

## Dataset tools

### Description

**Dataset tools** is a versatile 7-in-1 app designed to streamline dataset management. It offers powerful features to beautify, export, merge, remove, rename, and split datasets, ensuring data is organized, readable, and easy to work with.

This app simplifies complex data operations, making it a must-have for anyone working with JSON, XML, JDF, and similar file types.



### Use case

- **Beautify** dataset, to be nice and human readable
- **Export** dataset(s), from a list
- Get **info** about datasets
- **Merge** datasets (filtered on name with regex)
- **Remove** datasets (with blacklist white list)
- **Rename** dataset
- **Split** dataset (from multi-line order, to single line)

### Compatibility

Switch 2023 fall or higher.

### Output behavior and Connections

Each action will have their related connections and output behavior. See detailed section for detailed information.

# Beautify

**Dataset tools “Beautify”** automatically format your JSON, XML, XMP, or JDF files to improve readability and consistency.

## Use case

This feature ensures your data files are neatly structured, making them easier to interpret and work with.

## Flow element properties

- Act on**  
 Act on the actual file or dataset(s) attached to this file.
- Dataset**  
 List of dataset(s) you want to beautify.  
 "Automatic" default value will try to look after every valid XML JDF and XMP available for the actual job.

## Connections

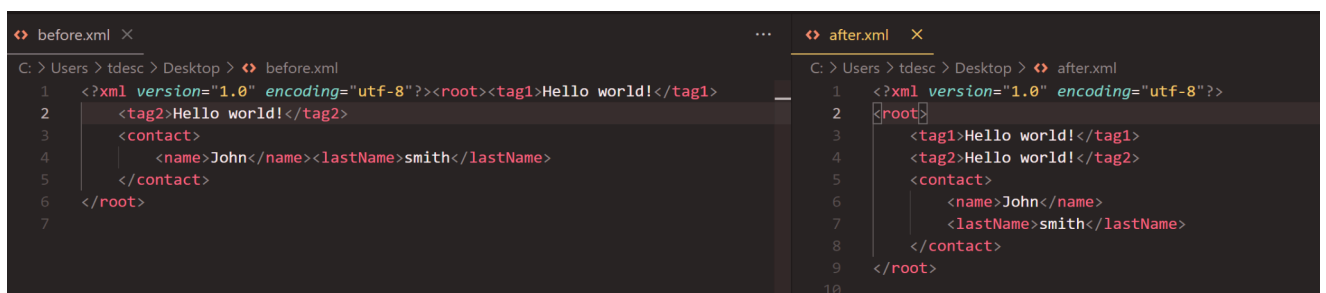
**Incoming connection** : At least one data connection.

**Outgoing connection** : At at least one data connection.

## Output behavior

Case	Output level
Everything went as intended. The file or dataset(s) have been successfully processed	Success
Some files or datasets can't be read (may not be valid XML, JDF, XMP or JSON files)	Warning
None of the the dataset(s) or files can't be read (may not be valid XML, JDF or XMP files)	Error

## Example



```

before.xml
1 <?xml version="1.0" encoding="utf-8"?><root><tag1>Hello world!</tag1>
2   <tag2>Hello world!</tag2>
3   <contact>
4     <name>John</name><lastName>smith</lastName>
5   </contact>
6 </root>
7

after.xml
1 <?xml version="1.0" encoding="utf-8"?>
2   <root>
3     <tag1>Hello world!</tag1>
4     <tag2>Hello world!</tag2>
5     <contact>
6       <name>John</name>
7       <lastName>smith</lastName>
8     </contact>
9   </root>
10
  
```

## Export

**Dataset tools “export”** allows you to export multiple datasets at once, specified with multi-line text or with automatic detection.

### Use case

Exporting multiple datasets to a hierarchy can be hard with Switch basic tool if you have a lot of datasets, or if you don't know their names in advance.

### Flow element properties

- **Dataset(s)**  
List of datasets you want to export. "Automatic" default value will try to look after every datasets available for the actual job.

### Connections

- **Incoming connection :**  
At least one connection.
- **Outgoing connection**  
At least one data and one log\* outgoing connection.  
*\*Multiple datasets will be exported as a folder.*

### Output behavior

Case	Output level
All specified dataset(s) successfully exported	Success
Some of the specified dataset(s) could not be exported	Warning
None of the specified dataset(s) could be exported	Error

## Info

**Dataset tools “info”** will gather all information about the actual job datasets and store them in a XML dataset log.

## Use case

In some cases, you want to get the path of datasets present alongside your job.

## Flow element properties

No properties.

## Connections

- **Incoming connection**  
At least one connection.
- **Outgoing connection**  
At least one data and one log outgoing connection.

