



# Generalized Cleansing Component for Pentaho®

## Clean Any Type of Data Using 6 Cleansing Operations

The Melissa Generalized Cleansing Component for Pentaho empowers you to build data cleansing scripts for data suffering from a wide range of errors and inconsistencies. This component can be beneficial to users in many different industries, giving them the ability to standardize and validate inventory lists, correct automobile data and much more.

## What Does Generalized Cleansing Offer You?

- **Cleanse any type of data**, programmatically, with a just a click
- **Save time and effort** combining any of six operations and save operations (simple or complex) for future projects
- **Highly customizable** so you can create rules (triggering) to standardize data when expressions come out true
- **Multiple Rules** can be defined for each field of data to allow for more complex and nested generalized cleansing procedures

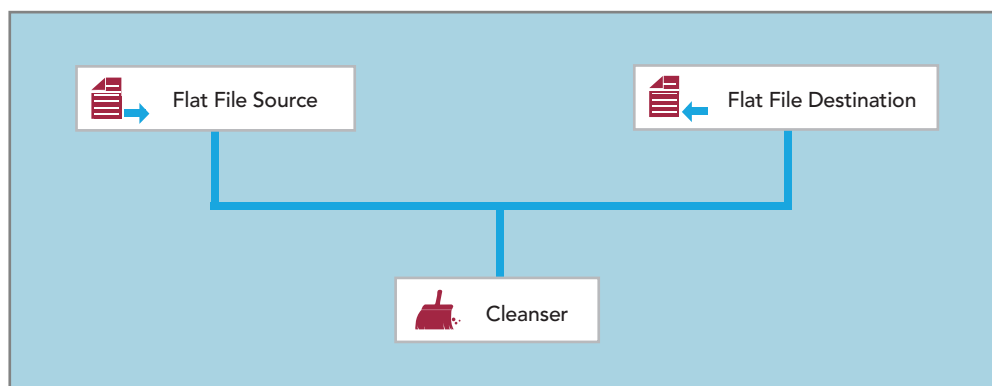
## Cleansing Operations:

- **Casing** – change the casing of data from capital to lowercase, etc.
- **Punctuation** – add or remove punctuation
- **Abbreviation** – expand or contract abbreviations, for example: CA to California
- **Search and Replace** – replace portions of a string
- **Expressions** – create programmatic expressions to make sense of data values
- **Regex** – use regular expressions to extract, validate, etc.

The Melissa cleanser offers a whole collection of operations in one centralized component to simplify the process of applying different cleansing rules. Programmatic expressions and regular expressions can also be used to trigger operations – giving users the ability to self-validate non-contact data to their own specifications.

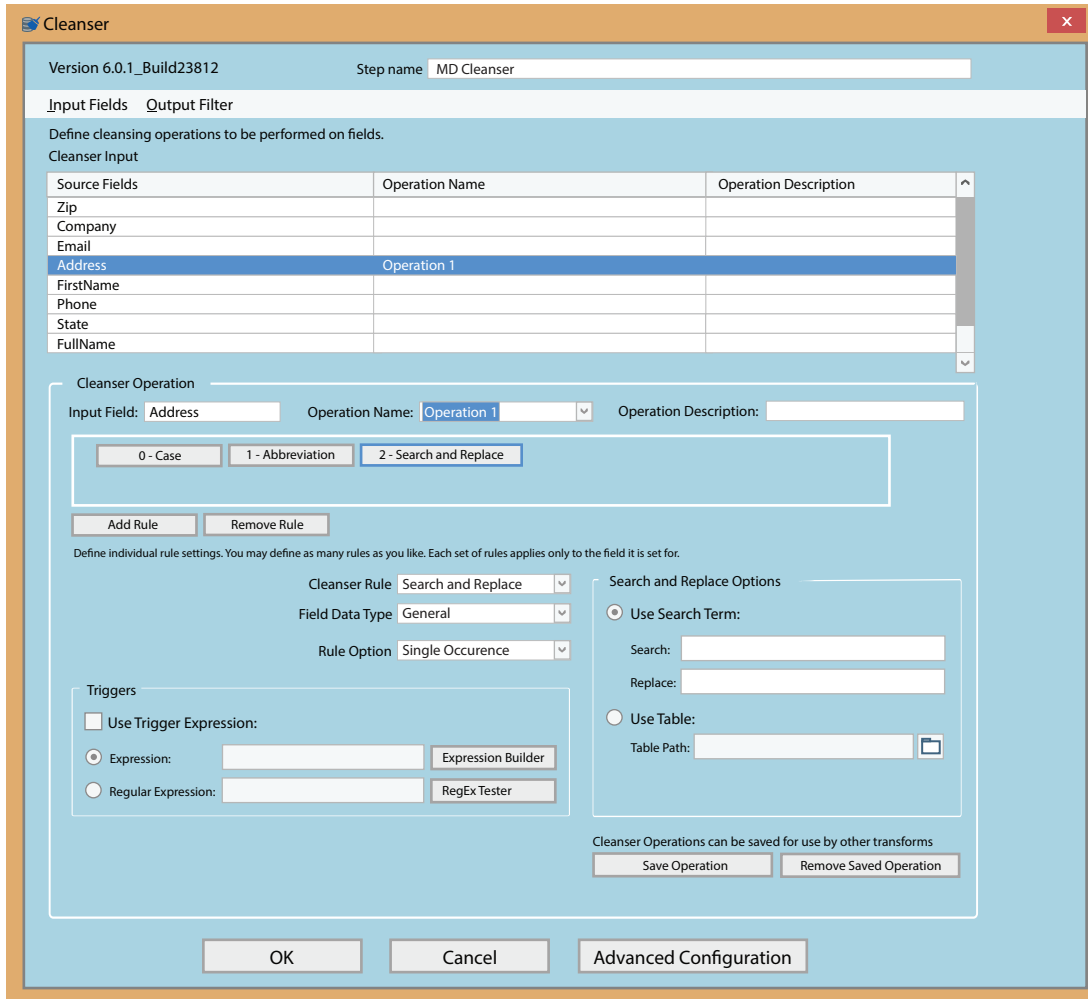
## Knowledge Bases

Leverage search and replace tables, also known as knowledge bases, to perform either text or regex cleansing. The use of a knowledge base table allows for multiple search and replace values, for example, a company synonym table to expand company name abbreviations. Make use of existing knowledge bases, or generate your own to satisfy specific and unique business rules.



# Get Started Today!

>> Request a Free Trial: [www.melissa.com/gcc-SSIS](http://www.melissa.com/gcc-SSIS)



*Cleansing Operation Chaining*

## About Melissa

Since 1985, Melissa has specialized in global intelligence solutions to help organizations unlock accurate data for a more compelling customer view. More than 10,000 clients worldwide in arenas such as retail, education, healthcare, insurance, finance, and government, rely on Melissa data quality and ID verification software, including data matching, validation, and enhancement services to gain critical insight and drive meaningful customer relationships.

### Melissa

22382 Avenida Empresa RSM, CA 92688-2112 | 1-800-MELISSA | [www.melissa.com](http://www.melissa.com)

Find us on: