Information Builders

iWay

Omni-Gen™ Personal Edition Getting Started Guide

Release 3.9

July 23, 2019

Contents

Preface

This documentation describes how to install Omni-Gen™ Personal Edition and how to work with data. This documentation is intended for all users that want to profile, standardize, and cleanse data.

How This Manual Is Organized

This manual includes the following chapters:

	Chapter/Appendix	Contents
1	Introducing Omni-Gen™ Personal Edition	Describes Omni-Gen™ Personal Edition and highlights of the product.
2	Installing Omni-Gen™ Personal Edition	Describes the requirements for Omni-Gen™ Personal Edition and how to install Omni-Gen™ Personal Edition on a Windows platform.
3	Navigating the Omni-Gen™ Personal Edition User Interface	Identifies and describes key areas of the Omni-Gen™ Personal Edition user interface.
4	Working With Data	Describes how to import data from Excel spreadsheets, comma-separated values files (CSV), and database tables.
5	Data Profiling	Describes data profiling, which enables you to examine data from an existing data source or file to collect statistics or summaries about the data.
6	Data Quality Rules	Describes the Data Quality rules and Rules functionality.
A	Managing the Windows Service	Describes how to manage the Windows service that is installed with Omni-Gen™ Personal Edition.
B	Additional Functionality	Describes the additional functions of Omni-Gen™ Personal Edition Getting Started.

Documentation Conventions

The following table lists and describes the documentation conventions that are used in this manual.

Convention	Description
THIS TYPEFACE or this typeface	Denotes syntax that you must type exactly as shown.
this typeface	Represents a placeholder (or variable), a cross-reference, or an important term. It may also indicate a button, menu item, or dialog box option that you can click or select.
underscore	Indicates a default setting.
Key + Key	Indicates keys that you must press simultaneously.
{}	Indicates two or three choices. Type one of them, not the braces.
|	Separates mutually exclusive choices in syntax. Type one of them, not the symbol.
...	Indicates that you can enter a parameter multiple times. Type only the parameter, not the ellipsis (...).
. . .	Indicates that there are (or could be) intervening or additional commands.

Related Publications

Visit our Technical Documentation Library at *http://documentation.informationbuilders.com*. You can also contact the Publications Order Department at (800) 969-4636.

Customer Support

Do you have questions about this product?

Join the Focal Point community. Focal Point is our online developer center and more than a message board. It is an interactive network of more than 3,000 developers from almost every profession and industry, collaborating on solutions and sharing every tips and techniques. Access Focal Point at *http://forums.informationbuilders.com/eve/forums*.

You can also access support services electronically, 24 hours a day, with InfoResponse Online. InfoResponse Online is accessible through our website, *http://www.informationbuilders.com*. It connects you to the tracking system and known-problem database at the Information Builders support center. Registered users can open, update, and view the status of cases in the tracking system and read descriptions of reported software issues. New users can register immediately for this service. The technical support section of *http://www.informationbuilders.com* also provides usage techniques, diagnostic tips, and answers to frequently asked questions.

Call Information Builders Customer Support Services (CSS) at (800) 736-6130 or (212) 736-6130. Customer Support Consultants are available Monday through Friday between 8:00 A.M. and 8:00 P.M. EST to address all your questions. Information Builders consultants can also give you general guidance regarding product capabilities. Be prepared to provide your six-digit site code (*xxxx.xx*) when you call.

To learn about the full range of available support services, ask your Information Builders representative about InfoResponse Online, or call (800) 969-INFO.

Help Us to Serve You Better

To help our consultants answer your questions effectively, be prepared to provide specifications and sample files and to answer questions about errors and problems.

The following table lists the environment information that our consultants require.

Platform	
Operating System	
OS Version	
JVM Vendor	
JVM Version	

The following table lists additional questions to help us serve you better.

Request/Question	Error/Problem Details or Information
Provide usage scenarios or summarize the application that produces the problem.	

Request/Question	Error/Problem Details or Information
When did the problem start?	
Can you reproduce this problem consistently?	
Describe the problem.	
Describe the steps to reproduce the problem.	
Specify the error messages.	
Under what circumstance does the problem *not* occur?	

User Feedback

In an effort to produce effective documentation, the Technical Content Management staff welcomes your opinions regarding this document. Please use the Reader Comments form at the end of this document to communicate your feedback to us or to suggest changes that will support improvements to our documentation. You can also contact us through our website, *http://documentation.informationbuilders.com/connections.asp*.

Thank you, in advance, for your comments.

iWay Software Training and Professional Services

Interested in training? Our Education Department offers a wide variety of training courses for iWay Software and other Information Builders products.

For information on course descriptions, locations, and dates, or to register for classes, visit our website, *http://education.informationbuilders.com*, or call (800) 969-INFO to speak to an Education Representative.

Interested in technical assistance for your implementation? Our Professional Services department provides expert design, systems architecture, implementation, and project management services for all your business integration projects. For information, visit our website, *http://www.informationbuilders.com/consulting*.

Chapter 1

Introducing Omni-Gen™ Personal Edition

Omni-Gen™ Personal Edition is a browser-based, self-service application that enables you to import, profile, inspect, and cleanse data from various data sources in a unified, user-friendly interface. The resulting trusted data can be quickly exported to provide insightful analytics and decision-making, or made easily available for other data integration initiatives and solutions.

In this chapter:

Why Omni-Gen™ Personal Edition?

For most organizations, data is the key to effective operations and analytics. But since digital data growth is over 40 percent per year, it is becoming exceedingly difficult to acquire the resources to keep up. Organizations often need additional developers to import and prepare this data, so workers can use it successfully. Coping with changes in a data environment can slow the process down dramatically.

Omni-Gen™ Personal Edition helps you profile, standardize, and cleanse your own data, without involving developers, so you can immediately get the high-quality data you need.

Highlights of Omni-Gen™ Personal Edition

Omni-Gen™ Personal Edition enables you to:

- Quickly import, understand, and cleanse personal data sets.
- Create cleansing and validation rules without the need for developers.
- Experiment with ways to improve data quality.
- Add multiple data sources from CSV, Excel, and database tables.
- Export individual data sources after cleansing.
- Export cleansing rules to Omni-Gen™ Enterprise Master Data and Data Quality Editions.

Chapter 2

Installing Omni-Gen™ Personal Edition

This section describes the system-level requirements and system configuration for installing Omni-Gen™ Personal Edition on a Windows platform. It also provides the steps for installing Omni-Gen™ Personal Edition.

In this chapter:

System Requirements

The following are the system-level requirements and system configuration for installing Omni-Gen™ Personal Edition.

- 16 GB RAM (Installed Memory) is recommended; 8 GB RAM is minimum.
- 2 GB disk space for the baseline product. In addition, 700 MB of disk space for the installer.
- Oracle JDK Release 8.
- Appropriate JDBC drivers for loading data from data sources. The drivers are required during the configuration of the JDBC data sources.
- Microsoft Visual C++ 2010 SP1 Redistributable Package (x64). For more information, see *Microsoft Visual C++ 2010 SP1 Redistributable Package (x64)* on page 12 under Considerations and Limitations.

Note: The installation process requires access to the internet to obtain the proper license for the product. The installation will access the iwse.ibi.com public host on port 8092. Ensure that you are online during the installation process and that there are no firewall rules in your domain, which may prevent the proper access to the iwse.ibi.com public host.

Considerations and Limitations

This section provides an overview of the considerations to be taken into account while using Omni-Gen™ Personal Edition. You are reminded that this product is designed for personal use and may carry limits for data load, data processing, and other limitations. For more information on the Enterprise Level Omni-Gen™ Master Data Management and Omni-Gen™ Data Quality Editions, contact iWay Customer Support.

Note:

- Installation into the Program Files directory is not supported. It is recommended that you use any other location of the root drive.
- Omni-Gen™ Personal Edition supports Microsoft Excel data sets (spreadsheets) up to 3 MB. If you need to work with larger data sets, it is recommended that you export your Microsoft Excel data set to CSV format. If you require any additional support with your data sets, contact iWay Customer Support.

Microsoft Visual C++ 2010 SP1 Redistributable Package (x64)

Omni-Gen™ Personal Edition Release 3.9 and higher requires Microsoft Visual C++ 2010 SP1 Redistributable Package (x64) to be installed on the target machine. The current Omni-Gen™ Personal Edition installer is not able to automatically install this package. If the target machine does not have this package installed, then you must download and install this package before running the installer. You can download Microsoft Visual C++ 2010 SP1 Redistributable Package (x64) from the following Microsoft website:

https://www.microsoft.com/en-ca/download/details.aspx?id=13523

Data Source Names

The product currently supports file-based data source names that do not contain special characters. Supported data source names must contain the following characters:

- A-Z
- a-z
- _
- 0-9

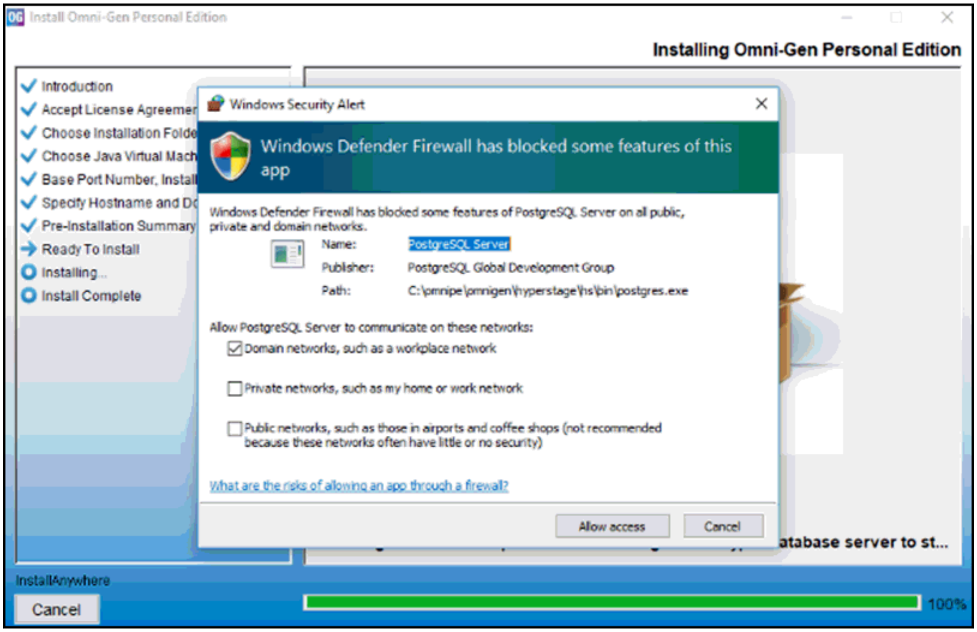
The maximum length for a data source name is 35 characters.

SQL Server Non-dbo Schemas

To load data from a table on SQL Server that is outside of the default *dbo schema*, you must have the appropriate permissions configured on SQL Server to access another schema (other than *dbo schema*). This setting is controlled by SQL Server configuration settings for a specific user, and not by Omni-Gen™ Personal Edition.

Windows Defender

During installation of Omni-Gen™ Personal Edition, you may encounter a message from Windows Defender or similar protection applications to allow access for PostgreSQL server and other Omni-Gen™ Personal Edition services. The following image is a sample that you may encounter from Windows Defender, where you are advised to select the *Allow access* option.



Excel Data Support

Excel data sources can contain many types of data, including formulas, charts, pictures, and others. This release of Omni-Gen™ Personal Edition supports only traditional simple Excel data and does not provide support for formulas. If the data source contains formulas or any other unsupported component, the data cell will be skipped during import, and a corresponding error message will be generated for you to review.

Support for the Datetime Datatype

During the data import process, the Datetime datatype columns may experience a nullification of milliseconds in the Date Time value. The milliseconds value may be turned into .000 when viewing the data in the Data Profiling screen.

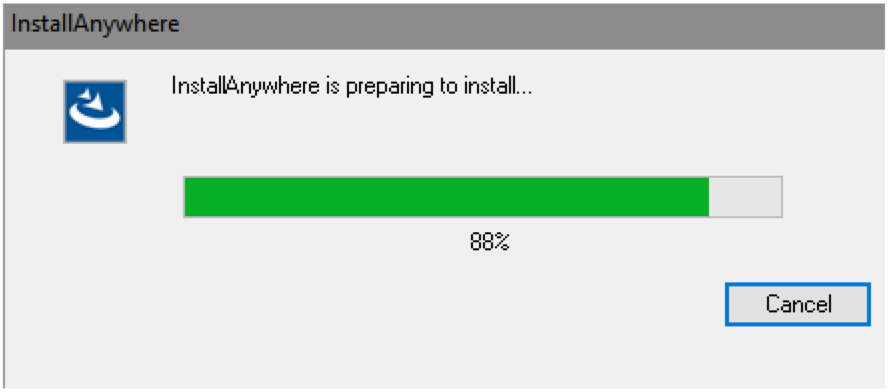
Installing Omni-Gen™ Personal Edition

The following procedure describes how to install Omni-Gen™ Personal Edition.

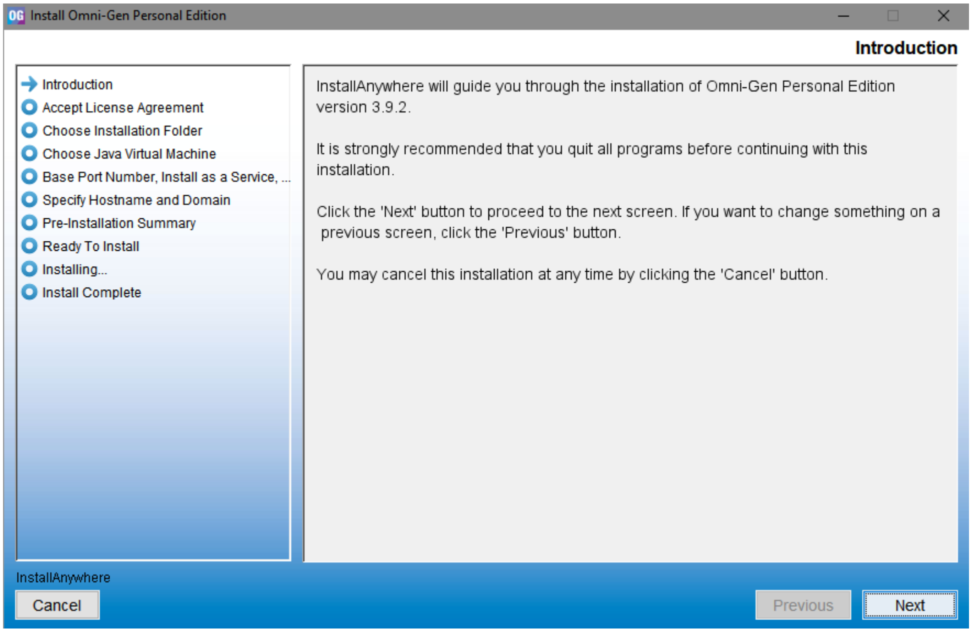
Procedure: How to Install Omni-Gen™ Personal Edition

1. Double-click the *omnigen-installer-pe-3.9.2-Windows-PE.exe* installation file to run the installer.

   The InstallAnywhere program opens, as shown in the following image.

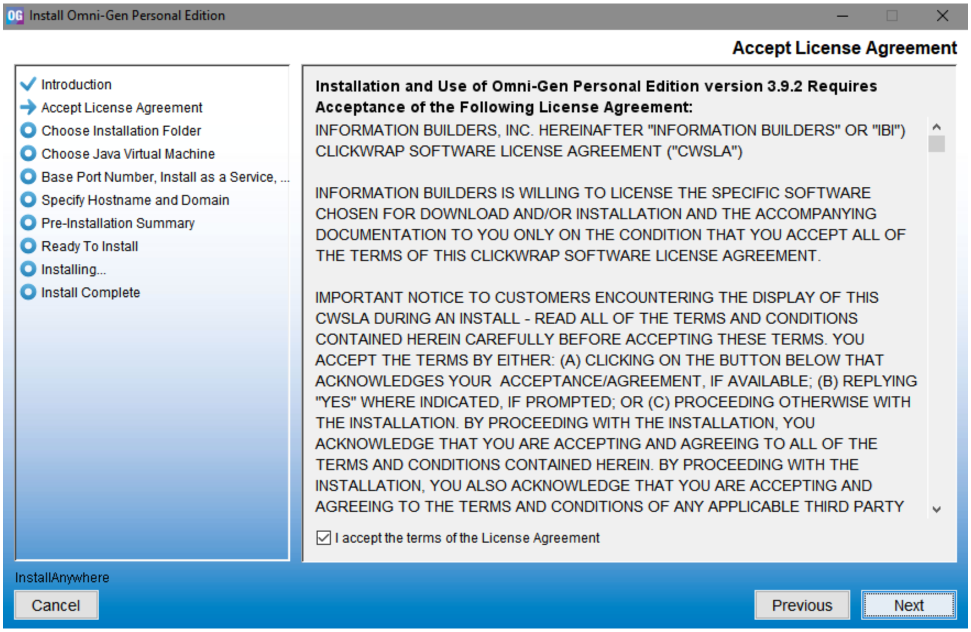


When the installation is ready to begin, the Introduction window opens, as shown in the following image.



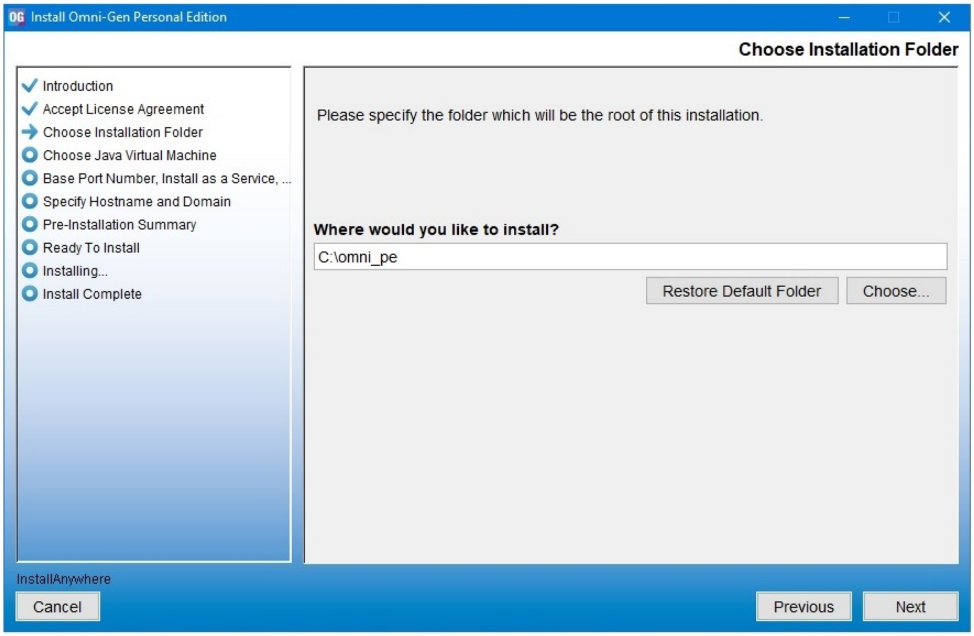
2. Click *Next*.

The Accept License Agreement window opens, as shown in the following image.



