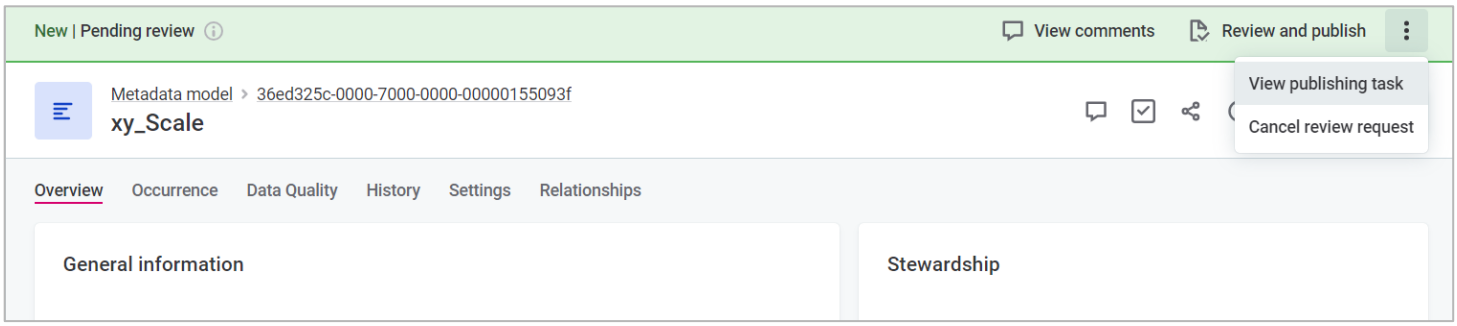
NOTE *New assets have green color, changes are displayed with orange color and deletions are displayed as red.*

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

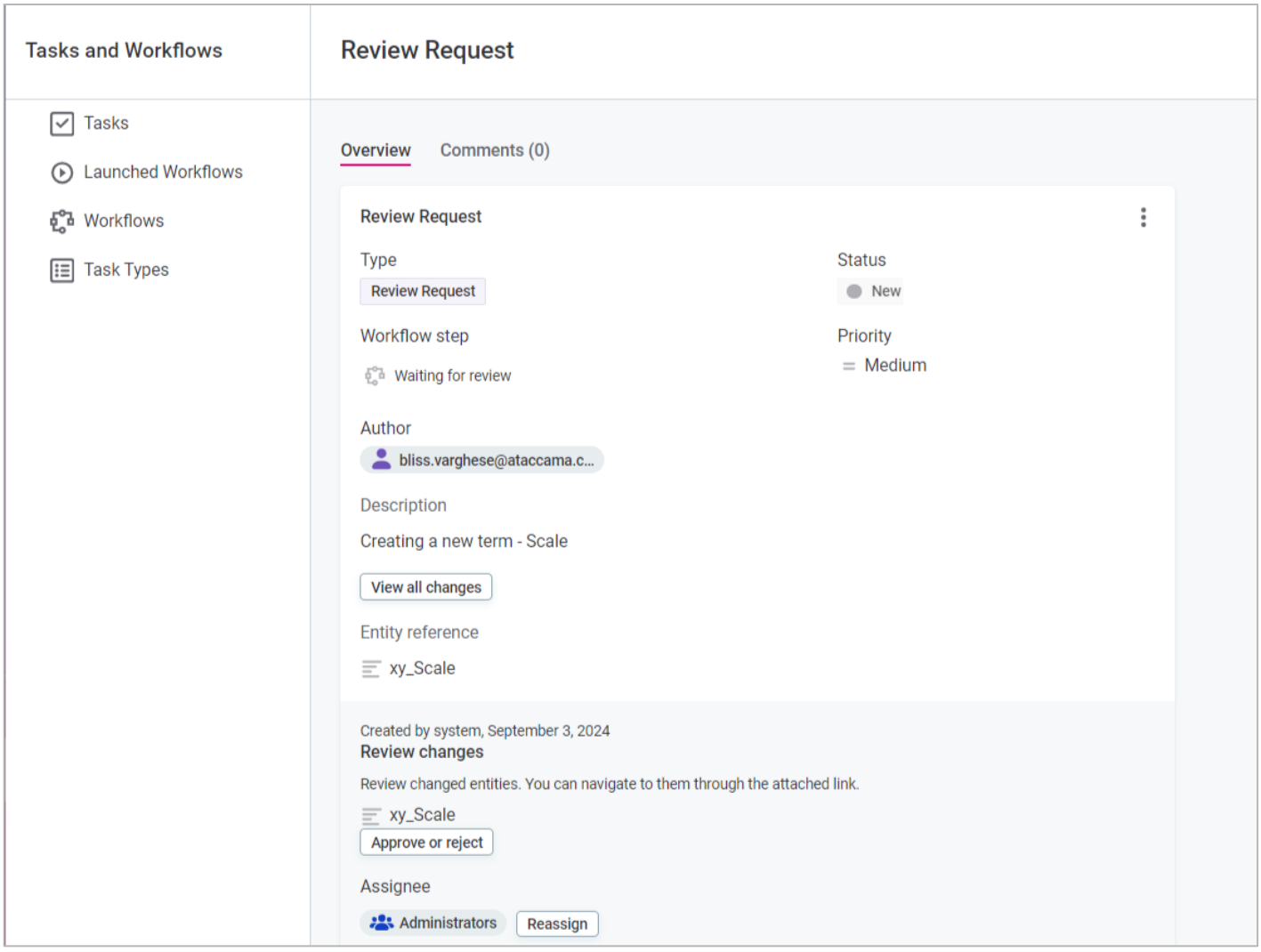


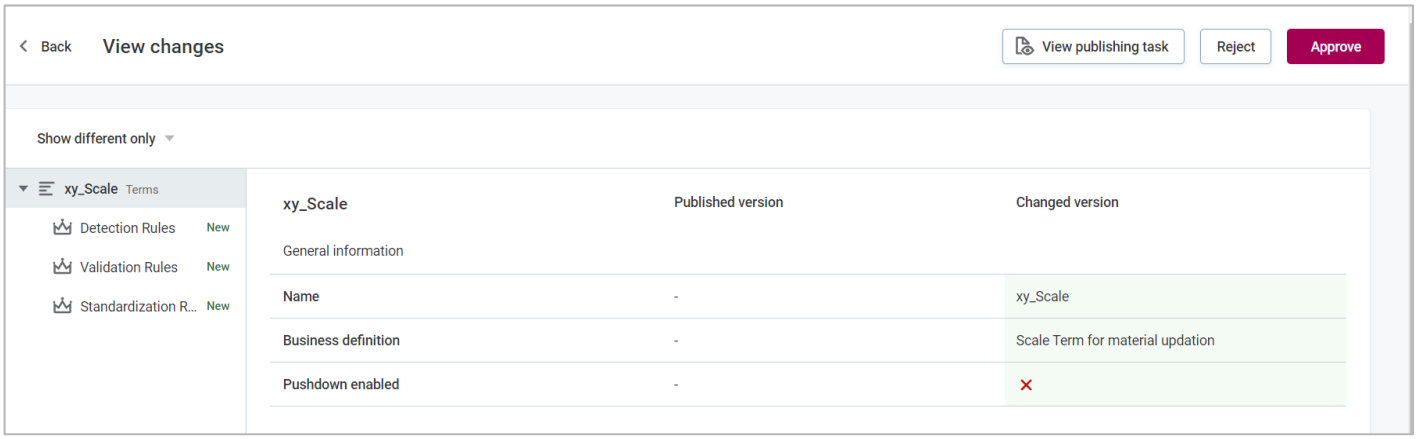
› If you are satisfied with the changes, write a short comment describing your change (the description is mandatory at this step) and click the **'Submit'** button and choose **View publishing task**.





