

Data Import Wizard

Data import wizard is used to load data from various sources to standard cubes or time series databases. Data import wizard is available in various tools of Prognoz Platform 8.

The importer enables the user to load data using minimal settings and create all the objects that are required to access this data (cubes or time series databases, dictionaries, and tables).

This manual contains examples of import:

- [From table](#)
- [From crosstab](#)

The correct import of data to cubes or time series databases requires observing certain requirements and recommendations for the structure and organization of data in the source file.

General Recommendations

- Values of all dates must be in the same format. This format must be reproducible using the codes reserved in data format settings, otherwise the dates are identified as string values. The following codes are available to specify the format:

Data format		Displayed value
Russian	English	
М	M	Month displayed as a number between 1 and 12
ММ	MM	Month displayed as a number between 01 and 12
МММ	MMM	Month displayed as Jan-Dec
ММММ	MMMM	Month displayed as January-December
Д	D	Day displayed as a number between 1 and 31
ДД	DD	Day displayed as a number between 01 and 31
ДДД	DDD	Day displayed as Mon-Sun
ДДДД	DDDD	Day displayed as Monday-Sunday
ГГ	YY	Year displayed as 01-99
YYYY	YYYY	Year displayed as 1900-9999

- Do not use different formats to specify same-type data. For example, one delimiter should be used to specify all real values in the file. Violating this condition may cause incorrect identification of data, in which not all data will be imported.
- To import data from Microsoft Excel and Microsoft Access files in 64-bit operating systems, Prognoz Platform 8 and Microsoft Office also must be 64-bit.
- If import is executed to existing consumer, its structure must correspond to imported data. If they differ, the existing consumer structure will be changed. In this case the data contained in the consumer is lost.

IMPORTANT. For repositories based on Teradata DBMS, data import is supported only for Teradata version 14 or later.

File Requirements

- Importing to a cube requires that the source should have at least two data fields.
- Importing to a database requires that the source should have at least three data fields. One of the fields must contain data for a calendar dimension.
- When importing from a source that contains a crosstab, the crosstab must have at least two or more rows.

Requirements and Recommendations for Microsoft Excel Files

- Data under hidden rows or columns will be shown on identification. If this data should not be imported, move it to other sheets or remove from file.

	A	C	D	E		A	B	C	D	E	
1	Date	Shop	Indicator	Value		1	Date	State	Shop	Indicator	Value
2	01.01.2013	Hartford Store	Refund, USD	1192		2	01.01.2013	Connecticut	Hartford Store	Refund, USD	1192
3	02.01.2013	Hartford Store	Refund, USD	546		3	02.01.2013	Connecticut	Hartford Store	Refund, USD	546
4	03.01.2013	Hartford Store	Refund, USD	900		4	03.01.2013	Connecticut	Hartford Store	Refund, USD	900
5	04.01.2013	Hartford Store	Refund, USD	1326		5	04.01.2013	Connecticut	Hartford Store	Refund, USD	1326
6	05.01.2013	Hartford Store	Refund, USD	1078		6	05.01.2013	Connecticut	Hartford Store	Refund, USD	1531
7	06.01.2013	Hartford Store	Refund, USD	1418		7	06.01.2013	Connecticut	Hartford Store	Refund, USD	1078
8	07.01.2013	Hartford Store	Refund, USD	1130		8	07.01.2013	Connecticut	Hartford Store	Refund, USD	1418
9	08.01.2013	Hartford Store	Refund, USD	549		9	08.01.2013	Connecticut	Hartford Store	Refund, USD	1130
10	09.01.2013	Hartford Store	Refund, USD	1847		10	09.01.2013	Connecticut	Hartford Store	Refund, USD	549
11	10.01.2013	Hartford Store	Refund, USD	1847		11	10.01.2013	Connecticut	Hartford Store	Refund, USD	1847

- Do not calculate any totals because this data will be imported as a separate record.

	A	B	C	D	E		A	B	C	D	E	
1	Date	State	Shop	Indicator	Value		1	Date	State	Shop	Indicator	Value
2	01.01.2013	Connecticut	Hartford Store	Refund, USD	1192		2	01.01.2013	Connecticut	Hartford Store	Refund, USD	1192
3	02.01.2013	Connecticut	Hartford Store	Refund, USD	546		3	02.01.2013	Connecticut	Hartford Store	Refund, USD	546
4	03.01.2013	Connecticut	Hartford Store	Refund, USD	900		4	03.01.2013	Connecticut	Hartford Store	Refund, USD	900
5	04.01.2013	Connecticut	Hartford Store	Refund, USD	1326		5	04.01.2013	Connecticut	Hartford Store	Refund, USD	1326
6	05.01.2013	Connecticut	Hartford Store	Refund, USD	1531		6	05.01.2013	Connecticut	Hartford Store	Refund, USD	1531
7	06.01.2013	Connecticut	Hartford Store	Refund, USD	1078		7	06.01.2013	Connecticut	Hartford Store	Refund, USD	1078
8	07.01.2013	Connecticut	Hartford Store	Refund, USD	1418		8	07.01.2013	Connecticut	Hartford Store	Refund, USD	1418
9	08.01.2013	Connecticut	Hartford Store	Refund, USD	1130		9	08.01.2013	Connecticut	Hartford Store	Refund, USD	1130
10	09.01.2013	Connecticut	Hartford Store	Refund, USD	549		10	09.01.2013	Connecticut	Hartford Store	Refund, USD	549
11	10.01.2013	Connecticut	Hartford Store	Refund, USD	1847		11	10.01.2013	Connecticut	Hartford Store	Refund, USD	1847
12				Sum	11517							
13				Average	2094							

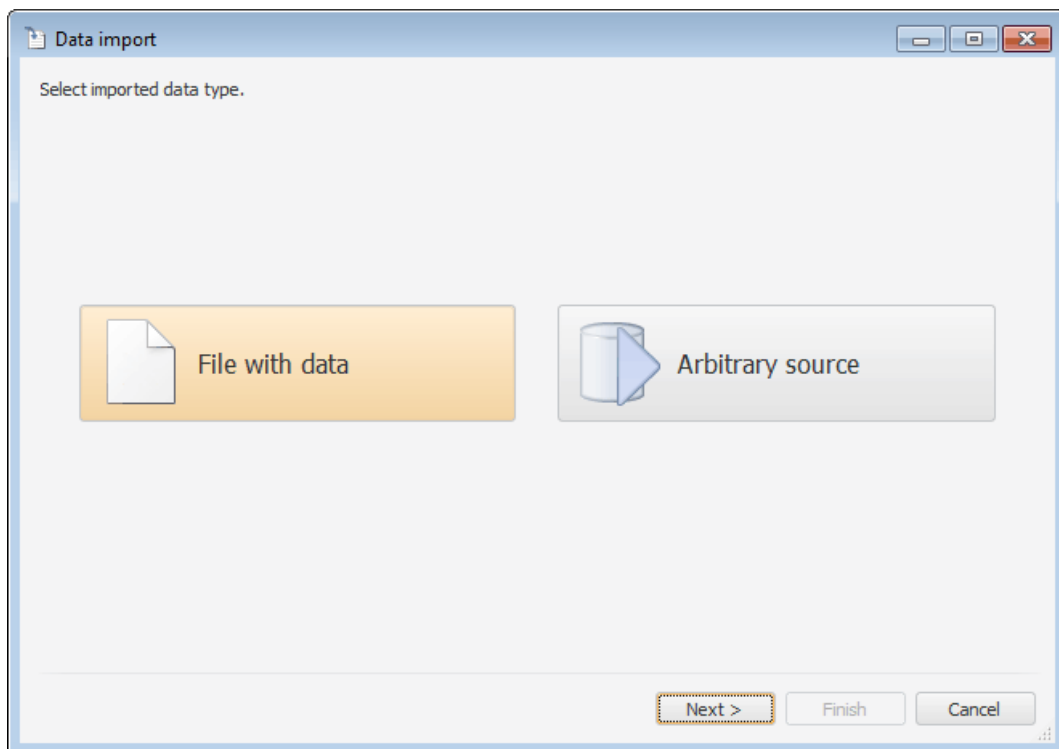
- Data import from large Excel files may require more virtual memory use. To avoid such situation, transfer data from an Excel file to the text file that will be used as a data source.

Import from Table

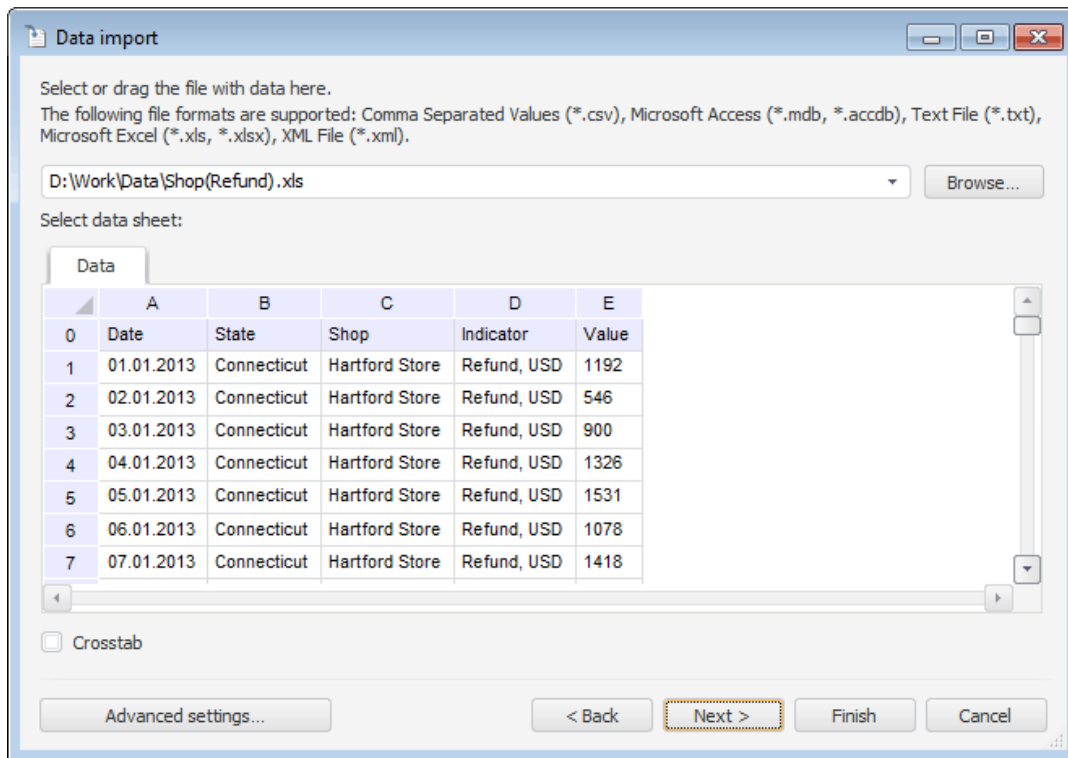
Consider the example of data import from a Microsoft Excel file to a new cube repository. Assume the following file structure and contents:

	A	B	C	D	E
1	Date	State	Shop	Indicator	Value
2	01.01.2013	Connecticut	Hartford Store	Refund, USD	1192
3	02.01.2013	Connecticut	Hartford Store	Refund, USD	546
4	03.01.2013	Connecticut	Hartford Store	Refund, USD	900
5	04.01.2013	Connecticut	Hartford Store	Refund, USD	1326
6	05.01.2013	Connecticut	Hartford Store	Refund, USD	1531
7	06.01.2013	Connecticut	Hartford Store	Refund, USD	1078
8	07.01.2013	Connecticut	Hartford Store	Refund, USD	1418
9	08.01.2013	Connecticut	Hartford Store	Refund, USD	1130
10	09.01.2013	Connecticut	Hartford Store	Refund, USD	549

