



ONE – Workshop

Data Transformation Plans

Prepared for: v15.4

Prepared by: Ataccama

Dated: November 2024

Contents of the Document

Introduction	3
1 – Transforming a Catalog Item with Data Transformation Plans	4
Task 1.1- Create a Transformation Plan	4
Task 1.2 – Set Up Data Input and output	6
Task 1.3 – Configure transformation steps	8
Task 1.4 – Preview Data	13
Task 1.5 – Validate Plan	14
Task 1.6 – Configure Output Step	15
Task 1.7 – Run transformation plan	17
Conclusion	20


Introduction

Data Transformation Plans are visual tools for defining and executing data processing jobs. After configuring the plan to handle input data, defining the data flow, and specifying the output format, you can run the plan. Each step in the plan involves fundamental algorithms and processes data record-by-record.

There are two main use cases:

- Standalone transformation of catalog items.
- Transformation of data quality monitoring project results.

Data Transformation Plans can include steps like filtering, splitting, joining, transforming, and managing attributes. They also allow for nesting plans through the Embedded Transformation Plan step. After transformation, you can export the data to ONE Data, store it in ONE Object Storage, or use it in another plan. Features like real-time data preview and expression validation are also available.



In this example, We will create a plan that inputs the catalog item Customers, adds attributes, removes other attributes, filters out selected records, and then writes the output to a ONE Data table.

Available Steps

- Catalog Item input
- Input
- ONE Data reader
- File Output
- ONE Data writer
- Output
- Add attributes
- Condition
- Delete attributes

- Filter
- Embedded component
- Join
- Embedded transformation plan
- Split
- Transform data
- Union streams
- Monitoring project post processing input

1 – Transforming a Catalog Item with Data Transformation Plans

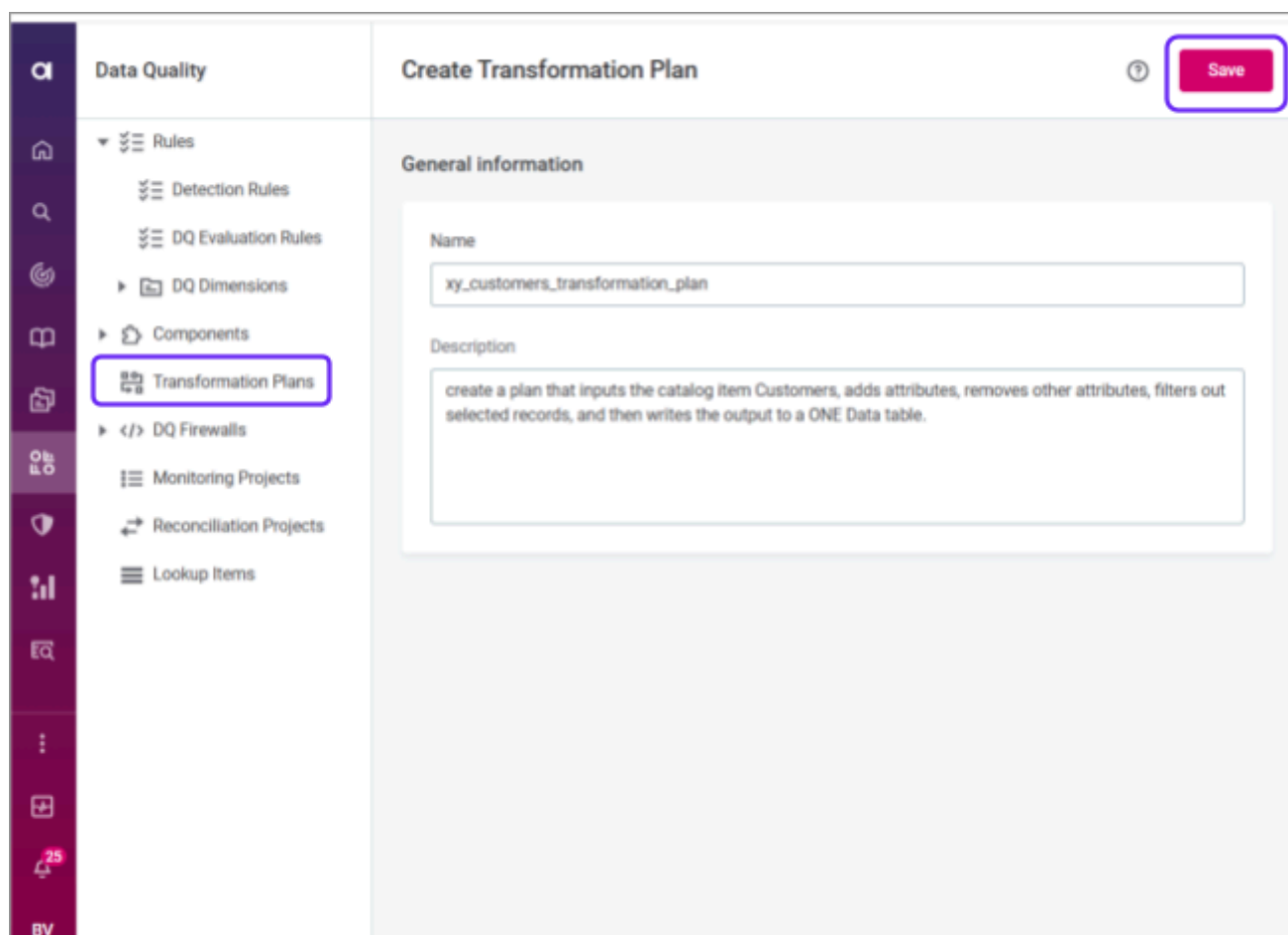
To transform data using Transformation plans:

- › Select your input data.
- › Build your plans using steps that can, for example, filter, transform, or join the data. (Note: These steps are similar to those in ONE Desktop, which leverage ONE Expressions. Each step represents a data processing job.)
- › Draw connections between steps to control the data flow.
- › Select your data output.
- › Run the transformation to trigger the flow of data through the plan.