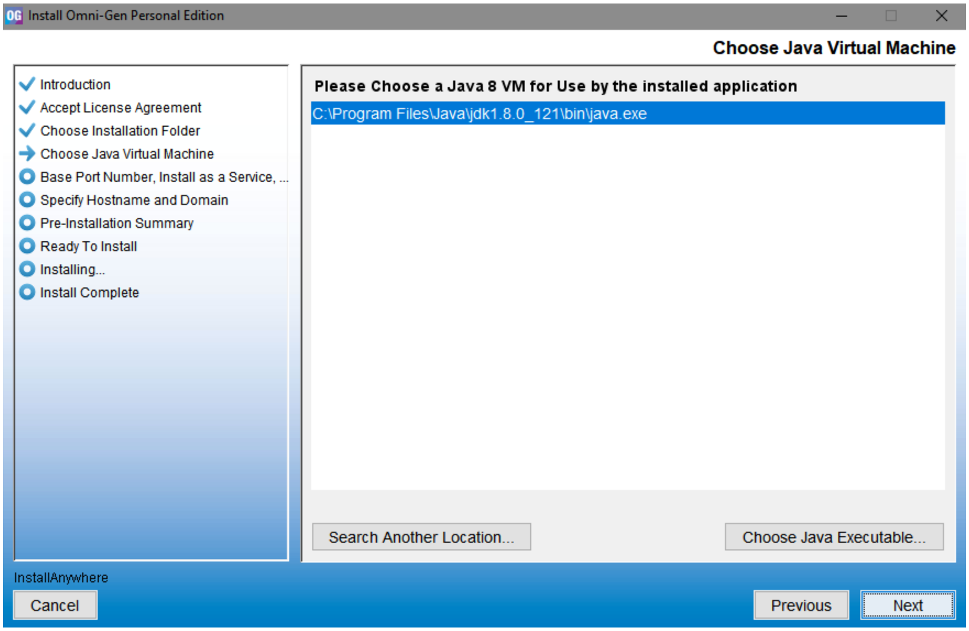
3. Read the License Agreement, select *I accept the terms of the License Agreement*, and then click *Next*.

The Choose Installation Folder window opens, as shown in the following image.



4. In the *Where would you like to install* field, specify the location where you want to install the required files, or click *Choose* to browse to the location for the installation.
5. Click *Next*.

The Choose Java Virtual Machine window opens, as shown in the following image.

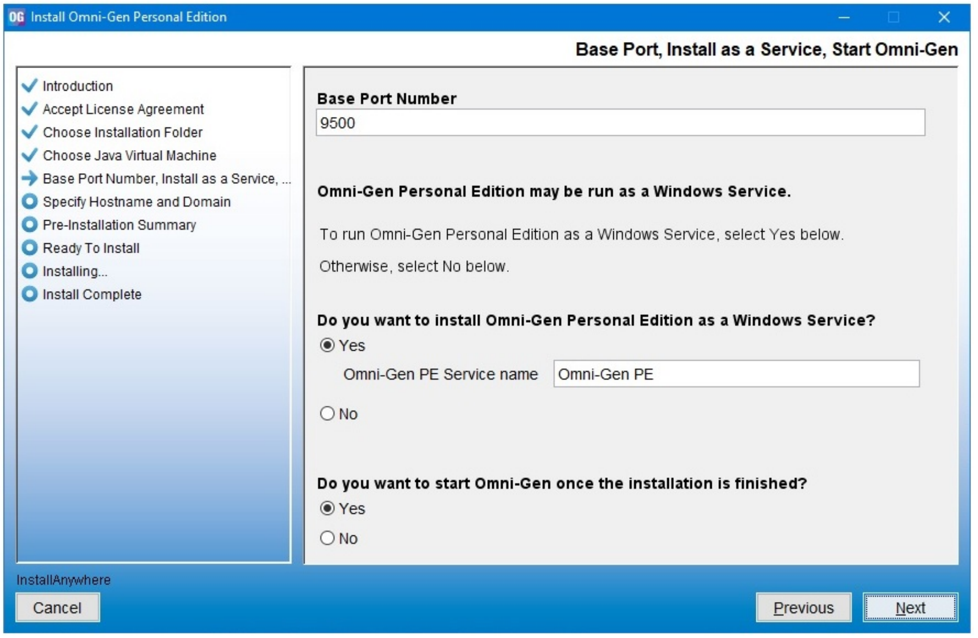


Note: The product will locate the available and suitable JVM.

6. If you want to use a different version, or if the product is not able to locate the JVM, click *Choose Java Executable* to navigate to and select your JVM.

7. Click *Next*.

The Base Port, Install as a Service, Start Omni-Gen window opens, as shown in the following image.



8. Provide the installation properties:

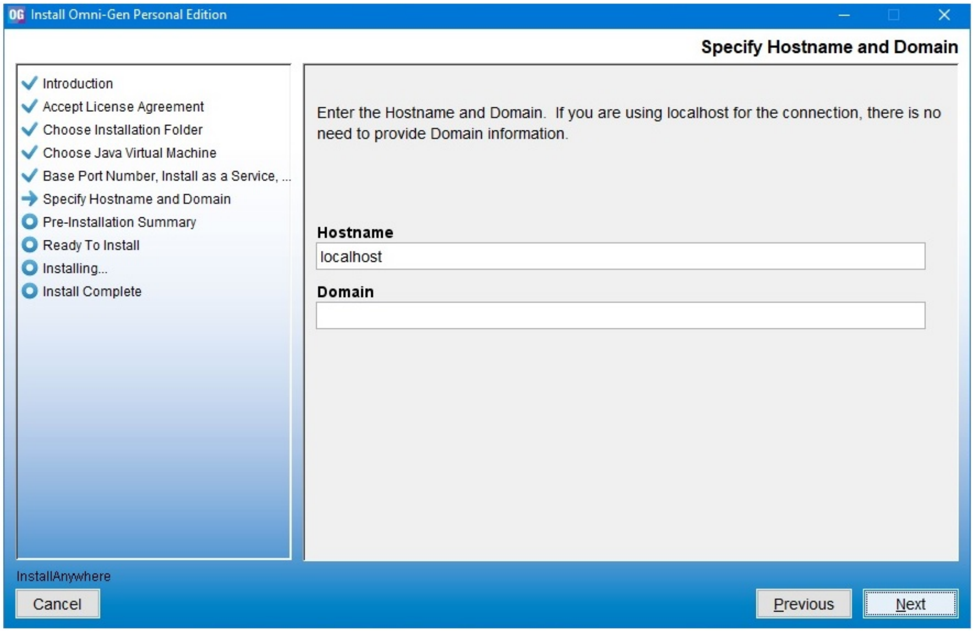
   - **Base Port Number.** A block of ports will be allocated to be used by the product, starting with the base port.

   - **Windows Service.** Choose whether to install Omni-Gen™ Personal Edition as a Windows Service. It is recommended to select *Yes*.

   - **Start Omni-Gen once the installation is finished.** Choose whether to start the product automatically once the installation has completed. It is recommended to select *Yes*.

9. Click *Next*.

The Specify Hostname and Domain window opens, as shown in the following image.

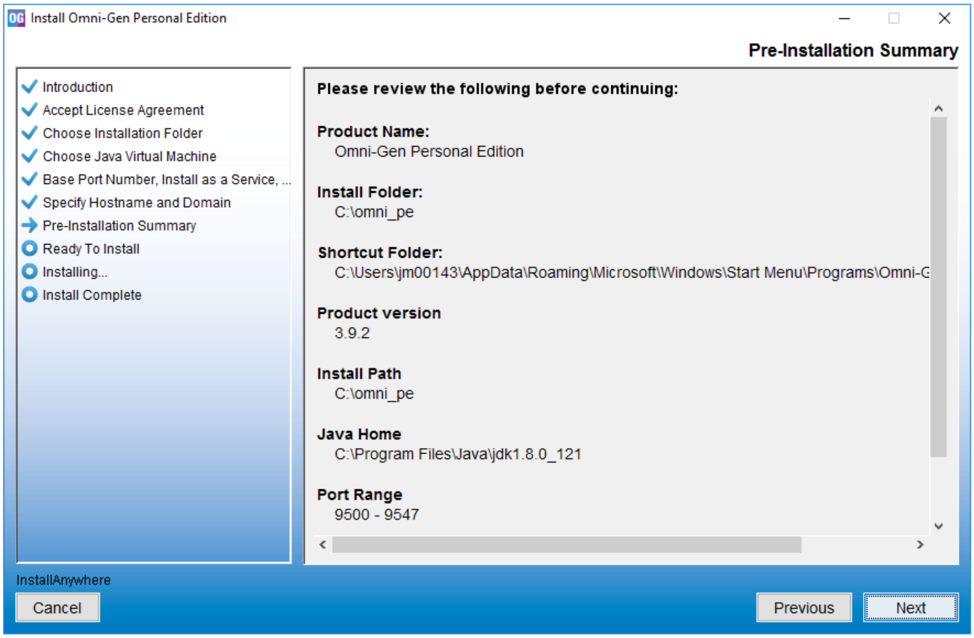


10. Provide the host name and domain for the system where the product is being installed. You may choose to continue with localhost as an option, which does not require domain knowledge or configuration.

    Note: The default host name is the machine on which you are currently installing.

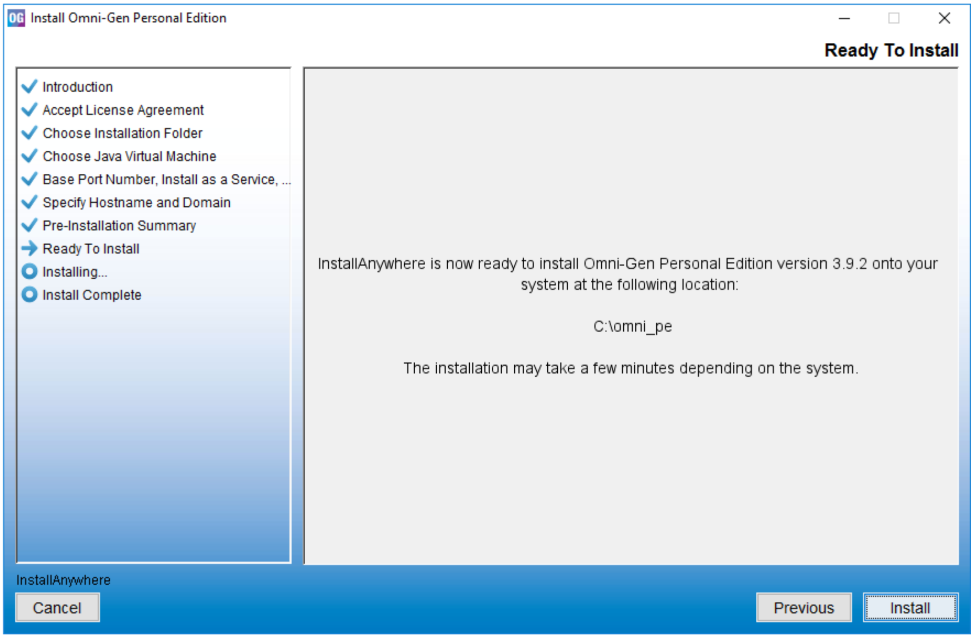
11. Click *Next*.

The Pre-Installation Summary window opens, as shown in the following image.



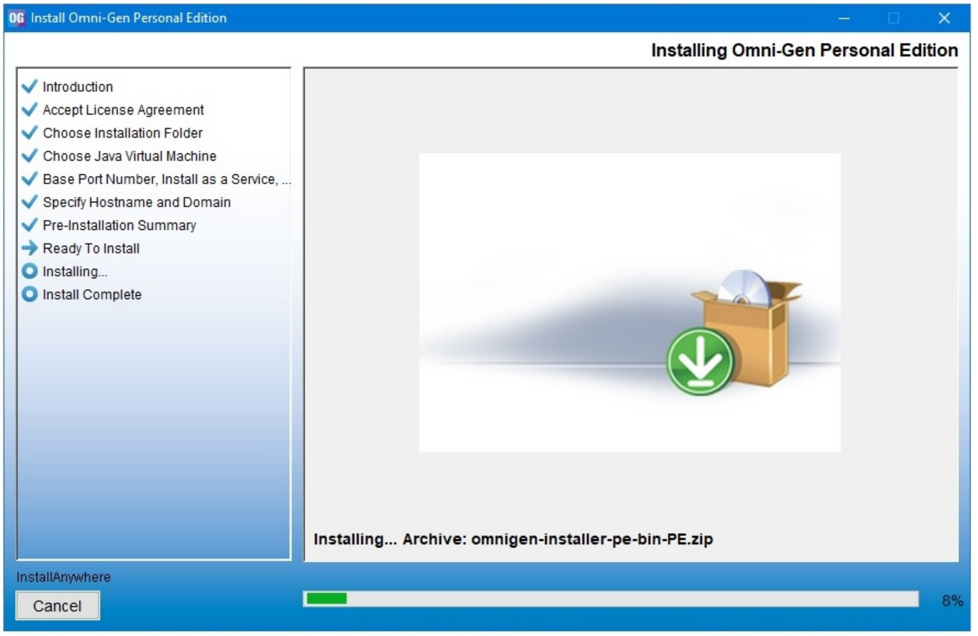
12. Verify that the information is correct, and then click *Next*.

The Ready To Install window opens, as shown in the following image.

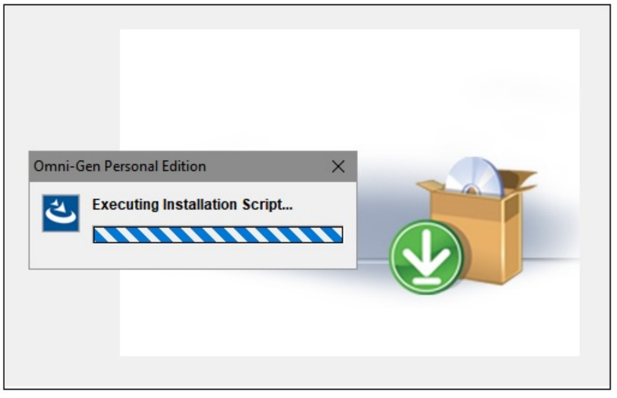


13. Click *Install*.

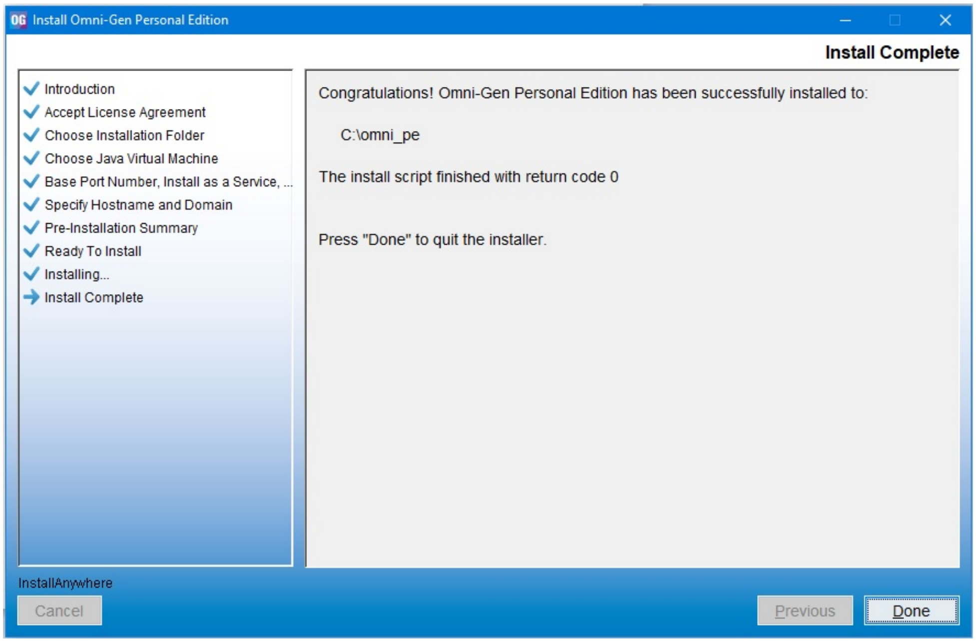
The installation starts and the product contents will be extracted onto the system. A progressive status displays, as shown in the following image.



Note: The Executing Installation Script extract, shown in the following image, takes the most time as it is responsible for configuring the product repositories, services, and starting the product components.

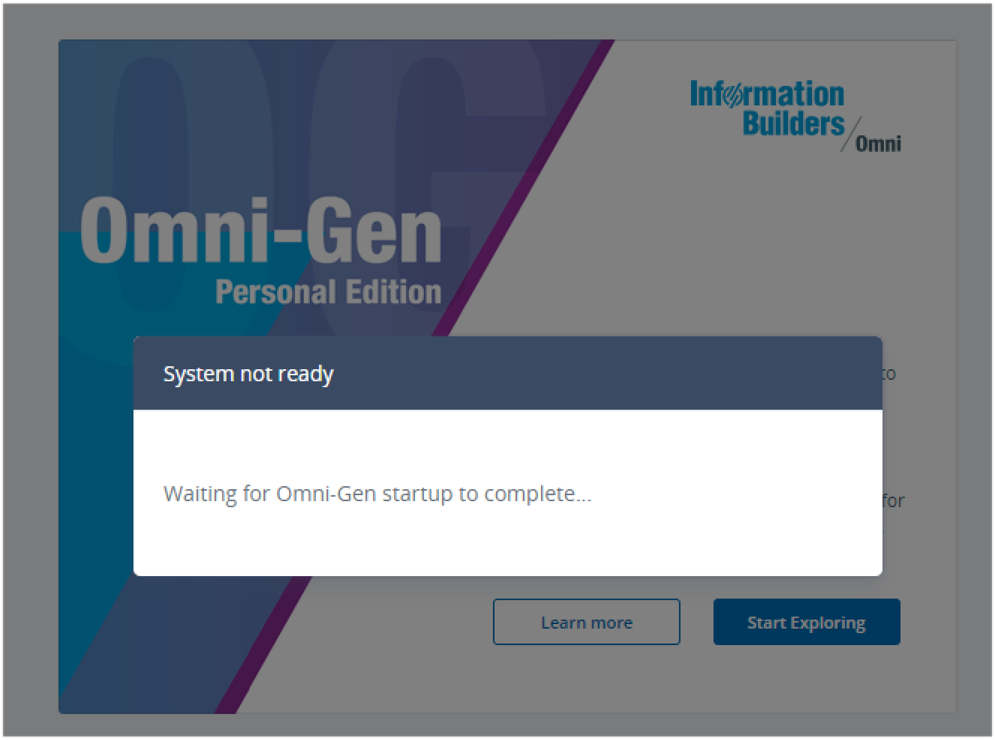


When the installation is complete. the Install Complete window opens indicating a successful status, as shown in the following image.

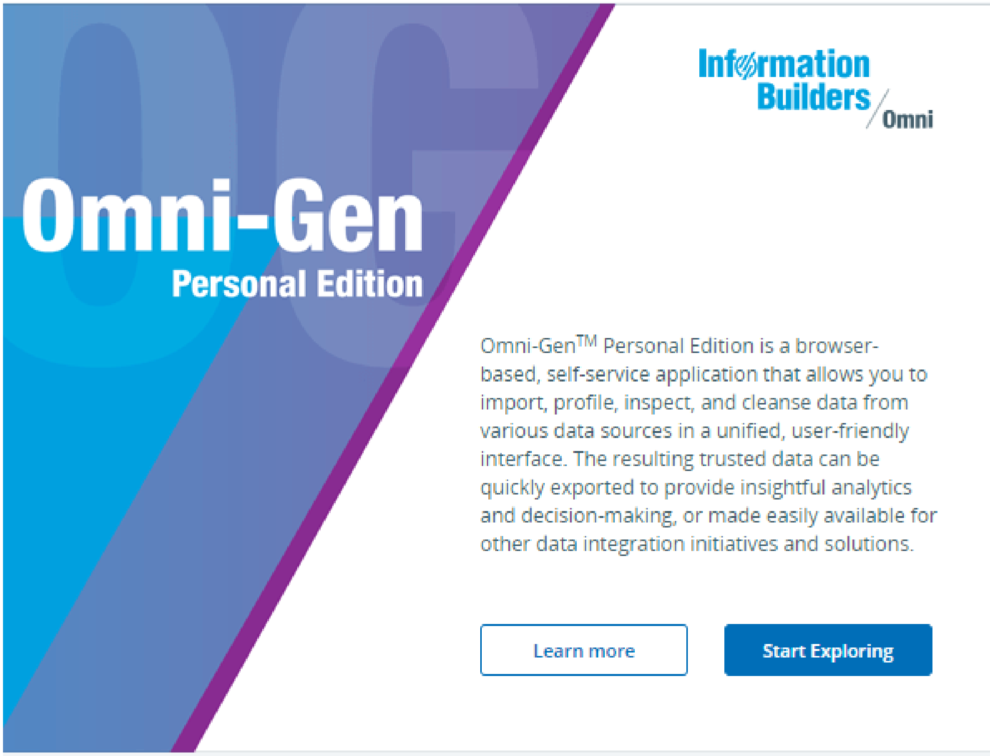


14. Click *Done*.

The web interface will be launched automatically, and the Omni-Gen™ Personal Edition home page opens. You may see a message that Omni-Gen™ is starting up, as shown in the following image.



Once Omni-Gen™ has started, the Omni-Gen™ wizard opens, as shown in the following image.



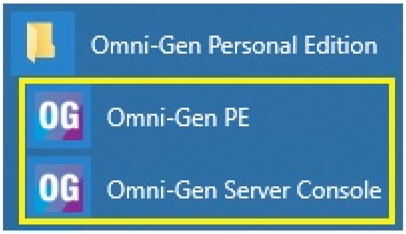
15. Click *Start Exploring* to explore the product and proceed with loading your data into the Omni-Gen™ Personal Edition.

    For more information on loading data, see *Working With Data* on page 35.

    For more information on the layout of the environment and user interface, see *Navigating the Omni-Gen™ Personal Edition User Interface* on page 29.

Accessing Omni-Gen™ Personal Edition From the Windows Application Menu

The Omni-Gen™ Personal Edition installer provides shortcuts from the Windows application menu that you can use at any time to access the product, as shown in the following image.



Clicking *Omni-Gen PE* will launch the main user interface (workbench) for Omni-Gen™ Personal Edition in your default browser.

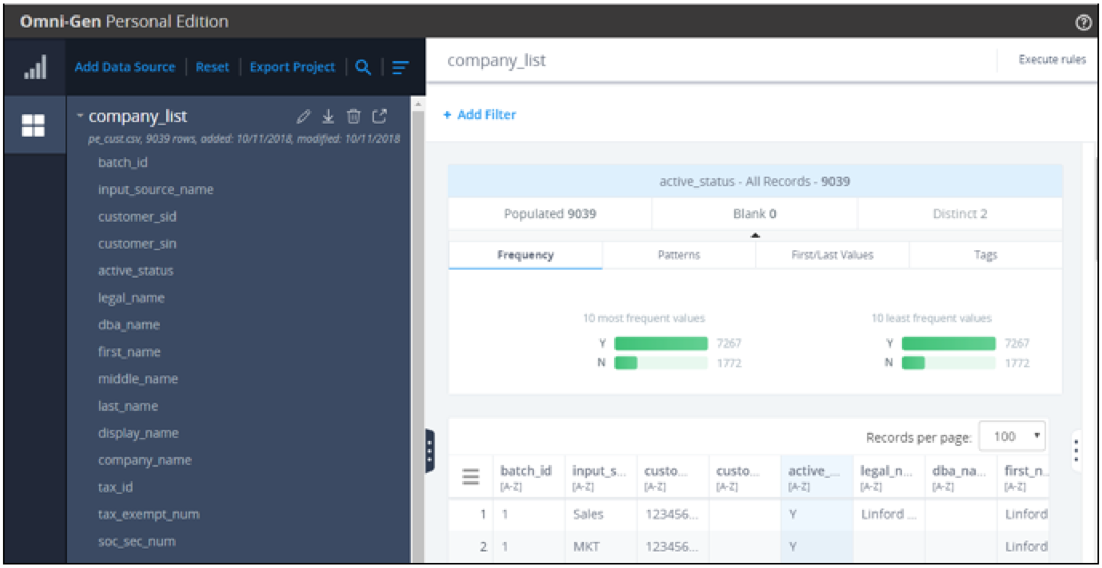
Clicking *Omni-Gen Server Console* will launch a console in your default browser that you can use for administration purposes. For example, you can manage (start and stop) internal services that are used by the Omni-Gen™ server, view system logs, and so on. For more information on using this console, see the *Omni Console User's Guide*.

Chapter 3 Navigating the Omni-Gen™ Personal Edition User Interface

This section identifies and describes key areas of the Omni-Gen™ Personal Edition user interface.

The following image shows the layout of the Omni-Gen™ Personal Edition user interface when a data source is loaded.

In this example, a comma-separated values (CSV) file named *pe_cust.csv* has been loaded into Omni-Gen™ Personal Edition.



In this chapter:

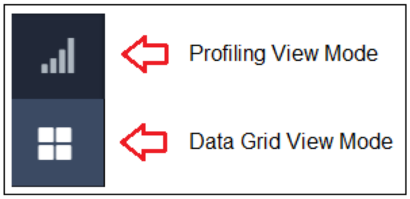
- Title Bar
- View Mode
- Menu Bar
- Loaded Data Summary View
- Workspace Header
- Filter Bar
- Workspace Area

Title Bar

The Title bar displays the product title on the left and provides a link to additional resources (About) on the right, as shown in the following image.

View Mode

Clicking the *Profiling* or *Data Grid* icon in the left pane changes how your imported data is viewed and presented in Omni-Gen™ Personal Edition. The Profiling and Data Grid icons are shown in the following image.



The Profiling view mode provides a data profile of your imported data that is currently selected. This view includes metrics that show the current state of the data (for example, subject completeness and completeness by attribute).

The Data Grid view mode provides an actual view of your imported data that is currently selected (as records on a page). Using this view, you can quickly see the relationship of one attribute in your data to the other attributes in the grid. The Data Grid view enables you to create Data Quality Rules and run experiments on the data.

Menu Bar

The menu bar, shown in the following image, provides a set of high-level options that enable you to quickly manage your data in Omni-Gen™ Personal Edition.

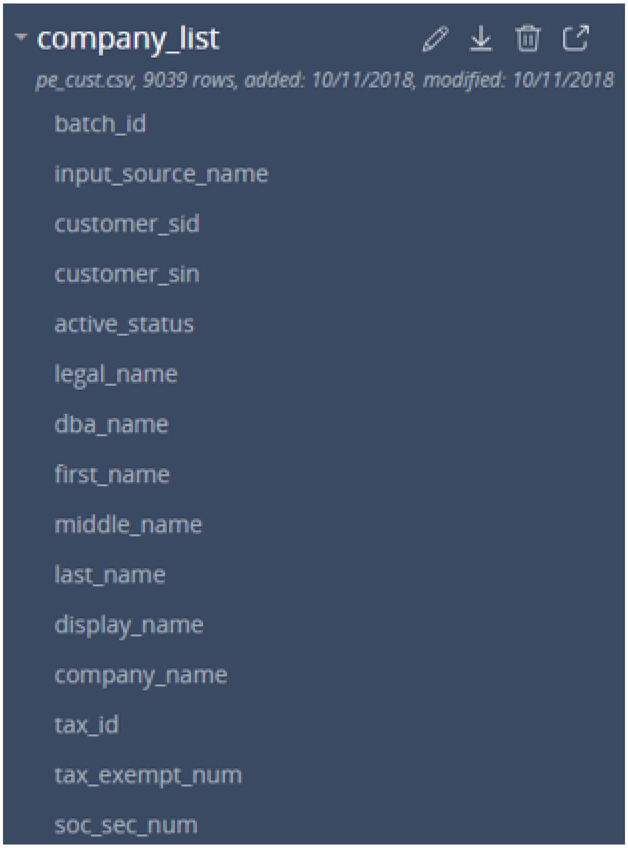
The set of options, from left to right, includes:

- **Add Data Source.** Adds (imports) a new data source into Omni-Gen™ Personal Edition. You can add multiple data sources and toggle between them for analysis.

- **Reset.** Removes all imported data sources from Omni-Gen™ Personal Edition.
- **Export Project.** Exports the current project from Omni-Gen™ Personal Edition for use in Omni-Gen™ Master Data and Data Quality Editions.
- **Search.** Searches across all of the data sources that have been imported into Omni-Gen™ Personal Edition.
- **Sort Data.** Sorts the available data sources in Omni-Gen™ Personal Edition.

Loaded Data Summary View

Shown in the left pane, the Loaded Data Summary View provides a summary of the imported data that is currently selected in Omni-Gen™ Personal Edition, as shown in the following image.



In this example (company_list, which is the header/title), a CSV file named *pe_cust.csv* has been imported. This CSV file, which contains 9039 rows was added on 10/11/2018 and also modified on 10/11/2018. The list of individual values (for example, *batch_id*, *input_source_name*, *customer_sid*, and so on) represent the attributes in this imported data. You can select each attribute to view information (for example, a data profile) about that attribute on a more granular level.

The Loaded Data Summary View also provides a set of options that you can apply to the data that is currently selected, as shown in the following image.

The set of options, from left to right, includes:

- **Rename.** Renames a data source in the system.
- **Reload.** Reloads a data source. This removes the existing data and reloads it from the same source.
- **Delete.** Deletes the selected data source from Omni-Gen™ Personal Edition.
- **Export.** Exports the selected data source into a selected format (CSV, Excel, or database table).

Workspace Header

The Workspace Header, shown in the following image, provides the name of the data source currently selected (for example, *company_list*) and the *Execute rules* button, which allows you to apply any of the defined data quality rules you have selected against your data source.

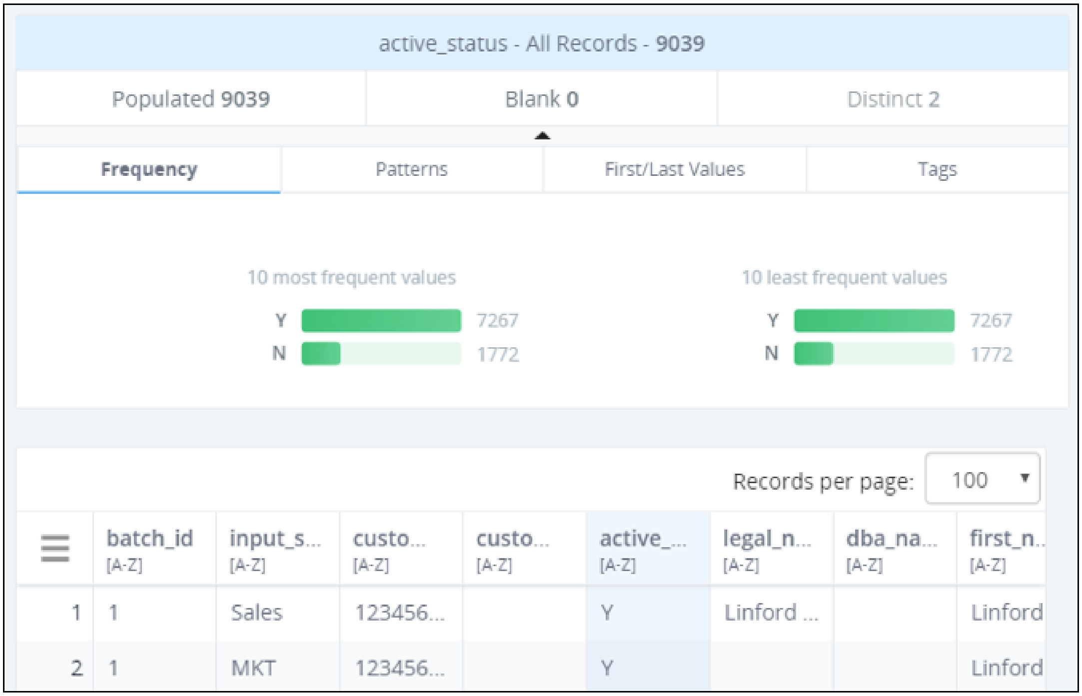
Filter Bar

The Filter bar, shown in the following image, enables you to specify and set a variety of filters to narrow the focus of your data and provide a subset of your data that you can analyze.

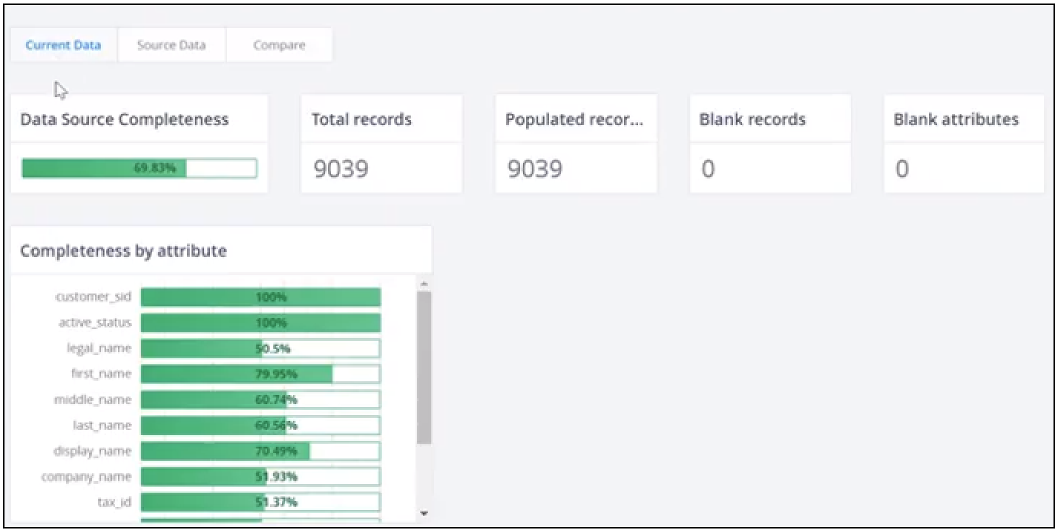
Workspace Area

Depending on the selected view mode (*Profiling* or *Data Grid*), the workspace area provides all of the data profiling information and data metrics related to the selected data source (as a whole) and/or any specific attribute that you may have selected.

The following example shows information for the *active_status* attribute in the Data Grid view mode.



The following example shows the data profile of a selected data source in the Profiling view mode.



After you apply data quality rules to your data in Omni-Gen™ Personal Edition, you can quickly compare the profiles of your source data with your current data to analyze the impact of the data quality rules. You can think of this as a *Before* and *After* scenario.

Chapter 4

Working With Data

Omni-Gen™ Personal Edition enables you to import data from Excel spreadsheets, comma-separated values file (CSV), and database tables.

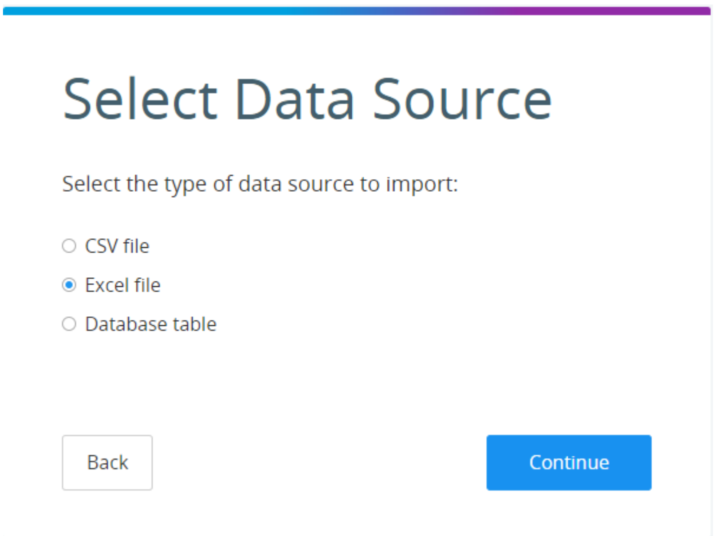
In this chapter:

Working With Excel Data

The following procedure describes how to load Excel data into Omni-Gen™ Personal Edition.

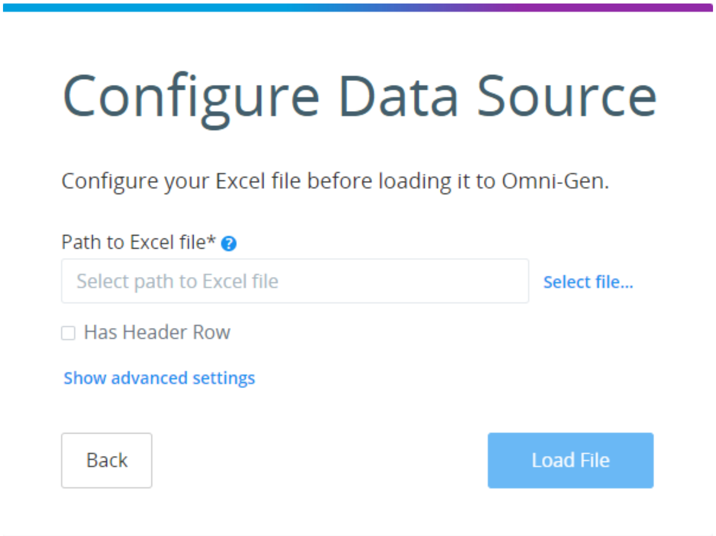
Procedure: How to Work With Excel Data

1. To load Excel data, select *Excel file* from the Select Data Source dialog box, as shown in the following image.

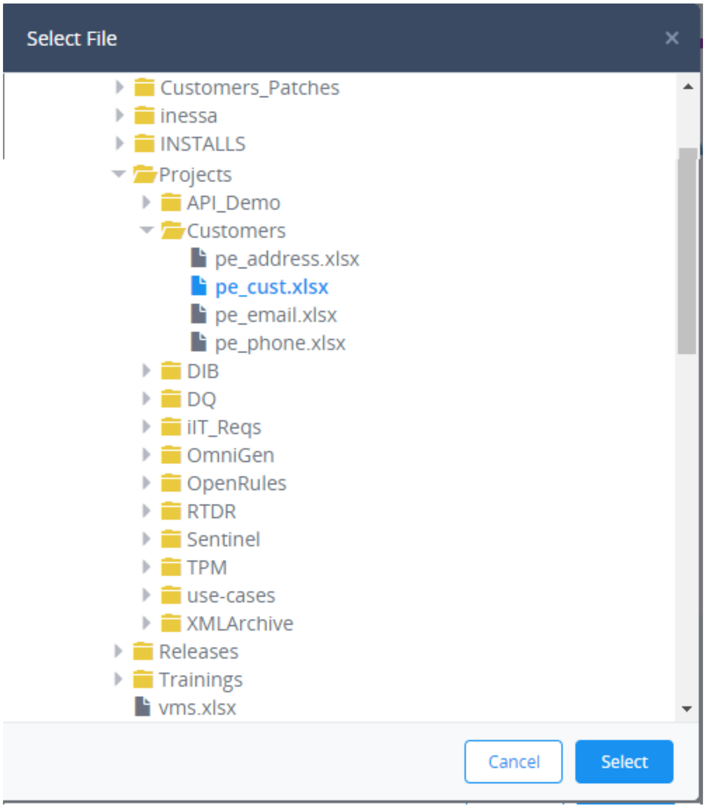


2. Click *Continue*.

The Configure Data Source dialog box opens, as shown in the following image.

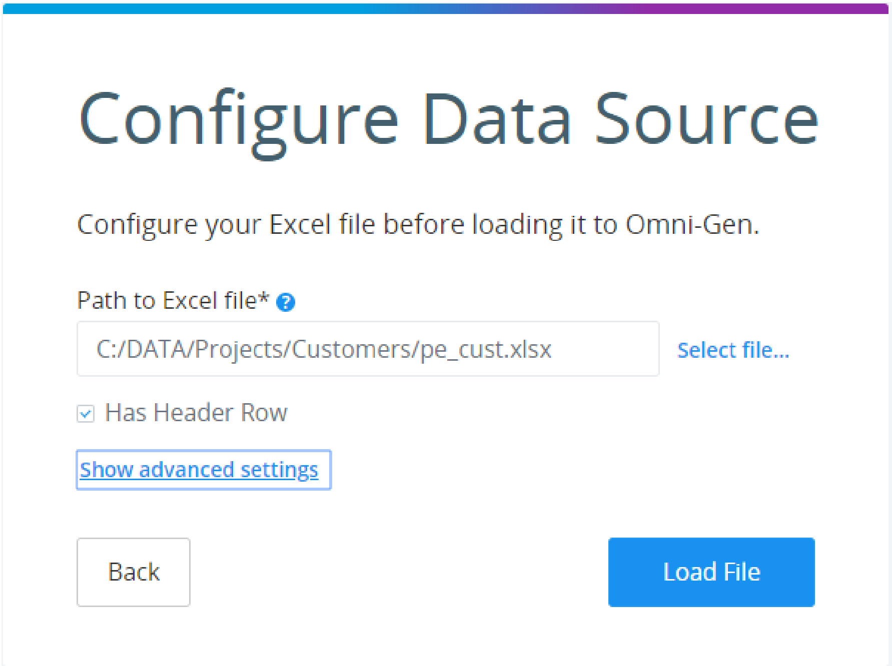


3. Click *Select File* to navigate to the Excel file to load, as shown in the following image.



4. Click *Select*.

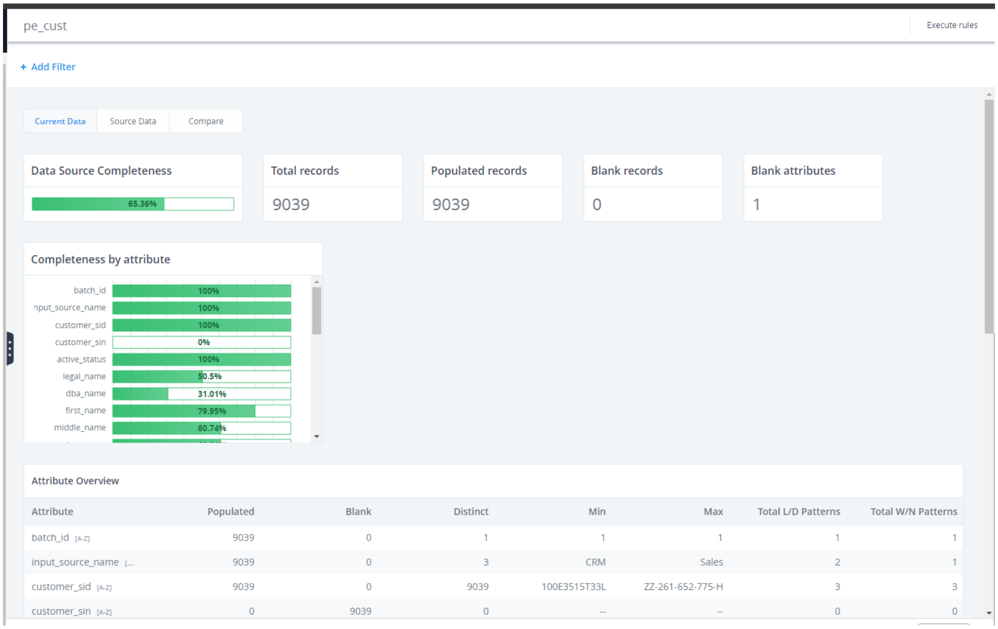
5. If your file has a Header Row, select *Has Header Row* on the Configure Data Source dialog box, as shown in the following image.



You can also access advanced settings, such as Header settings and Row count, from the *Show advanced settings* link.

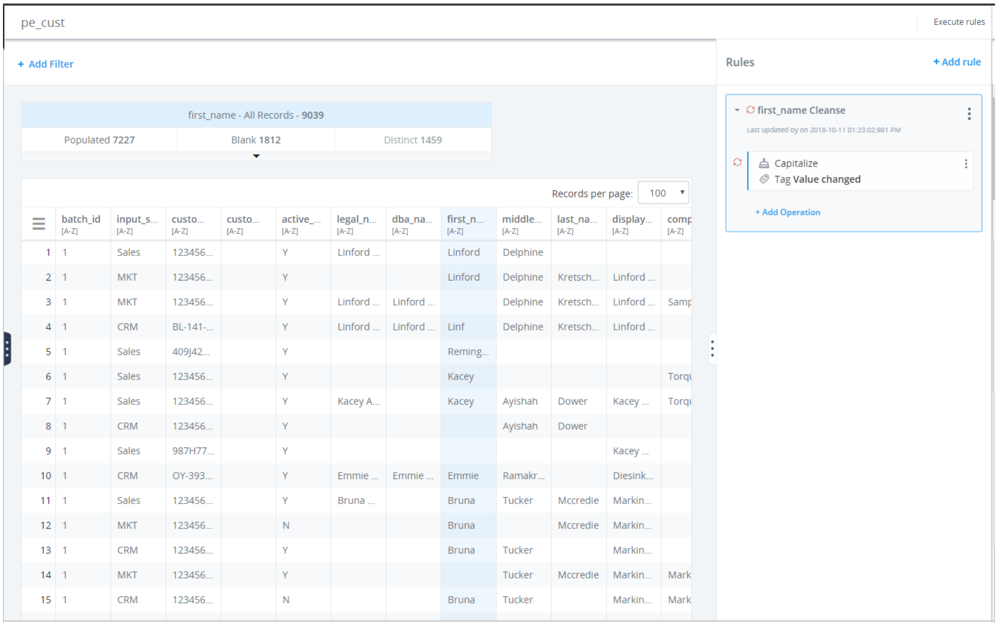
6. Click *Load File*.

The data is loaded and displays in Profiling view, as shown in the following image. For more information on profiling, see *Data Profiling* on page 51.



7. To access the Data Quality view for Rules creation and experiments, select the Data Grid view.

In the Data Quality section, you can create and execute the rules on your data, as shown in the following image.



Working With CSV Data

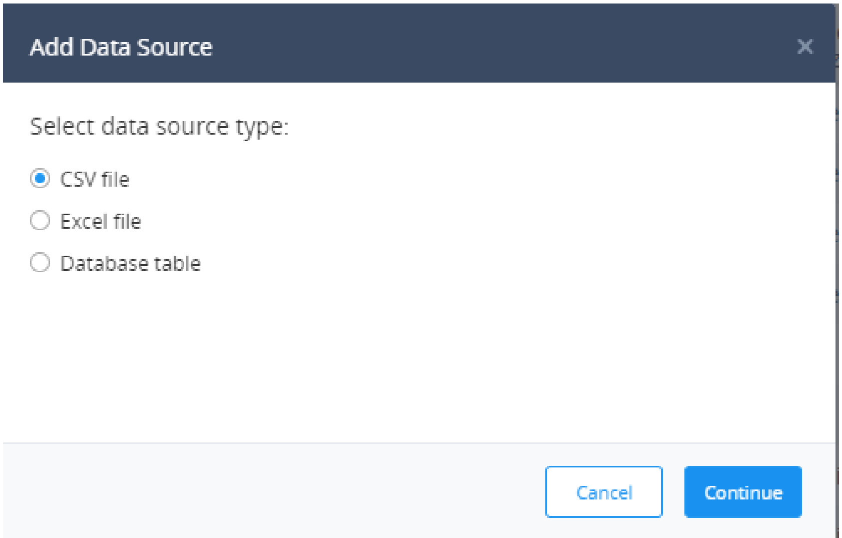
The following procedure describes how to load CSV data into Omni-Gen™ Personal Edition.

Note: The loading of CSV data is currently supported only for standard comma-separated format.

Procedure: How to Work With CSV Data

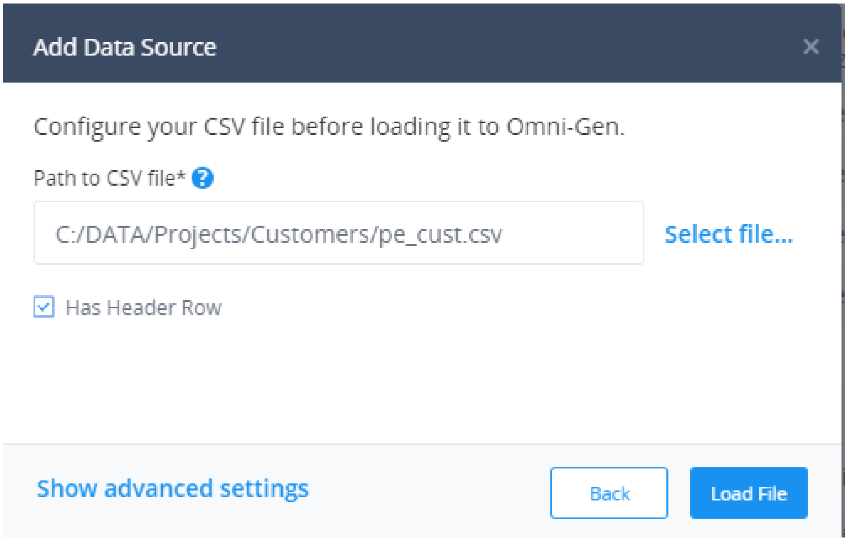
Note: The Omni-Gen™ Personal Edition Trial only supports data for which a comma (,) is the delimiter.

1. To load CSV data, select *CSV file* from the Add Data source dialog box, as shown in the following image.



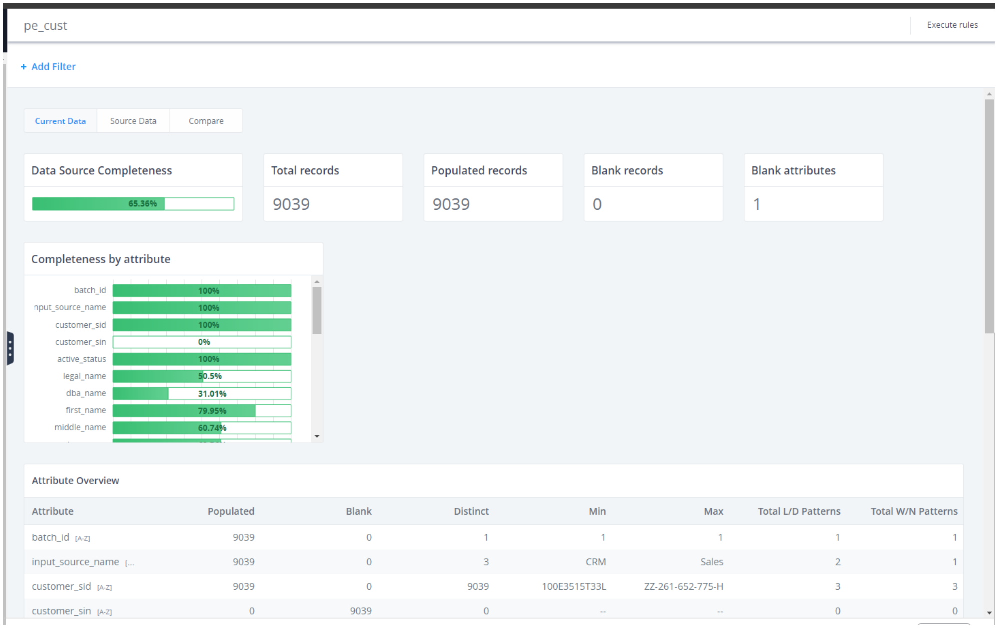
2. Click *Continue*.

The Add Data Source dialog box opens, as shown in the following image.



3. Click *Select file* to navigate to the CSV document to load.
4. If your file has a Header Row, select *Has Header Row* on the Add Data Source dialog box. You may also access advanced settings from the *Show advanced settings* link.
5. Click *Load File*.

The data is loaded and displays in Profiling view, as shown in the following image. For more information on profiling, see *Data Profiling* on page 51.

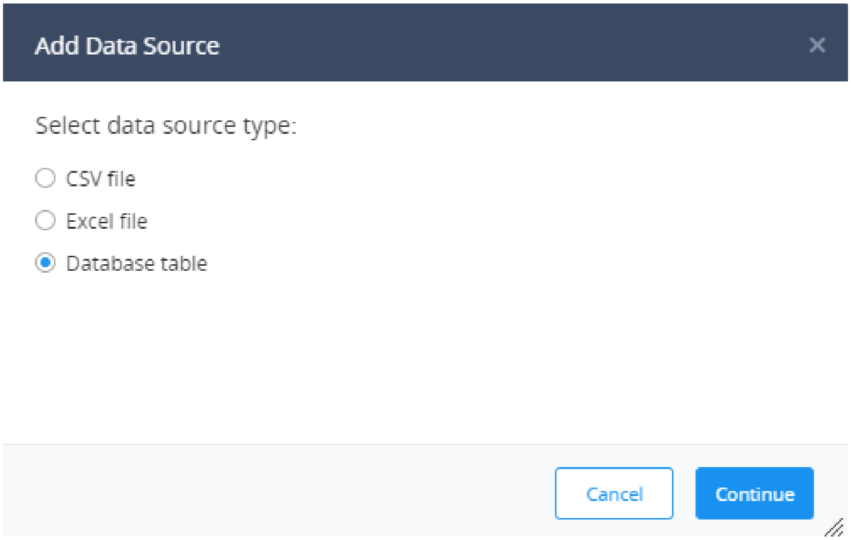


Working With a Database Source

The following procedure describes how to import data from a database into Omni-Gen™ Personal Edition.

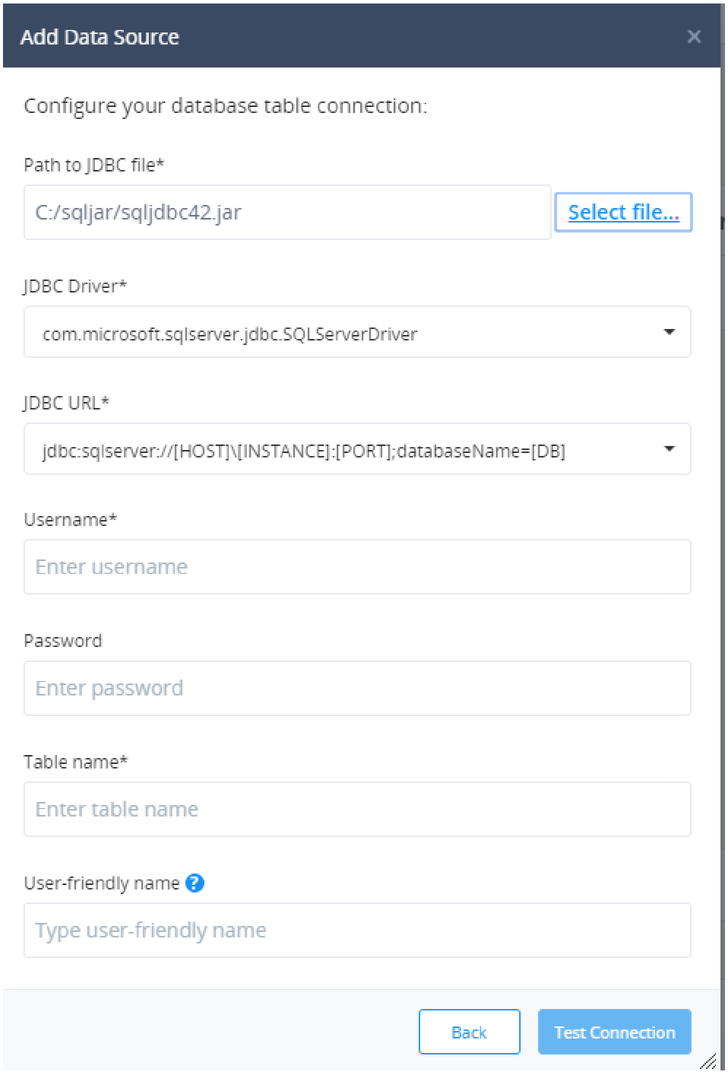
Procedure: How to Work With a Database Source

1. To import data from a data source, select *Database table* from the Add Data Source dialog box, as shown in the following image.



2. Click *Continue*.

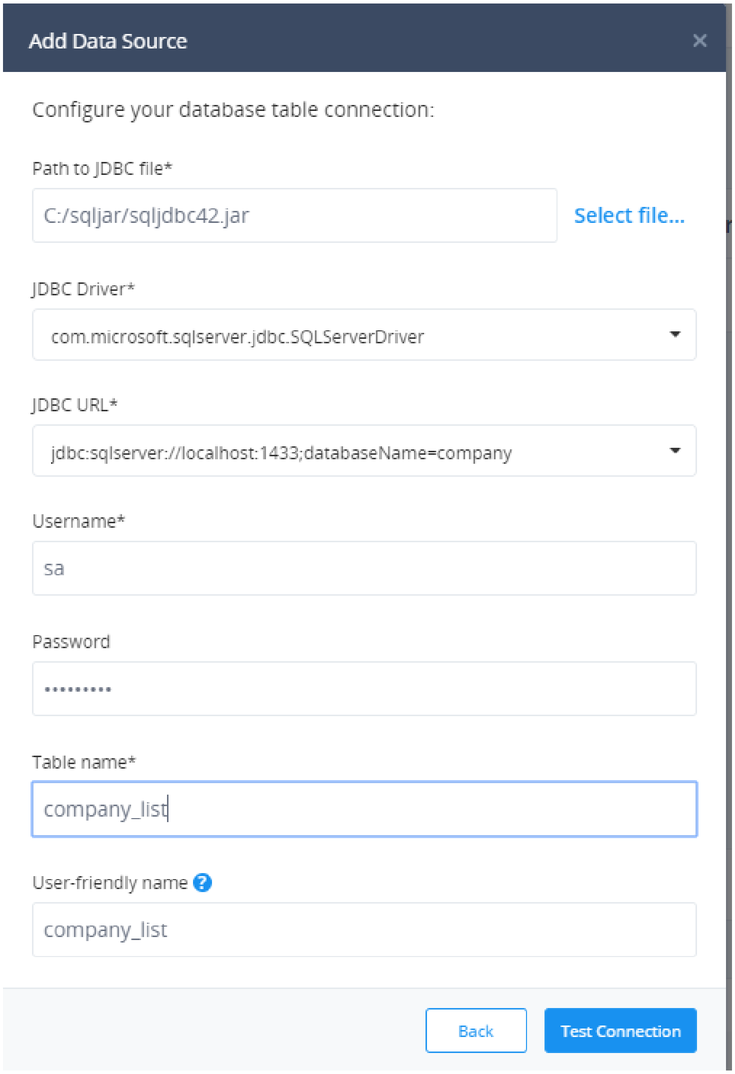
The Add Data Source Configuration dialog box opens, as shown in the following image.