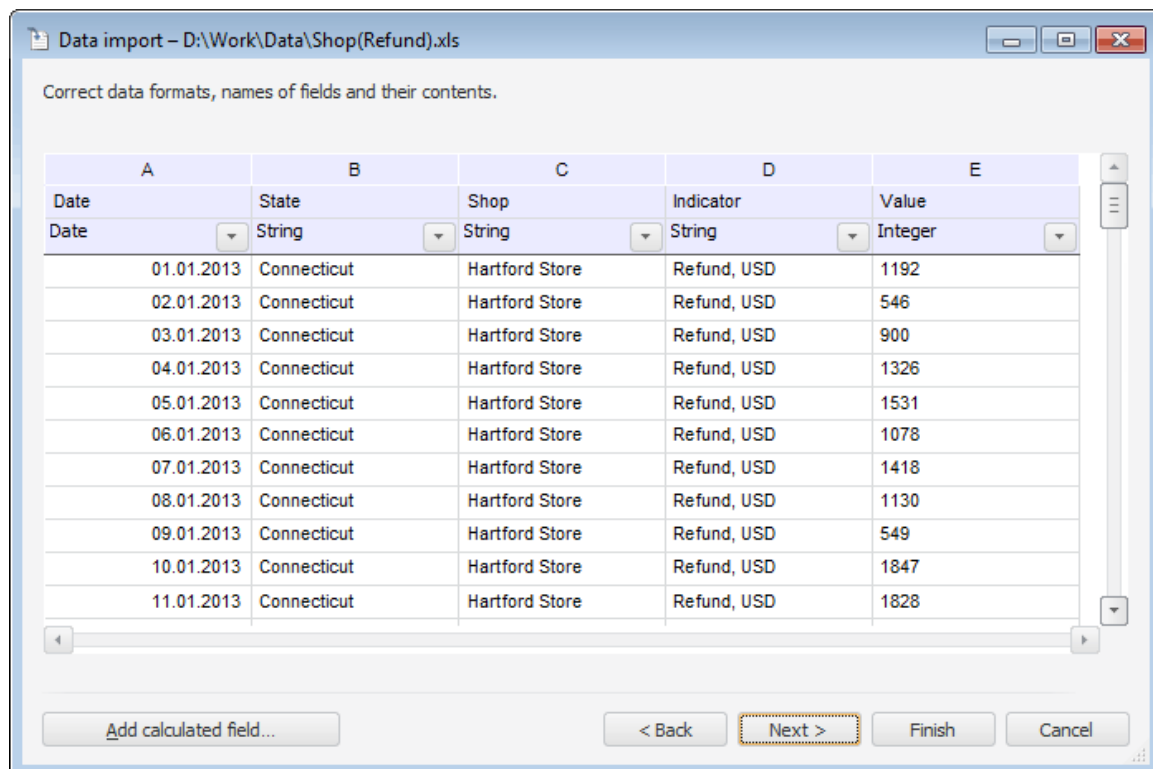
Start import wizard by selecting the Tools > Data Import > item in the object navigator's main menu. The following dialog box opens:



Select the File with Data option on the start page and click the Next button. Click the Browse button on the next page and select the file with the data:



After parsing file structure and data identification the first 100 strings are loaded and displayed from the selected file. Click the Next button.

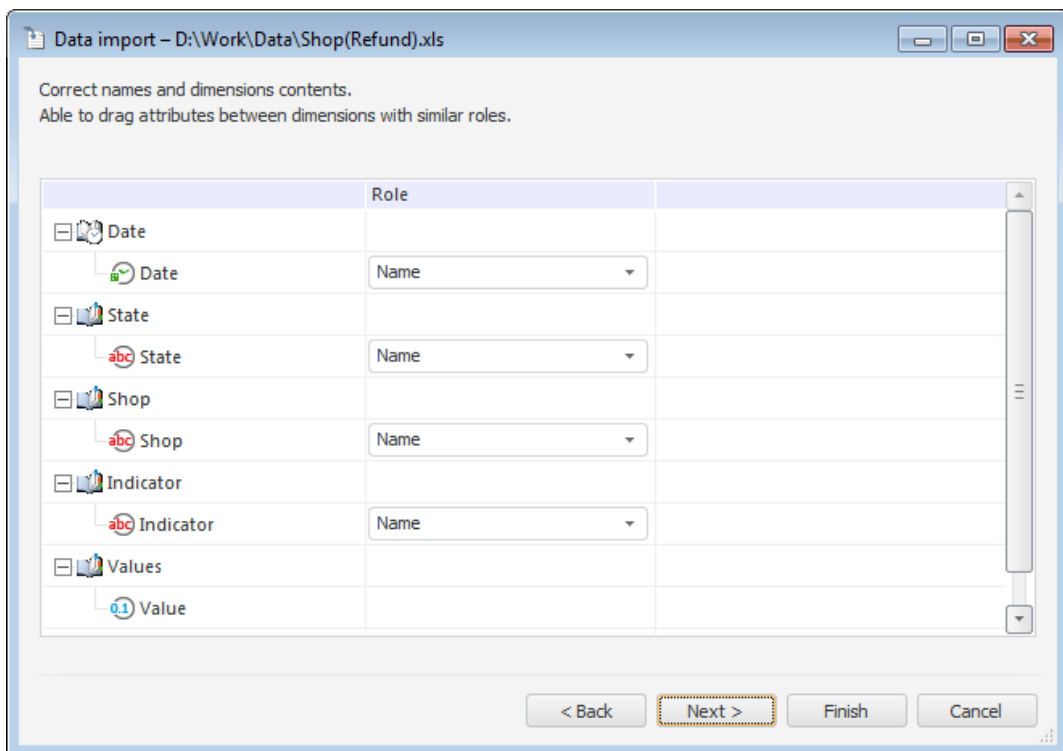


If required, correct the data:

- Edit field names.
- Exclude fields.
- Edit data format.

- Parse string fields.
- Add calculated fields.