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





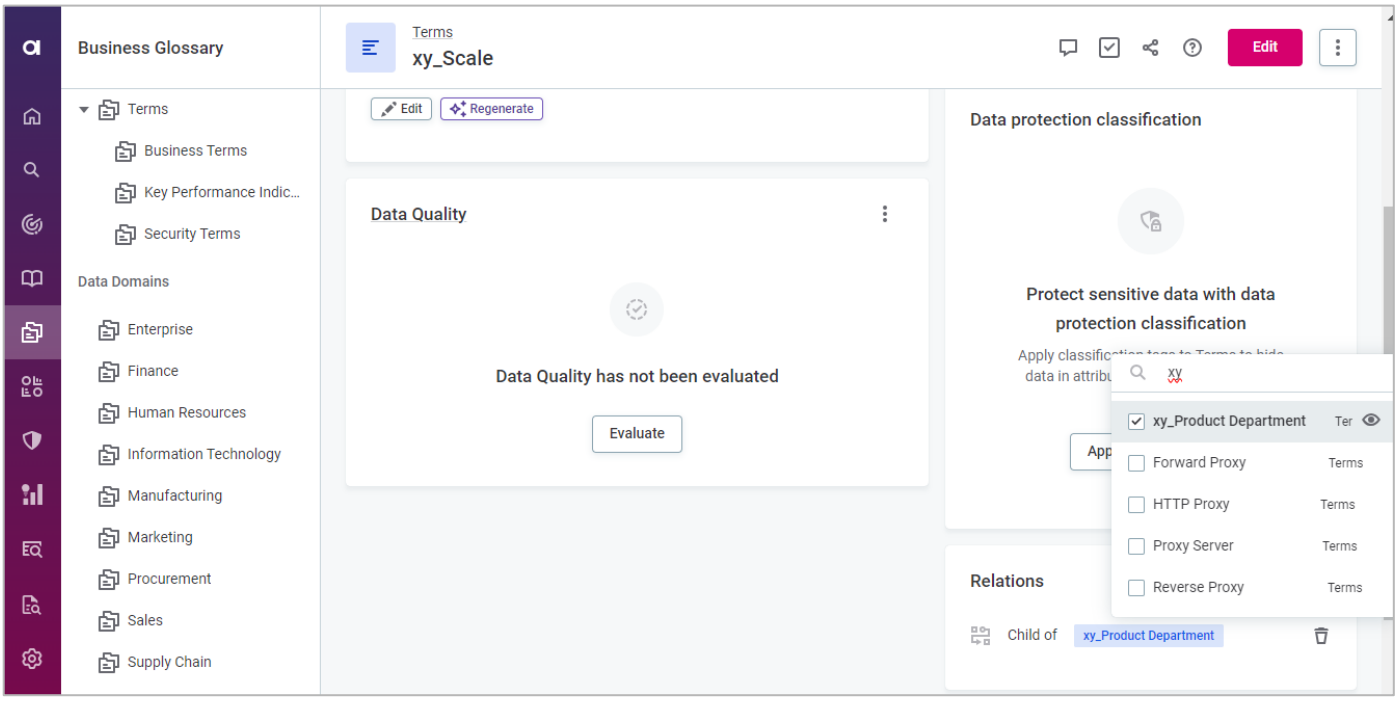
TASK

*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.

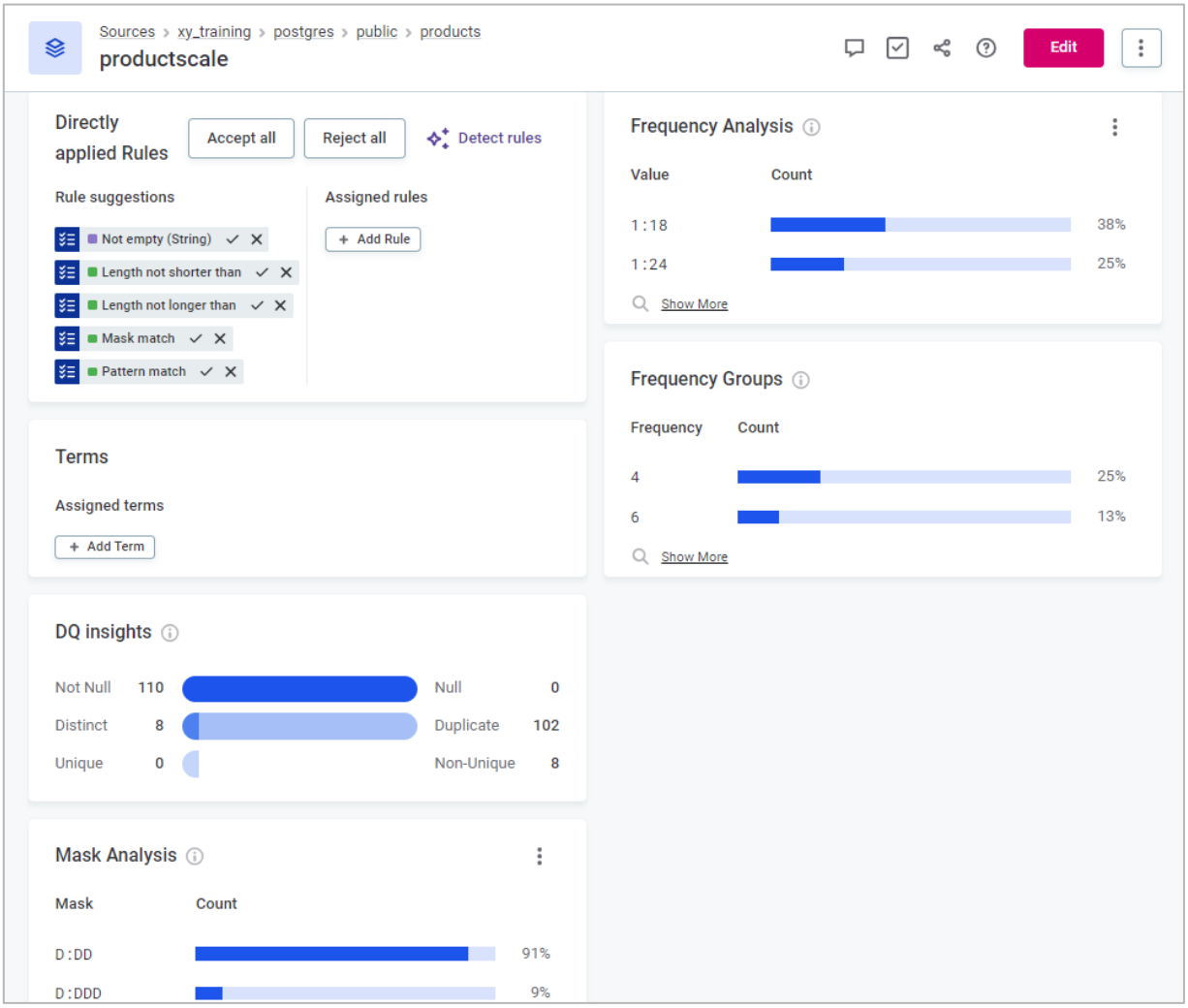


NOTE *Each **Term** can have multiple **Relations** defined. In each relation, multiple terms can be chosen.*

3. Creating a Detection Rule (Based on data)

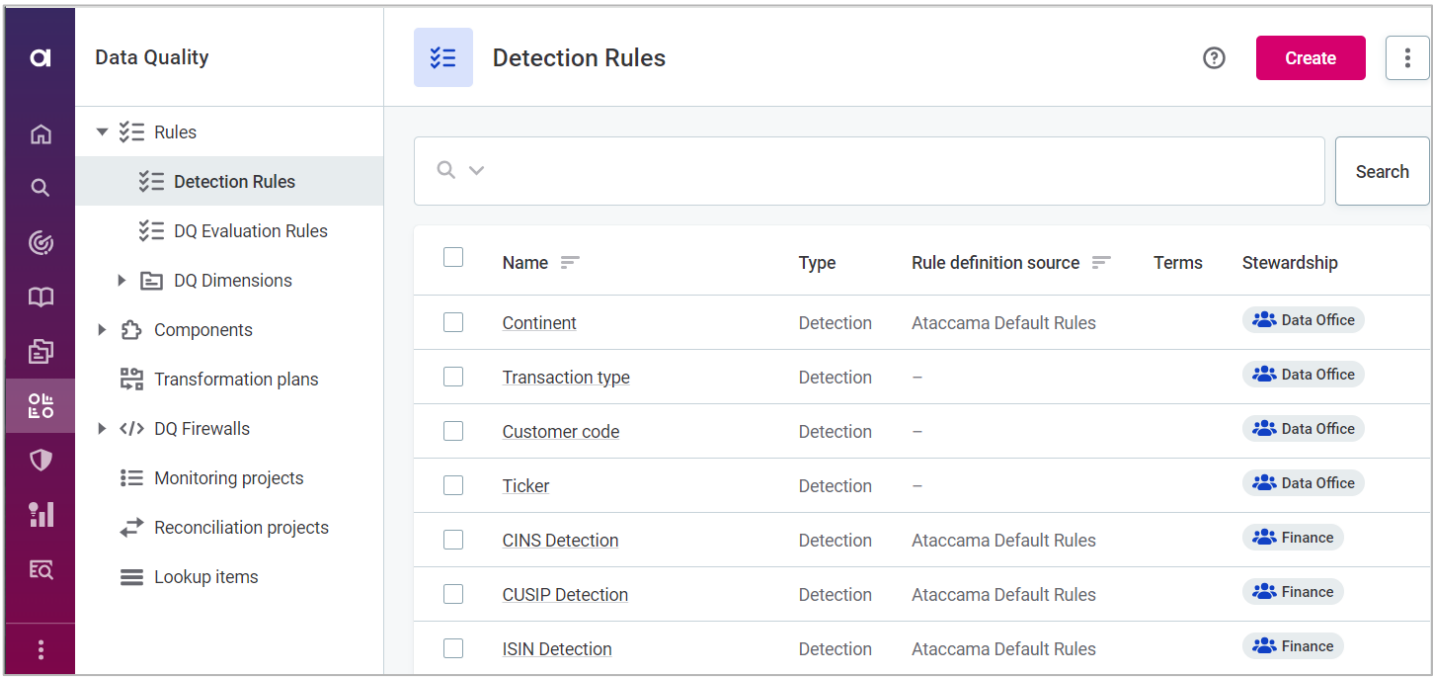
In order to make sure the **<prefix>_Scale** term gets assigned to the **productscale** attribute during the profiling, we need to create a Detection Rule. In this regard, we should have a closer look into the attribute.

- 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.

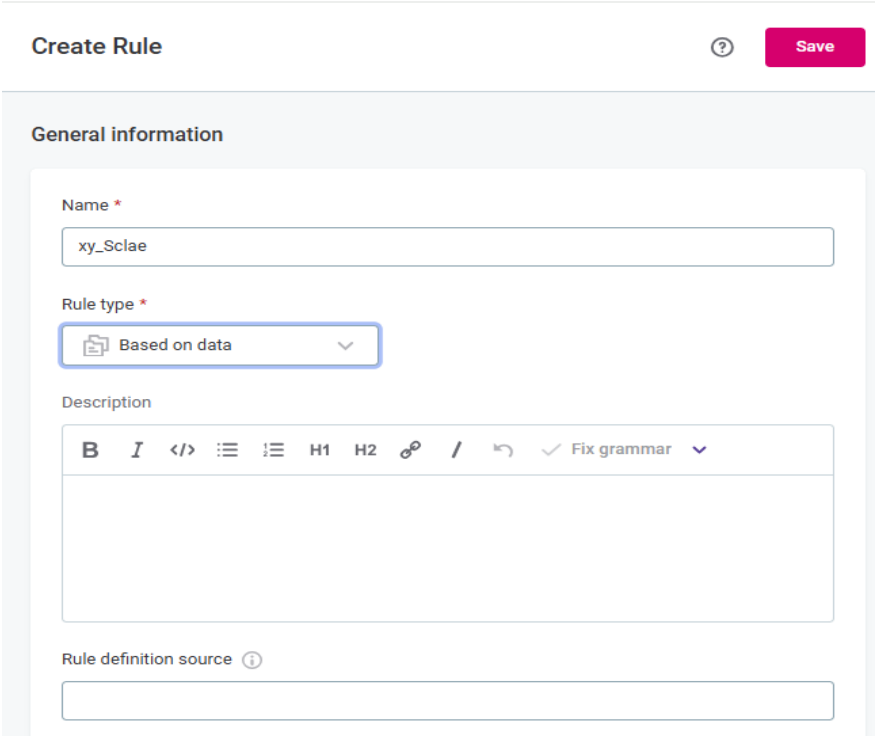


› Fill in the **Name** of the rule "**prefix_Scale**". Optionally, you can fill in a Description.

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.

› Select **Detection – Based on data** as the **Rule Type**.

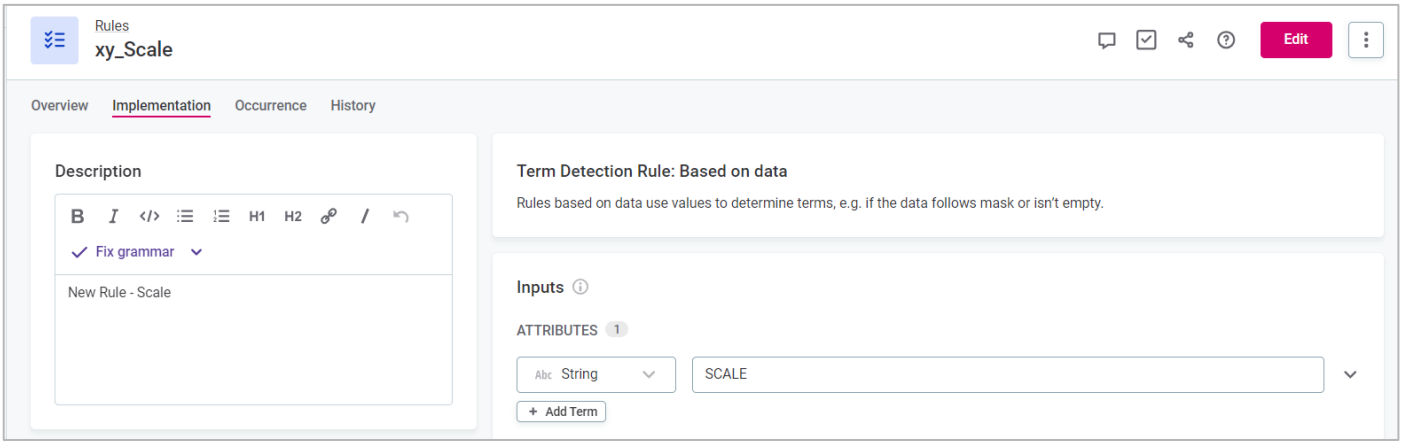


› **Save** the rule definition.

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

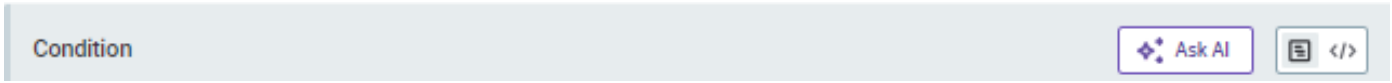
› Click on the Rule's **Implementation** tab to define the rule logic.

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



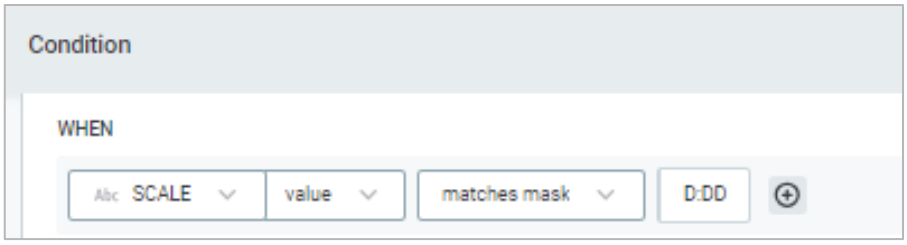
Based on the pattern we identified in our attribute, using **Regular Expression** is the best choice to construct our logic.

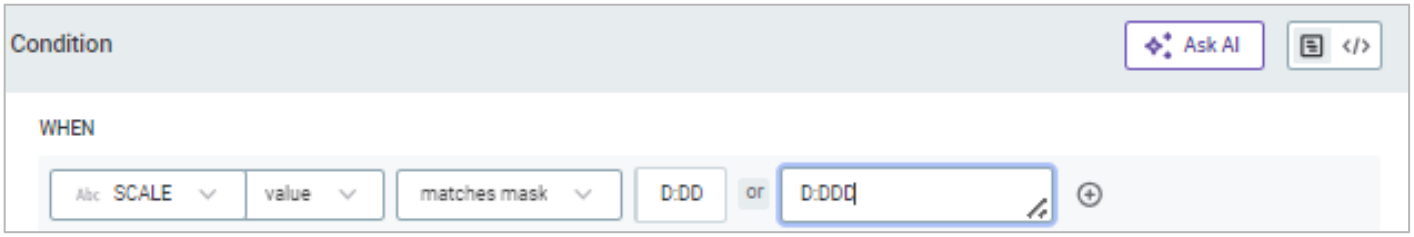
› Make sure the condition builder is selected.



