 **tFileExcelSheetOutput**

http://www.cimt-ag.de

**Purpose**

This component creates and writes into Excel sheets.

This component needs the components tFileExcelWorkbookOpen (open a file or creates a workbook)

and tFileExcelWorkbookSave (writes the workbook to the same or another file)

Advantages of this component:

* The columns to write can be set also with gaps
* The format can be set for numbers and dates
* Can create a copy of an existing sheet (templates)
* Can reuse existing styles (even alternating styles)
* Can write formulas
* Can reuse conditional formats for the written cells
* Uses always the latest Apache POI API
* Can include header names (not limited to the schema column names)
* Can set the with of the column fit to the content
* Can write comments
* Can freeze a region (e.g. to see always the header line)

**Talend-Integration**

This component can be found in the palette under File/Spreadsheet

This component provides several return values.

**Parameters for tFileExcelSheetOutput**

|  |  |
| --- | --- |
| **Property** | **Content** |
| Workbook | Choose the tFileExcelWorkbookOpen component holding the Apache POI Excel workbook |
| Schema | The schema of the output |
| Sheet name | The name or the index of the sheet. Please take care of a valid sheet name or simply type the index of the sheet your want to write.  If the sheet does not exist, it will be created automatically. |
| Create sheet as copy | This is very useful if you have a template sheet (e.g. with styles and conditional formats) and you want to create multiple sheets in the same way.  You can specify the source sheet with an index or a name (just like in sheet name)  *This option will not work if you use the Memory saving mode in the tFileExcelWorkbookOpen component.* |
| Exchange rows/columns | If this option is true the component writes every dataset in a transposed way. Every new row is at the end a new column and every input schema column fills a new row in the current excel column. |
| Append existing rows | The component detect the last row in the given workbook and starts writing at the next row. |
| Row start index | Enabled if Append existing rows is switched off.  This row is the first row where the first dataset (or the header row) will be written.  The number is 1-based (for a better understanding within excel row number) |
| Include header | At first row the header will be written. Normally the component takes the schema column names but you could also use self-defined names in the Column Configuration. |
| Use individual column positions | You can specify the columns in the Column Configuration in the column Sheet Column Name. Here you can use the Excel letter reference (“A” for the first column) or an index (0 for the first column). It is possible to have gaps between the different columns (unlike the build-in Talend components). |

|  |  |
| --- | --- |
| Column Configuration | **Column:** the schema column  **Header nam**e: an optional header name, if blank the schema name will be used  **Sheet column name:** the excel column where the schema column will be written at  **Date or Number format:** the format of the cell in excel. This is useful to set an appropriated date format or a number format. You can see all possible formats in Excel under custom formats. This option is only enabled of you do not use the existing styles!  **Auto size:** the column size will be set automatically depending on the largest size of the content  **Comment:** The content will be written as comment in the excel cell. The comment will be shown with a default width: 3 column and a height: 2 rows. The cell value will not be changed. Therefore it is possible to use for such columns an already used cell column without losing the formally written value.  **Hyperlink:** The content will be written as hyperlink url. The actual value of this cell will not be changed. If you do not have an explicit value for the cell set the url also as String typed value for this cell.  Hyperlinks starts with an URL protocol followed by :// will be handled as URL.  Hyperlinks starting with “mailto:” will be handled as email link.  All other hyperlinks will be handled as file type hyperlink.  **Group rows by:** If checked all rows, which has in this column the same value will be grouped. It can have side effects if you check more than one column for grouping when the value ranges overlaps. The grouping could be build unlike you expect it. Refer to the scenario 3. |
| Overwrite existing cell content with null allowed | If you want to keep existing content of a cell and avoid to clear it with null values, switch of this option |
| Remove last empty rows | After the processing of the input flow, the component can delete all rows after the last written row. This is useful if you read an existing excel file and update an existing sheet. |
| Freeze row | Here you can enter a row number (starting with 1), which should always be visible if the user scrolls the document in Excel. |
| Group columns | Configure the columns you want to group. Add groups separated with comma of column ranges separated with minus.  Example: “A-D, AB-AF” will create 2 groups of columns A-D and AB-AF.  Refer scenario 3. |
| Freeze column | Here you can enter the column name or index (0-based) which column should be keep visible at the left side if the user scrolls vertical. |
| Reuse existing style from the first written row | This works if you read an existing excel file and fill an existing sheet (even if it a template).  The component memorizes the styles of all written columns and applies them to all new created rows. The first written row means the first written row with data and does not mean the row containing the header!  It is recommended to design you Excel file as a template file (not a Excel template – it means a normal Excel file!) and define header and styles and so on and reuse them in your target document.  *In the memory saving mode of tFileExcelWorkbookOpen this function creates sometimes a malformed content (found in Apache POI 3.10 final)* |
| Reuse existing style alternating | This option enhances the previous option and keeps the styles from the first and the second written data row and applies them to the even and odd new created rows. This is useful if you want to have alternating styles to better separate the rows e.g. with alternating background colours. |
| Reuse conditional formats from the first data row | You have to specify conditional formats in your template sheet and these conditional formats will be extended in its cell range to the written cells.  The component clones the existing conditional format as new format with the full cell range.  If the conditional format is already defined as full range, the component does nothing on it). |
| Reuse the row height from the first data row | The height from the first data row will be applied to all new created rows. |
| Extend cell range for written tables | If the component writes rows into a so-called Table (special Excel concept) the component can extend the initially configured cell range to the cell range of the written cells. Only the row number will be changed, the width of the Table keeps unchanged.  This feature is very helpful to update Pivot-tables. See the chapter about Updating Pivot.  Refer scenario 3. |

**Typical examples of Date or Number formats:**

|  |  |
| --- | --- |
| **Format** | **Example value** |
| YYYY-MM-DD | 2014-04-28 |
| DD.MM.YYYY hh:mm:ss | 28.04.2014 14:36:59 |
| #,##0.00 | 1,234.23 |
| #### | 1234 |

Please take care you use always the English formats. Excel usually shows a language depending format but internal it use always the English format and the Apache POI API does not translate the formats from language depend formats to English.

**Return values of the component:**

|  |  |
| --- | --- |
| Value | Content |
| NB\_LINE | Number of lines written |
| ERROR\_MESSAGE | Error message if something went wrong |
| LAST\_ROW\_INDEX | Index of the last written row in this sheet. This can be used to append in a next sub job to the same sheet. This is more secured than use the Append rows option because you know exactly where to start with the next row. |

**Writing formulas:**

It is possible to write formulas with this component. To do this define a String typed schema column and start the content with “=”. This will switch the cell type from a simple text type to a formula cell.

All functions must be written in English language. It can be sometimes a bit cumbersome to get the original English name for the functions – sorry there is currently no way to translate them automatically.

The row reference can be set with the expression “{row}” in the formula. The component will replace it with the current row number.

Examples:

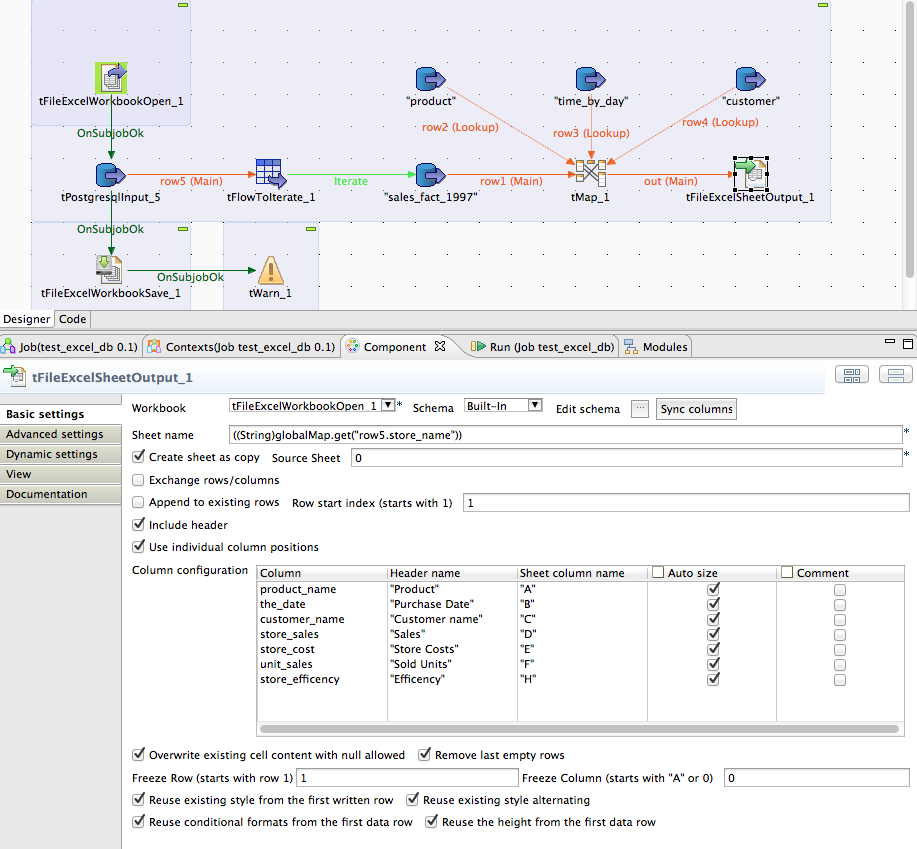
“=A{row}+D{row}”

“=SUM(E{row}:H{row})”

It is generally a good idea if possible to install the English language package for Office.

**Scenario 1: Creating multiple sheets according the input data**

Read a template excel file and create a lot of new sheets base on a template sheet:



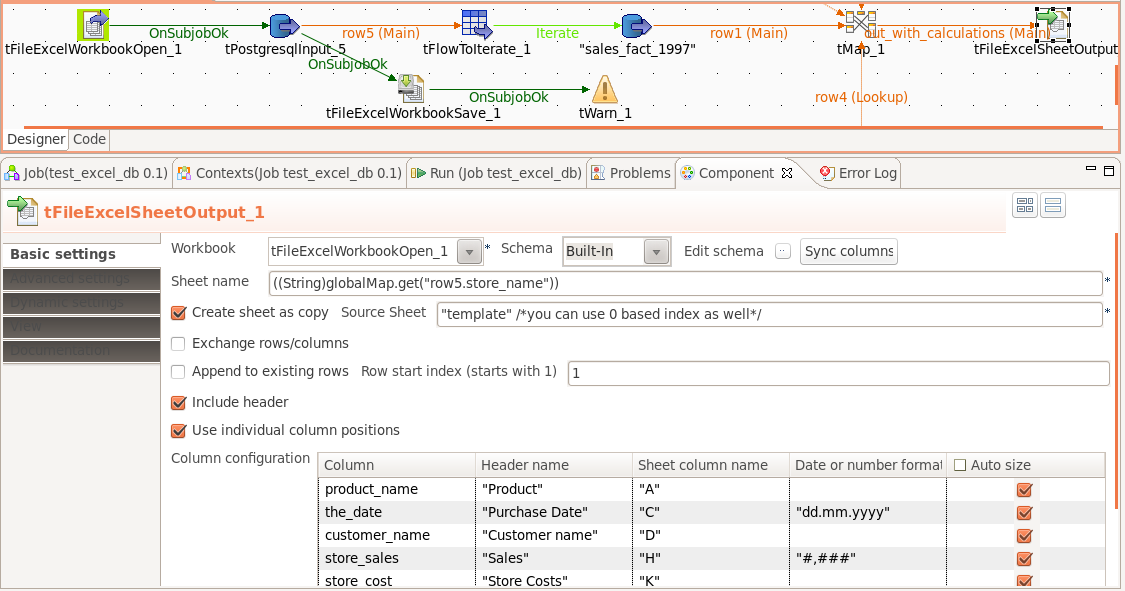
In this example the tFileExcelWorkbookOpen reads an existing file.

This file will get as much as iterations happens new sheets as a copy of the first sheet.

At the end a new excel file will be written with tFileExcelWorkbookSave.

**Scenario 2: Write Excel file with self defined header and formats**

Create a new Excel file without using a template and define the data formats for columns.



As template for the new sheets a sheet with the name “template” will be used here. You can name it as you want.

You will get the column Date or number format if you switch off the option “Reuse existing style from the first row”.

**Scenario 3: Updating Pivot-tables**

A normal Pivot-table is based on a preconfigured cell range. Unfortunately we cannot configure or create the Pivot-table directly. If a Pivot-table is based on an Excel-table the Pivot-table use always all data in this Excel-table even when the Excel-table has more rows.

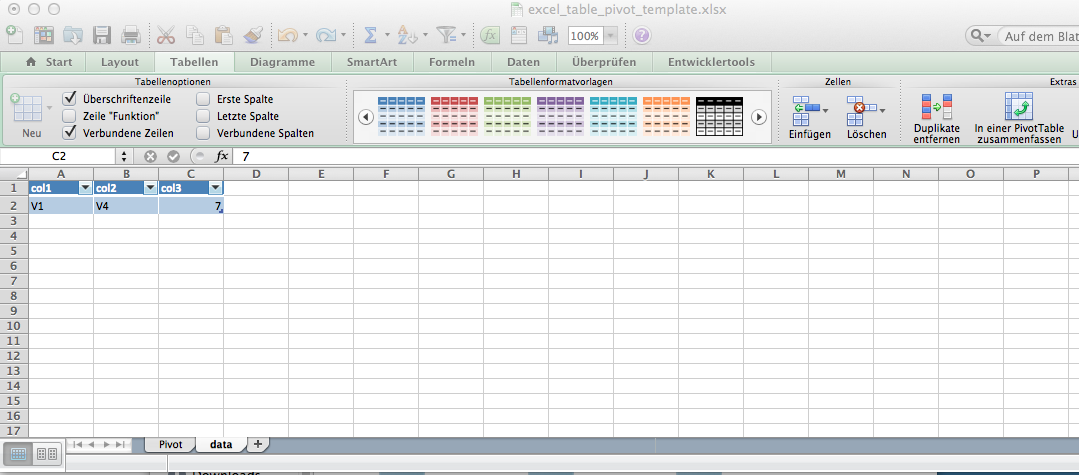
The way to update a Pivot-table is to update and reconfigure the underlying Excel-table.

The Pivot-table should be configured as “Update data when the document will be opened”

Following steps are necessary to update a Pivot-table:

1. Create an Excel file and add an Excel-table
2. Fill this Excel-table with some example data and create a Pivot-table based on this Excel-table
3. Use this file (we now call it the excel template) in your job in the tFileExcelWorkbookOpen component
4. Configure the tFileExcelSheetOutput to write with the first data row (a header line does not make sense if you write into an Excel-table) in the first data row of the Excel-table.
5. Check the option “Extend cell range for written tables”

**Step 1 Create the Excel file with the Excel-table**

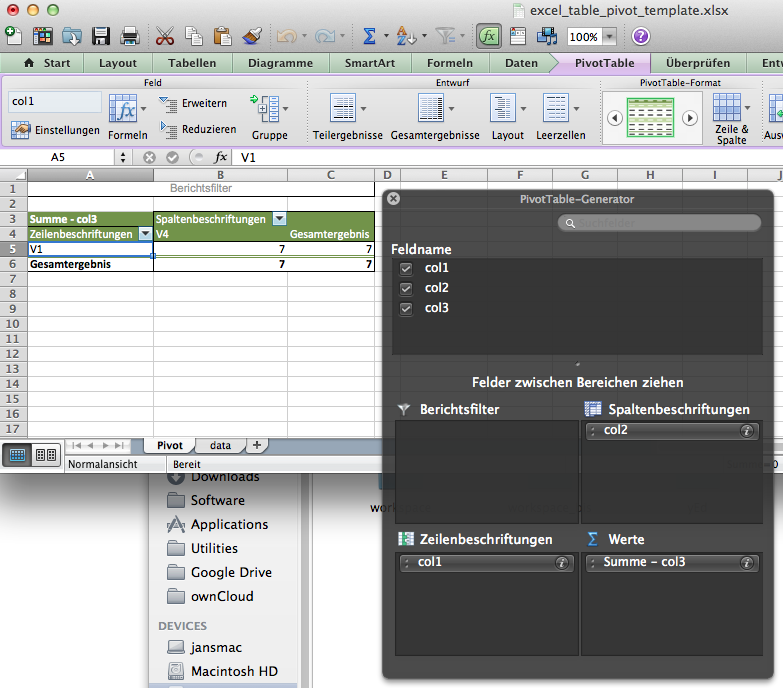


This is the sheet containing the Excel-table. Address this sheet in the tFileExcelSheetOutput component.

All overdue rows will be automatically deleted.

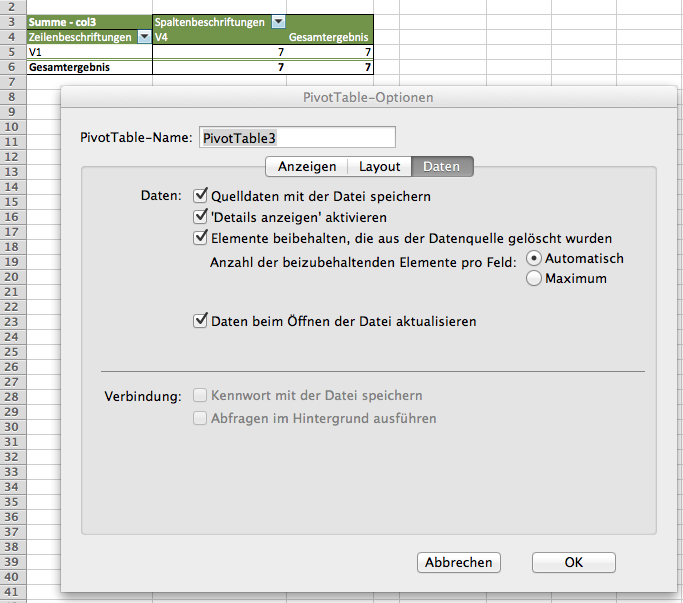
*In workbook with the memory saving mode of tFileExcelWorkbookOpen this function creates sometimes a malformed content (found in Apache POI 3.10 final) or simply does not work!*

**Step 2 Create Pivot based on the Excel-table**



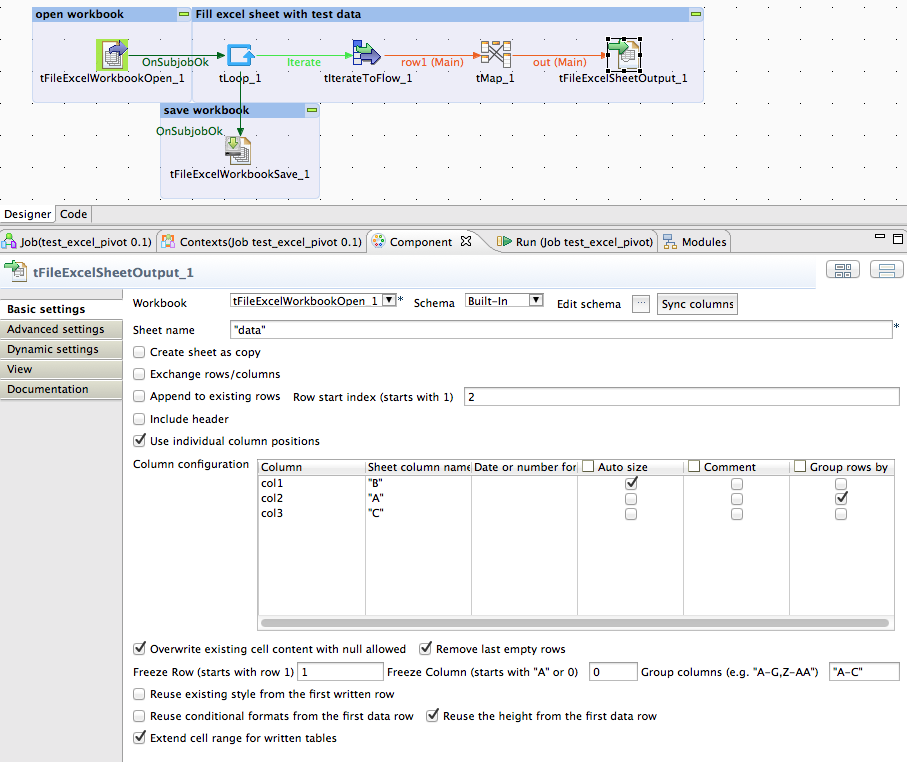
Configure the Pivot-table with the columns from the Excel-table.

In the options of the Pivot-table setup the table to “Update pivot when document opened”



“Daten beim Öffnen der Datei aktualisieren” -> Refresh data while opening the file…

**Step 3-5 will be done in Talend in the Job**

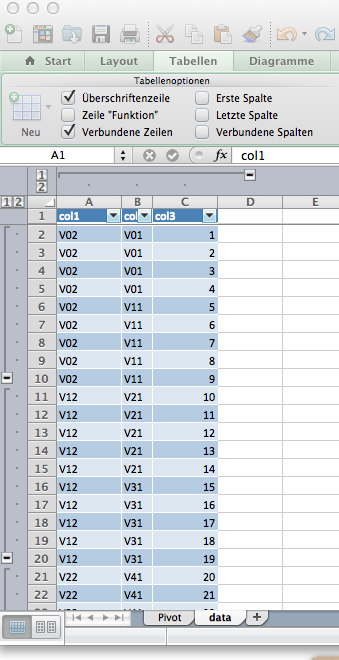


In the component tFileExcelWorkbookOpen choose your just created excel template file.

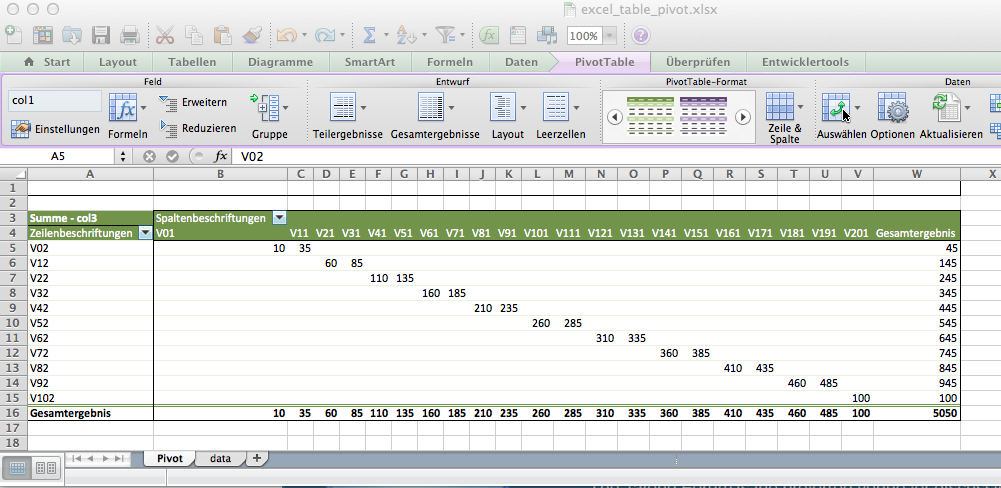
In the tFileExcelSheetOutput set the option Extend cell range for written tables – means all Excel-tables affected by the writing of this component will be extended in its cell range to the amount of the written rows.

The tLoop component will be used to create dummy example data. Normally you will find here a database input component or similar inputs from another data source probably also processed with a tMap.

As the result of this job the Excel-table is filled with data and the Pivot-table refreshes it self when Excel open this result file. This output file shows also the grouping function.



… here the result pivot table:



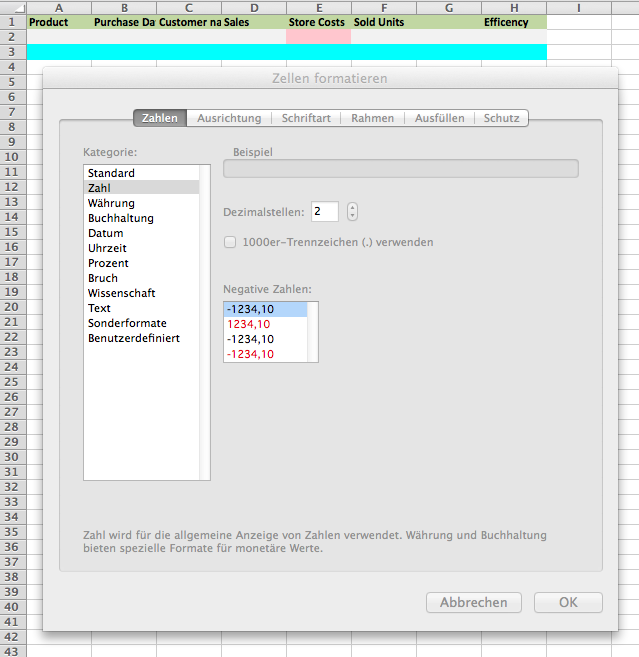
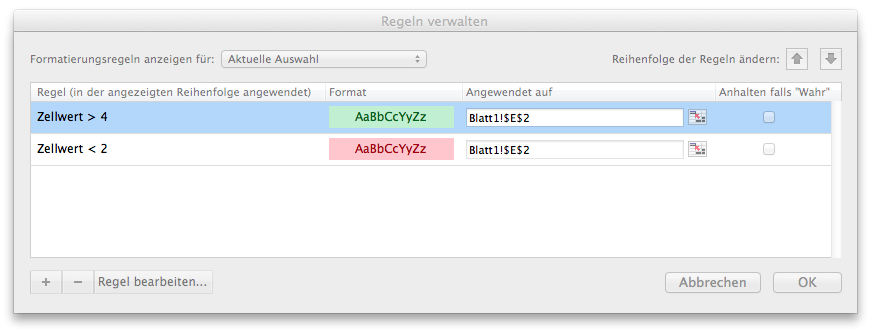
**Scenario 4: Reuse conditional formats:**

In this scenario we use also a template excel file in which all formats will be configured.

It is recommended to create for every related excel column its own conditional format.

If you want alternating colours define them in the first two data rows in excel.

Here an example template file screenshot:



Here the result after filling the sheet with the option “Reuse existing style from the first row” and “Reuse the existing style alternating” and “Reuse the conditional formats from the first data row”:

