

# CALUMMA ETL Tool

## User Manual



# Table of Contents

<b>Part I Introduction</b>	<b>3</b>
<b>Part II Spoon Basics</b>	<b>4</b>
<b>Part III Calumma Steps</b>	<b>9</b>
1 Master Data Importer.....	10
2 Data Importer.....	10
3 Enumeration Mapping.....	12
4 Hierarchy-Mapping.....	15
5 Crosslink Importer.....	16
6 Deletion Mode.....	17

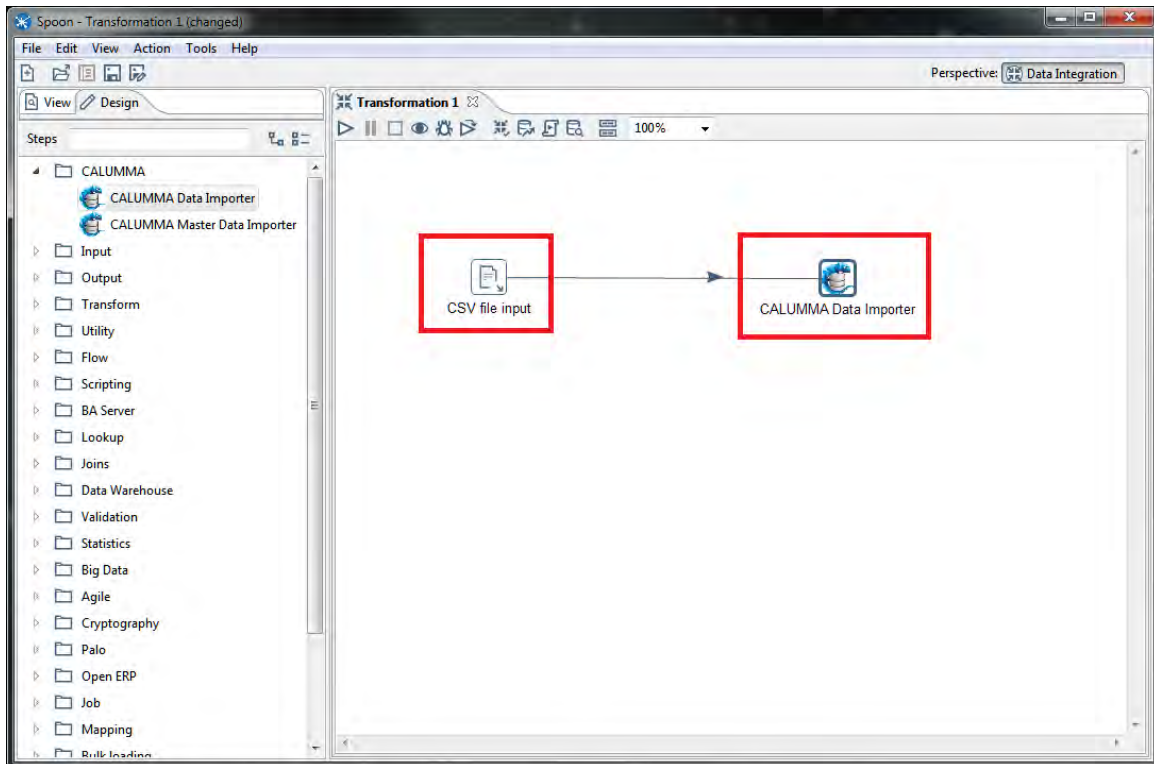
# 1 Introduction

The CALUMMA ETL (Extract-Transform-Load) consists of different Plug-ins for Pentahos Open Source Software Spoon (Kettle Data Integration) and can be used in addition to the CALUMMA Management Tool to import data from different sources into CALUMMA. This manual used Spoon version 7.1.

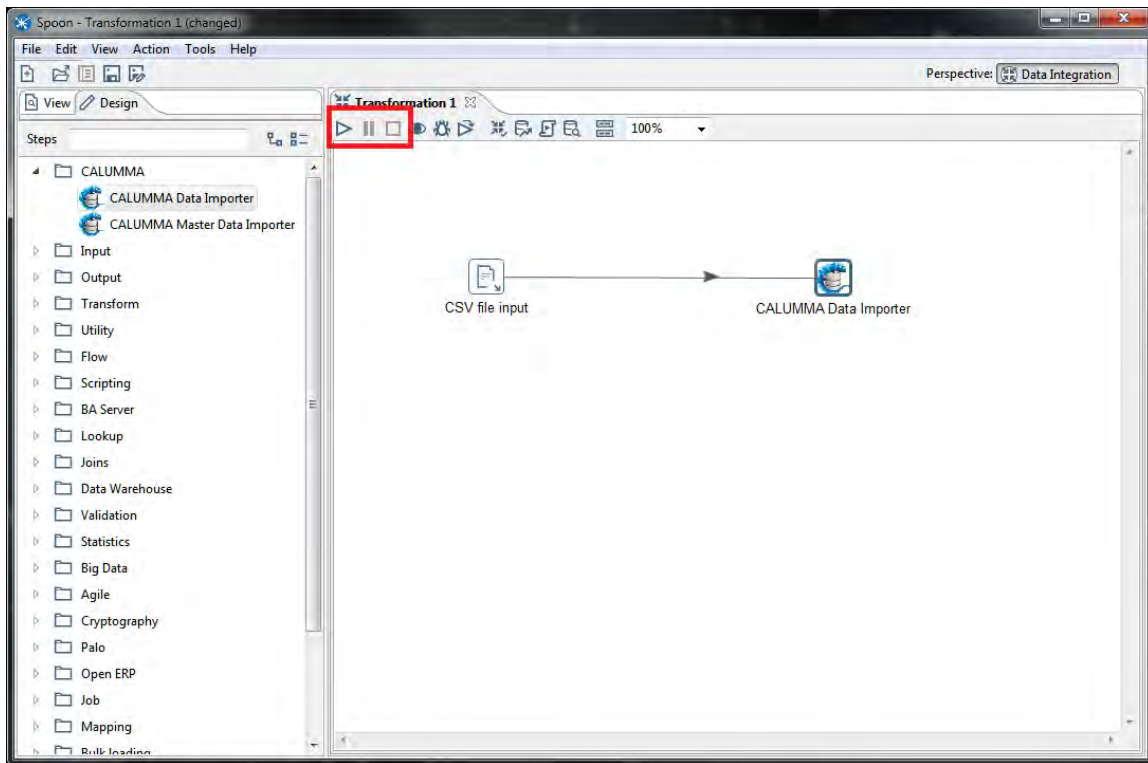
The following chapters explain the basics of Spoon to use the CALUMMA Plug-ins. However, Spoon consists of much more additional features which are not mentioned here. Please visit <http://community.pentaho.com/projects/data-integration/> for more information about Spoon.