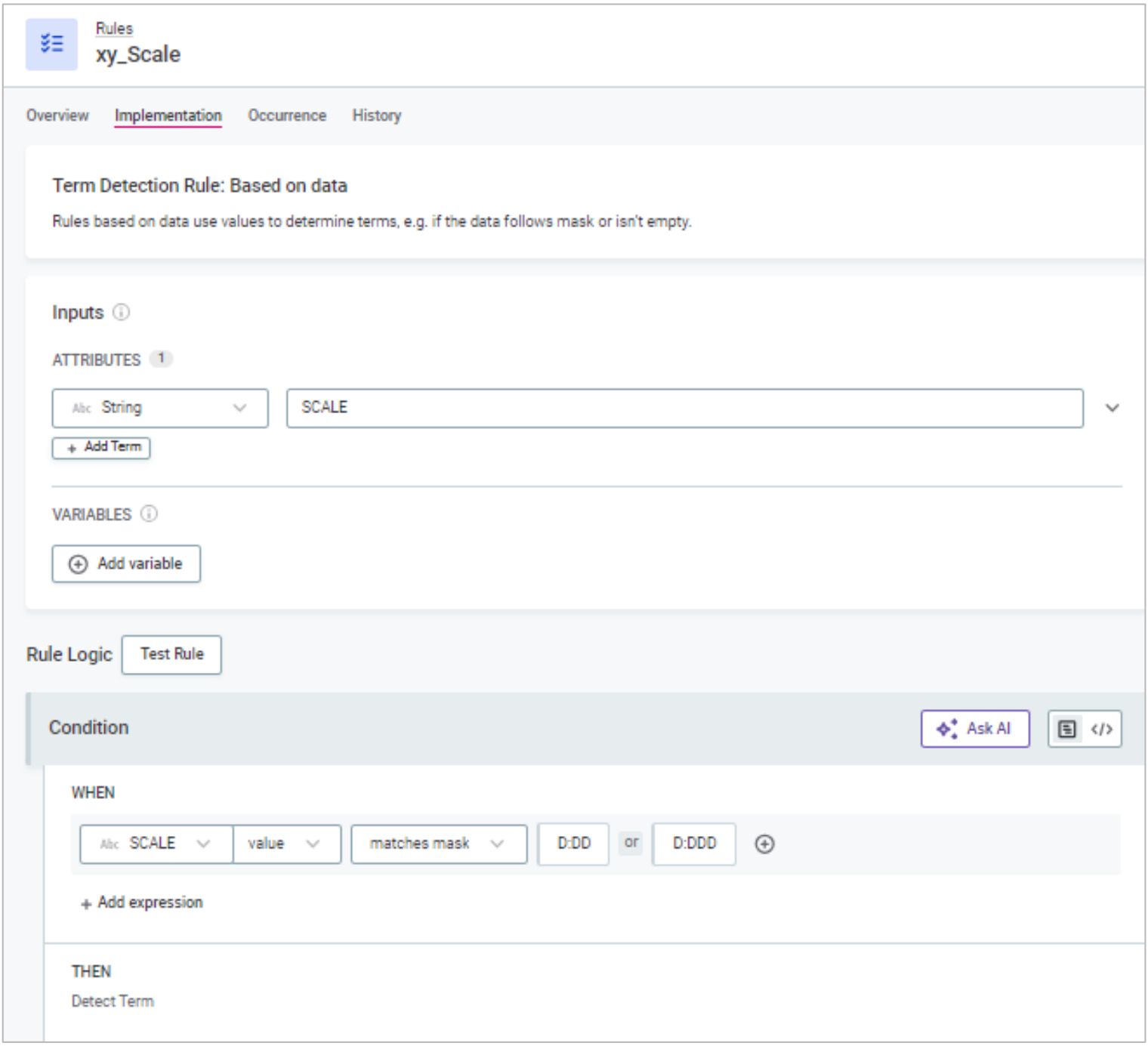
› Select the "**matches mask**" from the drop down list and provide both of the recognized patterns using the OR operator.

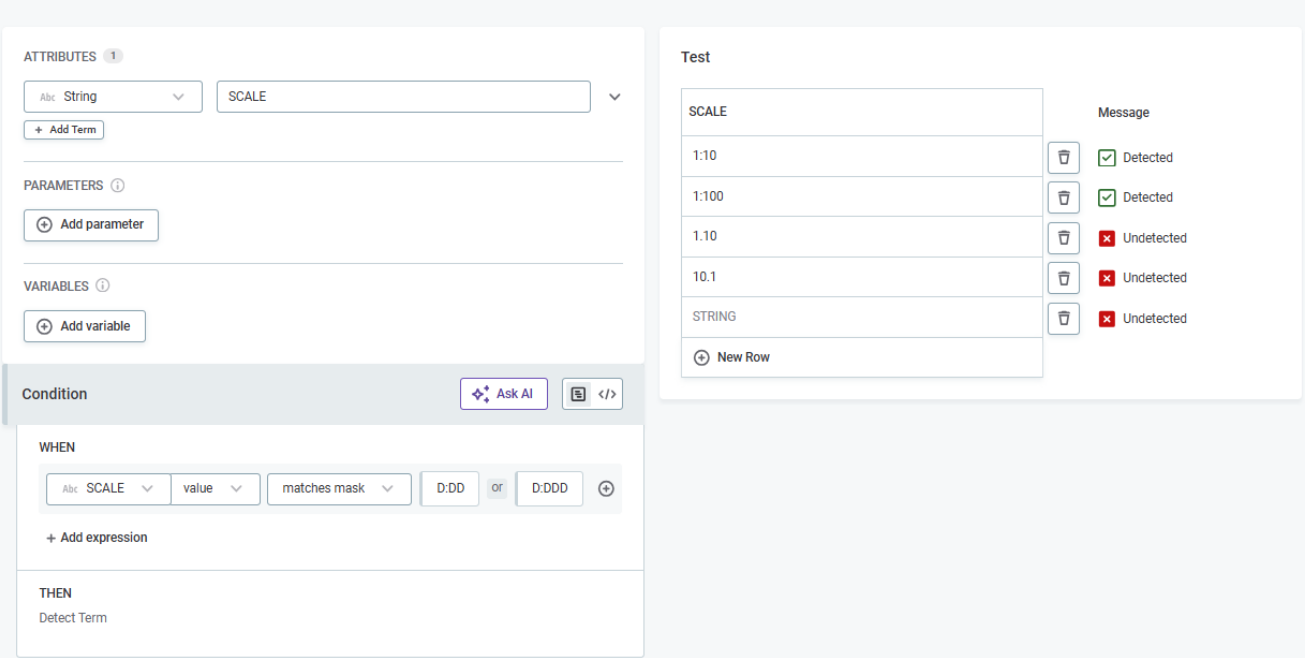
D:DD or D:DDD





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 1
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



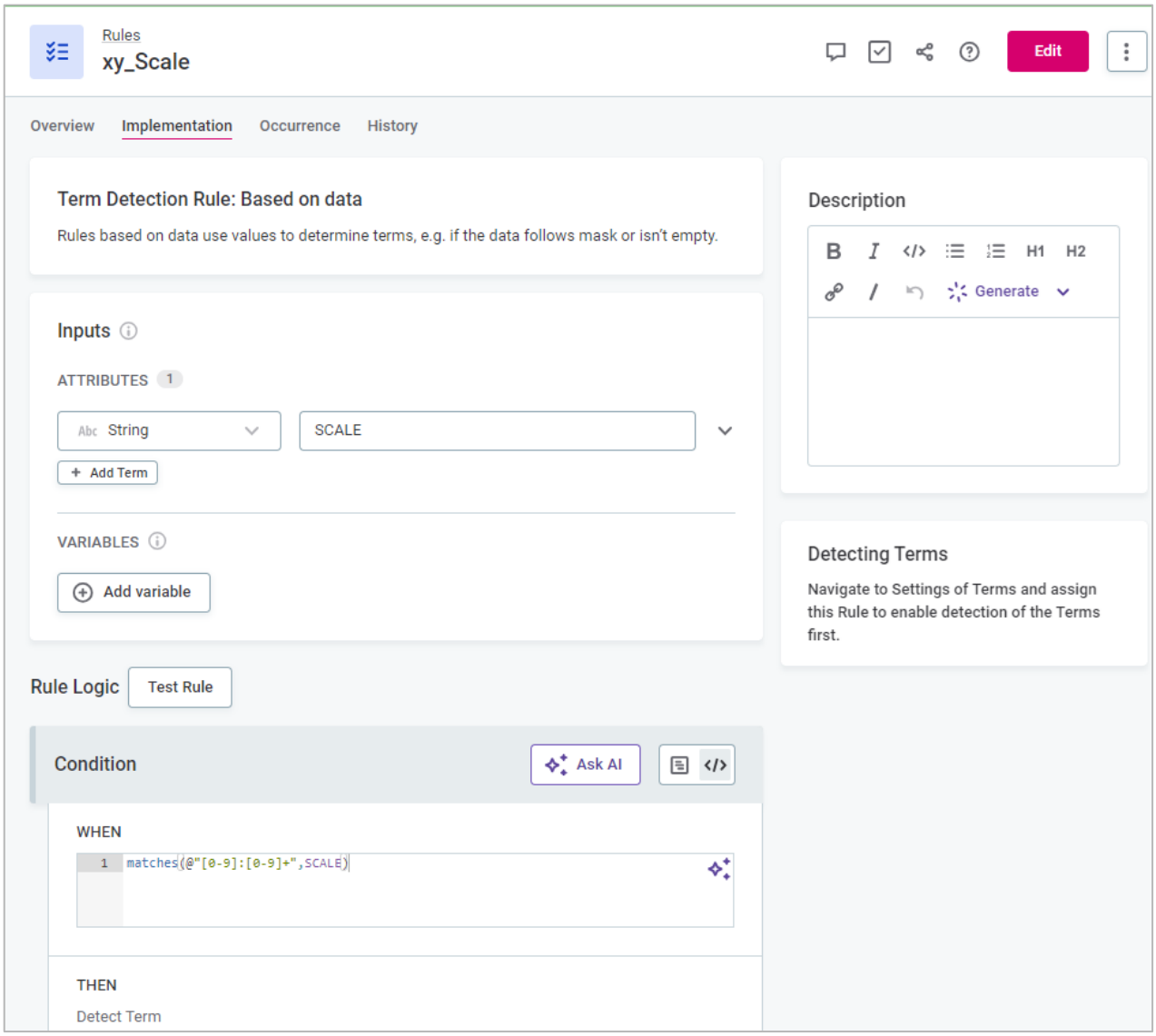
- Close the Test Rule window and **publish** the changes if you are happy with the expression.
- If you are happy with the test results, publish the rule.

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

NOTE

***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)



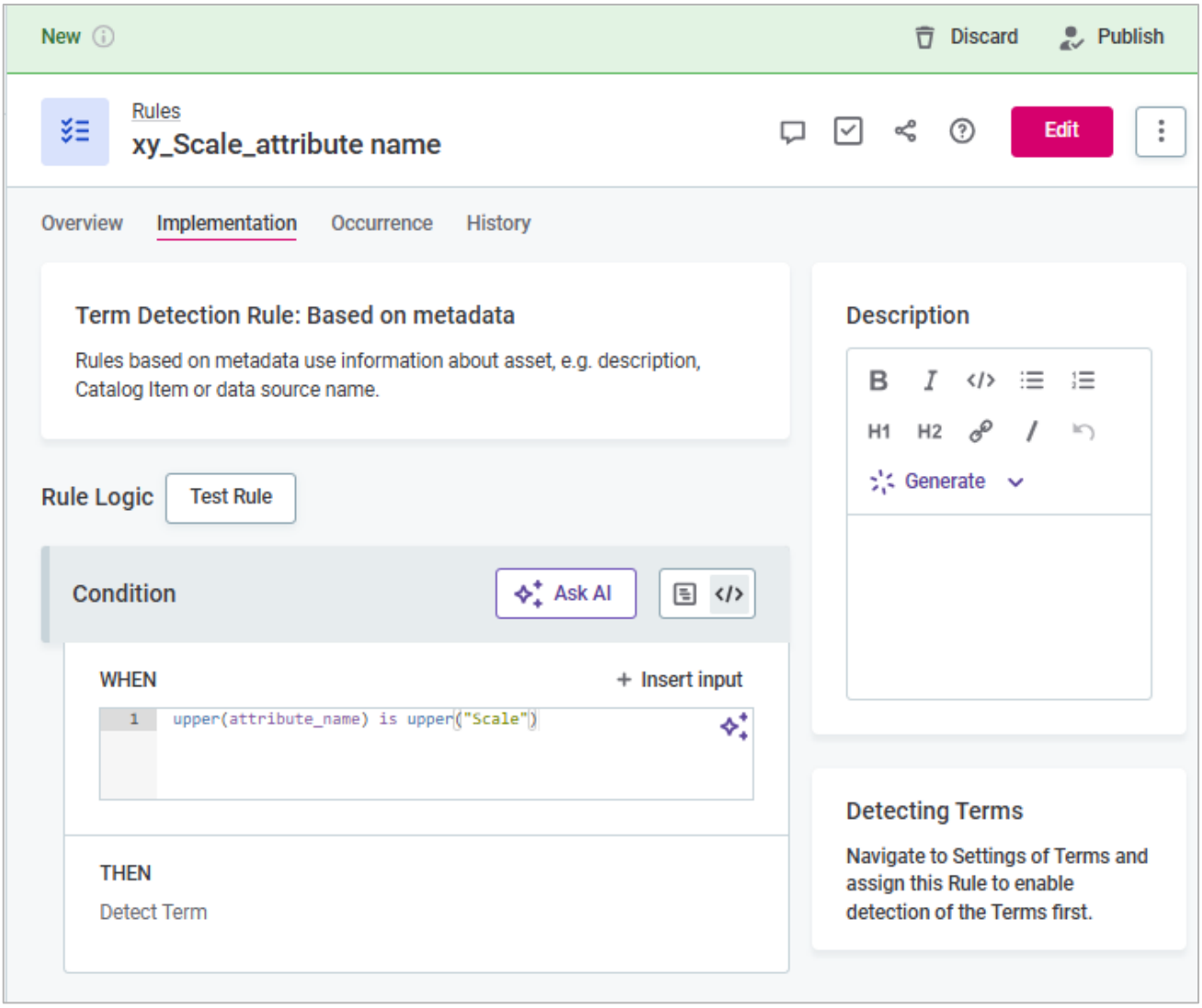
NOTE

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.



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**.

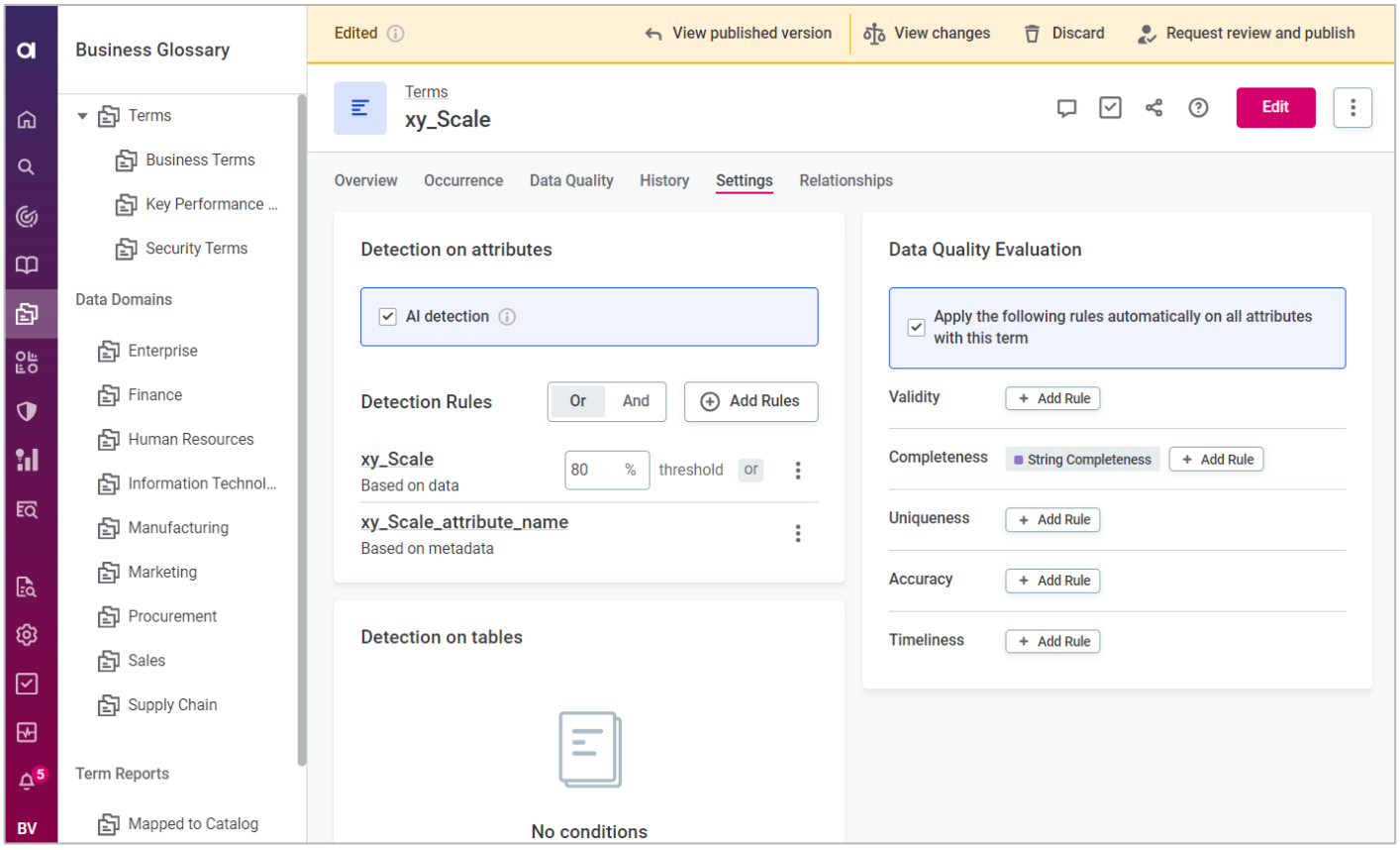
- Navigate to the **Business Glossary** and find your term **<prefix>_Scale**.
- Go to the **Settings** tab, the **Detection on attributes** section.
- Add the new rule **Scale** and set the threshold to 80%.

> **CAUTION**
>
> *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).



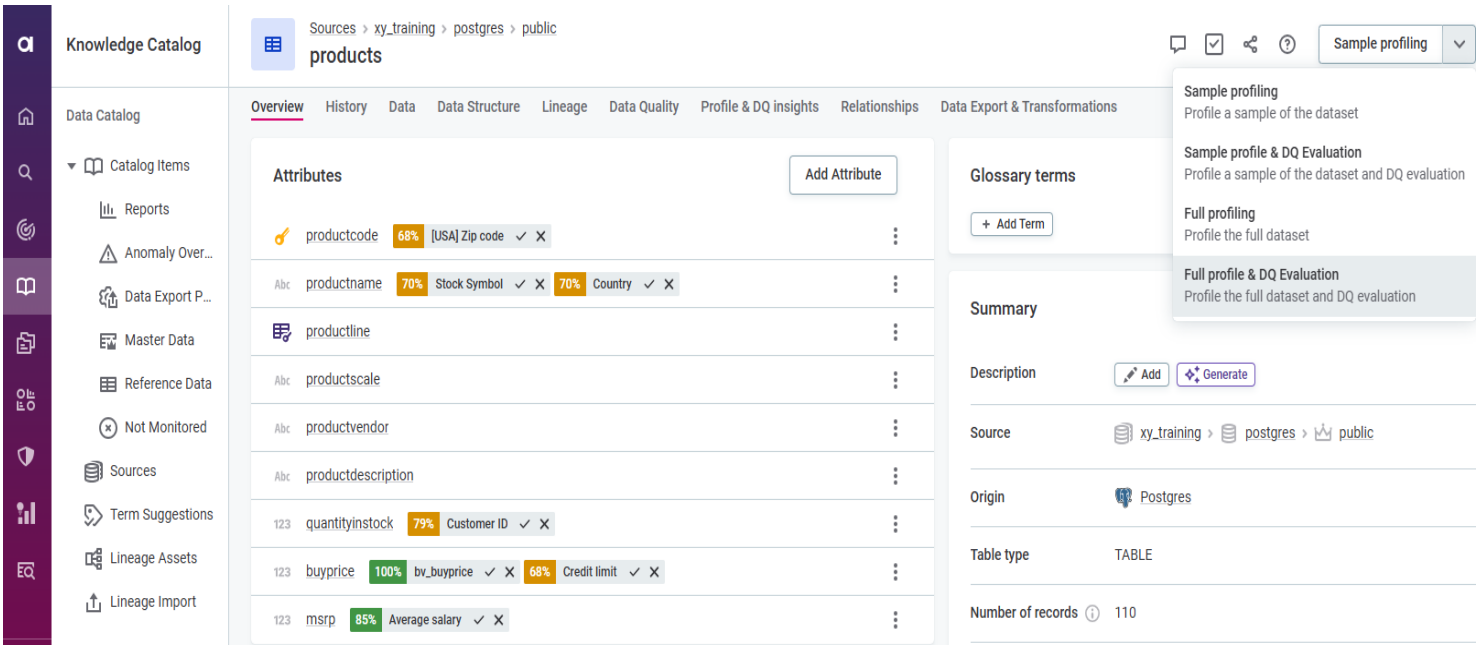
- Press the **Send for Approval** button and follow all the required steps until the changes are published.

