

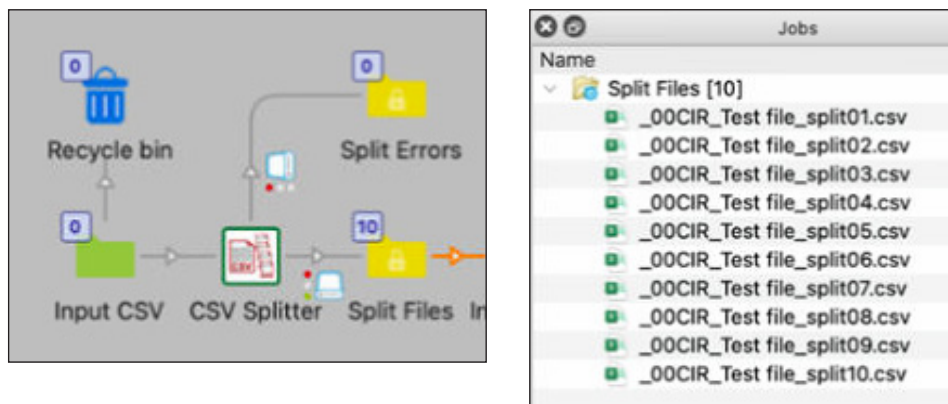
CSV Splitter

Description

CSV Splitter splits CSV files into multiple files, each containing the common header and “n” number of data rows. The user can specify the number of rows per file, and a unique suffix (appended with an appropriately padded number).

An external XML metadata set with the Input CSV file information is generated and attached to each split file (see **Metadata** below).

Especially useful where large CSV files used in flows, fail with processing errors.



Compatibility

Switch 2022 Fall and higher. Windows and Mac OSX.

Connections

CSV Splitter only accepts CSV files from one or many incoming connections. The CSV input file must pass a simple validation, requiring a non-empty file, minimum of 2 rows, and no empty header row values (validation error files are logged and sent to the Error connection). Files that need no splits are simply passed to the Success connection, with the original file names. See **Outgoing connections** below.

Flow elements properties

Property	Value
Element type	CSV Splitter
Name	CSV Splitter
Description	
Number of rows per split file	100
Split file suffix	_split

- **Number of rows per split file**
 - Number of rows per split file, each with the common header row. Default: 100
- **Split file suffix**
 - The suffix for the split files. This will be followed by an appropriately padded number. Default: _split

Operation

Simply set the properties as indicated above.

The incoming connection should have a filter for only the file type “Comma-separated text (*.csv)”.

Outgoing connections

CSV Splitter uses outgoing data Traffic Light connections. Flows should be set up with a Success and Error outgoing Data connection, as illustrated in the flow above.

Note that flows without an outgoing Data Error connection will log and send Error jobs to the flow’s Problem jobs folder. This is not suggested.

Metadata

Metadata with the Input CSV file information is generated as illustrated below:

The screenshot displays the configuration window for the CSV Splitter. At the top, there is a checkbox labeled "Show sample jobs" which is checked. Below this is a list of generated CSV files:

- _00CIR_Test file_split01.csv
- _00CIR_Test file_split02.csv
- _00CIR_Test file_split03.csv

The "Metadata dataset" section has two radio buttons: "Embedded" (unselected) and "External" (selected). Below the "External" option is a text field containing "inputFileInfo".

The "Location path syntax" section has five radio buttons: "XMP location path" (unselected), "XML location path" (selected), "JDF location path" (unselected), "XPath expression" (unselected), and "JSON Pointer" (unselected).

Below these sections is an empty text field for the "XML location path".

The "XML data tree" section shows a table with the following data:

Name	Value
<ul style="list-style-type: none"> Data <ul style="list-style-type: none"> @datasetCreator inputCsvFileName inputCsvFileNameProper numRowsPerSplit <ul style="list-style-type: none"> @comment splitCsvFileCount <ul style="list-style-type: none"> @comment 	Develop Mint, Inc. CSV Splitter Test file.csv Test file 100 Number of rows per split CSV files 10 Number of split files from input CSV file