



# ONE - Workshop

DQ Advanced – Data Slices & Post-Processing

Prepared for: v15.4.x

Prepared by: Ataccama

Dated: November 2024

## Contents of the Document

Introduction	3
1. Enabling Data Slices in your Monitoring Project	3
2. Exporting the Monitoring project results for post-processing	7
2.1. Create a Post-processing plan to export Monitoring Results	7
2.2. Create a Transformation Plan to Export Monitoring Results	12
Conclusion	16

# Introduction

In this workshop, you will practice running monitoring on data slices. In addition you will explore how to export monitoring project results.

## Tasks

We have two tasks in this workshop. The first one focuses on updating an existing Monitoring project by adding a new Catalog Item and monitoring its data quality on a specific slice. The second one is about exporting results of monitoring for further post-processing.

- Installing the ONE Desktop application and connecting to the ONE Web Application is a prerequisite for task 2\_1.



*The connection and integration with the ONE Desktop application were already covered in the previous workshop focusing on the **ONE Desktop Integration**.*

### 1. Enabling Data Slices in your Monitoring Project

In the first task of this lab, we will update an existing Monitoring Project by adding a new item and configuring data slices for it.

Specifically, we want to add the '**orders**' item and run monitoring on the records with '**Shipped**' status.

Let's start by navigating to your **Monitoring Project**:

- › Navigate to the **Data Quality** section - **Monitoring Projects**.
- › Select your existing monitoring project called '<prefix>\_Training Project'.

Name	Terms	Items	Latest run	Stewardship
KS_Training project		2	about 23 hours ago	

- › Navigate to the **Configuration & Results** tab.
- › Click the **Add** button located on the right-hand side.
- › Select the '<prefix>\_Orders' table that you created in the "Desktop Integration" workshop.

If done, let's start enabling **Data Slices** on it:

- › In the **Configuration & Results** tab, open the **<prefix>\_Orders** CatalogItem.
- › Locate the “**Run Monitoring on Data Slices**” toggle and switch it **on**.

Structure ⓘ  
⚙️ No checks applied

Anomaly Detection ⓘ  
⚙️ No checks applied

Data Quality ⓘ  
⚙️ No checks applied

☒ Run Monitoring project on data slice

⏏️ Filter Attributes and Rules

All Attributes ▾

Standard view ▾ Hidden columns ▾

Name	Terms	Filter by ⓘ	Structure	Anomaly Detection	Overall Quality	Applied Rules
🔑 ordernumber		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add
📅 orderdate		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add
📅 requireddate		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add
📅 shippeddate		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add
Abc status		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add
Abc comments		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add
📄 customernumber		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add
✕ isactive		🔵 ▾	+ Make Mandatory	+ Enable Detection		+ Add

- › Hit the **select** button.

☒ Run Monitoring project on data slice → **Select**

Select data slice

📄

There aren't any data slices yet

Create a data slice to focus on Data Quality of a specific part of a Catalog item.

**Create slice**

As we haven't previously defined data slices for this catalog item, we need to do it on the fly.

- › Hit the “**Create slice**” button and fill the fields in the opening window as follows:
  - **Name:** e.g. “State”
  - **Slice by:** “status”
  - **Value:** “Shipped”

Create data slice

×

Name \*

State

Slice by ⓘ

Value \*

Contains date

Abc status ▾

= ▾

Shipped

☐

+

Save and select

Cancel

- › Hit **“Save and Select”**.

With Data Slices configured, it's time to enable data evaluation on our catalog item.

- › Enable the **Anomaly Detection** for this Catalog item for an attribute of your choice.
- › Add some **DQ checks** – add a **String Completeness** on **comments** or **Day Completeness** to any DATE type attributes (**orderdate**, **requireddate**, etc.). Use the filter to search through the configured rules:

Filter Attributes and Rules

All Attributes ▾

[Check for Rule suggestions](#)

Standard view ▾

Hidden columns ▾

Name	Terms	Filter by ⓘ	Structure	Anomaly Detection	Overall Quality	Applied Rules
123 ordernumber		<input type="radio"/> ▾		Enabled		+ Add
orderdate		<input type="radio"/> ▾		+ Enable Detection		+ Add <span>New</span> Day Completeness
requireddate		<input type="radio"/> ▾		+ Enable Detection		+ Add <span>New</span> Day Completeness
shippeddate		<input type="radio"/> ▾		+ Enable Detection		+ Add <span>New</span> Day Completeness
Abc status		<input type="radio"/> ▾		+ Enable Detection		+ Add
Abc comments		<input type="radio"/> ▾		+ Enable Detection		+ Add <span>New</span> String Completeness
123 customernumber		<input type="radio"/> ▾		+ Enable Detection		+ Add
isactive		<input type="radio"/> ▾		+ Enable Detection		+ Add
123 processing_time		<input type="radio"/> ▾		+ Enable Detection		+ Add

- › Once you have finished configuring the item, **Publish** the project to Save the changes.
- › Click on **Run monitoring** to see the updated information in the Configuration and Results tab.

Items to monitor

[+ Add](#)

Q ▾

Search

Name	Structure	Anomalies	DQ Rules	Overall	Origin
<div> <div>📄</div> <div>Slice of orders</div> <div>≡</div> </div> <div> <div>BV_Session1_Training &gt; postgres &gt; public</div> <div>8</div> <div>2 checks</div> <div>4 DQ Checks</div> <div><a href="#">Check for Rule suggestions</a></div> <div>Not calculated yet</div> <div>postgres</div> <div>⋮</div> </div>					

- › Switch to the **Report** tab and check the results for **sliced data**.

The screenshot displays the Data Quality monitoring dashboard for the 'xy\_Training project'. The left sidebar contains navigation links for Rules, Detection Rules, DQ Evaluation Rules, DQ Dimensions, Components, Transformation Plans, DQ Firewalls, Monitoring Projects, Reconciliation Projects, and Lookup Items.

The main area shows the 'Report' tab under 'Configuration & Results'. It features a tree view on the left with 'Overall' expanded, showing 'BV\_Session1\_Training' at 20% quality. A red box highlights this entry. Another red box highlights the filter bar above the chart, which includes 'Data Quality by data slice → status = Shipped'.

The central chart displays 'Overall Quality' as a donut chart at 20%. To its right, a summary table shows 'Completeness' at 20%.

Category	Status	Count
Overall Quality	Passed	62
	Failed	241
<b>Total</b>		<b>303</b>

A line chart below the donut chart shows quality trends over time, with a single data point at 12 PM on Fri 06.

The bottom section, 'Detailed results', includes a filter bar and a table listing specific rules:

Rule name	Data quality	Applied on	Dimension
String Completeness	⚠️ 20%	Slice of orders  > comments	Completeness
Day Completeness	✅ 100%	Slice of orders  > orderdate	Completeness

## 2. Exporting the Monitoring project results for post-processing

We have already learned how to review and display the results of monitoring projects in the One web application. However, you might need to use the results from your Monitoring Projects for further **post-processing**. For example, you may need to write the results to a flat file or use them for integration with DQIT or some external 3<sup>rd</sup> party tool. Therefore, you would need to set up post-processing jobs for your project.

There are three post-processing options available for this purpose( Post-processing, Transformation, and Remediation), two of which are covered as tasks in this workshop( Post-processing and Transformation).

### 2.1. Create a Post-processing plan to export Monitoring Results

We can export the results of our monitoring project by creating a **post-processing plan**. This method is currently the most comprehensive method, as it allows for different types of further actions. However, for our training purposes, we will simply create a CSV output containing the monitoring results of a catalog item.

To create these types of plans we require to be connected to the **ONE Desktop** application as the **post-processing components** are created there and then deployed to the ONE Web application.

To practice this method we will export monitoring project results of our VCI (orders\_transformed), Before doing so, we need to add it to our project:



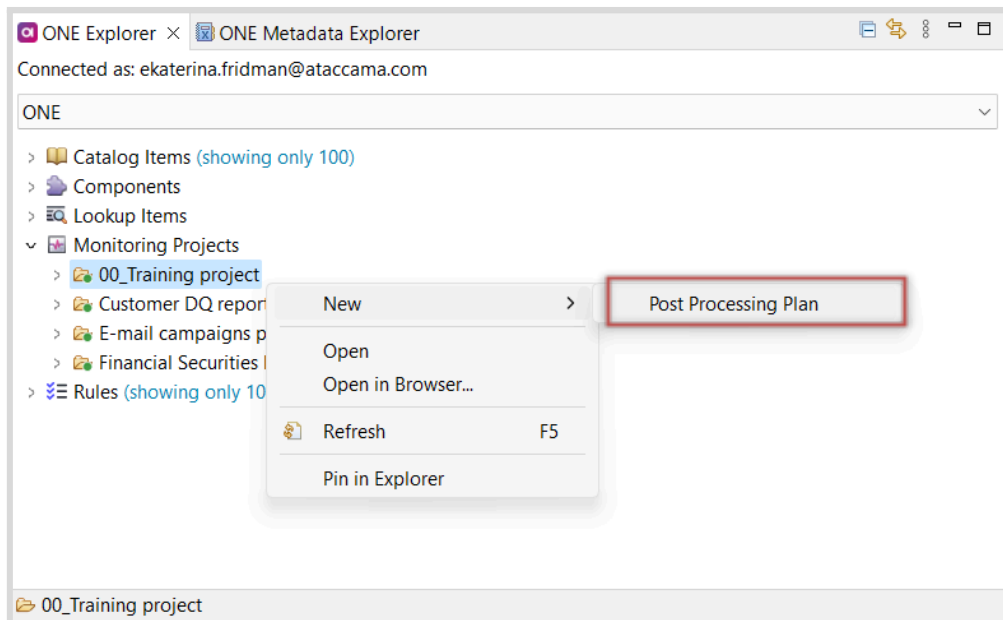
- This VCI (Virtual Catalog Item) was created as part of the **ONE Workshop - Integration of ONE Desktop**. If you don't have it, choose the **orders** item instead.
- The process of adding and configuring a VCI in a project is similar to regular items.

- › Go to the '**<prefix>\_Training Project**' and update the Configurations and Results tab to include the **orders\_transformed** VCI and apply some DQ rules. You can apply the rules you applied previously to the Orders item; 'Day Completeness' to the date columns ('order\_date', 'shipped\_date', 'required\_date') and 'string completeness' to the 'comments' column.
- › Go to your **ONE Desktop** application and connect to your ONE Web application.



*The connection and integration with the ONE Desktop application were already covered in the previous workshop focusing on the **ONE Desktop Integration**.*

- › In the **ONE Explorer** tab, find your project in the **Monitoring Projects** section.
- › Right-click on your project name and choose '**New**' and '**Post-processing Plan**':



This will open a new configuration window. The new post-processing plan will require a few details to be populated:

- › Fill in the necessary information:
  - **Name** of the plan (e.g. “**export\_dq\_customers**”),
  - Select a **Catalog Item** on which you want to run the export (let’s start with **customers**),
  - Specify the **Output Location** as **export/\${date}/\${catalogItemName}/**
  - Optionally you can fill in a plan’s **Description**.

- › Click **Finish** to complete the creation of your plan.

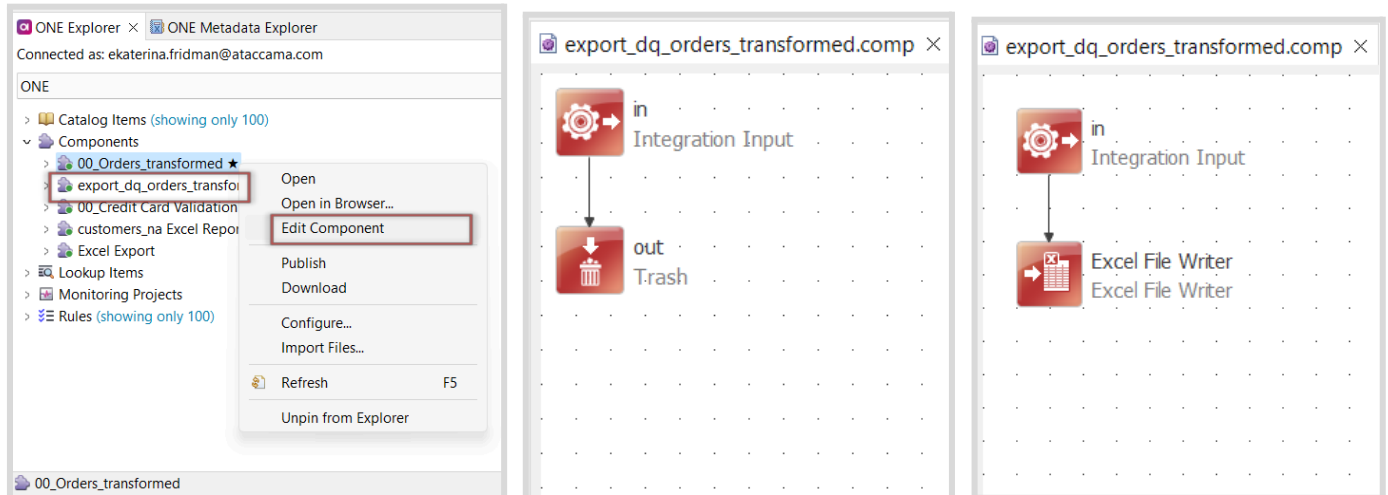


Every Post Processing Component contains one **Integration Input** step for one Catalog Item. Therefore, you must create Post Processing Components for each Catalog Items in your Monitoring Project separately.