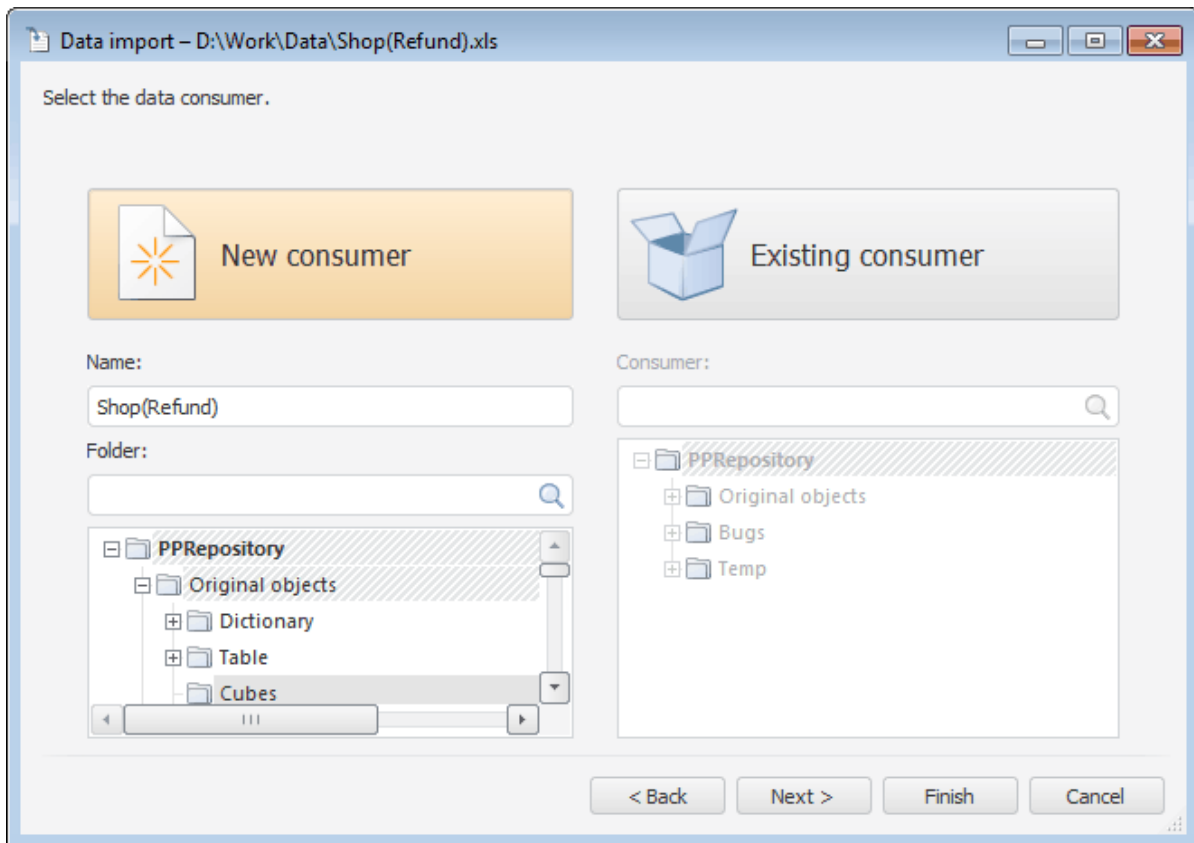
After correcting the content of the data click the Next button.



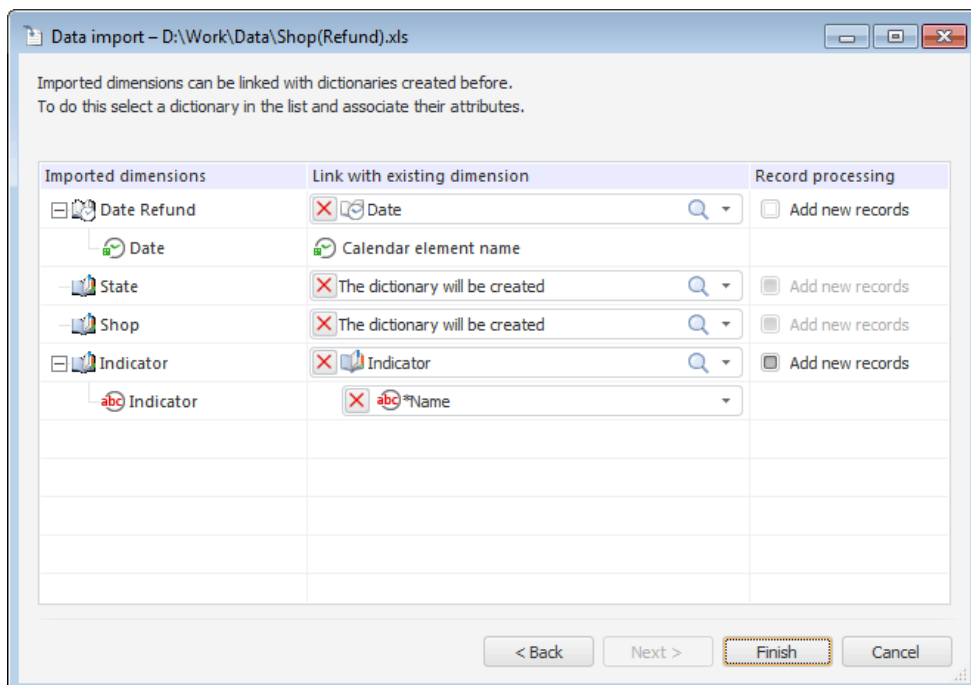
The fields are automatically divided into dimensions, values of text fields are imported as values of the Name attribute of dimension elements. All fields having a real data type are imported as values. A fact in the cube dimension is created for each fields with value.

If required, the contents of dimensions and their combination can be changed. To do this, drag attributes to the desired dimension contents (attributes are displayed as child elements of the dimensions, so that each attributes has its role). If several attributes within one dimension have the Name role, the dimension is created as hierarchical one.

After specifying dimension contents click the Next button.



Click the New Consumer button, name a new cube and select the folder to save the cube in the repository. Click the Next button.



If the repository already contains dictionaries having the same purpose and structure as the imported dimensions, they can be selected from the drop-down lists. Only table or composite MDM table dictionaries and calendar dictionaries are available to choose from.