6. Reprofiling after term creation

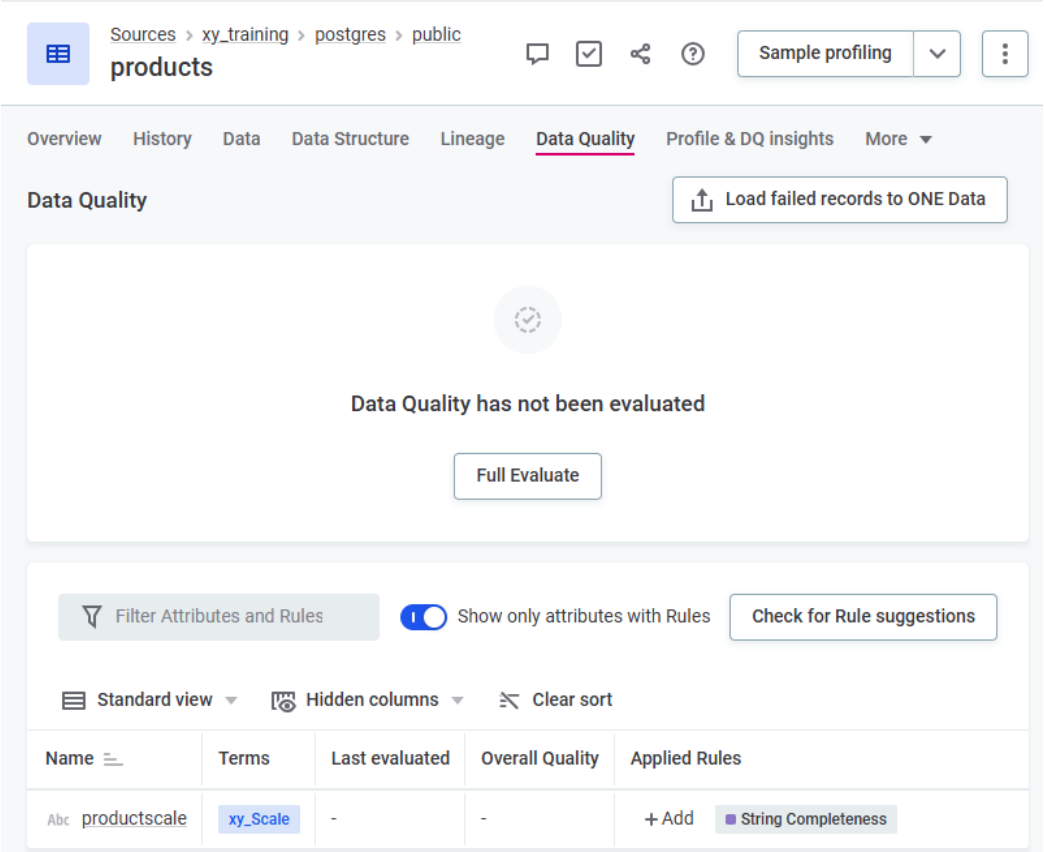
Now, it's time to reprofile your Catalog Item to see if the new term gets properly assigned.

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

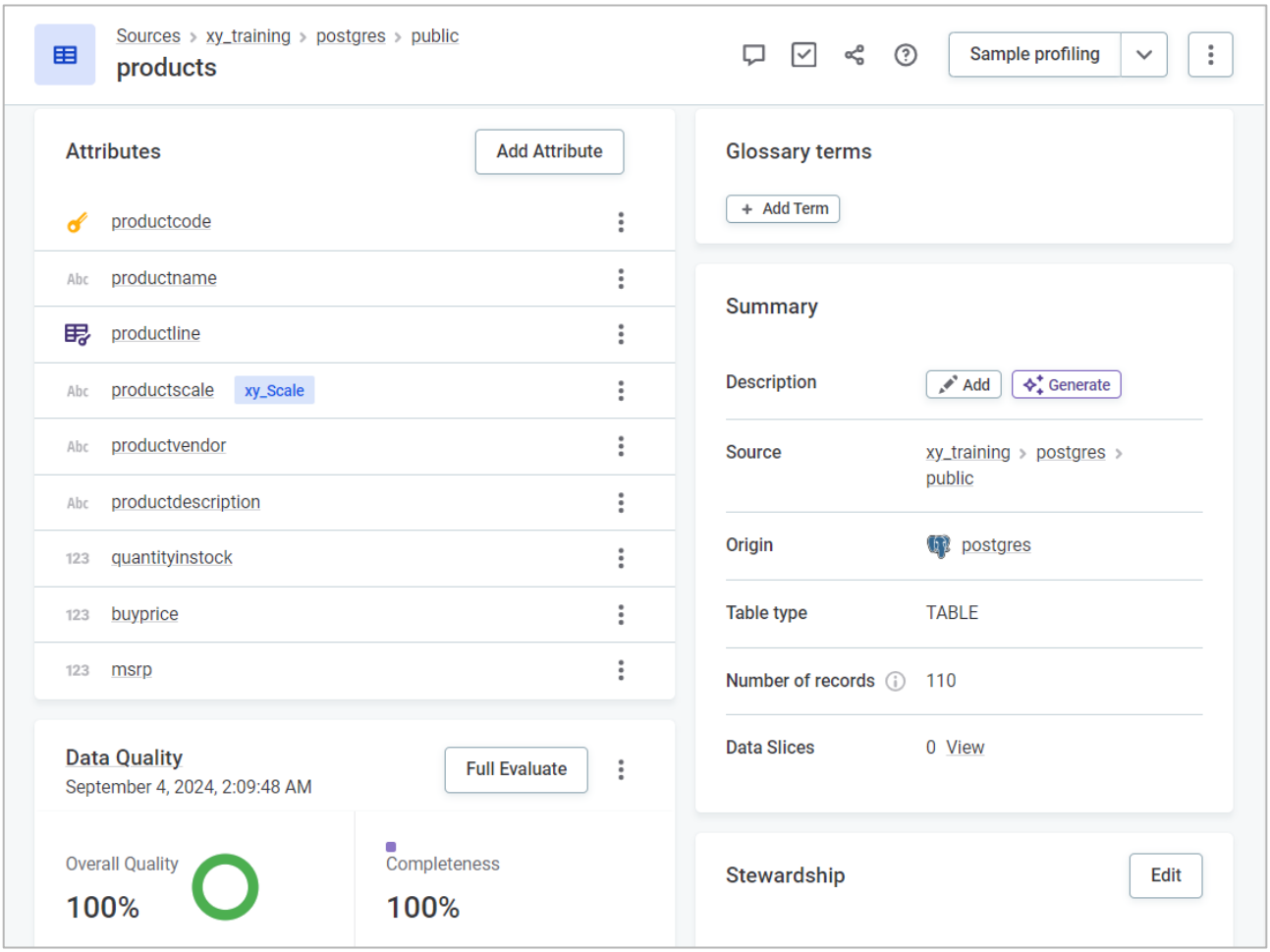
> NOTE
>
> *As you can see, there is no data quality available even though the Catalog Item has already been profiled; the data quality is only available for catalog items that have attributes with associated data quality rules.*



- 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.



› 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.



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.