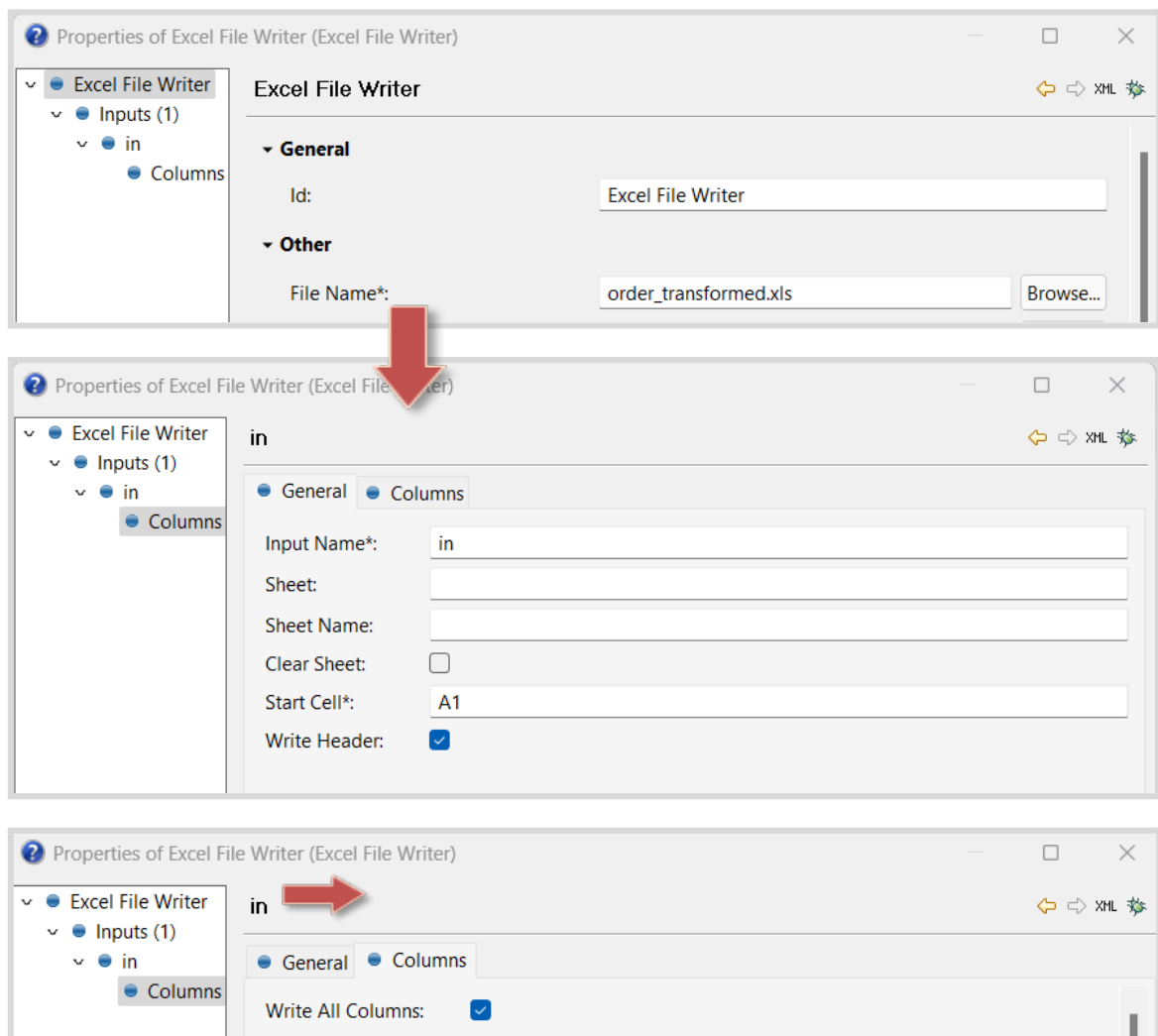


Your newly created component can now be found under the **Components** in the **ONE Explorer** window.