3. Click *Select file* to navigate to and locate the applicable JDBC driver for your data source.

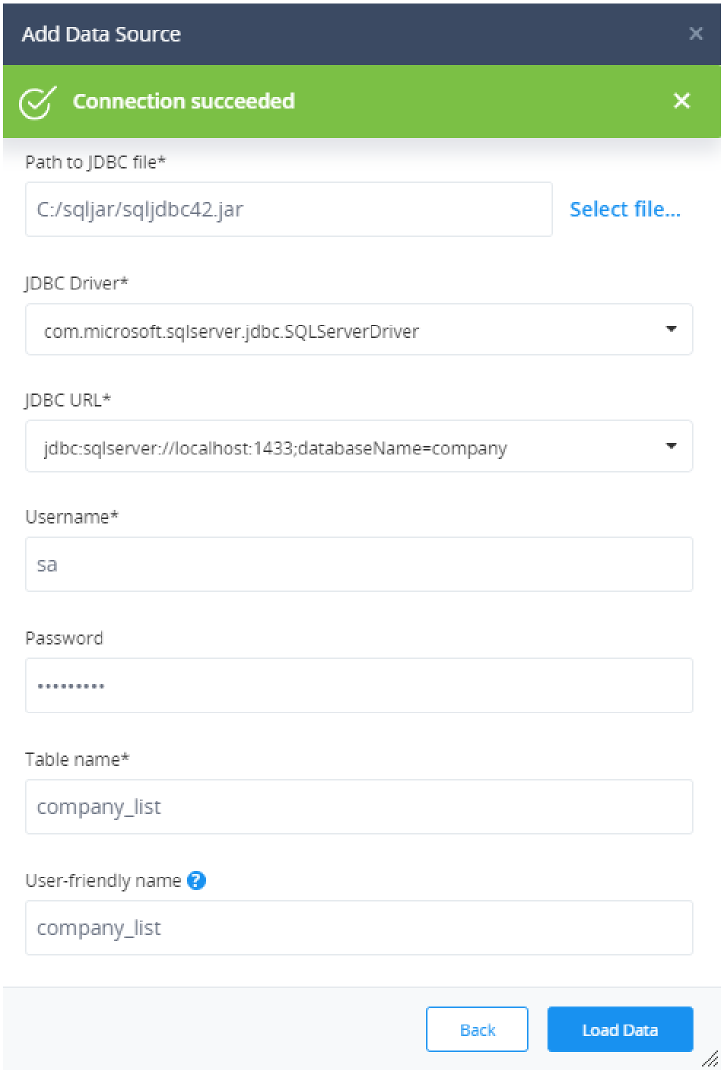
   Based on the selection of the JDBC driver, and if it is available in its metadata, the JDBC driver and JDBC URL parameters will be populated.

4. Provide the connectivity information to your data source and a table name to load, as shown in the following image.



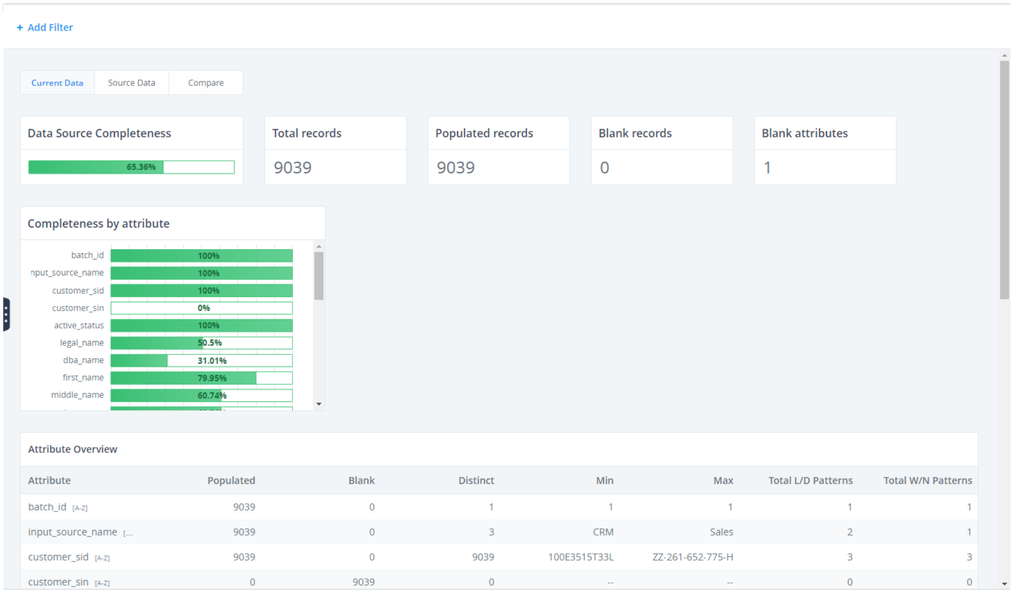
5. Click *Test Connection*.

If the connection is successful, the Load Data option is available, as shown in the following image.



6. Click *Load Data*.

The data is loaded and displays in Profiling view, as shown in the following image.

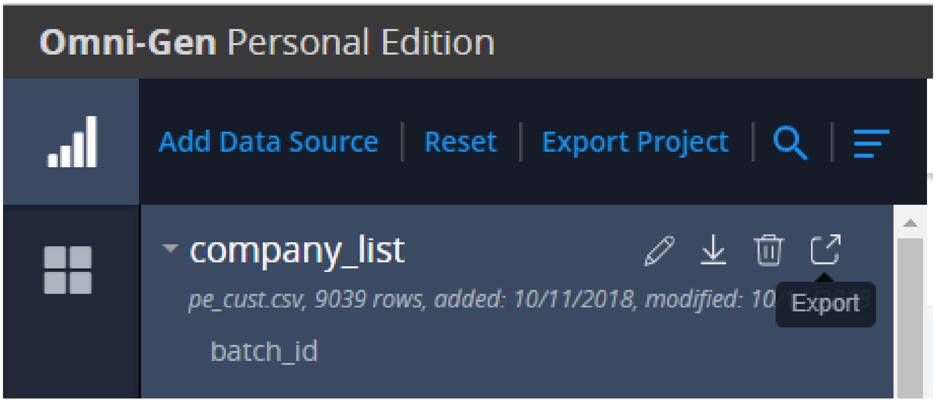


Exporting Data

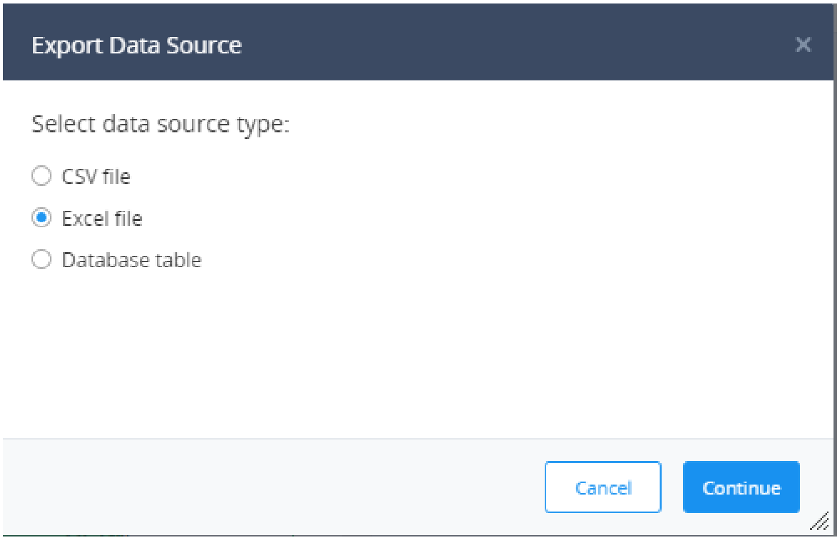
Upon completion of the Data Quality operations, you can export the data for external consumption, in file or table format.

Procedure: How to How to Export Data

1. To export a data source, choose the *Export* icon, as shown in the following image.

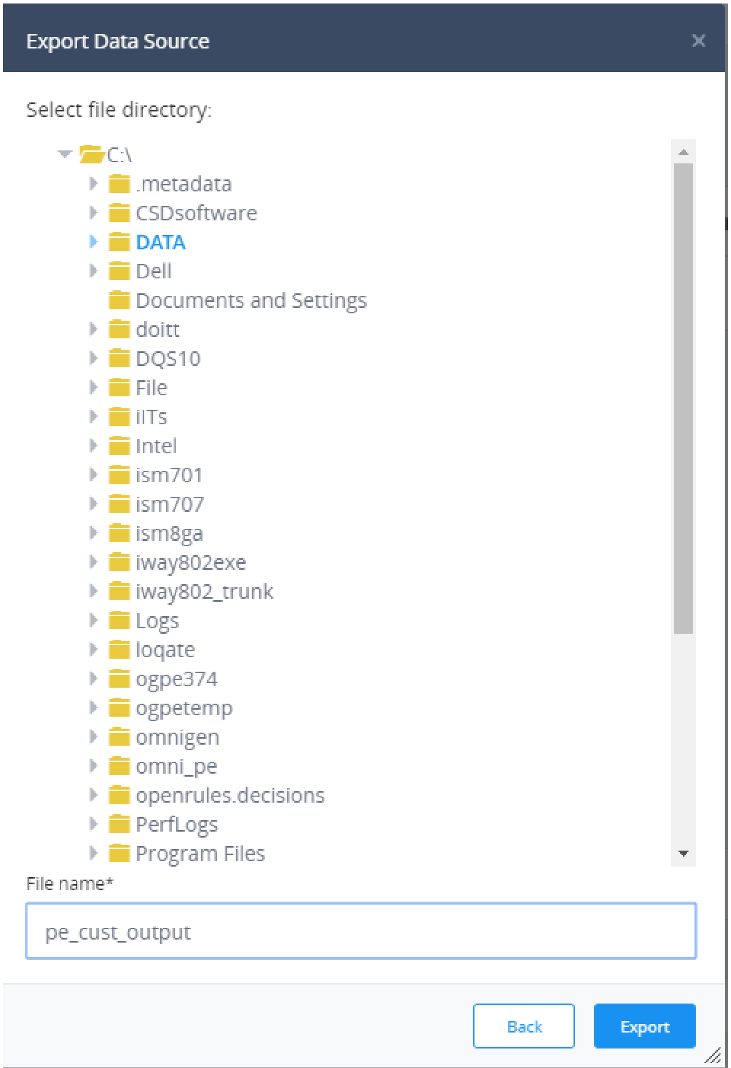


The Export Data Source dialog box opens, as shown in the following image.



2. Select the Export format, for example, *Excel file*.
3. Click *Continue*.

4. Navigate to the location and file to use for the export, as shown in the following image.



5. Click *Export*.

You will receive a message with the status of the export, as shown in the following image.

Chapter 5

Data Profiling

Data profiling enables you to examine data from an existing data source or file to collect statistics or summaries about the data.

The Profile Subject and Attribute pages can be displayed in three modes:

- Current
- Source
- Compare

In general, all of the modes display the same page elements, but the Current mode displays the current state of the records that you are currently working with. The Source mode displays the source state of the records and Compare mode displays a comparison of Current and Source records.

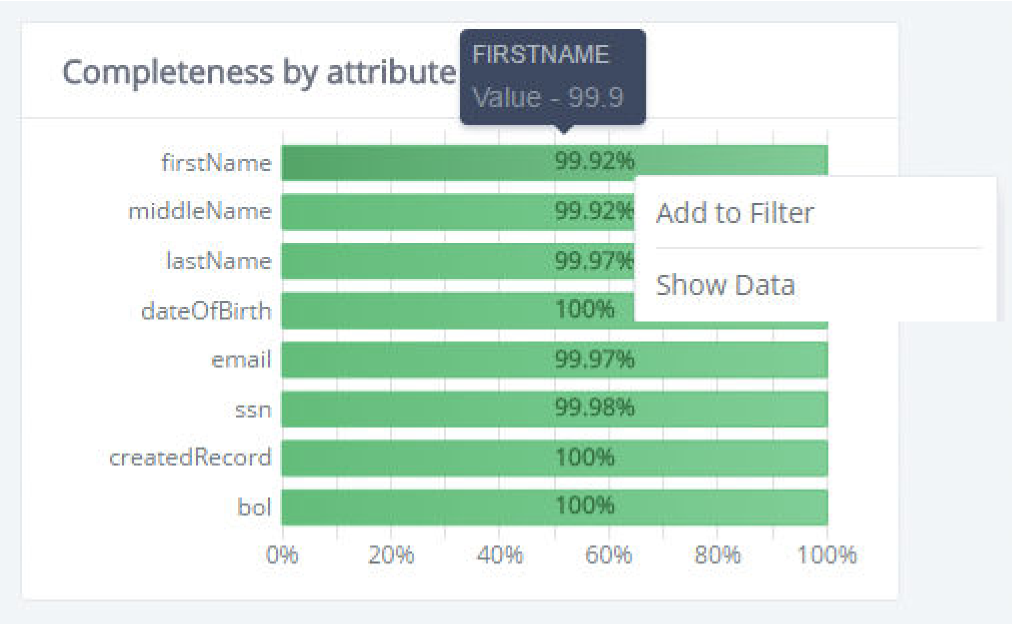
In this chapter:

Subject Profile

The Subject Completeness bar chart displays the percentage of Subject data completeness.

Total records, Populated records, Blank records, and Blank attributes tiles display the number of records corresponding to the name of the tile (Blank/Populated, and so on).

The Completeness by attribute bar chart displays the percentage of completeness for each attribute. It shows you a visual representation of the subject you are working with and helps identify where the possible data problem might be.



The Attribute Overview table displays attribute data characteristics from the following categories:

Populated, Blank, Distinct, Min, Max, Total Word/Number Patterns and Total Letter/Digit Patters.

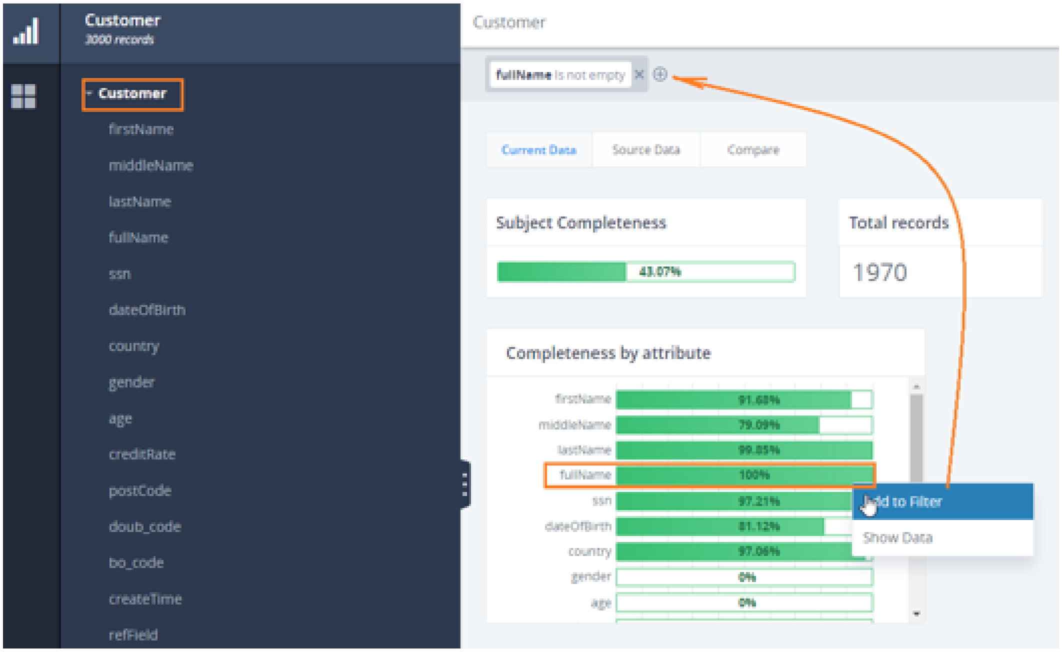
Attribute Overview

Attribute	Populated	Blank	Distinct	Min	Max	Median	Total W/N Patterns	Total L/D Patterns
firstName [A-Z]	5996	5	6046	0	999	18097	1	5
middleName [A-Z]	5996	5	6046	0	999	18097	1	5
lastName [A-Z]	5999	2	6048	0	9997	18097	1	5
dateOfBirth	6001	0	2	05/22/2017	05/24/2017	05/24/2017		
email [A-Z]	5999	2	6048	10000	9997	18100	1	5
ssn [A-Z]	6000	1	6049	0	9997	18097	1	5
createdRecord	6001	0	917	2017-05-23 07:05:4...	2017-05-25 06:33:5...	2017-05-25 06:07:0...		
bol [true/false]	6001	0	1	true	true	true		

Using the Add to Filter Option From the Profile Subject Page

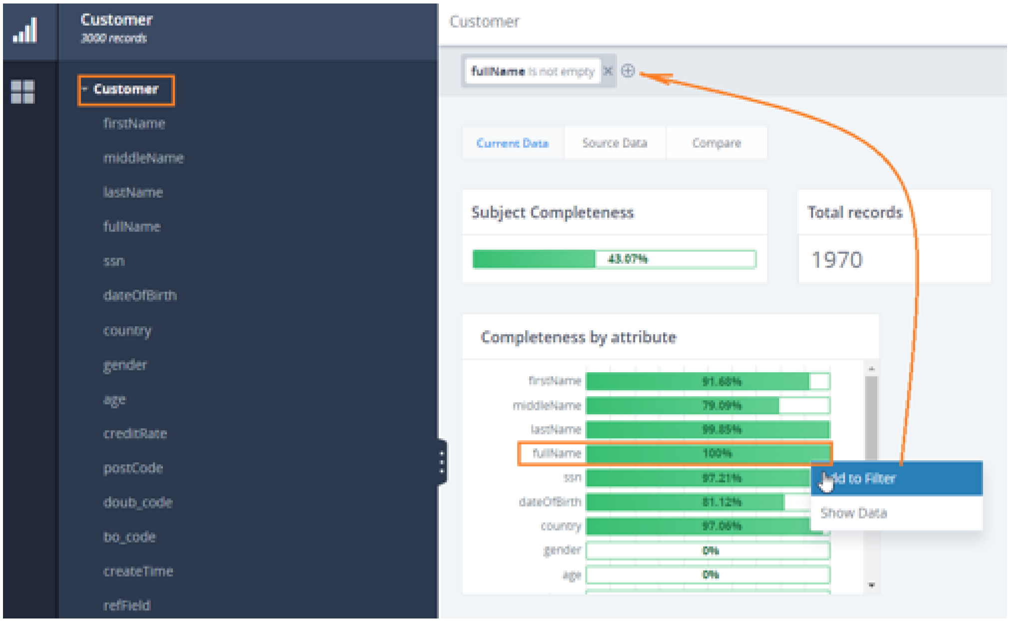
The Add to Filter functionality allows you to add a filter tag from the bar chart, tile, or grid. The Add to Filter option is present in the context menus which are available by right-clicking on most of the following components: tile, bar chart, and table cell. There is no context menu available for the Blank Attributes tile.

After selecting the *Add to Filter* option, a global filter will be added containing the appropriate criteria.



The following list describes the specifics for adding a filter from the different context menus:

- When adding a filter for the Populated records tile or Subject completeness bar chart, the global filter contains every attribute in that subject, and the operator indicates that it is not empty, as shown in the image below. Filter tags will be separated with the OR operand.



- You can right-click the Blank records tile and select the *Add to Filter* option. A global filter will be added containing a series of Filter tags for every column of the subject. Each tag will have property = <column name> and operand is *empty*. Filter tags will be separated with the AND operand.

- There is no *Add to Filter* option for Total records. Adding that filter makes no sense since it will show all the records in the system, and they are already displayed on the page.

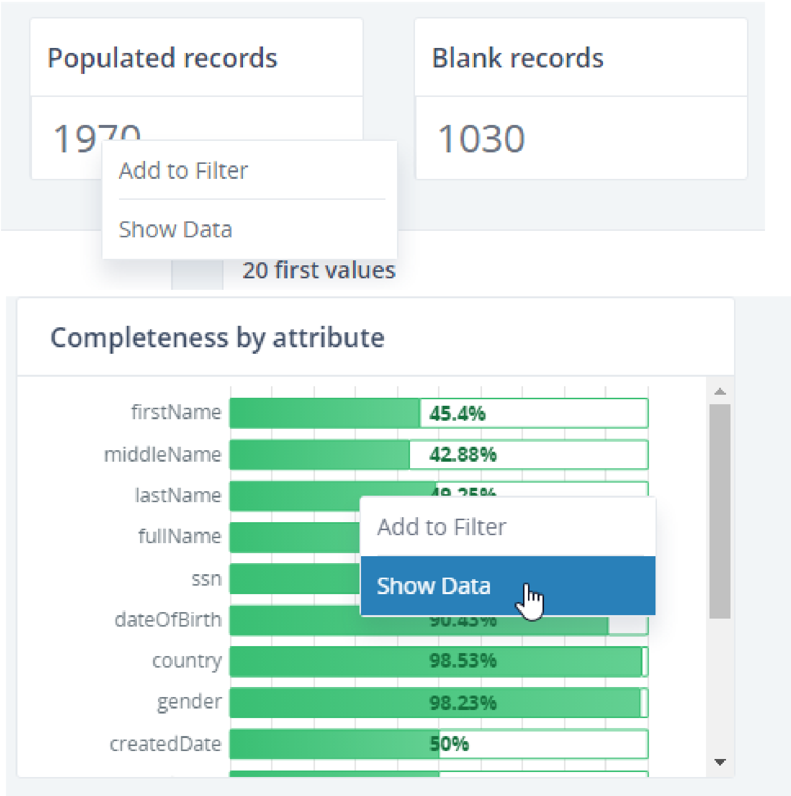
- You can right-click the *Completeness by Attribute* bar charts and select *Add to Filter*. As a result, the Filter tag will be added to the global filter with property = <column name> and operand *is not empty* (or *is empty*, if subject is empty).

- You can right-click the Attribute name for a specific column in the Attributes Overview table and select the *Add to Filter* option from the context menu. The filter tag will be added with property = <column name you clicked on>, operand *Equals*, and value = <empty>. You will need to enter values for them.
- You can right-click the Populated value for a specific column in the Attributes overview table and select the *Add to Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on> and operand *is not empty*.
- You can right-click the Blank value for a specific column in the Attributes overview table and select the *Add to Quick Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on> and operand *is empty*.
- You can right-click the Min value for a specific column in the Attributes overview table and select the *Add to Quick Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on>, operand *Equals*, and value = <the min value you clicked on>. The Attribute Overview table will be filtered according to the filter criteria.
- You can right-click the Max value for a specific column in the Attributes overview table and select the *Add to Quick Filter* option from the context menu. The filter tag will be added with property = <column name of the bar you clicked on>, operand *Equals*, and value = <the max value you clicked on>. The Attribute Overview table will be filtered according to the filter criteria.
- There is no *Add to Filter* option for the Distinct, Total W/N Patterns, Total L/D Patterns columns, and the Attribute Overview table header.

Using the Show Data Option From the Profile Subject Page

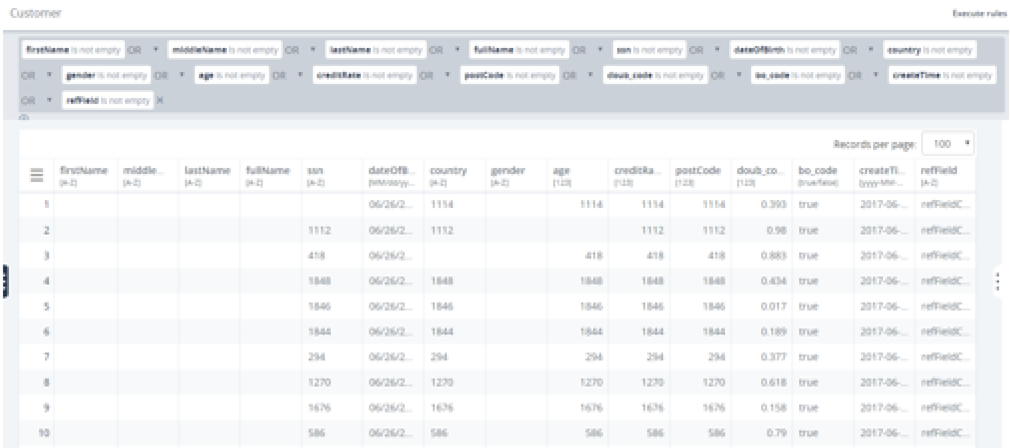
You can navigate to the Data table and see data according to the selected item in the on-screen filter or global filter data by selecting *Show Data* from the context menu. The presence of an On-Screen filter and global filter on the Data table page depends on the selected bar chart on the Profile page.

You can right-click any tile, table cell, or bar chart and select the *Show Data* option, as shown in the following image.



The following list shows how the Show Data option applies to each different component.

- For the Populated records tile, the same filter will be set for the *Add to Filter* option, and the Data table page will open with Populated records. None of the table columns will be selected.



- The behavior will be the same for Blank records, but the Data table page will open with Blank records.
- For the Total tile, the Data table page will open with all records shown. There is no *Show Data* option for the Blank attributes tile.
- For the Subject Completeness bar charts, results will be the same as the Populated records tile. None of the table columns will be selected.
- For the Completeness by attribute bar chart:
  - If you right-click the bar representing the Populated values, the Data table page for the selected column opens. The Column Profile pane has *Populated* selected.
  - If you right-click the bar part representing the Empty values, then the Data table page for the selected column opens. The Column Profile pane has *Empty* selected. You can right-click and select *Show Data* from every column on the Attribute Overview table.

  For example:

  - **Populated/Blank.** The Data table page opens with Populated/Blank records displayed for the selected column and with the Populated/Blank tab in On-Screen filter selected.

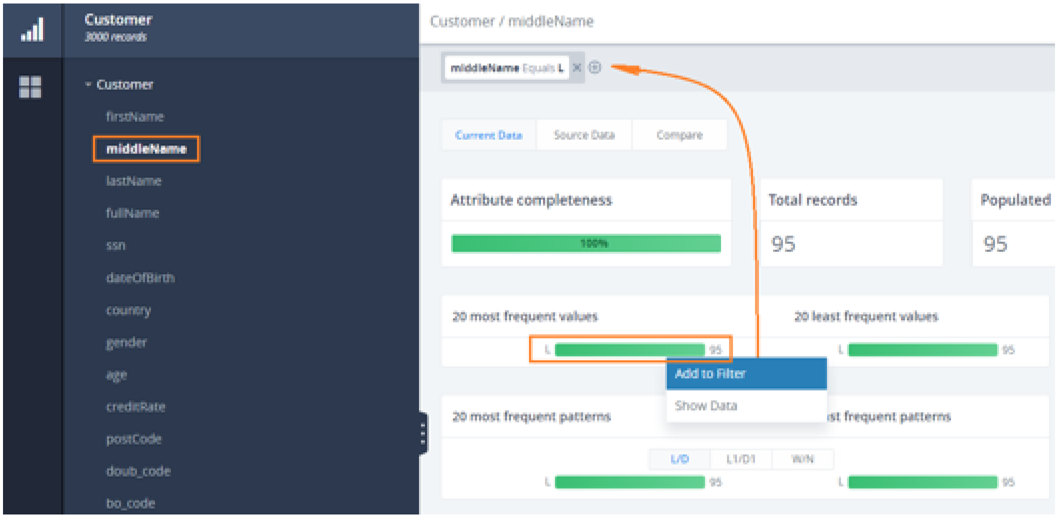
- ❑ **Min column.** The Data table page opens with the selected column. The On-Screen Column Profile is switched to the First/Last values tab and the first value in the First tab is selected.

- ❑ **Max column.** The Data table page opens with the selected column. The On-Screen Column Profile is switched to the First/Last values tab and the last value in the Last tab is selected.

❑ When you right-click anywhere else in the Attributes overview table row and select *Show Data* from the context menu, the Data table page opens but none of the grid columns will be selected.

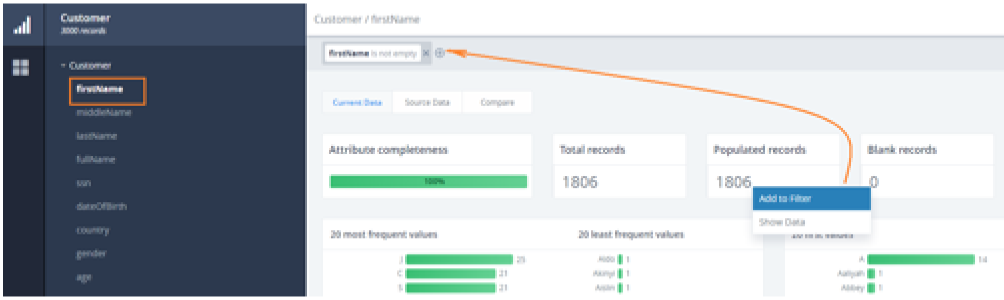
Using the Add to Filter Option From the Profile Attribute Page

The *Add to Filter* option can be selected in the Profile/Attribute page from the context menus by right-clicking the tile and bar chart components. After clicking *Add to Filter*, the filter tag will be added to the global filter, as shown in the following image.

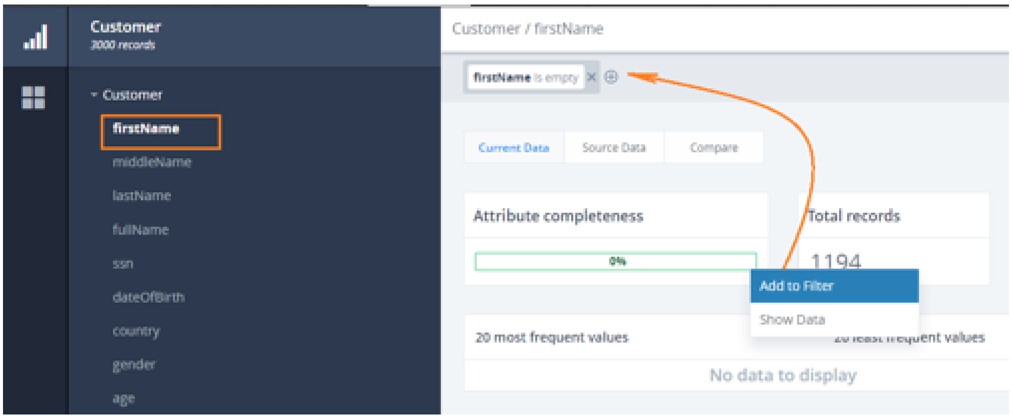


The following list describes the results of adding filters from different context menus.

- ❑ When adding a filter for the Populated records tile or Completeness by Attribute (green bar), the behavior will be the same. The global filter will be added for the selected attribute in the tree and contains the *is not empty* operator, as shown in the following image.

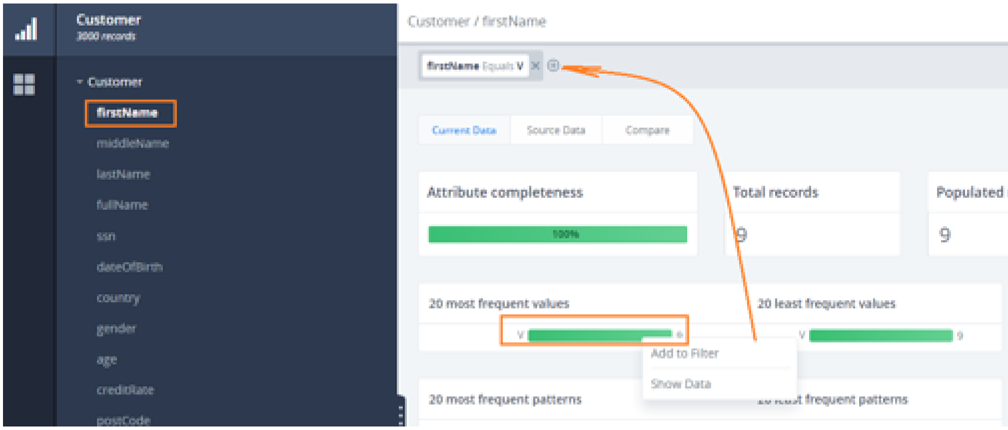


- ❑ When adding a filter for the Blank records tile or Completeness by Attribute, the global filter is added for the selected attribute in the tree and contains operator *is empty*, as shown in the following image.

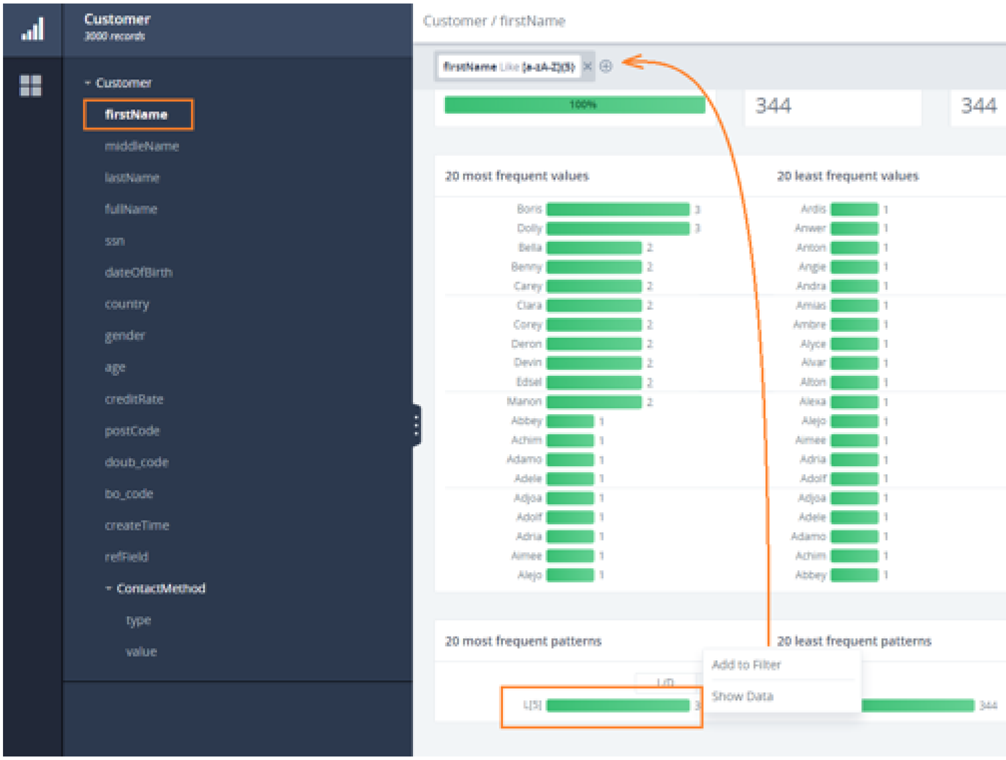


- ❑ There is no *Add to Filter* option for Total records and Distinct records tiles.
- ❑ You can also open the context menu by right-clicking either the bars, name of the bars, or the value in the 20 most/least frequent values, 10 first/last values, 20 most/least frequent patterns, or tags.

- You can right-click on the 20 most frequent values, 20 least frequent values, 10 first values, and 10 last values charts, and select *Add to Filter*. The filter tag will be added with property = <column name of the bar chart you clicked on>, operand *Equals*, and value = <the value you clicked on>. The attribute page will be filtered according to the filter criteria, as shown in the following image.



- When you right-click the L/D, L1/D1, W/N bar charts pattern and select *Add to Filter*, the filter tag will be added with property = <column name of the bar chart you clicked on>, operand *Like*, and value = <the value you clicked on transformed to RegEx format>. The Attribute page will be filtered according to the filter criteria, as shown in the following image.



- You can select *Add to Filter* for the Tags section, and select either the Distinct or Individual tab. The filter tag will be added with property = <column name of the bar chart you clicked on>, operand *Tagged with*, and value = <tag label>. The Attribute page will be filtered according to the filter criteria, as shown in the following image.

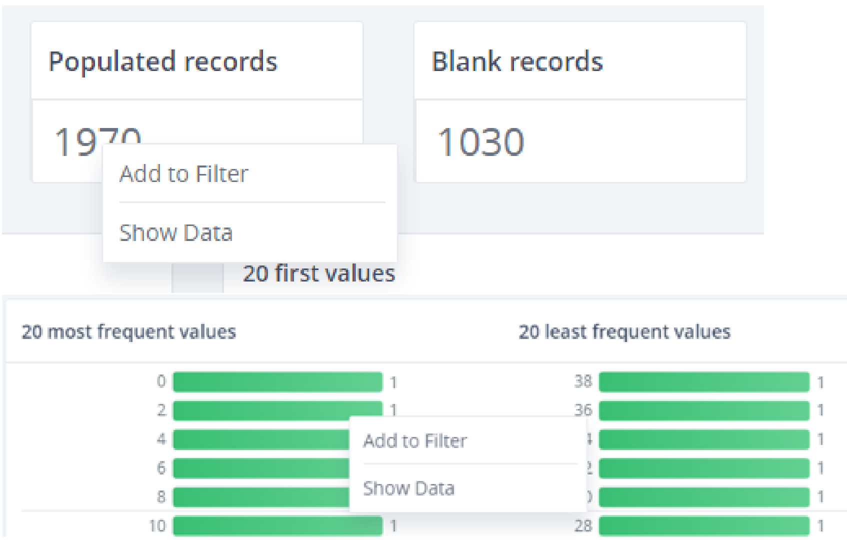
There is a delimiter line between each of the bar charts for each bar chart section.

Once a filter is added, the Attribute page will be filtered according to the filter criteria.

Using the Show Data Filter Option From the Profile Attribute Page

You can navigate to the Data table and see data according to the selected item in the filter or global filter data by selecting the *Show Data* from the context menu. The presence of a filter and global filter on the Data table page depends on the selected bar chart on the Profile page.

You can right-click on any tile, table cell, or bar chart and select *Show Data*, as shown in the following image.



Chapter 6

Data Quality Rules

If you identify problems or issues with the data and wish to change it, you can use the Rules functionality.

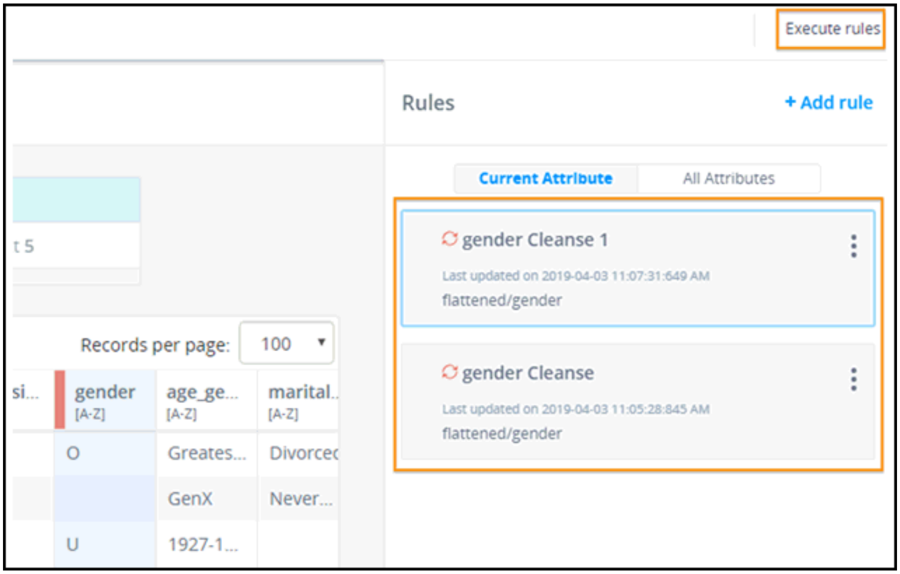
The Rules functionality allows you to edit, mark, and remove data from the interface without using the Data Quality Server.

It is also used to work with data records (update/remove) by creating data quality rules (Cleansing).

In this chapter:

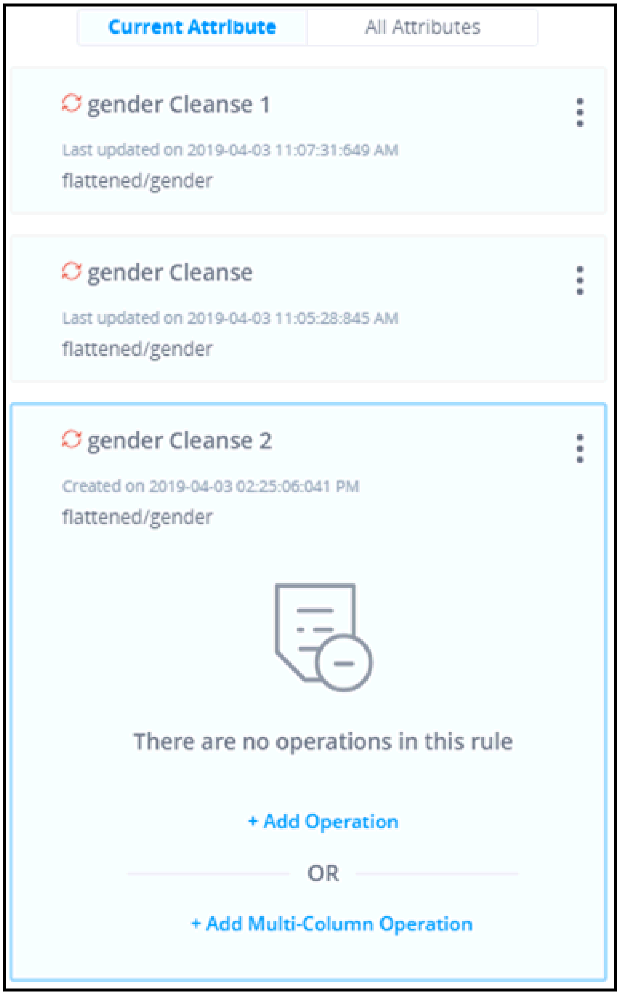
Rules Functionality

The Rules functionality is represented with the Rules panel and the Execution process, as shown in the following image.

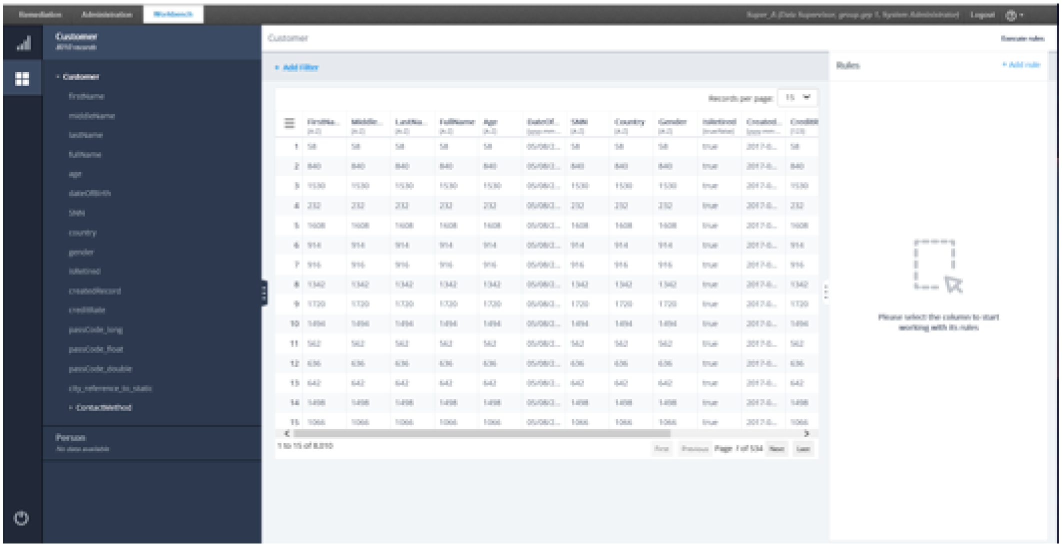


The Rules panel is available for the selected column in the Data table and is collapsed, by default. The Rules panel displays a list of rules associated with the selected column. The Rules are collapsed, by default, when navigating to the Rules panel after working with filters or sorting.

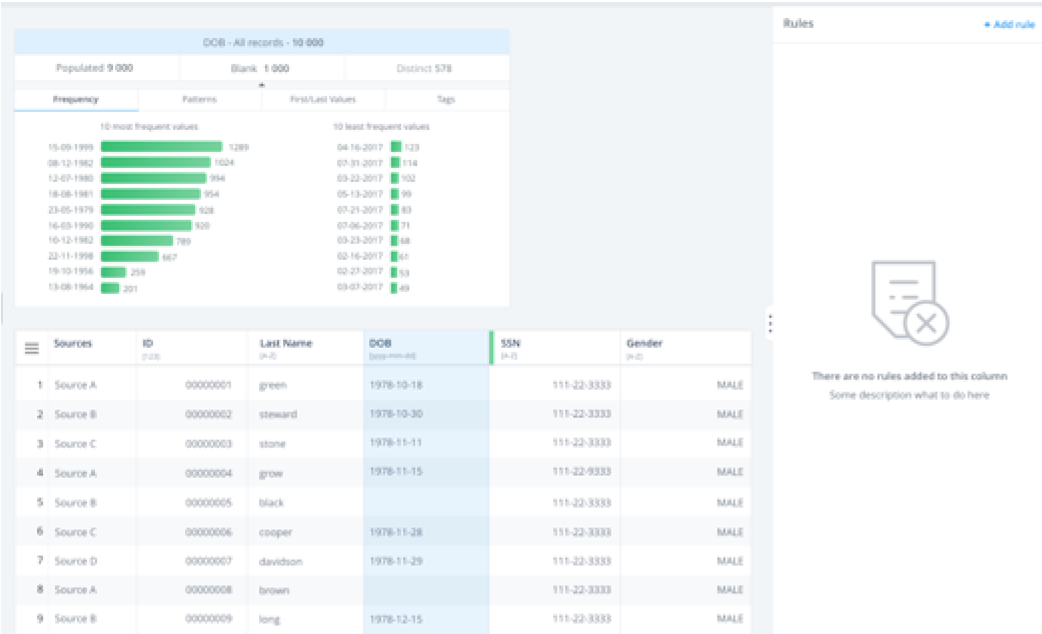
If you are adding a new rule, it is added in the expanded state, as shown in the following image.



If the column is not selected, the corresponding information is displayed in the Rules panel, as shown in the following image.

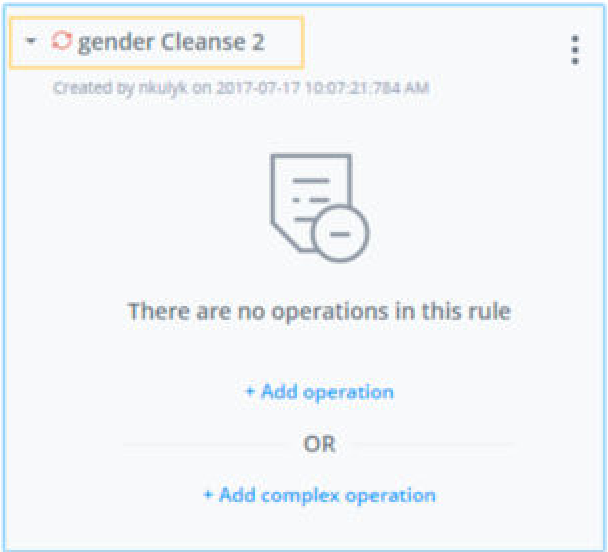


If there are no rules associated with the selected column, the corresponding information displays, as shown in the following image.

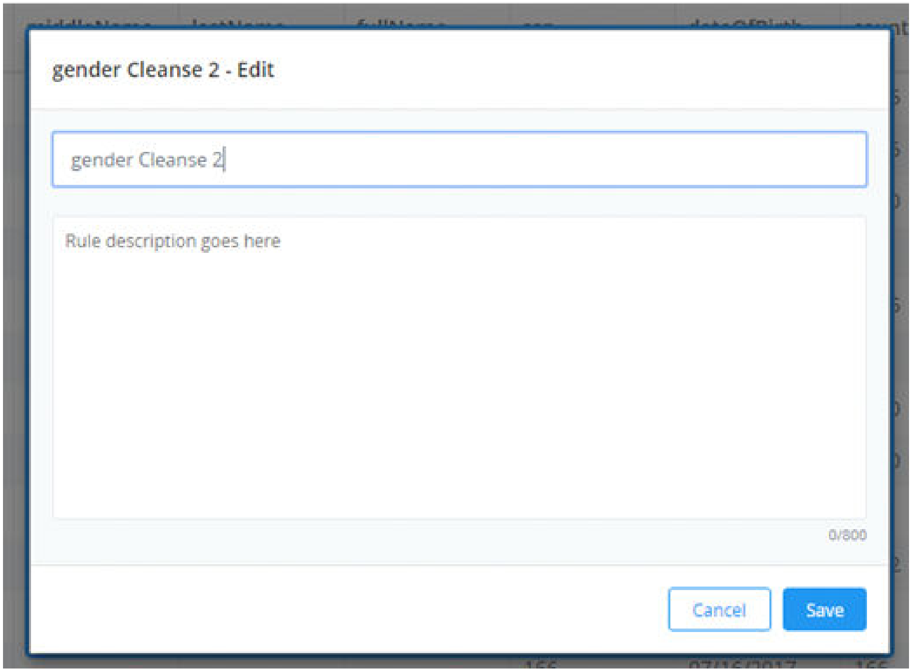


Adding Rules

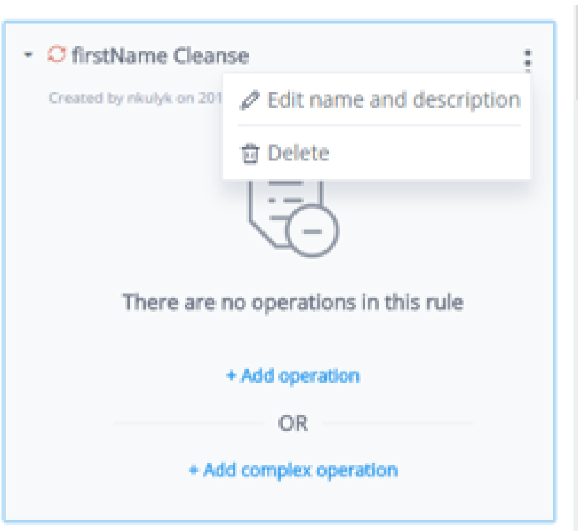
To add a new rule, select the column, expand the Rules panel and click *Add rule*. A new rule is created with a predefined name, for example, *<Column name> Cleanse* [number representing the last existing rule with this name +1], as shown in the following image.



The Rule description is empty, by default, as shown in the following image.

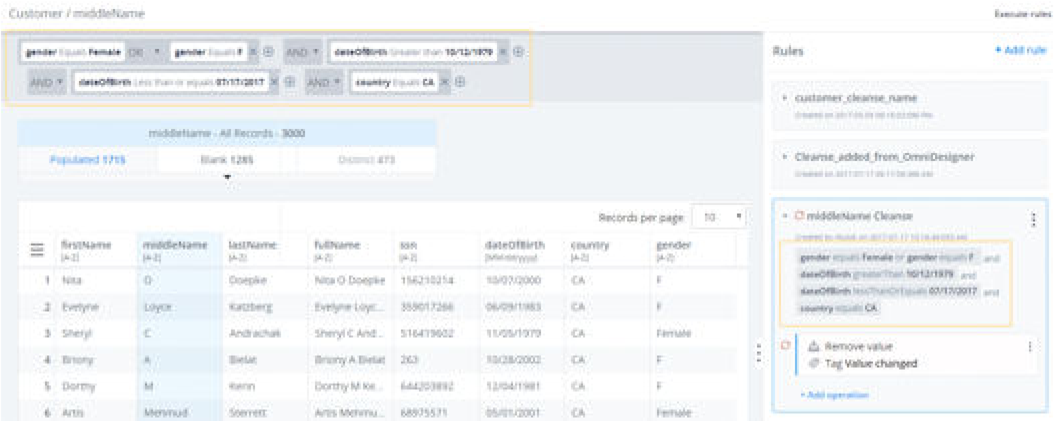


You can edit the Rule name and description, as shown in the following image. A simple validation is added to the rule edit mode.

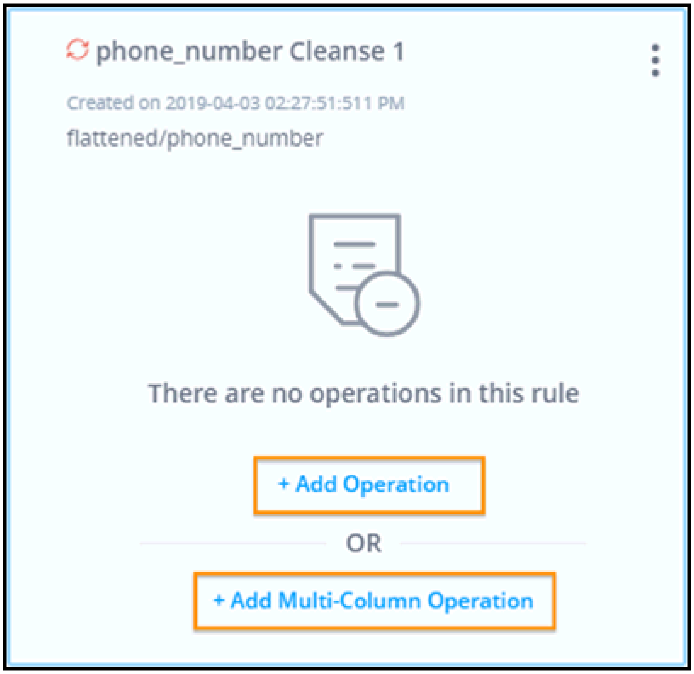


You can also delete rules. A confirmation message appears when you click *Delete*, as shown in the following image.

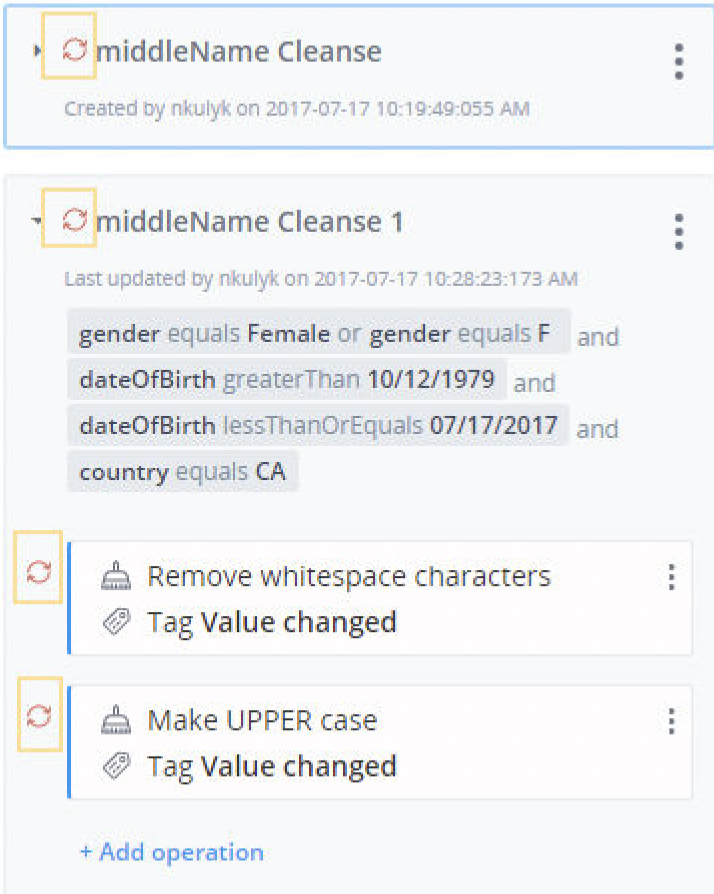
The rule inherits the global filter of the data set that is added at the moment of adding the rule, as shown in the following image.



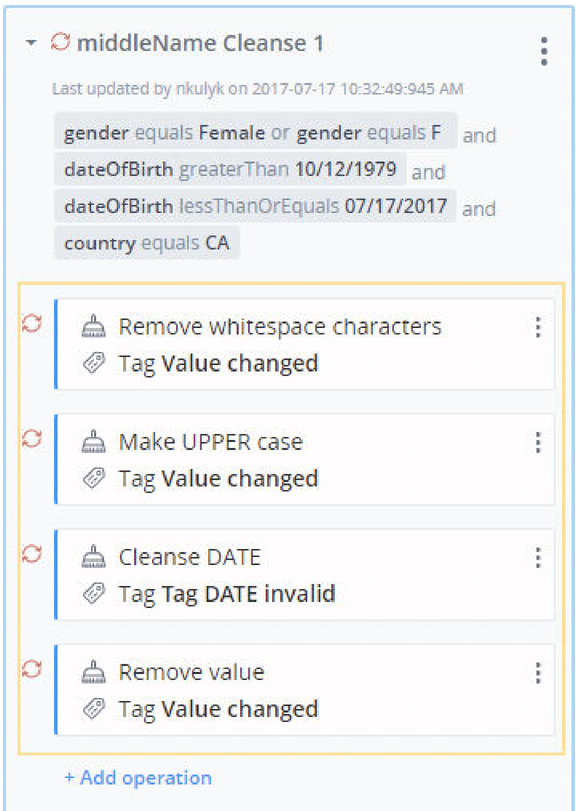
You can click *Add operation* to add a new operation, as shown in the following image.



Once you select the rule or operation, it becomes highlighted. The new rule is selected as an uncommitted (orange color) icon, as well as a newly added operation, as shown in the following image.



You can expand the rule to view a list of operations you can choose from to add to the rule, as shown in the following image.

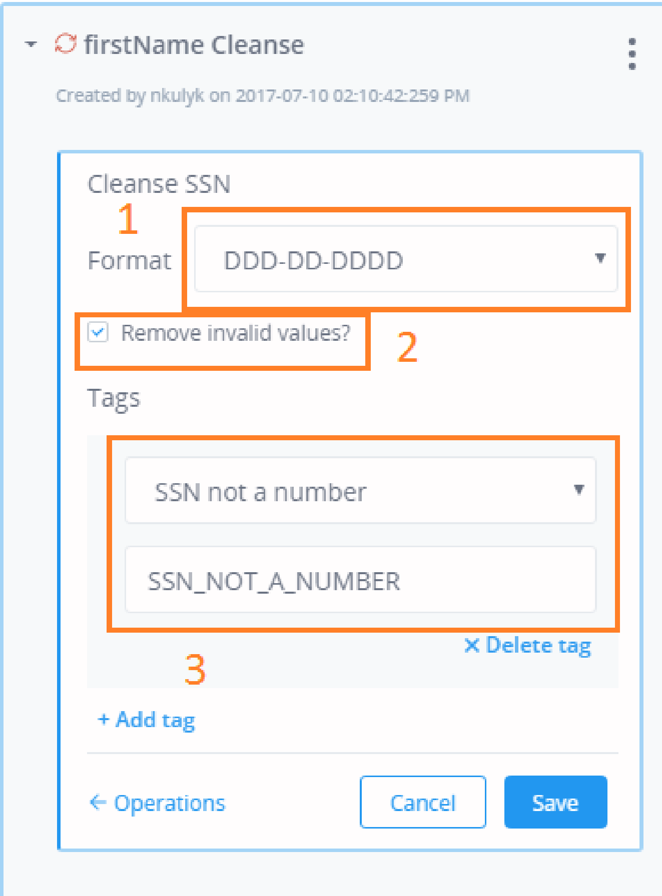


Working With Operations

The Operation is a single action that allows you to customize the way you view data.