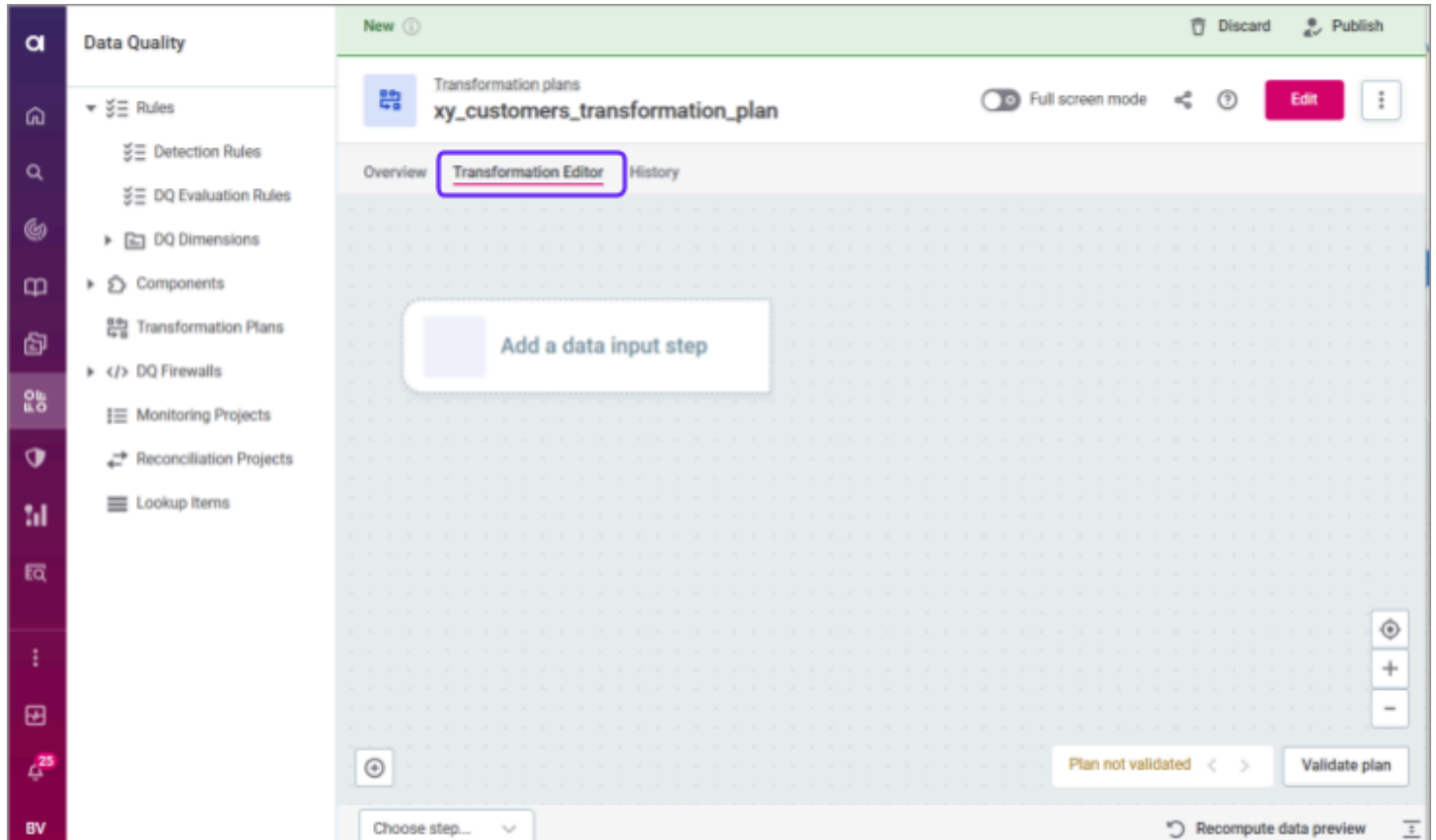
Task 1.1- Create a Transformation Plan

Let's start with creating a new transformation plan and add data input:

- › Navigate to the **Data Quality > Transformation Plans** section.
- › Create a new plan and fill in its **Name** – '**<prefix>_customers_transformation_plan**' and optionally some Description if needed.



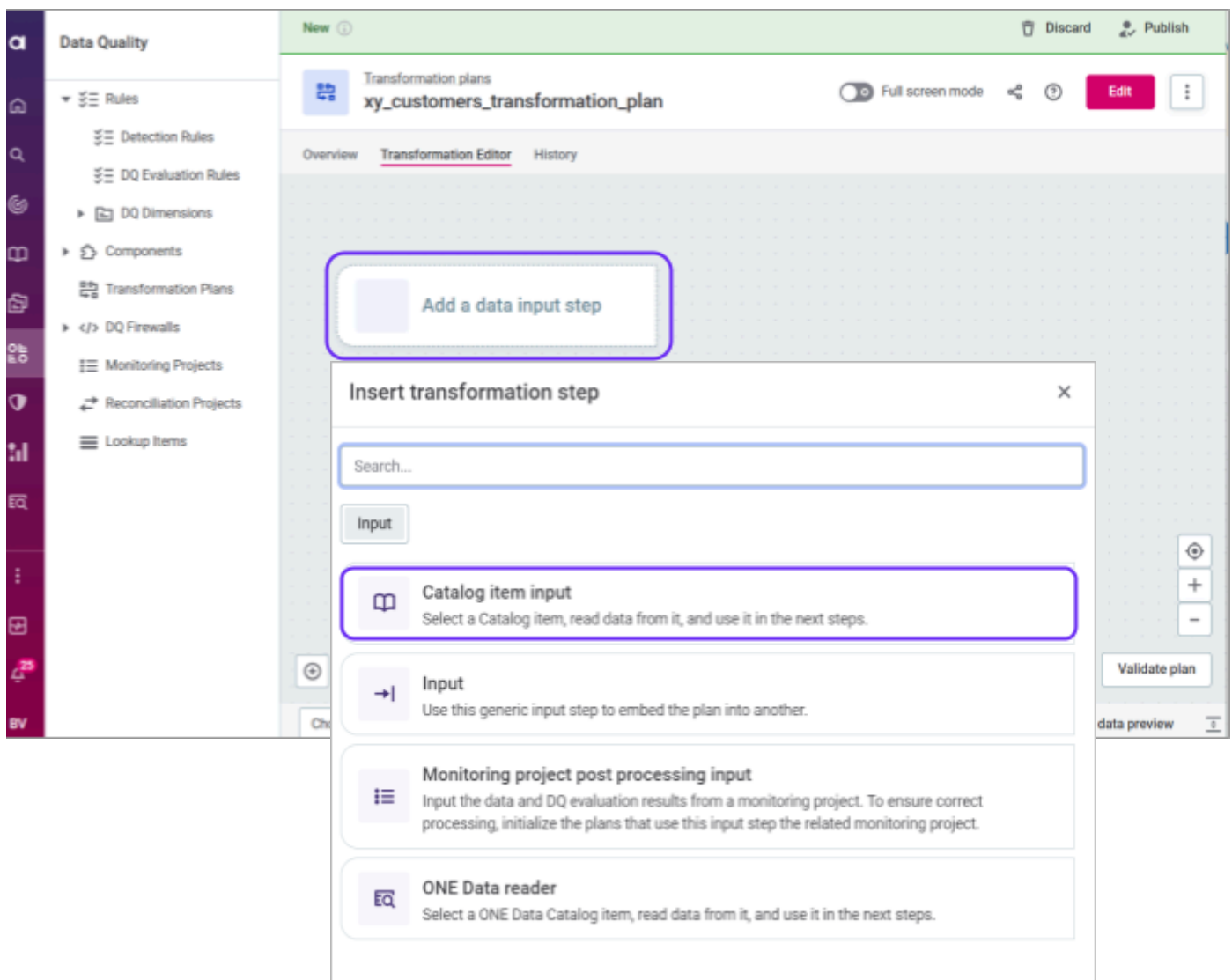
- › **Save** your changes and you will be redirected to the **Transformation Editor** tab.



Task 1.2 – Set Up Data Input and output

Once the plan is created, we need to add data input:

- › In **Transformation Editor** tab and Click on the Add a data input step and add Catalog item input.



- › Click anywhere on the blue banner to maximize or minimize a step and access the three dots menu.
- › In the **Catalog item**, select the catalog item you would like to use as an input. In our case, this is '**customers**'.

New

DiscardPublish

Transformation plans

xy_customers_transformation_plan

Full screen mode

Edit

Overview

Transformation Editor

History

Catalog item input step

Catalog item input

Select catalog item

Plan not validated

Validate plan

Catalog item ...

Compare

Recompute data preview

Compute data preview

To see results of data transformation compute data preview.

Compute data preview

Select catalog item

Fulltext customers

Clear sort

	Catalog item	# Attributes	# Records	Origin	Location
<input type="radio"/>	ACTIVE_CUSTOMERS_SAMPLE	21	555	Oracle Demo Content	Sources > Oracle > INITDB > BUSINESS_OPERATIONS
<input type="radio"/>	CHURN_CUSTOMERS	14	407	DO_Snowflake	Sources > Finance DO > CONTENT_ENV > FINANCE_D
<input checked="" type="radio"/>	customers	15	122	postgres	Sources > xy_training > postgres > public
<input type="radio"/>	customers	15	122	postgres-testdata-svc	Sources > pgs_testdata > tcd > transactional_customer_data
<input type="radio"/>	CUSTOMERS	23		BusinessOPS	Sources > Analytics Snowflake DWH > DEMO_ENV > PUBLIC
<input type="radio"/>	customers_h	17	0	postgres-testdata-svc	Sources > pgs_testdata > tcd > transactional_customer_data
<input type="radio"/>	customers_na	17	1,003,505	tcd_north_america	Sources > pgs_testdata > tcd_north_america > tcd_large
<input type="radio"/>	customers_prime	16	1,059	tcd_north_america	Sources > pgs_testdata > tcd_north_america > tcd_large
<input type="radio"/>	CUSTOMERS_SAMPLE	23	555	DO_Snowflake	Sources > Finance DO > CONTENT_ENV > FINANCE_D

Select

Cancel

- › **To add more steps**, use the plus icon either in the connection or on the step output or select **"Add step."** You can also add steps through the three dots menu, and duplicate or delete them as needed.



NOTE You can utilize keyboard shortcuts (**delete and backspace**) to remove steps. Additionally, on Windows, use the **"Insert"** key to add a new step.

- › In this task, we need to add the following steps:
 - **Add attributes:** To create a new attribute, **"full_name,"** in the output by concatenating the existing input attributes first name and last name.
 - **Delete attributes:** To remove unnecessary attributes, retaining only the key attributes in the output.
 - **Filter:** To filter out all records not originating from a specified country; in this instance, everything except those from the United States.

Task 1.3 – Configure transformation steps

Initially, we aim to generate **a new attribute, "full_name,"** by concatenating the current input attributes, **"contactfirstname"** and **"contactlastname."**

- › To accomplish this, we employ the **"Add attributes"** step with the following configuration:
 - **Name:** A custom name for the step (optional).
 - **Name and Type:** Specifies the attribute name and its data type for the output. In our case, this is **"full_name"** and **"String."**

The screenshot displays the Data Quality Transformation Editor. The main workspace shows a transformation plan named "xy_customers_transformation_plan" with a "customers" catalog item input. A purple box highlights the plus icon next to the input. An "Insert transformation step" dialog is open, showing a search bar and three tabs: "Transform", "Output", and "All". The "Add attributes" step is highlighted with a purple box, showing the description "Add new attributes". Other visible steps include "Condition" and "Delete attributes".

- **Expression:** Utilizes a **ONE expression** to define the new attribute. In our scenario, the expression is **trim(contactfirstname+' '+contactlastname)**, which also trims the inputs before combining them.

The screenshot shows the 'Transformation Editor' for a plan named 'xy_customers_transformation_plan'. A 'customers' catalog item input is connected to a component labeled 'Adding new attributes'. This component has a table for defining new attributes:

Name	Type	Expression
full_name	String	1 trim(contactfirstname+' '+contactlastname)

Below the table is a button 'Add Expression attribute'. At the bottom of the editor, there is a 'Compute data preview' button and a status indicator 'Plan not validated'.

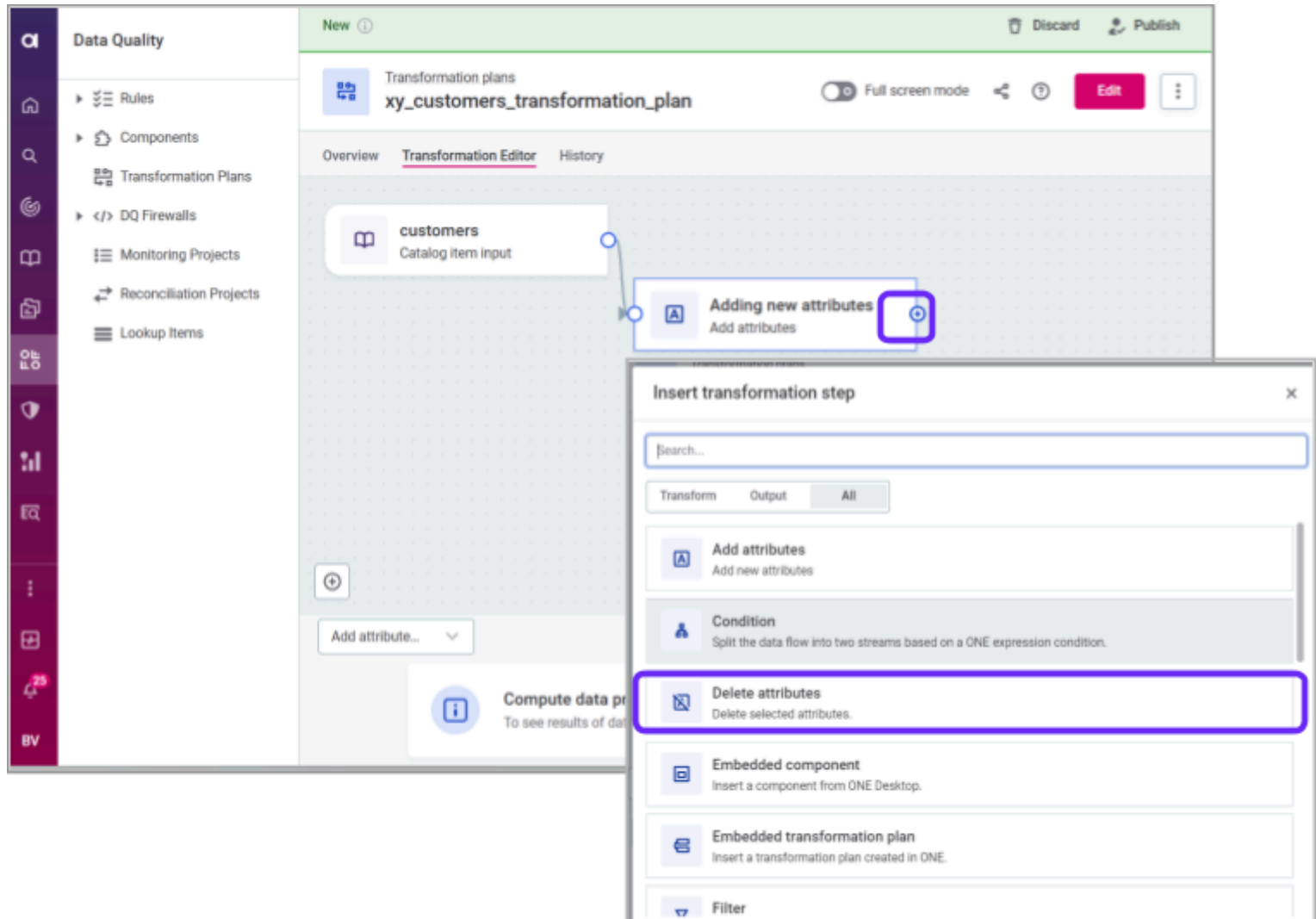
The updated plan should look like:

This screenshot shows the updated transformation plan. The 'customers' catalog item input is now connected to the 'Adding new attributes' component. The 'Adding new attributes' component is highlighted with a blue border. The 'Plan not validated' status is still present at the bottom.