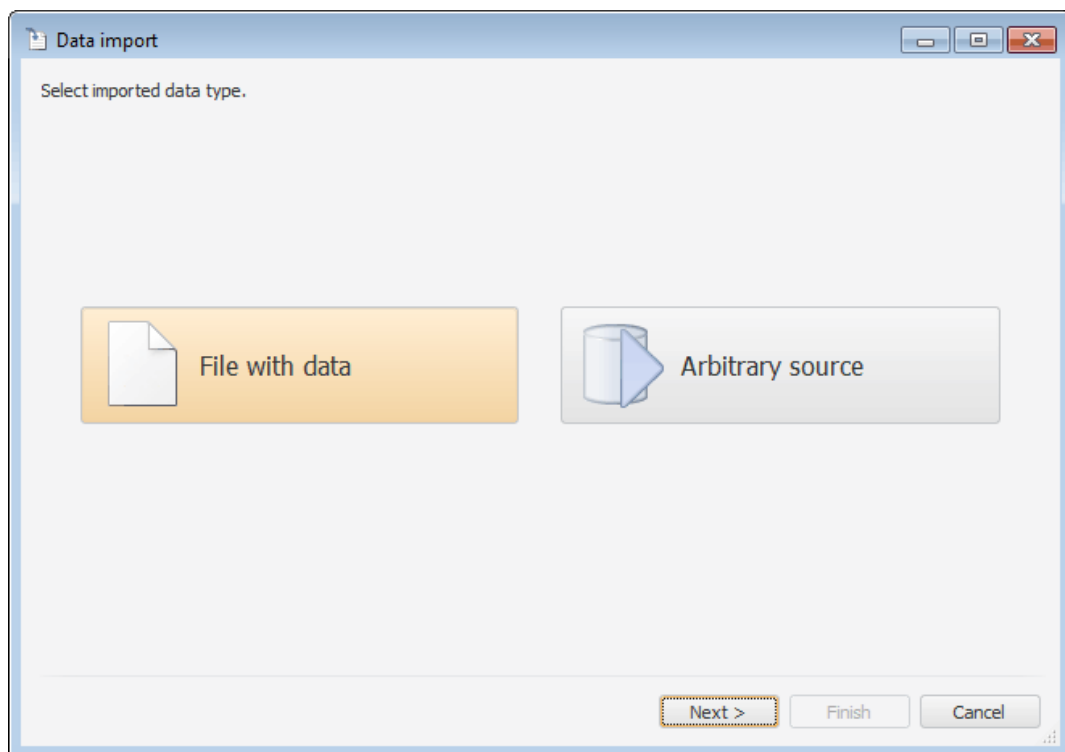
On clicking the Finish button the following objects are created and set up: a new cube, all required dictionaries and a table with imported data. Settings of data import to ETL task are also saved.

Import from Crosstab

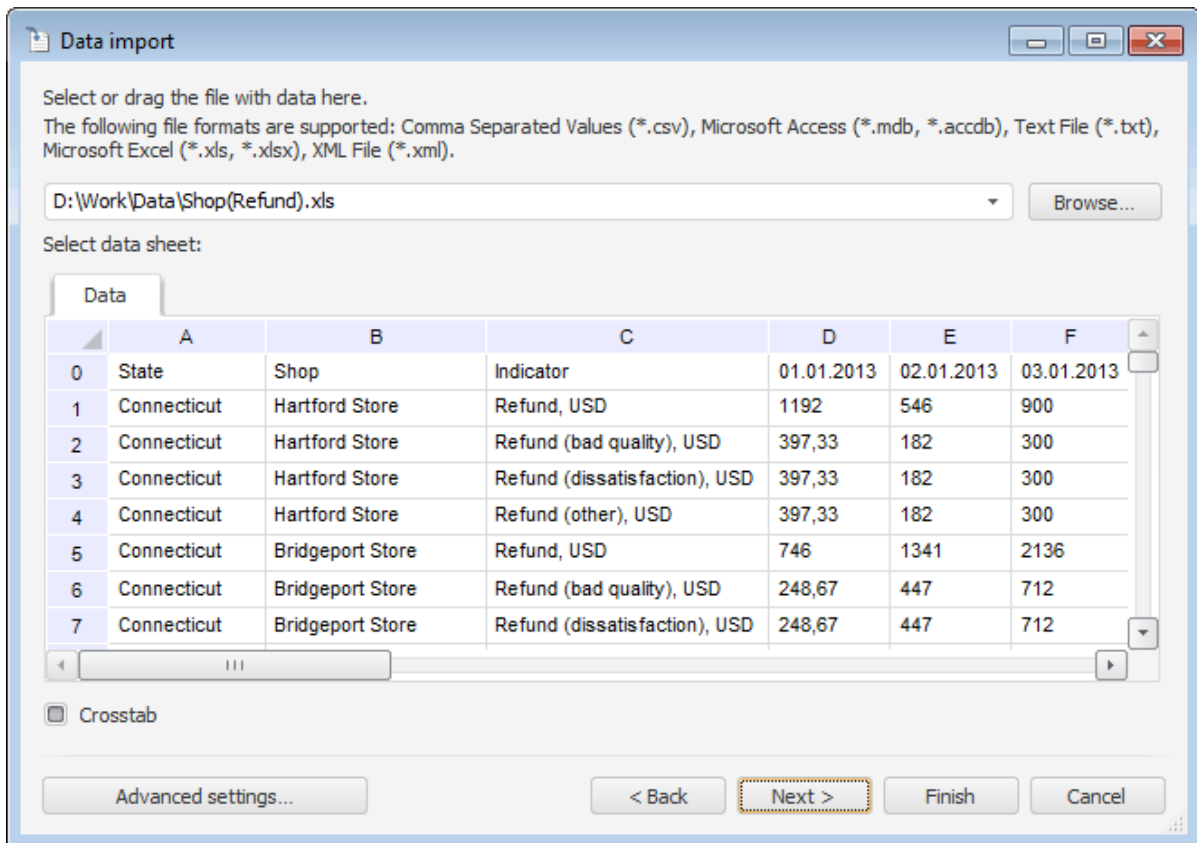
Consider the example of data import presented as a crosstab. Crosstab is a way of presenting data, where information about various factors is located in row and column headers, factual data is located in the intersection of rows and columns. Assume that the Microsoft Excel file to import contains the following structure and content:

	A	B	C	D	E
1	State	Shop	Indicator	01.01.2013	02.01.2013
2	Connecticut	Hartford Store	Refund, USD	1192	546
3	Connecticut	Bridgeport Store	Refund, USD	746	1341
4	Connecticut	Waterbury Store	Refund, USD	1579	924
5	Connecticut	Old Saybrook Store	Refund, USD	1087	453
6	Connecticut	Darien Store	Refund, USD	850	1239
7	Connecticut	New Haven Store	Refund, USD	1701	846
8	Connecticut	Litchfield County Store	Refund, USD	1575	2507
9	Connecticut	New London Store	Refund, USD	1104	870
10	Maine	Bangor Store	Refund, USD	3571	1233