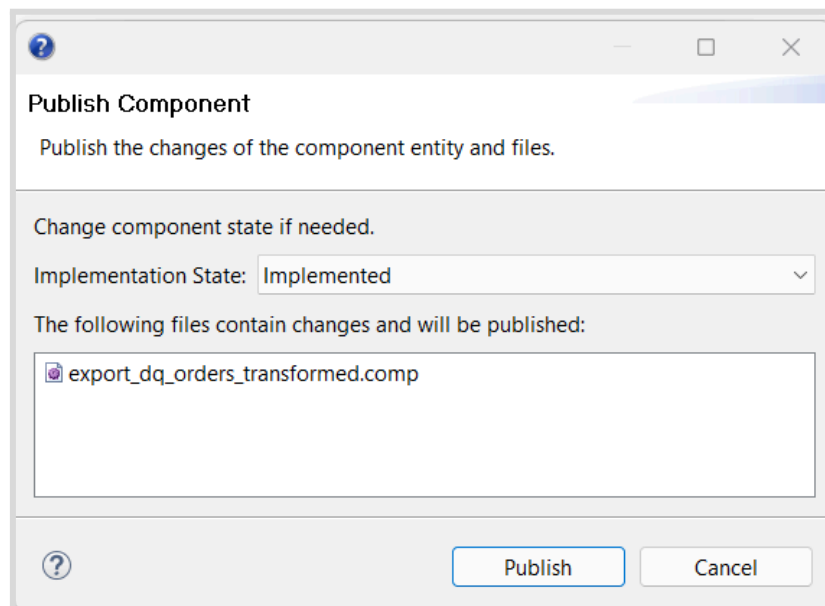
- › Right-click on the new component and choose the **Edit Component** option.
- › Replace the **Trash** step with the **Excel File Writer**.
- › Specify the **File Name** in the **Excel File Writer** as e.g. **"order\_transformed.xls"**.



Configure the Excel file writer step according to the following:

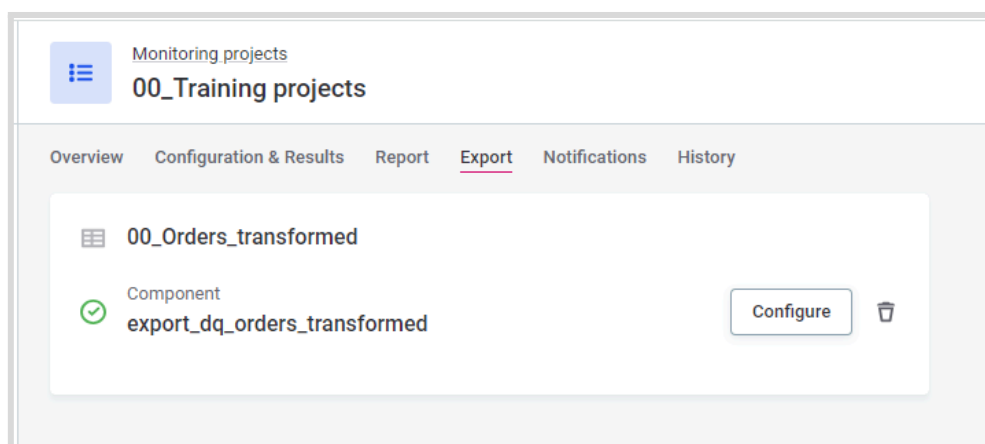


- › **Save** the component.
- › In the ONE Explorer tab, right click on the component and select the publish option.
- › Set the **State** as **Implemented** and publish it.




The updated component is in the components section of the application as well as the export tab of your Monitoring Project. However, the Excel output is not available until you execute the project's monitoring action again.

- › In the **ONE Web Application**, navigate to your Monitoring Project and check the **Export** tab:



To access and download the report output:

- › Press the **Run monitoring** button.
- › Press the **Download**  button to store the generated **orders\_transformed.xls** to your computer.