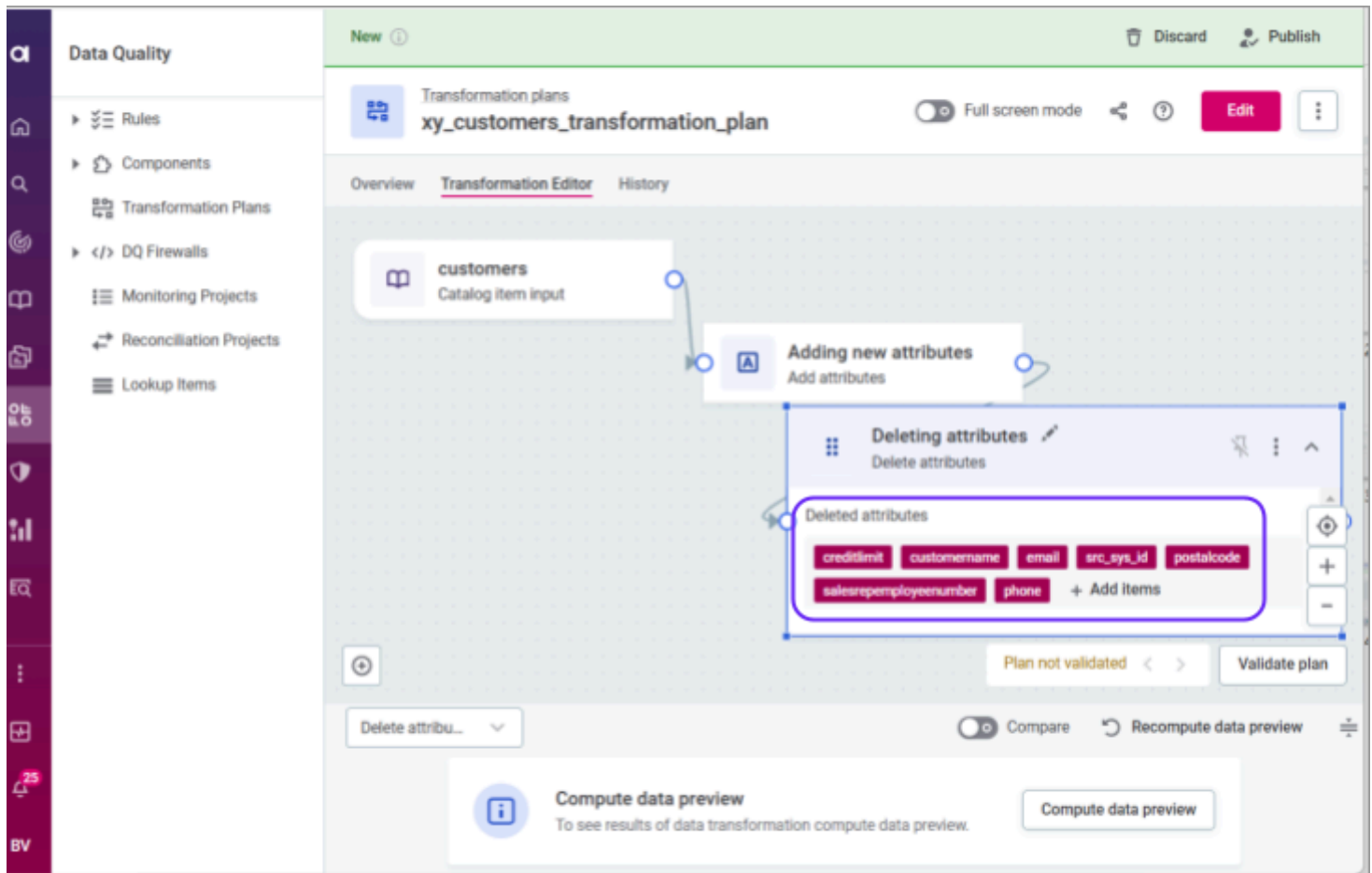
There are some attributes in our catalog item, **customers**, that are unnecessary in the output.

- To eliminate these attributes, we employ the **"Delete attributes"** step with the following configuration:

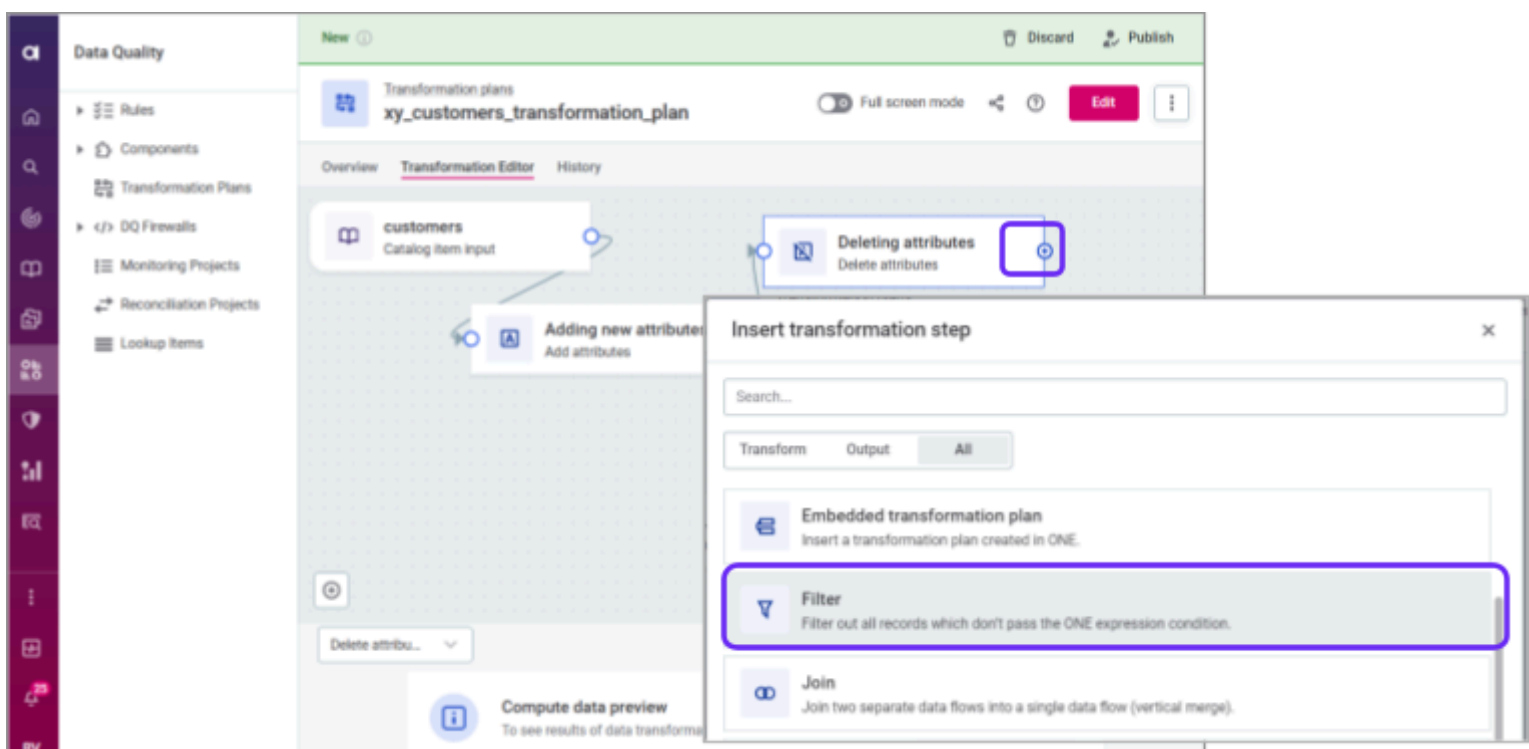
Name: A custom name for the step (optional).