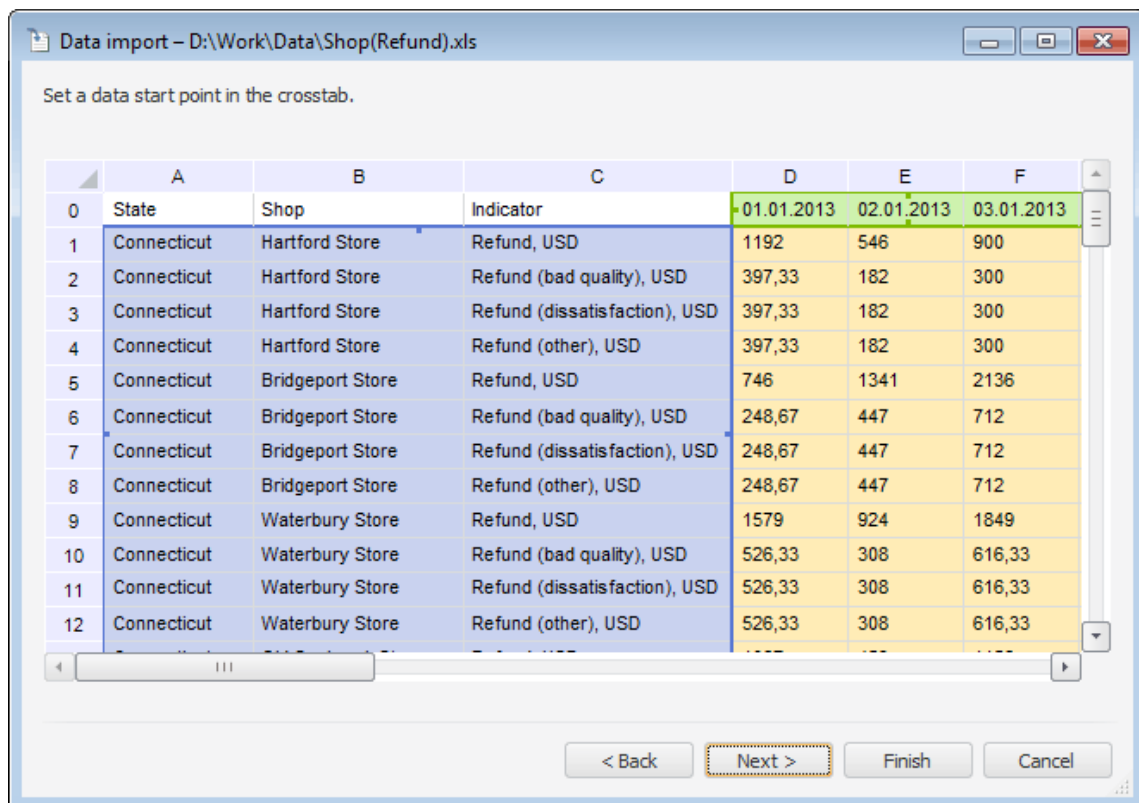
Start the import wizard by selecting the Service > Data Import main menu item.



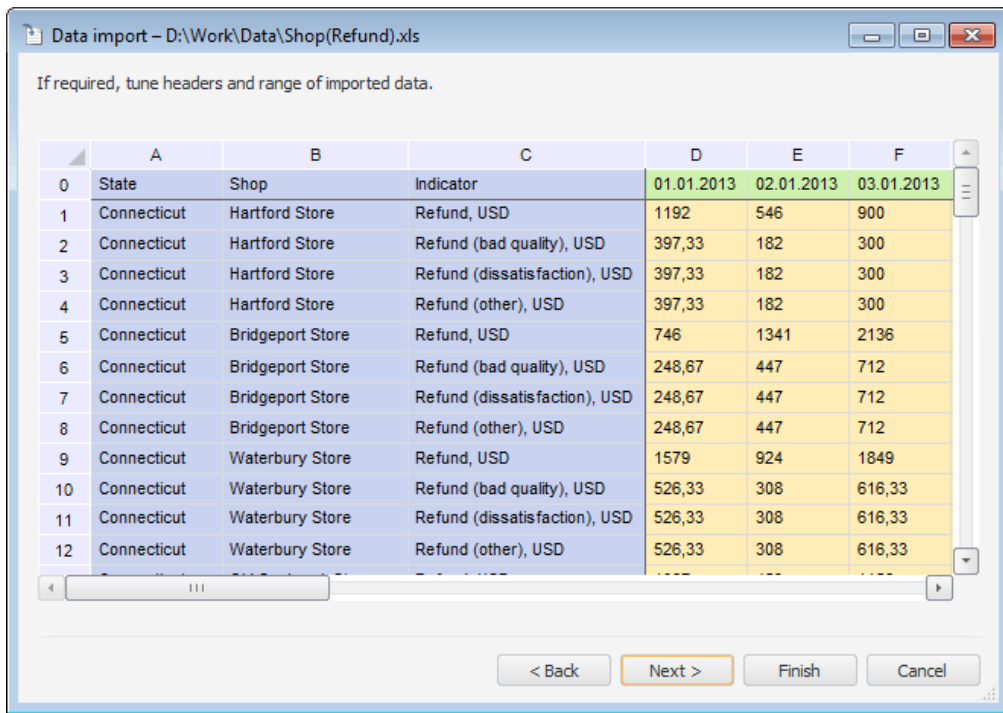
Select the File with Data option on the start page and click the Next button. Click the Browse button on the next page and select the file with the data:



Select the Crosstab checkbox and click the Next button.

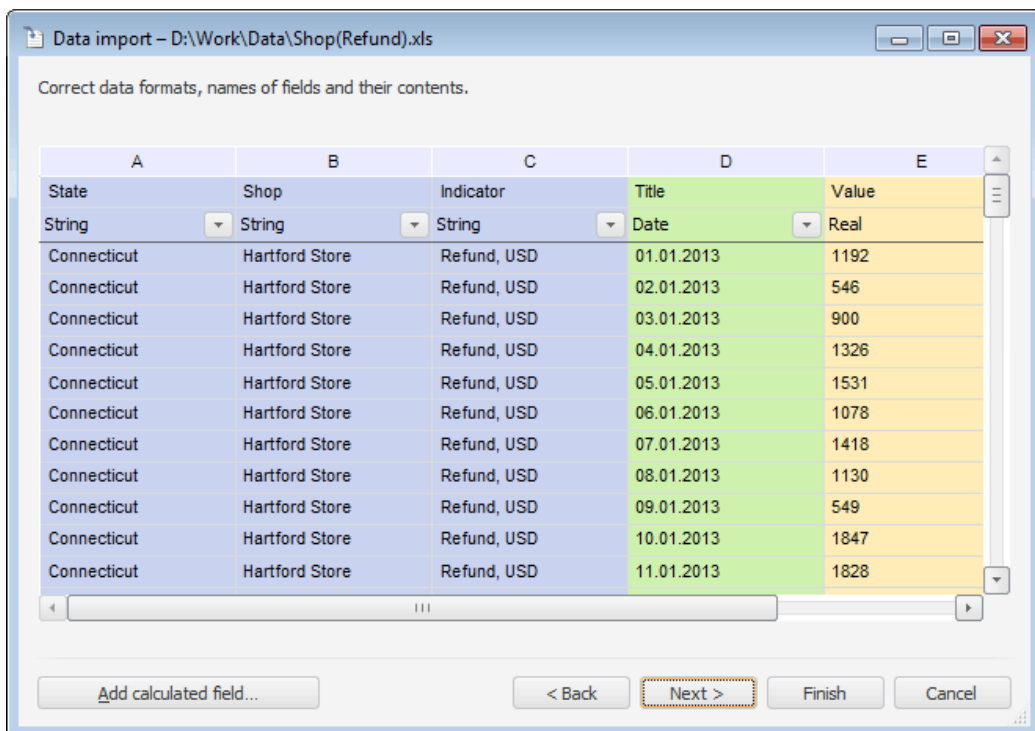


The import wizard will automatically determine the data start area. If it was determined incorrectly, change the data start area by selecting the corresponding cell. Click the Next button.



To exclude boundary columns from import, execute the Exclude context menu item. Data in these columns will not be imported.

Click the Next button.



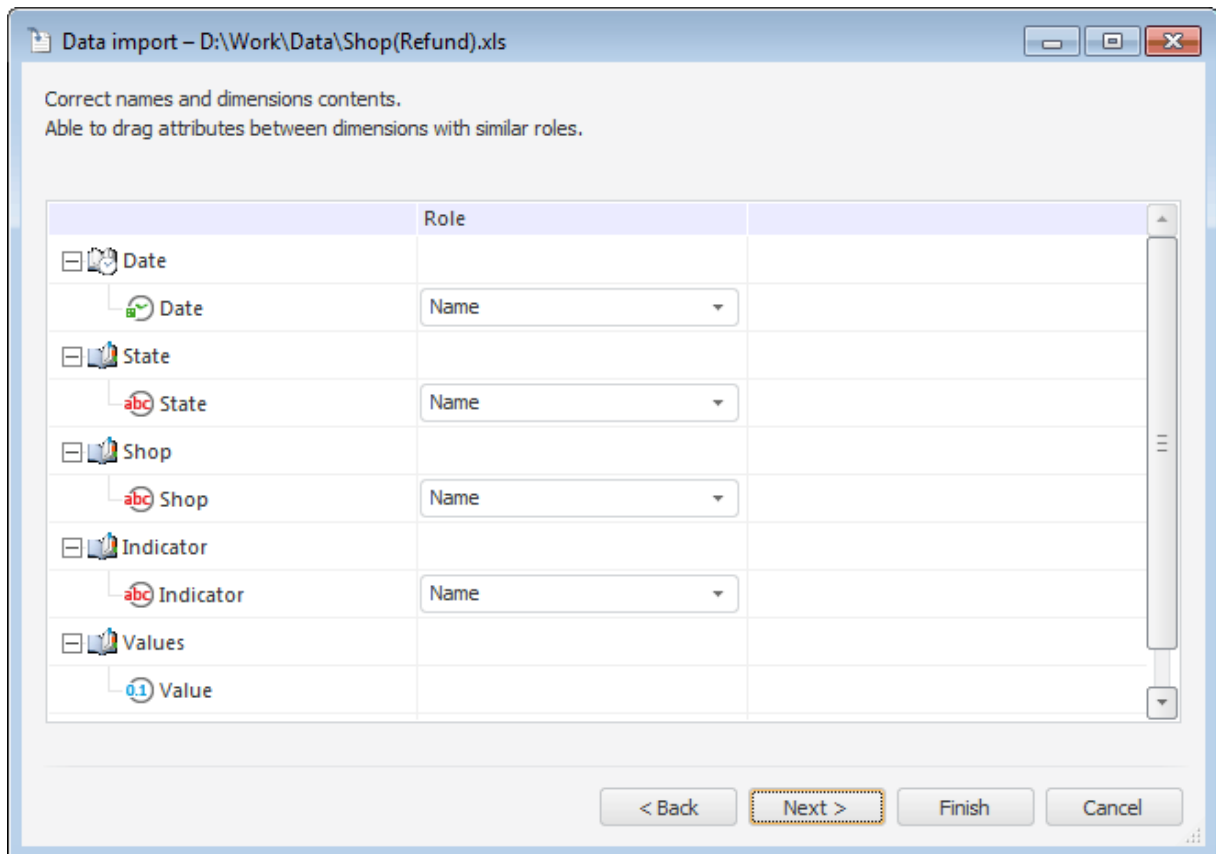
Identified data from the crosstab will be converted in the following way:

- Row and column headers are selected into separate fields.
- Numeric data, which start is denoted previously, is placed in the Value separate field.

If required, correct the data:

- Edit field names.
- Exclude fields.
- Edit data format.
- Parse string fields.
- Add calculated fields.

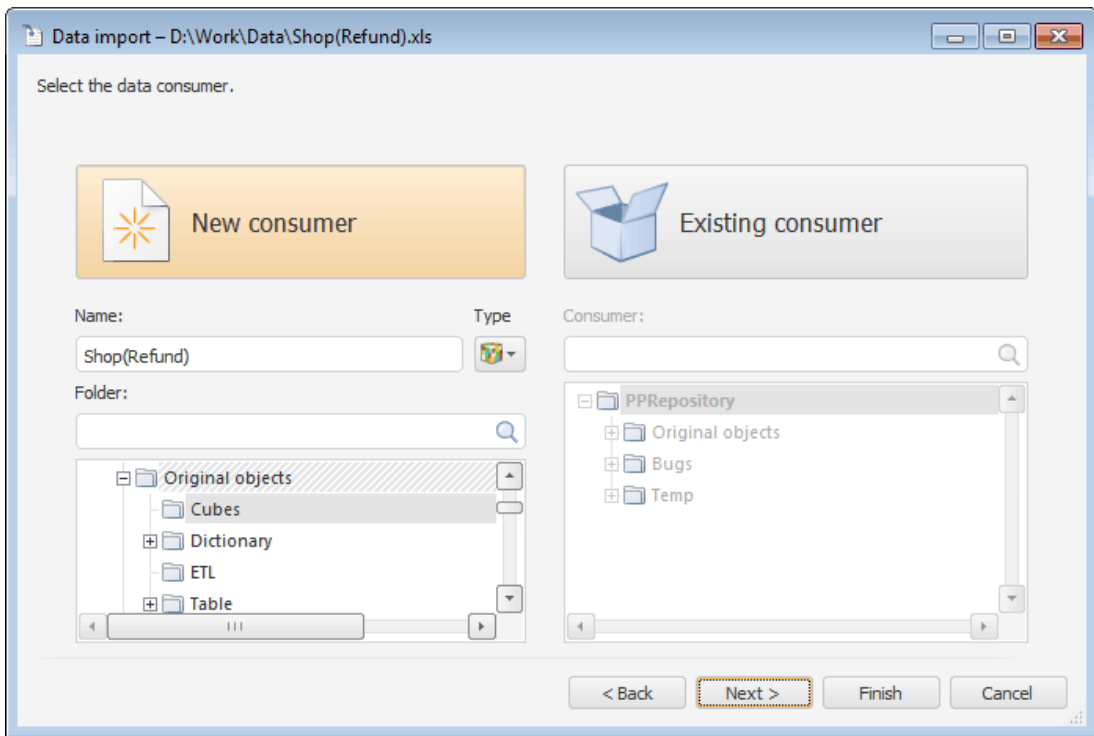
After specifying the content of dimensions click the Next button.



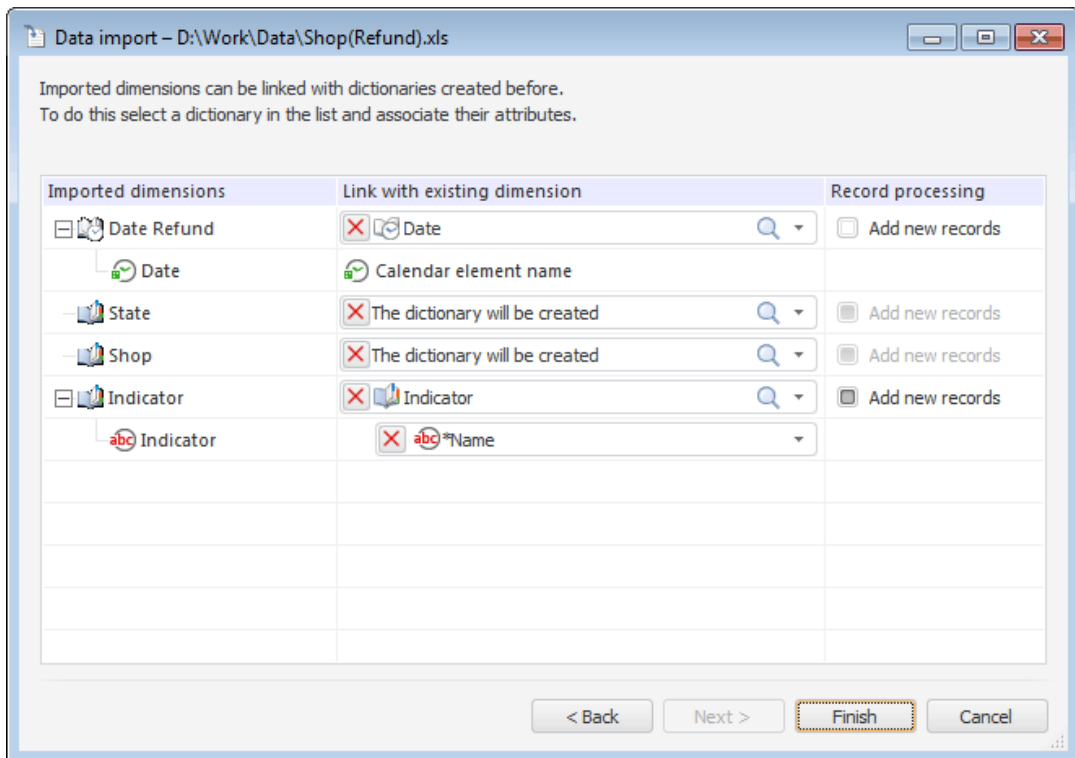
The fields are automatically divided into dimensions, values of text fields are imported as values of the Name attribute of dimension elements. All fields having a real data type are imported as values. A fact in the cube dimension is created for each fields with value.

If required, the contents of dimensions and their combination can be changed. To do this, drag attributes to the desired dimension contents (attributes are displayed as child elements of the dimensions, so that each attributes has its role). If several attributes within one dimension have the Name role, the dimension is created as hierarchical one.

After forming dimensions contents click the Next button.



Click the New Consumer button, name a new cube and select the folder to save the cube in the repository. Click the Next button.



If the repository already contains dictionaries having the same purpose and structure as the imported dimensions, they can be selected from the drop-down lists. Only table or composite MDM table dictionaries and calendar dictionaries are available to choose from.

On clicking the Finish button the following objects are created and set up: a new cube, all required dictionaries and a table with imported data. Settings of data import to ETL task are also saved.