## Output behavior

Case	Output level
All information about dataset(s) has been gathered.	Success
Something went wrong gathering dataset(s) information.	Error

```

1  <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
2  <datasets>
3    <dataset>
4      <name>toRename</name>
5      <model>XML</model>
6      <extension>xml</extension>
7      <datasetPath>D:/Enfocus/Enfocus_Switch_Dataroot/datasets/03INV000.xml</datasetPath>
8    </dataset>
9    <dataset>
10     <name>Log</name>
11     <model>XML</model>
12     <extension>xml</extension>
13     <datasetPath>D:/Enfocus/Enfocus_Switch_Dataroot/datasets/03INV0001.xml</datasetPath>
14   </dataset>
15   <dataset>
16     <name>Log2</name>
17     <model>XML</model>
18     <extension>xml</extension>
19     <datasetPath>D:/Enfocus/Enfocus_Switch_Dataroot/datasets/03INV0002.xml</datasetPath>
20   </dataset>
21 </datasets>

```

# Merge

**Dataset tools** “**Merge**” allows you to merge multiple files or datasets into one single file.

## Use case

If a system creates multiple dataset for the same order, and you want to merge them to send it all at once.

## Flow element properties

- **Act on**  
Either you want to act on datasets or on an incoming job folder of datasets.
  - **Datasets**
    - **Regular expression**  
The regular expression that will be used to filter datasets.
    - **Remove merged datasets**  
Delete all datasets that have been merged.
- **Dataset model**  
The model of the dataset you want to merge.

## Connections

- **Incoming connection**  
At least one connection.
- **Outgoing connection**  
At least one data and one log outgoing connection.

## Output behavior

Case	Output level
All specified dataset(s) successfully have been merged	Success
None of the specified dataset(s) could be merged	Error

## Remove

**Dataset tools** “Remove” allows you to quickly remove dataset(s) from the actual job.

### Use case

Sometimes, the automatic cleanup of Switch is not frequent enough, and can still lead to bottleneck issues in heavy load scenarios. This option can help you to get rid of unwanted dataset as soon as they are not useful anymore.

### Flow element properties

- **Mode**

Either you want to remove or to keep specific dataset(s).

- **Remove following dataset list**

- **Dataset(s) to remove**

List of dataset you want to remove.

Literal value :

- All jobs : All dataset will be removed
- No jobs : All dataset will be retained

- **Keep following dataset list**

- **Dataset(s) to keep**

List of dataset you want to keep.

Literal value :

- All jobs : All dataset will be retained
- No jobs : All dataset will be removed

### Connections

- **Incoming connection**

At least one incoming connection.

- **Outgoing connection**

At least one outgoing data connection.

### Output behavior

Case	Output level
All specified dataset(s) have been successfully removed	Success
Some of the specified dataset(s) could not be removed	Warning
None of the specified dataset(s) could be removed	Error

# Rename

**Dataset tools** “Rename” allows you to rename dataset.

## Use case

Some apps do not give you control on how the dataset they create are named. In such cases, you will be able to rename them to avoid conflict.

## Flow element properties

- **Dataset(s) to rename**  
Dataset to rename and their new names (one per line) separated with a separator (see property below).  
Due to the Enfocus Switch mode of operation, the app can give unexpected result if :
  - You rename a previously renamed dataset in the same execution of the app;
  - You rename several datasets with the same final name.
- **Separator**  
The separator is the keyword character that will be used to split the dataset and his index.
- **Duplicates**  
Determines what happens when a dataset is renamed to the same name as a pre-existing dataset for this job.
  - **Overwrite**  
Overwrite the previous one with the renamed dataset.
  - **Skip**  
Do not rename the dataset.
  - **Add a version number:**
    - **Separator**  
Separator added between dataset name and version number.
    - **Width**  
Width of the version number (not including separator).

## Output behavior

Case	Output level
All specified dataset(s) successfully renamed	Success
Some of the specified dataset(s) could not be renamed (see duplicates “skip” option)	Warning
None of the specified dataset(s) could be renamed (see duplicates “skip option”)	Error

# Split

**Dataset tools** “Split” allows you to split into several datasets a pre-existing one containing several times the same tag.

## Use case

A lot of ERP or Web to print will give one dataset by order, not separating them for easy use within Switch. Dataset split such dataset by component!

## Flow element properties

- **Dataset to split**  
The name of the dataset you want to split.
- **XPath**  
Your node's XPath expression.  
Note that this app support only a subset of the full XPath standard, only the following expression are supported :
  - Descendent selectors (root/parent/offspring).
  - Attribute selectors (root/parent/offspring[@attribute='value']).
  - Tag value selectors (root/parent[offspring='value']).
- **Keep to level structure**  
In the case of a nested XPath, you can discard top level structure to keep only the tag found with the XPath.
- **Separator**  
The separator you want to be used between the original dataset name, and the new ones.
- **Index width**  
The index width (padStart) to be used.  
Automatic will select the correct number of leading 0, for sorting purposes.

## Connections

- **Incoming connection** : At least one incoming connection.
- **Outgoing connection** : At least one outgoing data connection.

## Output behavior

Case	Output level
All specified dataset(s) successfully splitted	Success
Something went wrong, the dataset may not be readable XML, JDF or XMP	Error