- Use the **+Add option** to select attributes from the available options. In our instance, these include **creditlimit**, **customername**, **email**, **phone**, **postalcode**, **salesemployee number**, and **src_sys_id**.



- ▶ To filter out all records which are not from the United States, we can use the **Filter step**. It is configured as follows:
 - **Name:** A custom name for the step (optional).
 - **Condition:** The filter condition is specified using ONE Expressions. In our case, this is going to be **attribute name = 'value'**, resulting in **country = 'USA'**.



The screenshot shows the 'Transformation Editor' for a plan named 'xy_customers_transformation_plan'. The workflow consists of four steps connected by arrows: 'customers' (Catalog item input) → 'Adding new attributes' (Add attributes) → 'Deleting attributes' (Delete attributes) → 'Country USA' (Filter). The 'Country USA' filter step is expanded, showing a condition: '1 country = 'USA''. Below the workflow, there is a 'Filter data step' dropdown and a 'Compute data preview' button. The status bar at the bottom indicates 'Plan not validated' and a 'Validate plan' button.

Finally, add **data output**. In this example, we are using the **ONE Data writer** step.

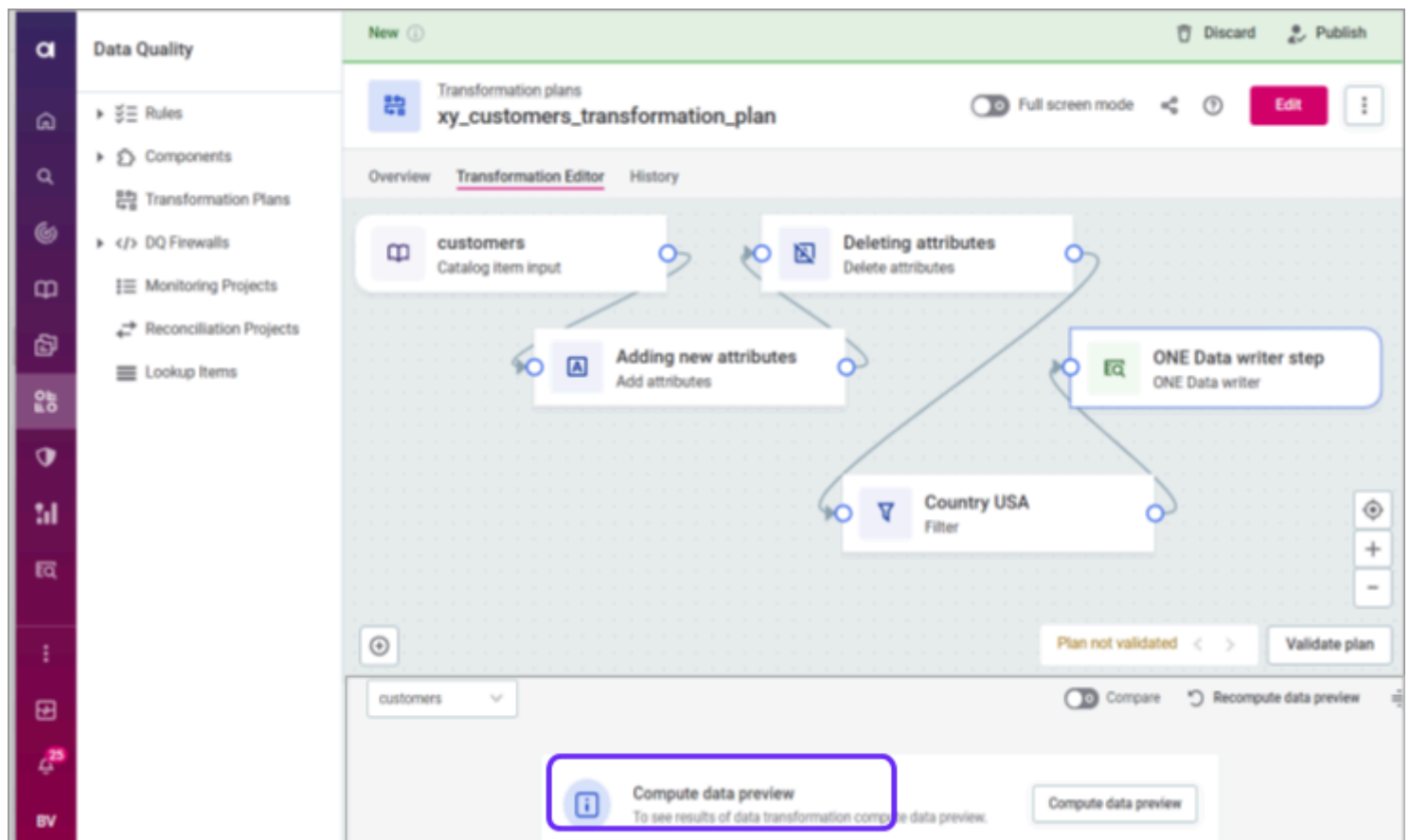
The screenshot shows the 'Insert transformation step' dialog. The dialog has three tabs: 'Transform', 'Output', and 'All'. The 'Output' tab is selected, showing three options: 'Data remediation output', 'File output', and 'ONE Data writer'. The 'ONE Data writer' option is highlighted with a red box.

Task 1.4 – Preview Data

At any point, you can preview the data resulting from the chosen step and confirm that it functions as intended.

Please note that the preview is not automatically recalculated after modifications, so it is necessary to trigger this manually.

- Click on '**Compute data Preview**'. You can also select the step either from the drop-down list or by clicking directly on the step on the canvas.



- Your result page looks like this:

The screenshot shows the 'Deleted attributes' result page. A dropdown menu in the top left is set to 'customers'. The table displays the following data:

123	creditlimit	Abc	addressline2	123	salesrepeployeeenumber	Abc	state	Abc	contactfirstname	Abc	phone
	21800				1378				Carine		48.32.2555
	71800				1166		NV		Jean		7025551838
	117300		Level 3		1611		Victoria		Peter		03 9520 4555
	118200				1378				Janine		48.67.8555
	81700				1584				Jonas		07-98 9555
	210500				1165		CA		Susan		4155551450
	0								Zbyszek		(26) 642-7555
	59700				1584				Roland		+49 69 66 90 2555
	64600				1165		CA		Julie		6505555787
	114900				1323		NY		Kwai		2125557818



Preview is not visible for output steps, but you can select the preceding steps and check their outputs instead.

Task 1.5 – Validate Plan

You can also validate the plan during the process by choosing **"Validate plan"** in the bottom right corner of the canvas.

id	phone	state	contactlastname	addressline1
1a	40.32.2555		Schmitt	54, rue Royale
	7825551838	NV	King	8489 Strong St.
	03 9520 4555	Victoria	Ferguson	636 St Kilda Roa
	40.67.8555		Labrun	67, rue des Cinq
	07-98 9555		Bergulfsen	Erling Skakkes g
	4155551450	CA	Nelson	5677 Strong St.
	(26) 642-7555		Piestrzeniewicz	ul. Filtrowa 68

Any errors, such as incorrect expression conditions, will be brought to your attention.

In case validation issues persist, selecting **"Run transformation"** will result in the failure of the data processing job.

try	Abc	phone	Abc	state	Abc	contactlastname	Abc	addressline1
		48.32.2555				Schmitt		54, rue Royale
		7825551838		NV		King		8489 Strong St.
ia		83 9528 4555		Victoria		Ferguson		636 St Kilda Roa
		48.67.8555				Labrune		67, rue des Cinq
		87-98 9555				Bergulfsen		Erling Skakkes g
		4155551458		CA		Nelson		5677 Strong St.
		(26) 642-7555				Piestrzeniewicz		ul. Filtrowa 68

Task 1.6 – Configure Output Step

We want to write the data to a new table in ONE Data, using the ONE Data writer. It is configured as follows:

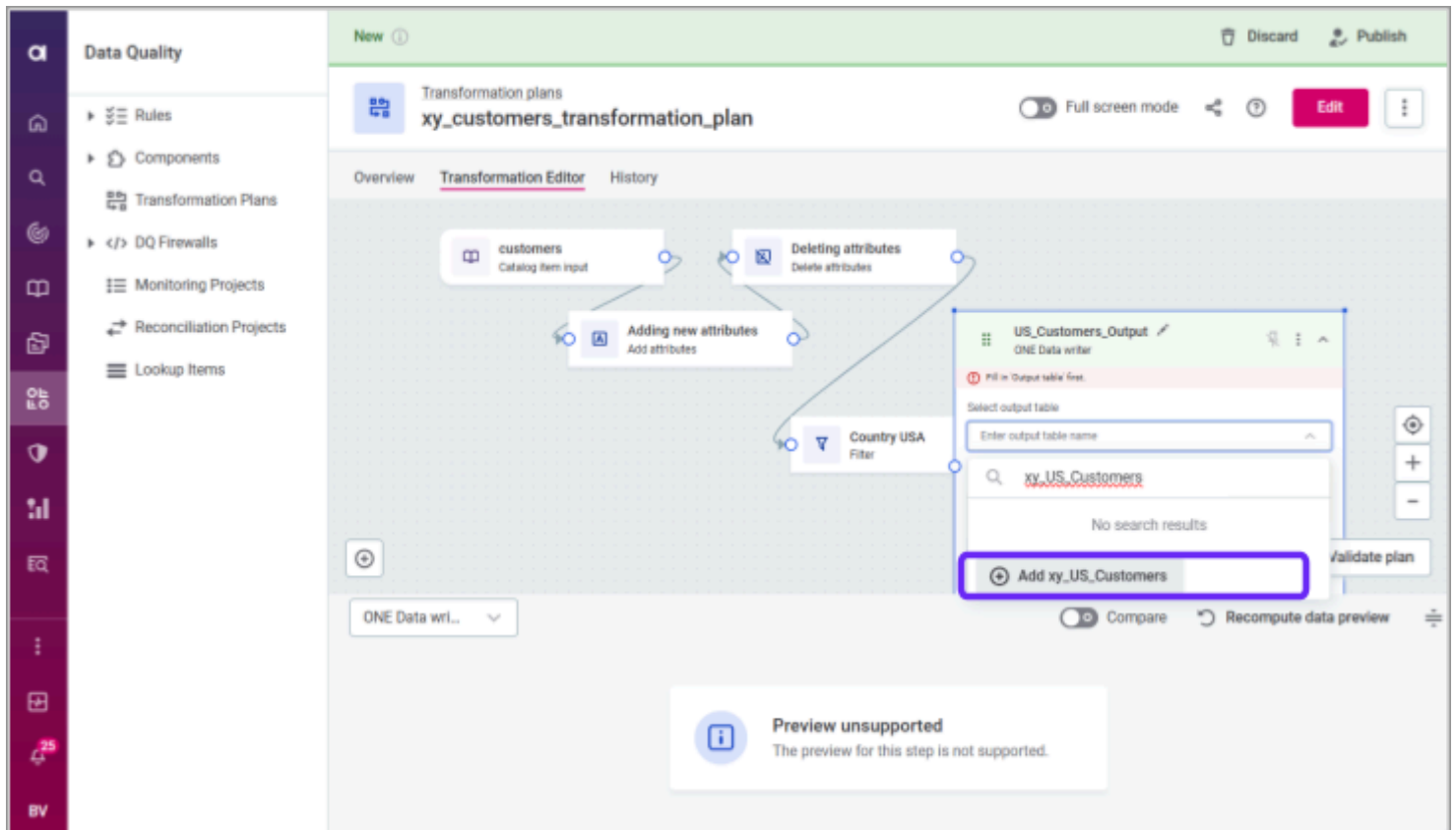
- › **Name:** A custom name for the step (optional).
- › **File Output Step :** Specify the desired name for the catalog item generated in ONE Data by this step.

Name :

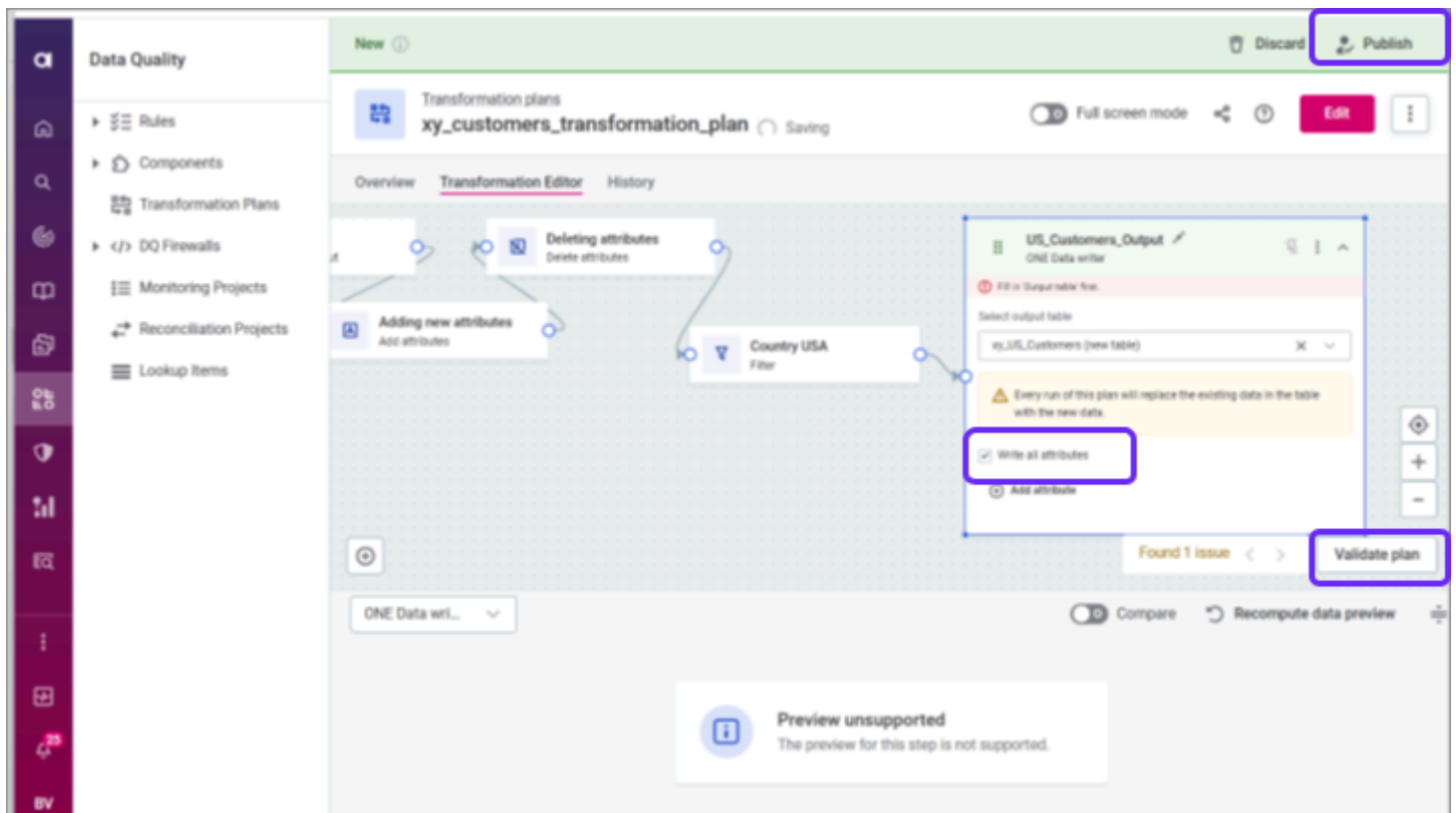
US_Customers_Output

ONE Data table Name :

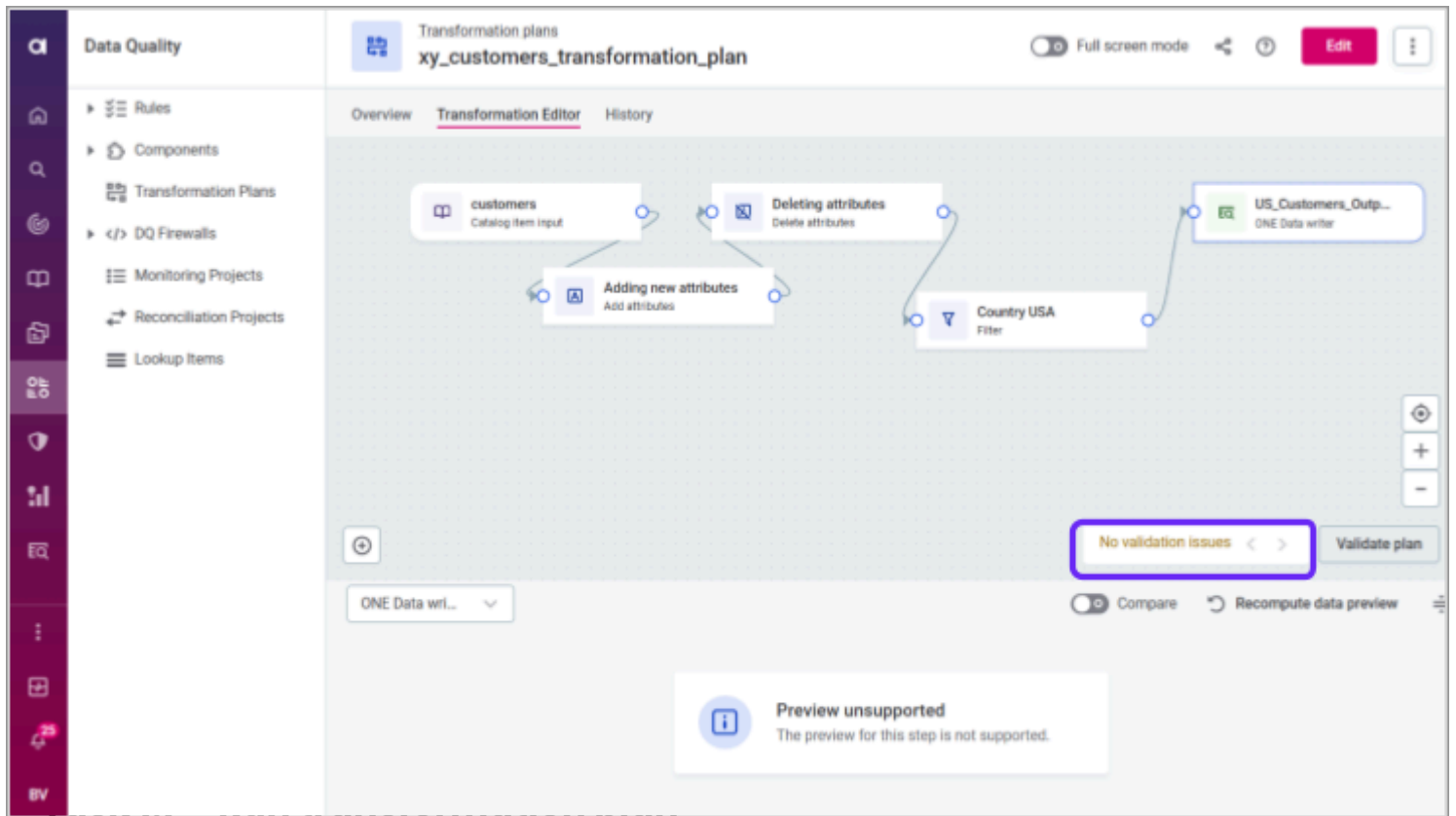
<name_prefix>_US_Customers



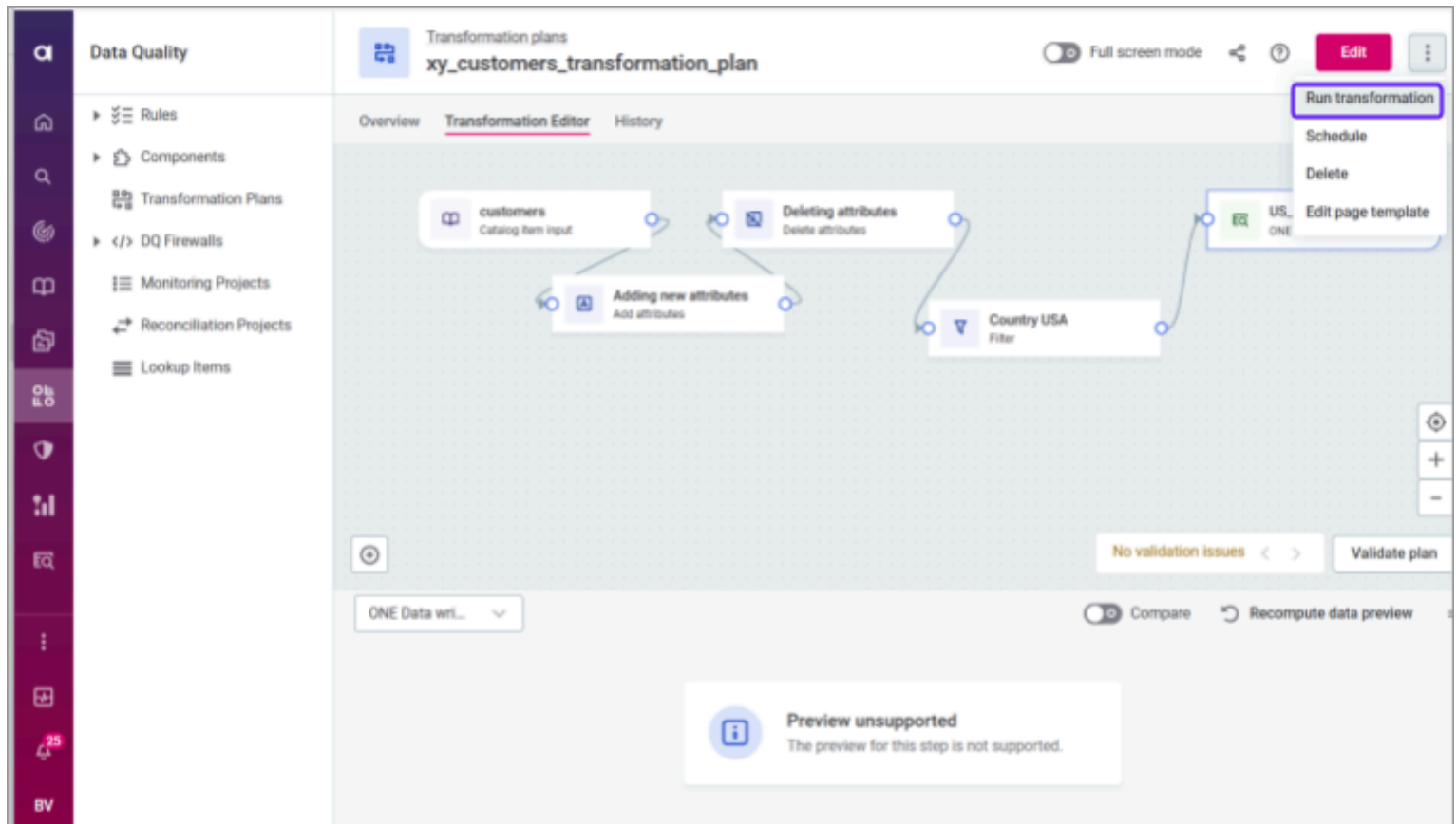
- › Select **Write all attributes** to include all attributes which are inputs to the step, without having to add them manually.

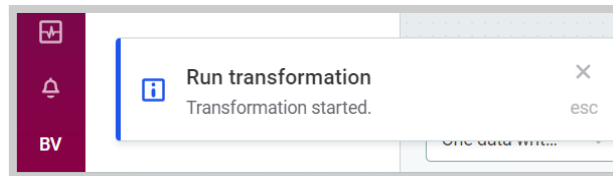



- › **Publish** the Changes and Click the **Validate Plan** button again to ensure there are **no validation issues**.



- › Navigate to the three dots menu on the top right corner and choose **"Run transformation."**





 Transformation plan jobs can be found by going to **Processing center > Base jobs > Transformation plan jobs**, or by searching for job Type DQC.

Latest jobs

Scheduled jobs 13

Mute ☐ i

less than a minute ago

DQC of DQ Plan 'xy_customers_transformation_plan'

Running

about 1 hour ago

DQC_PREVIEW of DQ Plan 'xy_customers_transformation_plan' (preview)

Finished

about 8 hours ago

SCHEMA_CHECK of Oracle Demo Content

Failed

about 8 hours ago

SCHEMA_CHECK of Oracle Demo Content

Finished

Open Processing Center

> Navigate to the **ONE Data** section on the left panel and search for the '**<name_prefix>_US_Customers**' table.

ONE Data Tables Create table

Standard view

Hidden columns

	Name	Terms	Anomalies	Overall Quality	# Attributes	# Records	Origin	Location
	xy_US_Customers	-		-	10		one-data	one-dat
	Turkey city	-		-	2	167	one-data	one-dat
	Saudi Arabia city	-		-	2	87	one-data	one-dat
	Airports	International Airport Code		-	2	1,942	one-data	one-dat
	HTTP Request	-		-	2	9	one-data	one-dat
	R&D	First name Surname E-mail +2		<div><div></div>23%</div>	7	1,000	one-data	one-dat
	HTTP Response Code	-		-	2	48	one-data	one-dat
	Italy city	ITA City		<div><div></div>>99%</div>	2	144	one-data	one-dat
	Conference Attendees	First name Surname E-mail +4		<div><div></div>52%</div>	11	199	one-data	one-dat
	Indonesia city	-		-	2	230	one-data	one-dat
	Country	Country ISO-2 Country Code +1		<div><div></div>66%</div>	5	240	one-data	one-dat

- Click on the '**<prefix_name>_US_Customers**' table and verify the results.

ONE Data Tables

xy_US_Customers

🗨️📧🔗🕒⋮

🏠

Data

Overview

History

Lineage

Data Quality

Profile & DQ insights

Relationships

Data Export & Transformations

🔍

Filter

🔄 Evaluate Data Quality

🔖

<input type="checkbox"/>	astname	⋮	Abc email	⋮	Abc addressline1	⋮	Abc city	⋮	Abc country	⋮	Abc full_name	⊕
1			signalgiftstores@gmail...		8489 Strong St.		Las Vegas		USA		Jean King	
2			minigiftsdistributorslt...		5677 Strong St.		San Rafael		USA		Susan Nelson	
3			miniwheelaco@gmail.co		5557 North Pendale Stre...		San Francisco		USA		Julie Murphy	
4			landoftoysinc@gmail.com		897 Long Airport Avenue		NYC		USA		Kwai Lee	
5			musclenachineinc@gmail...		4092 Furth Circle		NYC		USA		Jeff Young	
6			diecastclassicsinc.@gma...		7586 Pompton St.		Allentown		USA		Kelvin Leong	
7			technicsstoresinc.@gmai...		9408 Furth Circle		Burlingame		USA		Juri Hashimoto	
8			americansouvenirsinc@gm...		149 Spinnaker Dr.		New Haven		USA		Keith Franco	
9			cambridgecollectablesco...		4658 Baden Av.		Cambridge		USA		Jerry Tseng	
10			giftdepotinc@gmail.com		25593 South Bay Ln.		Bridgewater		USA		Julie King	
11			vitachromeinc@gmail.com		2678 Kingston Rd.		NYC		USA		Michael Frick	
12			automotoclassicsinc.@gm...		16780 Pompton St.		Brickhaven		USA		Leslie Taylor	
13			onlinemini collectables@...		7635 Spinnaker Dr.		Brickhaven		USA		Miguel Barajas	
14			toys4grownups.com@gmail...		78934 Hillside Dr.		Pasadena		USA		Julie Young	

⊕ Add record

Total records: 36

Conclusion

We've reached the conclusion of this workshop! Throughout this session, we've successfully created data transformation plans for catalog items. In this way, data transformation plans take precedence over post-processing plans.

Congratulations! You are now ready to embark on the development of your own Data Transformation Plan.