## 2 Spoon Basics

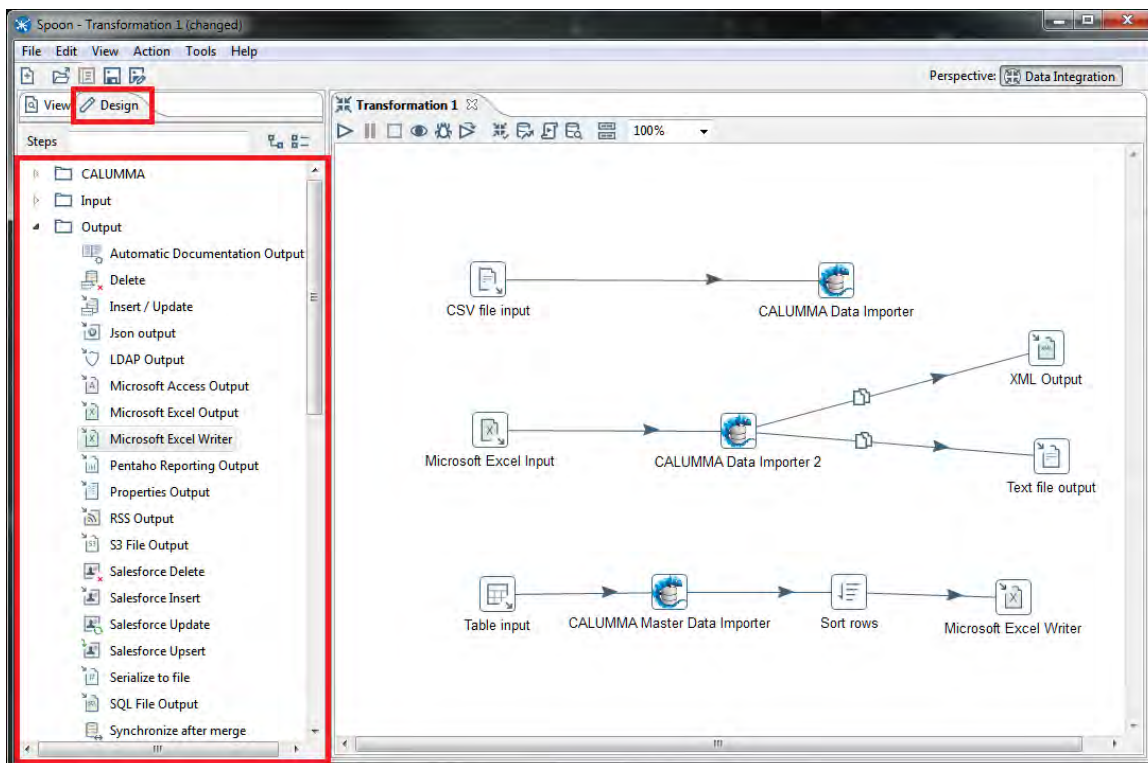
Spoon differentiates between jobs and transformations. A Job can run one or more transformations and a transformation consists of one or more steps which handle the data streams. For example, load, convert or save data. The following screenshot shows a transformation importing a CSV-File into CALUMMA with a CSV-Input step and the CALUMMA Data Importer.



Located on the upper left corner of the transformation is a toolbox where the transformation can be started, stopped and paused.

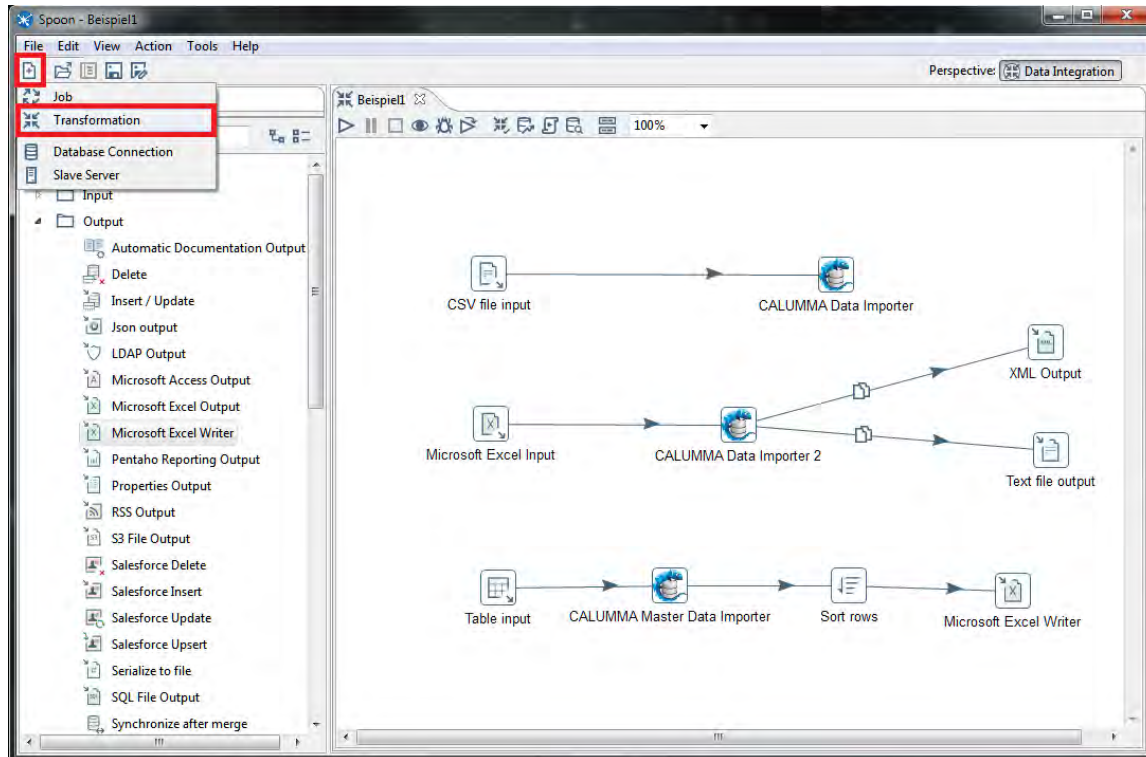


The "Design" tab on the left side contains all steps available for any transformation or job. Every transformation can contain any number of steps.

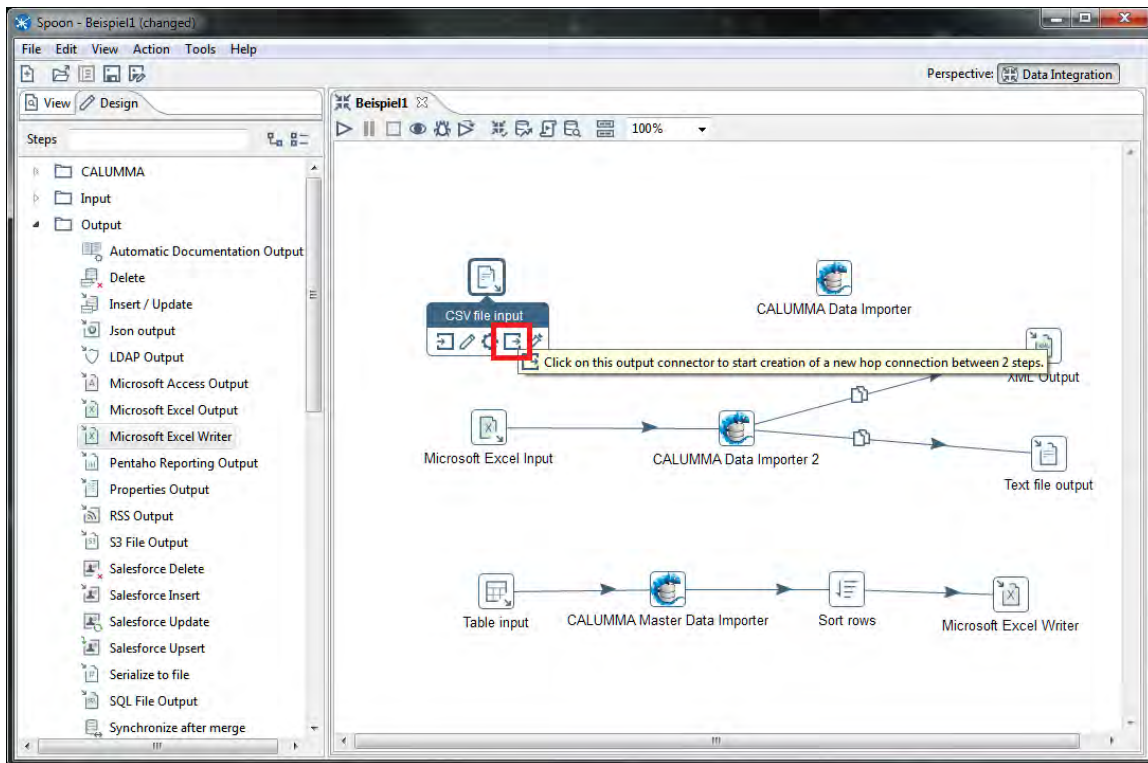


### Create a new Transformation

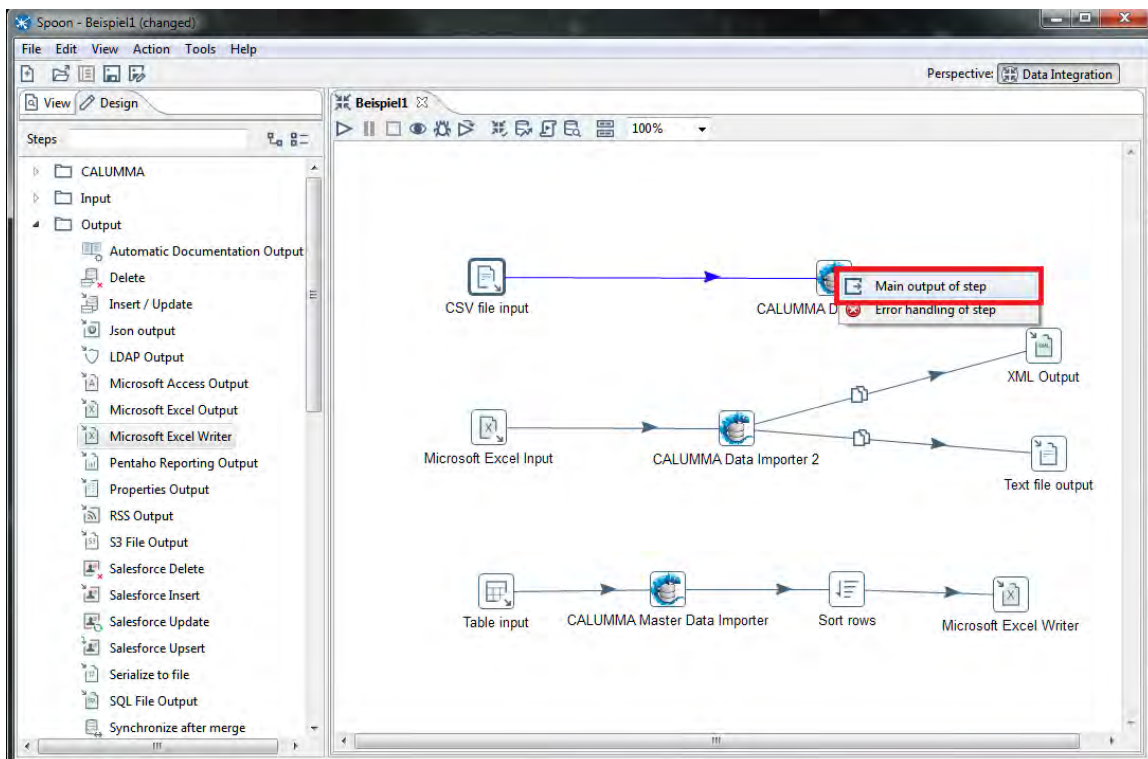
Select file, new in the menu bar and click on new "Transformation" or simply click on the new button on the toolbar and then "Transformation" to create a new transformation.



To add steps to a transformation, simply add them per Drag&Drop from the side "Design" tab to the transformation view. Next step is to connect the steps in the preferred/necessary order. There are two ways to connect steps. Either the middle mouse button is used on the first step and then the left mouse button on the second step or the step menu below the step can be used. Click on the left mouse button and drag the connection to the next step while holding the left mouse button down. The connections between steps are called hops.



After connecting the steps, the content of the destination step needs to be specified. The step can receive the main data or error messages.



## Delete steps

To remove a step from the transformation right click on it and select "Delete Step" or click on the step with the left mouse button and press the "DEL" button on your keyboard.

## CSV-Import

The following documentation has been created with Spoon Version "Kettle 7.1\_2017\_10\_25".

The CSV import step imports data from CSV files. Click on "Browse..." to add the path of the CSV file to import and specify the delimiter below. Finally, click "Get Fields" to load the fields. The data types are detected automatically, but should be checked manually anyway. Especially with date and time fields it is often necessary to state the date format, for example "yyyy/MM/dd" or "HH:mm:ss".

CSV Input

Step name: Kunden

Filename: C:\develop\calumma\EclipseRCP\_CalummaV2.1\EclipseRCP\_CalummaV2.1\workspace\calumma\Additional\doc\Har

Delimiter: ;

Enclosure: "

NIO buffer size: 50000

Lazy conversion?

Header row present?

Add filename to result

The row number field name (optional):

Running in parallel?

New line possible in fields?

File encoding: UTF-8

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	field	String							none

Buttons: Help, OK, Get Fields, Preview, Cancel

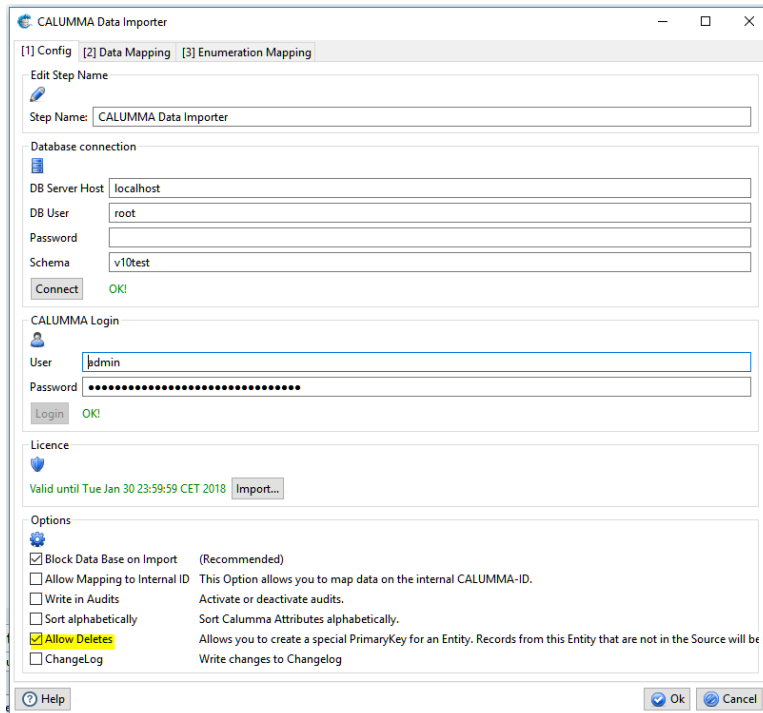


### 3 Calumma Steps

The following CALUMMA steps are available and should be executed in the following order. Master-Data-Importer, Data-Importer and Crosslink-Importer.

#### Configuration of the CALUMMA steps

The "Config" tab specifies the data base credentials and the user login data.



#### Options

The Options area allows you to activate/deactivate different functions.

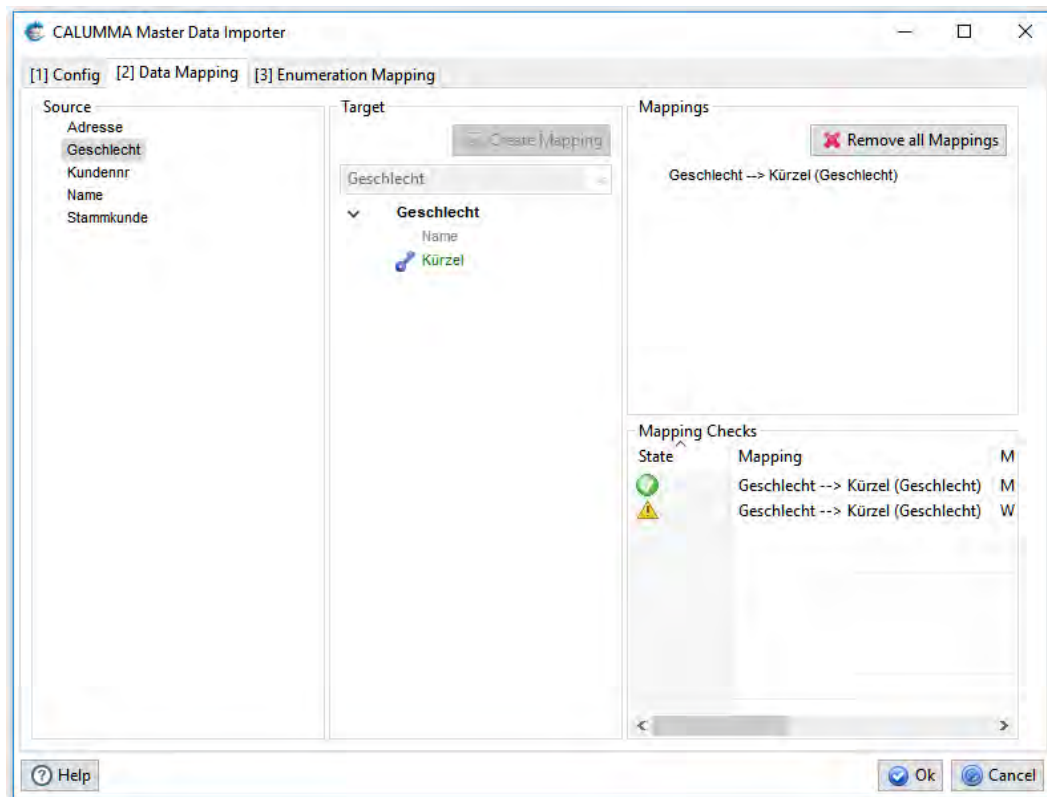
Option	Function
Block database on import	Activates CALUMMA's "Maintenance Mode" during the import. In this mode, all records are locked and can't be changed.
Allow mapping to internal ID	Activate this function to map on the CALUMMA ID of an entity.
Write in audits	When activated, the importer writes changes to the audits of CALUMMA.
Sort alphabetically	Changes the order of the CALUMMA attributes alphabetically.
Allow deletes	Activate/deactivate deletion mode.
Changelog	Activate/deactivate tracking of changes.

### 3.1 Master Data Importer

The Master-Data Importer can be used to create enumeration values in the CALUMMA Management Tool. Open the Master Data Importer with a double-click.

#### Data Mapping

The "Data Mapping" tab is used to create the mappings. In the "Source" group all fields from the input steps are shown, for example the headlines of a CSV file. First select the field from the source that contains the values for the enumeration, then select the entity from the combo-box in the target group. Below there are the attributes of the entity shown. Select an attribute and define a primary key and click on "Create Mapping".



For every enumeration entity a new step is required.

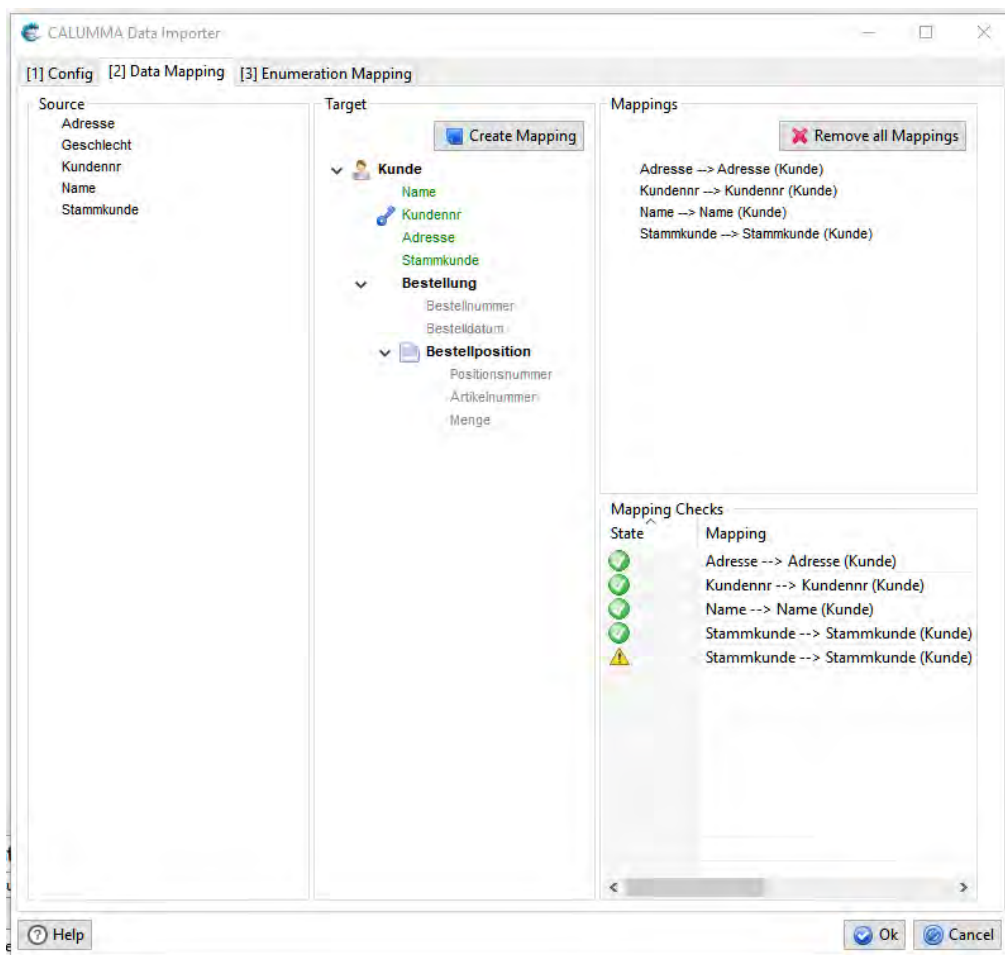
### 3.2 Data Importer

The data importer is used to import data from different kinds of sources to the CALUMMA Management Tool data base.

The following screenshot shows the "Data-Mapping" tab. The left side shows the fields from the source file and on the right side the entity-attribute tree from the Management

Tool is shown. To create mappings, click on a field from the source and on an attribute of the tree and then on "Create Mapping". The most right upper area lists all created mappings where you can delete them. For every (except compositions) entity that has a mapped attribute you need to create a primary key. Create a primary key by double-clicking on an attribute. There are three types of primary keys:

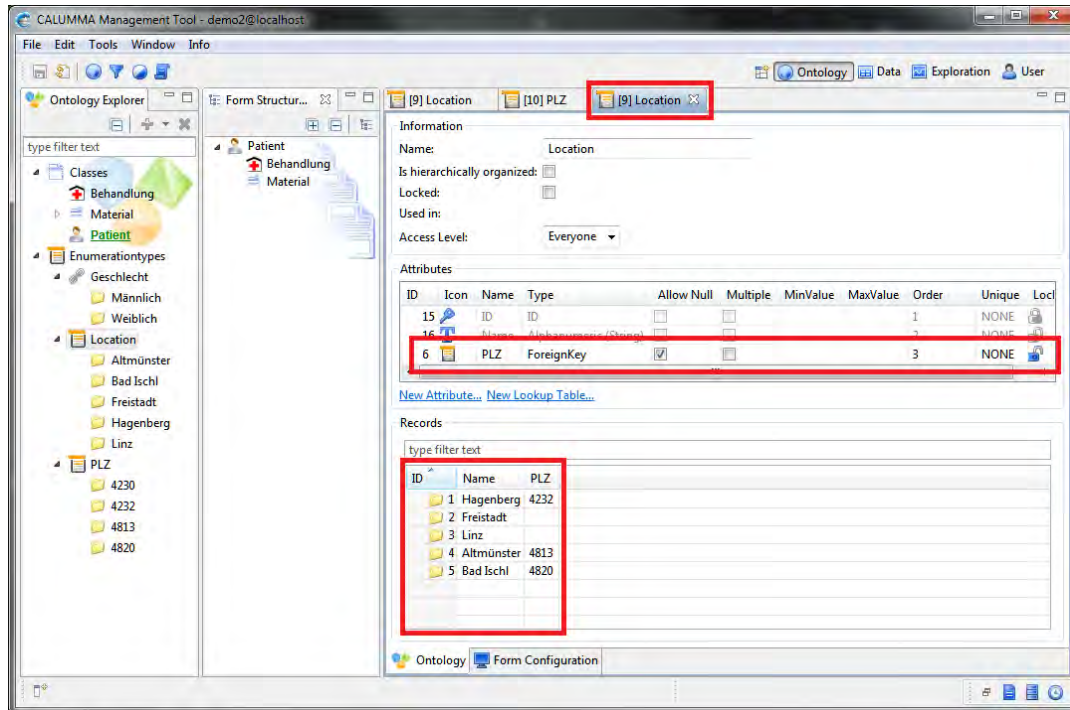
- Standard Primary Key (dark blue) -> double-click
  - allows create and updates of records
- Light Primary Key (light blue) -> 2x double-click
  - allows updates only
- Delete Key (black skull) -> 3x double-click
  - allows create, update and delete (Deletion mode see chapter: [Deletion Mode](#))



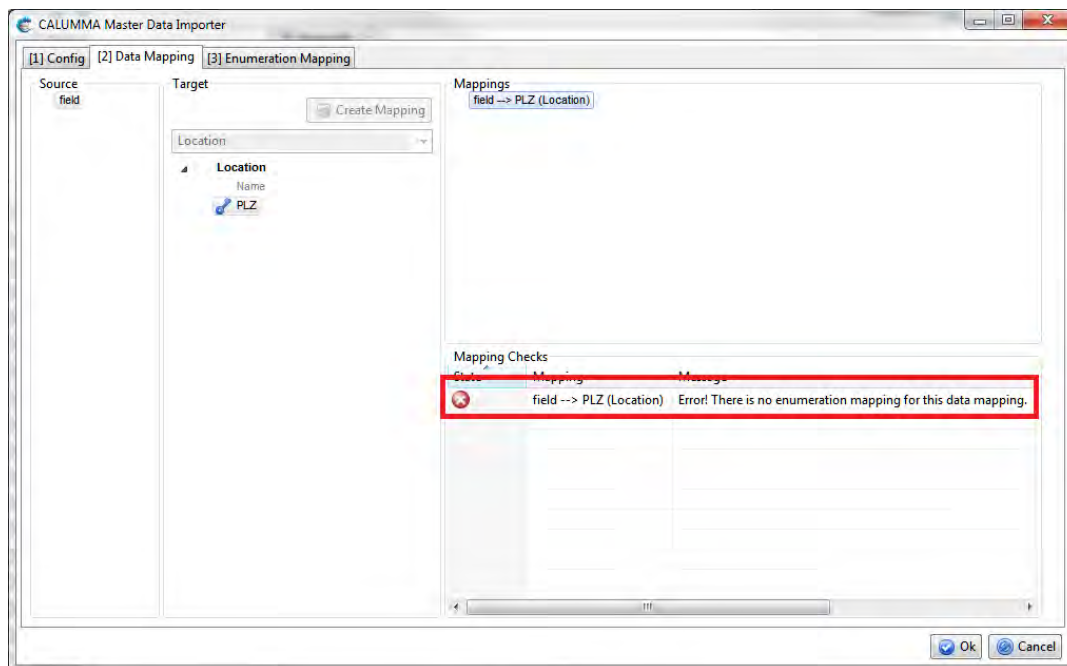
The "Mapping-Checks" area shows any error messages or information regarding your mappings.

### 3.3 Enumeration Mapping

The following screenshot from the CALUMMA Management Tool shows the enumeration "Location". A new Enumeration "PLZ" has been added to "Location". To map to this new enumeration a "Enumeration Mapping" needs to be created.

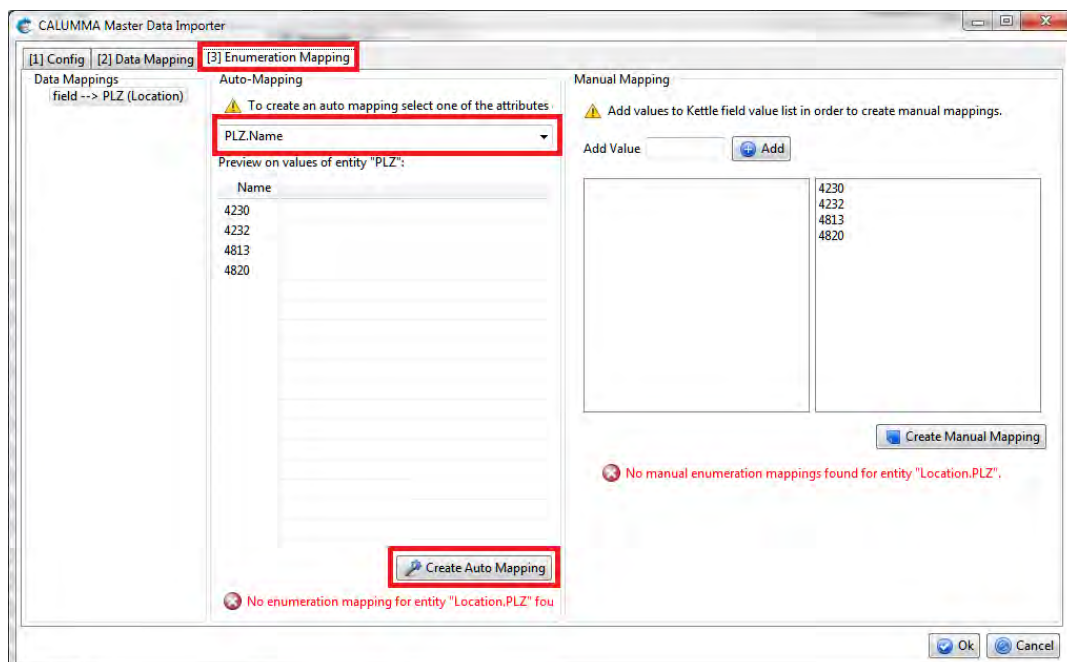


This mapping is always necessary when a data mapping on an enumeration Attribute is created. An info message is shown in the "Mapping-Checks" area.

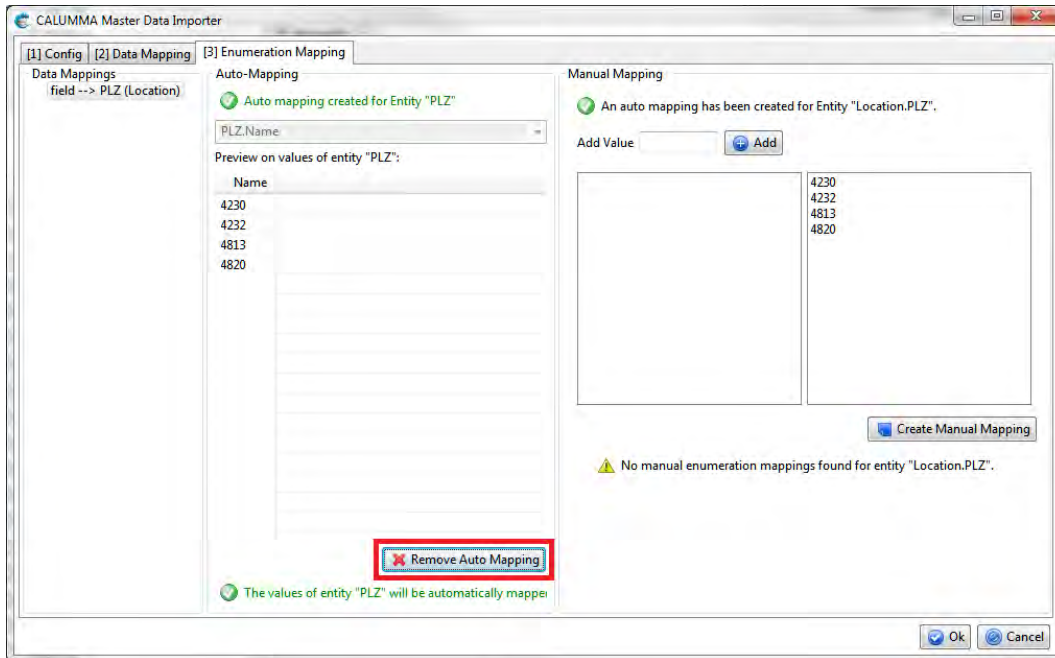


## Auto Mapping

There are two types of enumeration mappings, the auto mapping and the manual mapping. It depends on your data which one to use. If your data matches the items of the enumerations, you can use auto mappings. In the "Enumeration Mapping" tab on the left side, the data mappings that need an enumeration mapping are shown. To create an enumeration auto mapping, click on the data mapping on the left side and select an attribute from the combo box. Finally, click on "Create Auto Mapping."

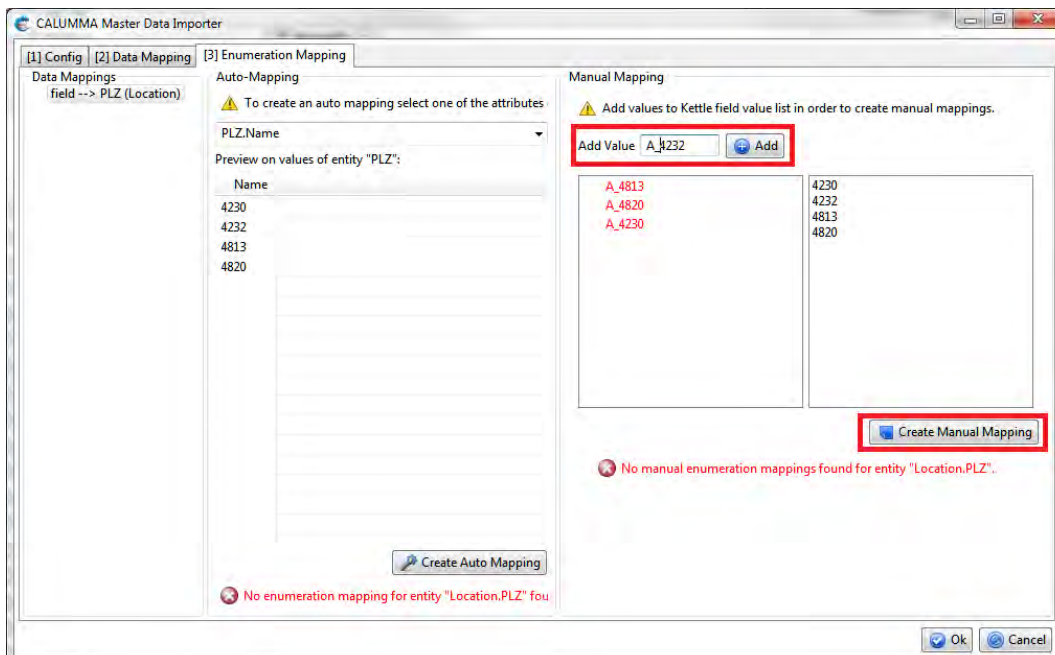


To delete an auto mapping, click on "Remove Auto Mapping".

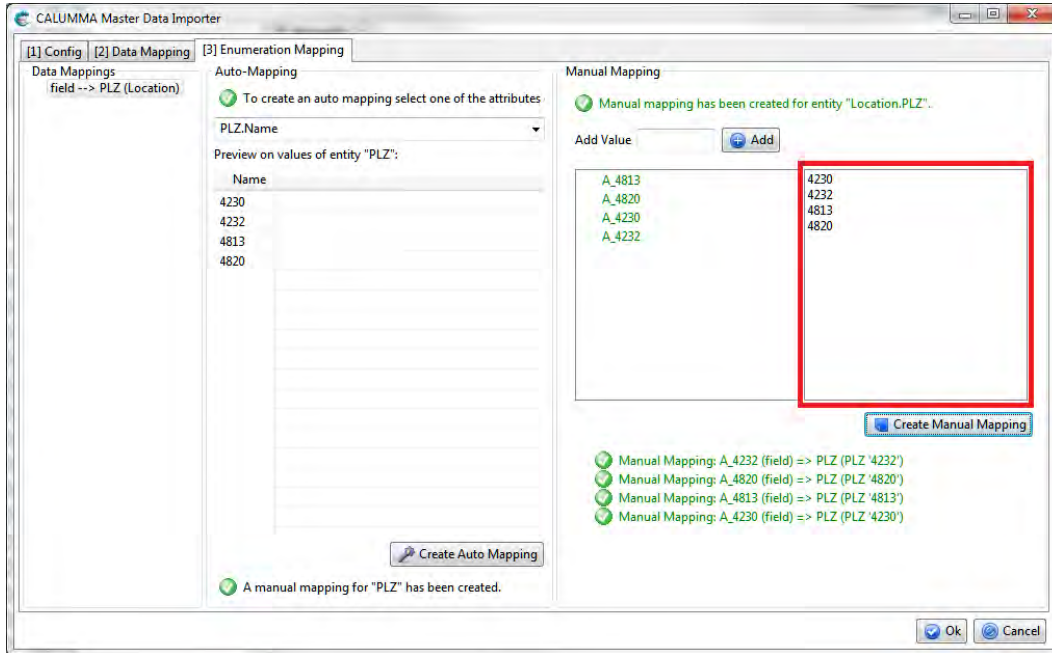


## Manual Mapping

In case your data does not match the items of the enumeration, manual mappings are needed. On the right side of the following screenshot is the area for the manual mappings. You need to add the values you want to map and then create the manual mappings by clicking on an item of the list of enumeration values. The example below maps postal codes manually.

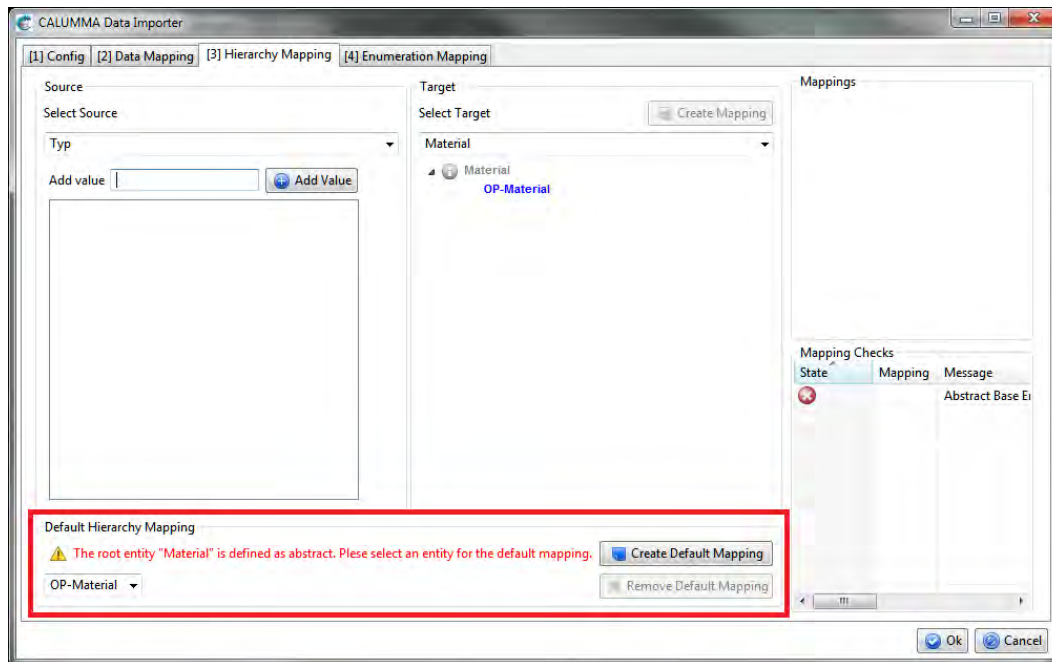


Mapped items are shown in green. The list below shows all mappings. A mapping can be deleted by clicking "Delete Mapping" in the context menu.



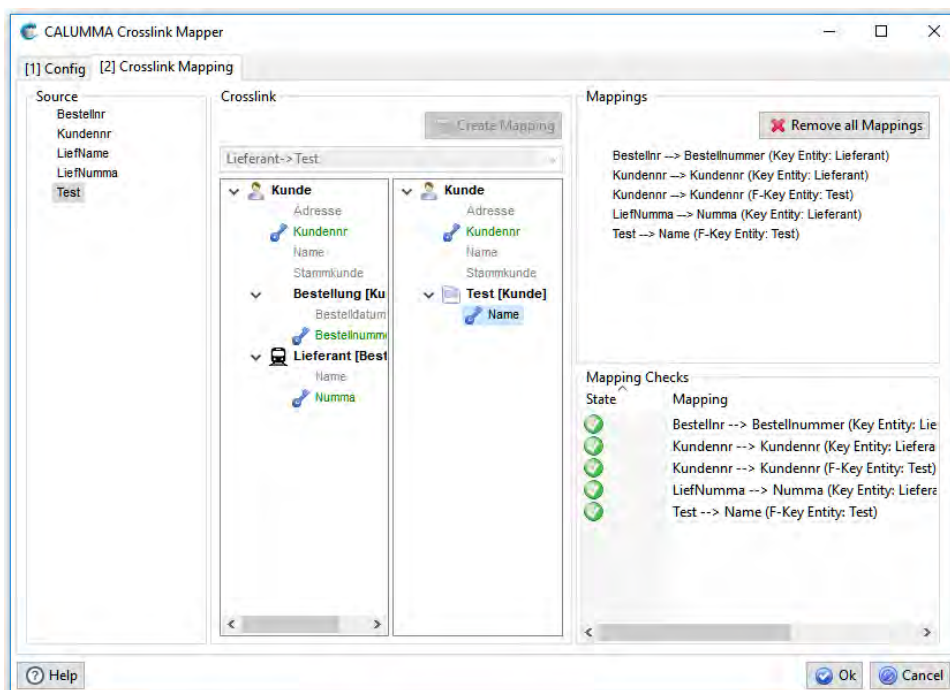
### 3.4 Hierarchy-Mapping

The "Hierarchy Mapping" is shown only when a "Hierarchy Mapping" is needed. When the root entity is abstract a default hierarchy mapping is necessary in case there is no hierarchy mapping defined. The Importer will then create records from this entity. In the example below is an abstract entity "Material", so no records from this entity can be created. Due to the default hierarchy mapping all materials are automatically added to "OP-Material".



### 3.5 Crosslink Importer

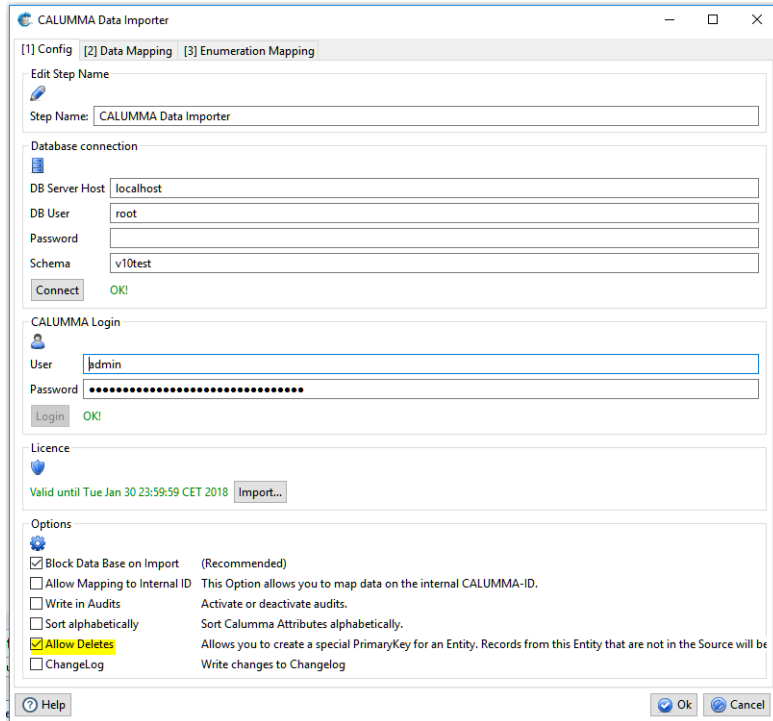
Is there a crosslink defined in the CALUMMA Management Tool, records can be linked with the crosslink importer. Select the crosslink from the combo box and map the attributes from the trees below. There is no need to specify key fields, every mapped field is a key-field.



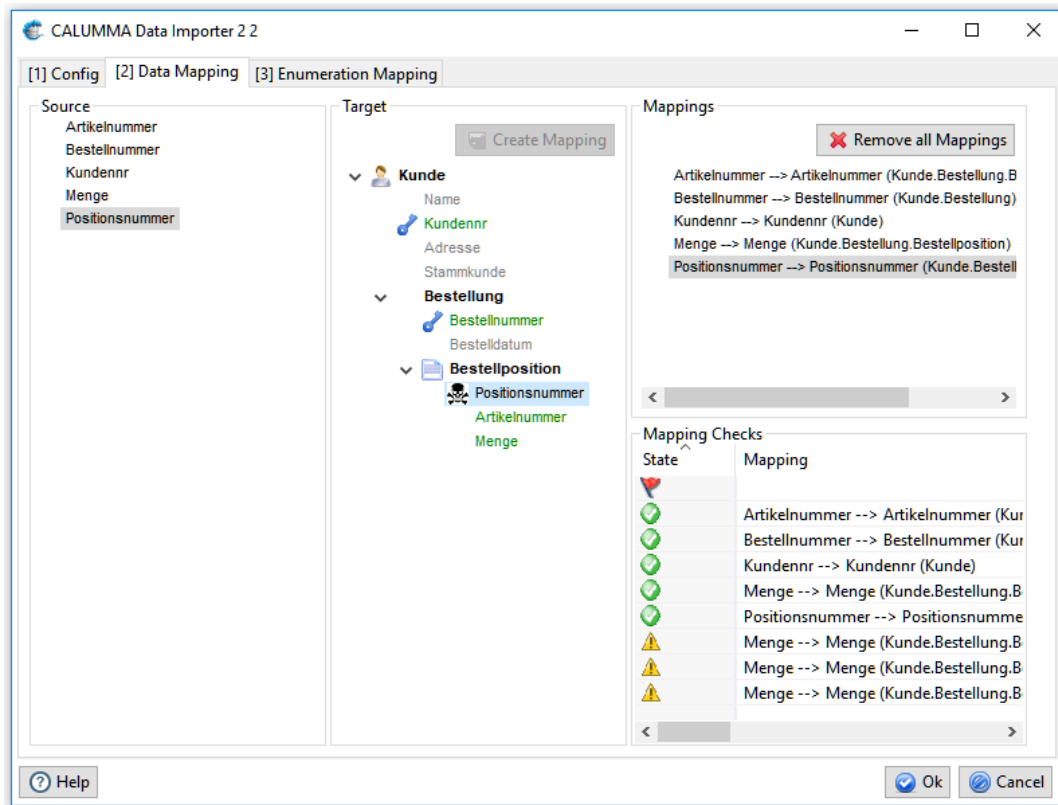


## 3.6 Deletion Mode

First, the deletion mode needs to be activated in the "Config" tab of the importer.



Then, add the delete key (black skull) to the entity that you want to delete. This key also allows create, so records will be created.



Data that is not in the source file will be deleted from the data base. The deletion mode only works within a hierarchy, for example: only the positions from this particular order and customer are deleted. All other positions stay in the data base.

Examine preview data

Rows of step: Positionen (22 rows)

#	Bestellnummer	Positionsnummer	Menge	Artikelnummer	Kundennr
1	456567	1	1	234	28345
2	456567	2	2	456	28345
3	4565678	1	1	678	38475
4	4565678	2	3	3454	38475
5	4565678	3	5	34656	38475
6	78789	1	3	23234	348759
7	34576	1	1	56757	349875
8	34576	2	1	345567	349875
9	34576	3	1	12343	349875
10	34577	1	1	4554768	28345
11	34578	1	1	324456	28345
12	34579	1	2	982754	28345
13	34580	1	1	8234795823	28345
14	34581	1	2	943759	28345
15	34582	1	1	9834702	28345
16	34583	1	3	93487523	28345
17	34584	1	2	903845	28345
18	34585	1	1	1234	28345
19	34586	1	1	394875	28345
20	34587	1	1	8723405	28345
21	34588	1	2	2034875	28345
22	34589	1	2	2873450	28345