00\_orders\_transformed

Component

export\_dq\_orders\_transformed

Configure


Post-processing results

shared:/export/2023-06-27/00\_orders\_transformed/order\_transformed.xls

75.5 KB

> Open the results and check them.

ordernumt	orderdate	requiredd	shippedda	status	comments	customern	isactive	processing_time	valid_rules	valid_rules_explanation	invalid_rules	invalid_rules_explanation	score
10100	2003-01-06	#####	2003-01-1	Shipped		363	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10101	2003-01-09	#####	#####	Shipped	Check on	128	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(36ed325c-0000-7000-0000-000001581a9a):OTHER,Day Com			C
10102	2003-01-10	#####	#####	Shipped		181	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10103	2003-01-29	#####	#####	Shipped		121	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10104	2003-01-31	#####	#####	Shipped		141	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10105	2003-02-11	#####	#####	Shipped		145	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10106	2003-02-17	#####	#####	Shipped		278	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10107	2003-02-24	#####	#####	Shipped	Difficult to	131	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(36ed325c-0000-7000-0000-000001581a9a):OTHER,Day Com			C
10108	2003-03-03	#####	#####	Shipped		385	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10109	2003-03-10	#####	#####	Shipped	Customer	486	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(36ed325c-0000-7000-0000-000001581a9a):OTHER,Day Com			C
10110	2003-03-18	#####	#####	Shipped		187	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10111	2003-03-25	#####	#####	Shipped		129	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(String Completeness:(comm	String Completeness:(comments)		25C
10112	2003-03-24	#####	#####	Shipped	Customer	144	TRUE	#####	Day Completeness	Day Completeness:(requireddate)(36ed325c-0000-7000-0000-000001581a9a):OTHER,Day Com			C



On your own now! Try to create exports for the other two Catalog Items of your Monitoring Project (products & customers) in formats (other than .xls).

## 2.2. Create a Transformation Plan to Export Monitoring Results

We can export the results of our monitoring project by creating a **transformation plan**. In contrast to the post-processing plan method, this one is **fully done** in the **ONE web application**. However, it comes with a limitation that using this method we can **only export to flat files or One Data**. Similar to post\_processing plans, using transformation plans you can apply changes such as filtering, splitting, joining, etc. to the results before writing them to output files.

To practice post-processing through transformation plans, we will develop a transformation plan that takes the results from a catalog item within a monitoring project, filters these results, and then outputs them to a .csv file for download.

To this end, we will exclude **'results not from the US'** and those **'without data quality issues'**, allowing us to appropriately **rename the output file** to **'US\_issues\_output'**.

- Still working with your monitoring project, navigate to the **Configuration & Results** tab.
- This time, we will work with the catalog item **customers**. From the ellipsis (three dots) menu, select **add post-processing transformation**.

Name	Structure	Anomalies	DQ Rules	Overall	VAL	COM	ACC	Origin
customers   DEV	1 check							
pgs_testdata > tcd > transactional_customer_data	15	2 checks	4 DQ Checks 14+ suggestions	76%	76%	Not calculated yet	84%	postgres-testdata-svc
products	9							
pgs_testdata > tcd > transactional_customer_data	303							
Slice of orders_transformed	8							
pgs_testdata > tcd > transactional_customer_data	303							
products	9							
pgs_testdata > tcd > transactional_customer_data	303							

- As before, choose a name for your transformation plan, such as **"<prefix>\_Transformation\_Plan\_Customer\_DQ\_Report"**, then click **Confirm** to continue.

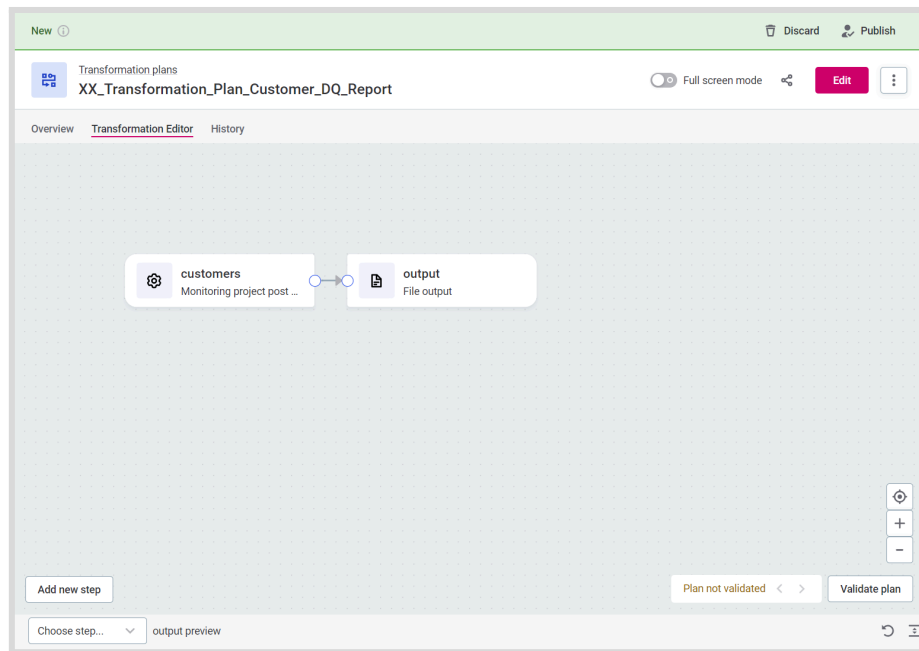
Add post-processing transformation

DQ plan name


XX\_Transformation\_Plan\_Customer\_DQ\_Report

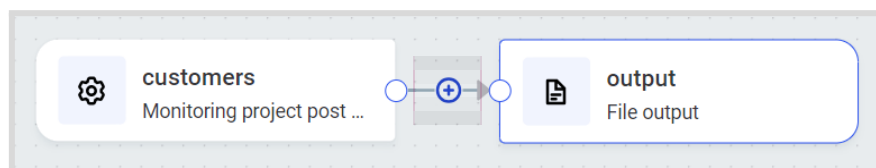
Confirm Cancel

- As we saw before, a simple **transformation plan** consisting of an **input** (**customers**) and an **output** is created.

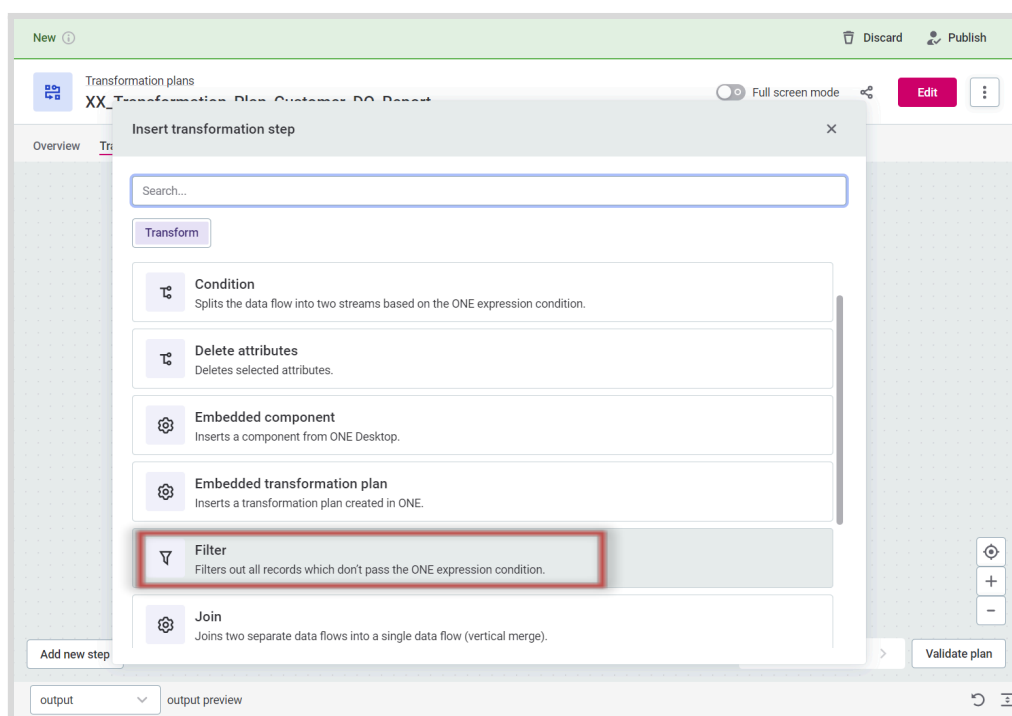


We will now be adding an additional step to filter our results before downloading them. As discussed earlier in **DQ Foundation**, there are two methods for adding a new step:

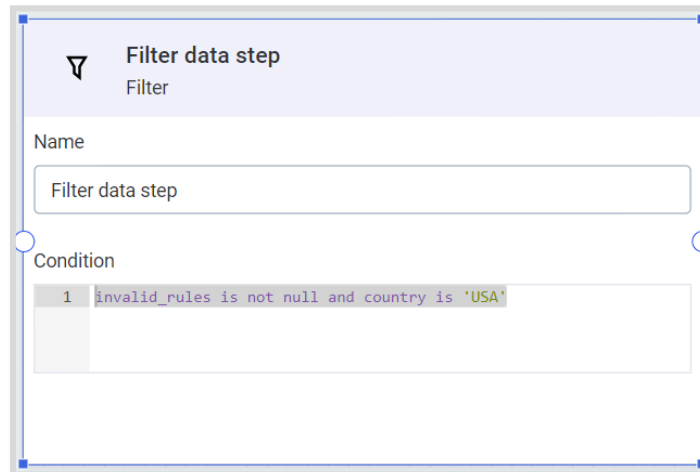
- Hovering the mouse over the link between two steps clicking the add step icon 
  - Using the **Add new step** button.
- › For simplicity, we will use the first method. Hover your mouse over the link between customers and output, then click the add step icon.



- › From the list of available steps, select the **filter**.



- › Click on the icon of the **Filter data step** to open it for configuration. Remember, we want to exclude all results that have **no data quality issues** and are **not from the US**.



The screenshot shows the 'Filter data step' configuration window. At the top, there is a funnel icon and the text 'Filter data step' and 'Filter'. Below this, there is a 'Name' field with the value 'Filter data step'. Underneath is a 'Condition' section with a table containing one row: '1' in the first column and 'invalid\_rules is not null and country is 'USA'' in the second column.

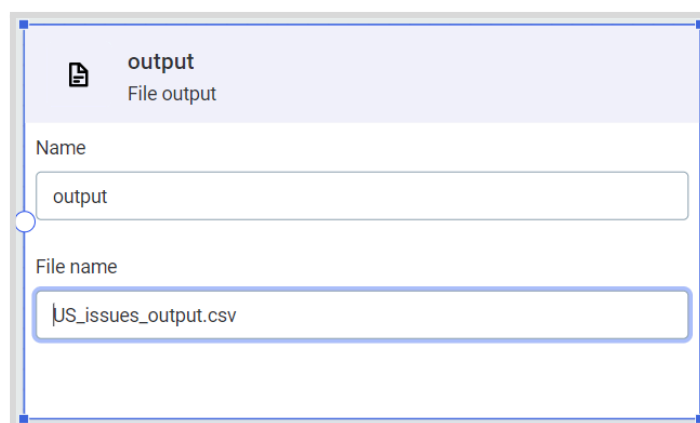
**Expression:** *invalid\_rules is not null and country = 'USA'*



NOTE

- Active **content assist** is available when writing your **Filter Condition**.
- Click on the step's icon to collapse it once you are done.

- › Lastly, we need to rename the output file. Open the output step and rename the File name to **'US\_issues\_output'**. **Publish** the transformation plan.



The screenshot shows the 'output' configuration window. At the top, there is a document icon and the text 'output' and 'File output'. Below this, there is a 'Name' field with the value 'output'. Underneath is a 'File name' field with the value 'US\_issues\_output.csv'.

- › Return to your **Monitoring Project**. As we once again have made changes to the project, we will need to **publish** them. Once that is done, **run monitoring** to populate results for export.
- › Upon completion, navigate to the **Export** tab. Click on the **download link** to download your post-processing results for **US\_issues\_output**.

Monitoring projects

KS\_Training project

Run monitoring

Overview

Configuration & Results

Report

Export

Notifications

History

customers

Component

ks\_export\_dq\_orders\_transformed

Configure

Post-processing results

shared:/export/2024-05-31/customers/order\_transformed.xls

67.5 KB

Transformation

KS\_Transformation\_Plan\_Customer\_DQ\_Report

Configure

Post-processing results

shared:/US\_issues\_output.csv

28.39 KB

shared:/US\_issues\_output.csv.metadata

5.53 KB

That's it! You have filtered and downloaded the data quality monitoring project results for one of the monitored catalog items through the transformation plans method. Do note that **post-processing transformation plans** must be created on a pre-monitored catalog item basis.

## Conclusion

Great! We have come to the end of another Ataccama ONE workshop!

We modified an existing monitoring project, and enabled data slices on it and interpreted the results. We also configured the exports of the monitored DQ results to be able to use them for post-processing through two different methods.

Congratulations! You are now ready to start developing your own Ataccama ONE project!