For example, the Uppercase operation turns all data in the selected column into Uppercase characters.

Each operation updates data for a selected column in a certain way defined by options (for example, inputs, values, radio buttons, check boxes, and so on) which are set in corresponding XML files and help to distinguish and update data, as shown in the following image.



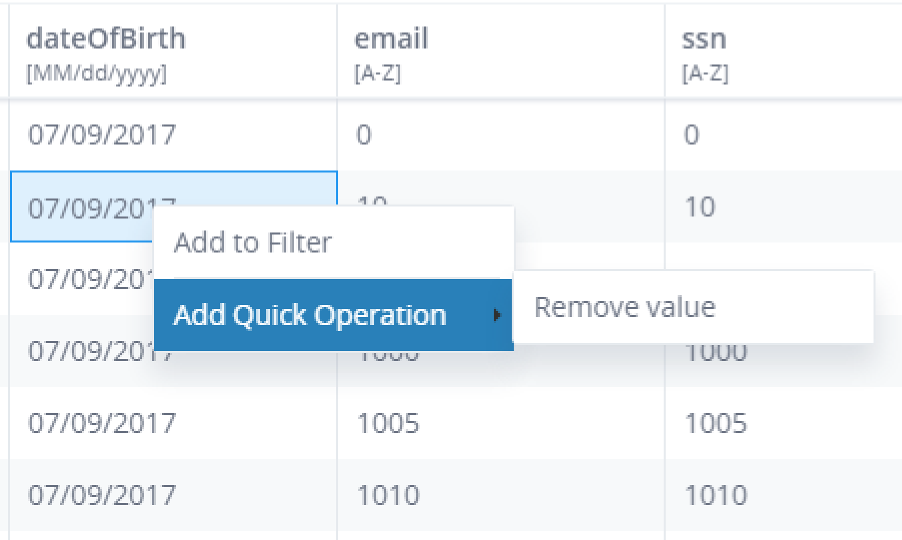
For example, the *Cleanse SSN* operation contains the following options:

1. **Format.** Indicates what format of data should appear after an operation is applied (nine digits will be displayed in a specified way).
2. **Remove invalid values?** Select this check box if you want to remove data that does not satisfy a specified format (less or more than nine digits, characters, or special symbols).
3. **Tags.** Added to data and characterizes it according to the set tag value.

The operation content corresponds to the column data type you select. You can set the parameters for any operation you choose. You can edit the following parameter types:

1. **Behaviour parameters.** What the operation has to execute.
2. **Tags.** Signs off the execution of an operation.
3. **Output format.** A set of possible output format options for the content type.

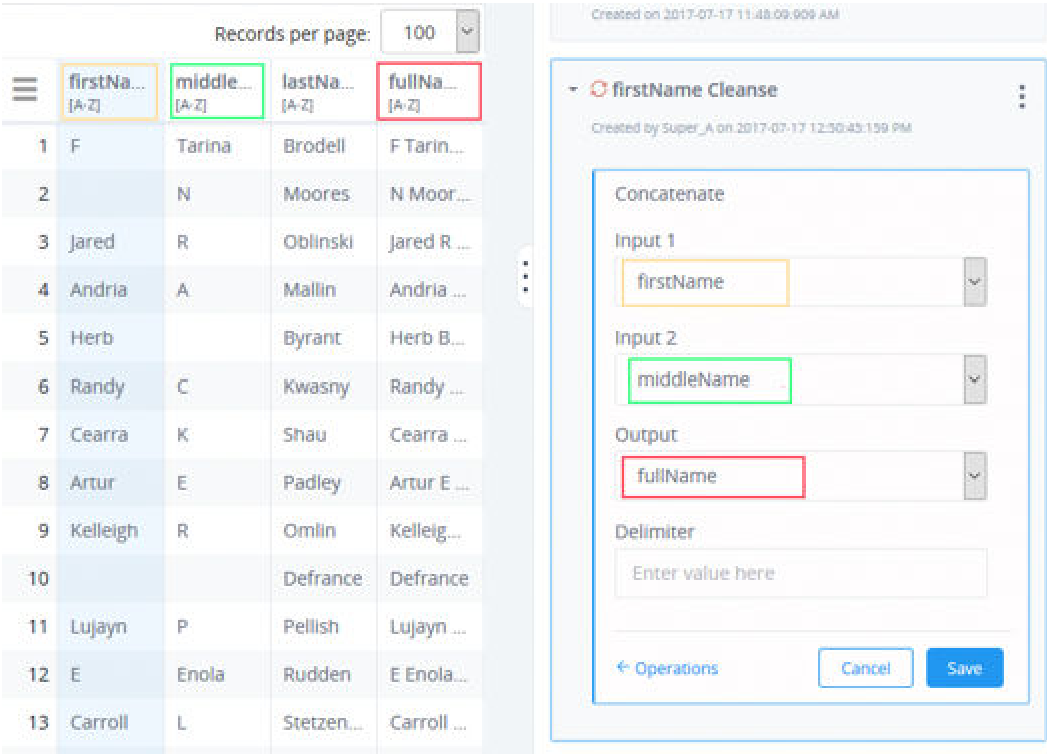
There are two ways to add an operation: add as quick operation from the table context menu and add from the rule. The Add Quick Operation is shown in the following image.



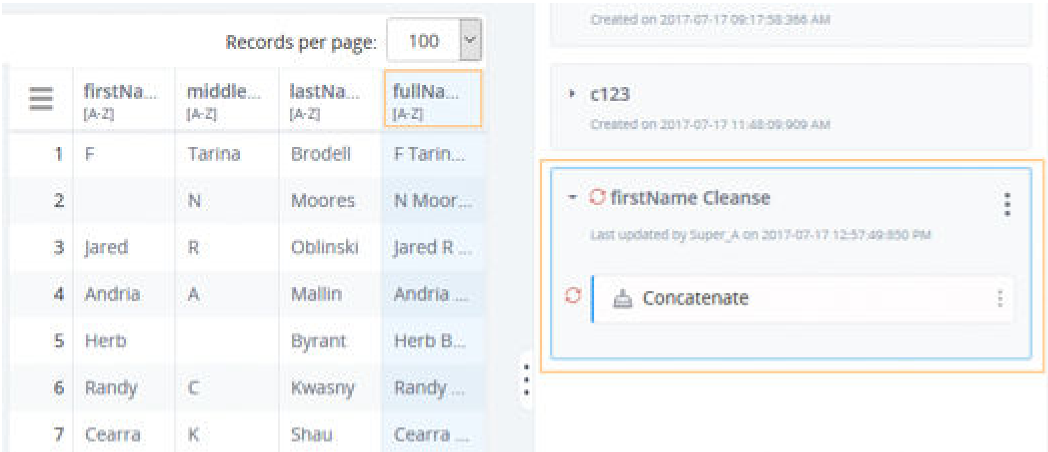
There are also two types of operations depending on the affected areas: Simple and Multicolumn.

Simple operations are related to one Attribute and after execution of the rule will affect data only for the selected attribute.

Mulitcolumn operations involve several attributes and work with data from those attributes that are listed in the inputs and outputs of the operations, as shown in the following image.

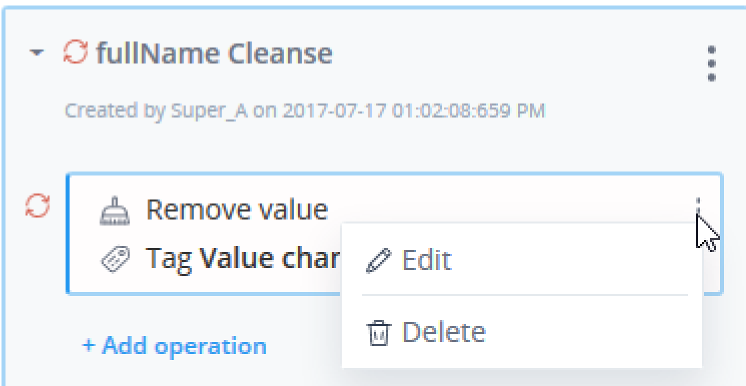


The rule that contains a mulitcolumn operation will be created for each involved attribute, as shown in the following image.

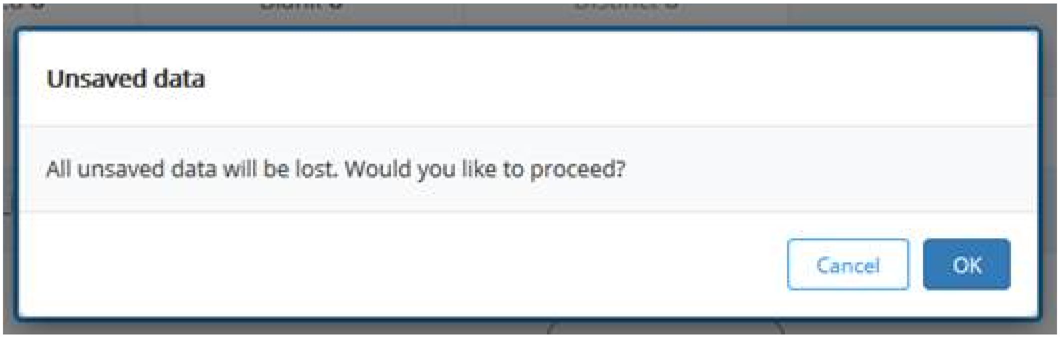


You can hover your cursor over the operation and see the name, parameters, and tags inside the corresponding tooltip, as shown in the following image.

You can edit and delete any operation, as shown in the following image.



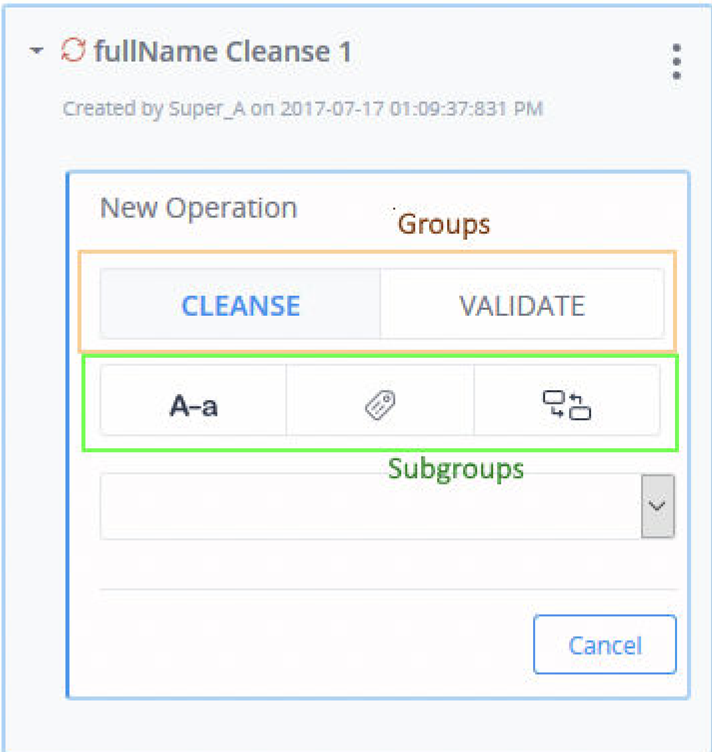
Only one operation at a time can be opened in Edit mode. If you try to open another operation or do another action while the operation is opened, you will be asked to save changes first, as shown in the following image.



Adding Operations From the Rules Panel

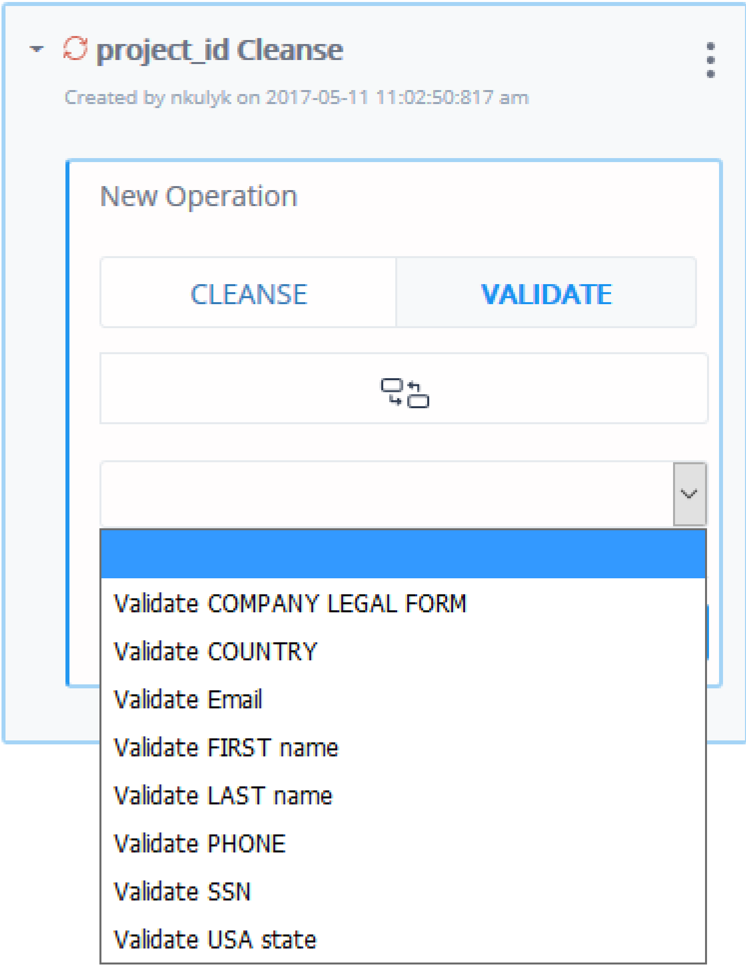
In order to add a rule from Rules panel, select either the *Add operation* or *Add multicolumn operation* buttons.

Currently, there are two groups (Cleanse and Validate), as shown in the following image.



The number of subgroups can be different. Groups and subgroups are configured in the corresponding XML file of the operation in a specific folder where the Omni-Gen™ system is installed.

If the number of operations is more than five, then the operations display in a drop-down list, as shown in the following image.

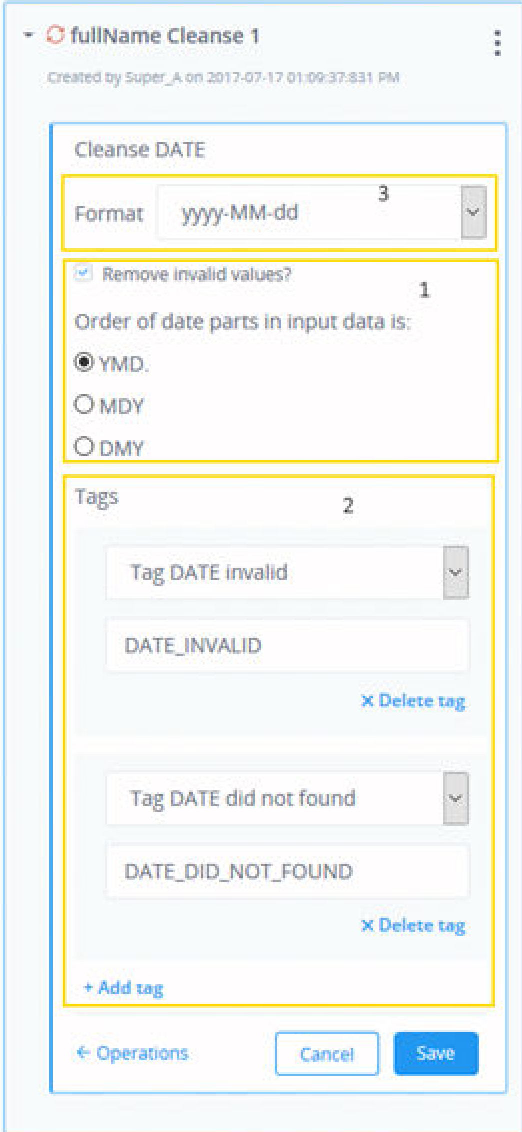


The operation content corresponds to the column data type you select. You can set the parameters for any operation you choose.

You can edit the following parameter types, as shown in the following image.

- **Behaviour parameters.** What the operation has to execute.
- **Tags.** Annotate positive or negative aspects of your data.

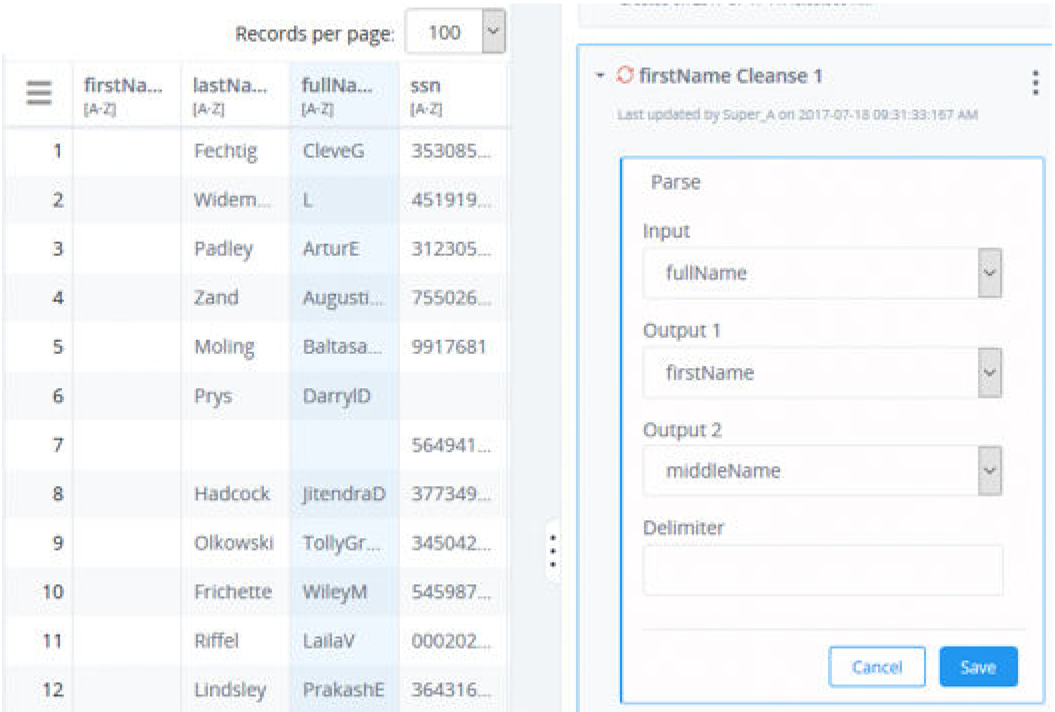
- **Output format.** A set of possible output format options for the content type.



Changes that were made with the operation/rule are marked with an uncommitted icon.

Adding Multicolumn Operations

Multicolumn operations affect more than one input attribute (for example, First name, Middle name, and Last name). This type of operation can be added to every involved attribute, as well as attributes that have been created, as shown in the following image.



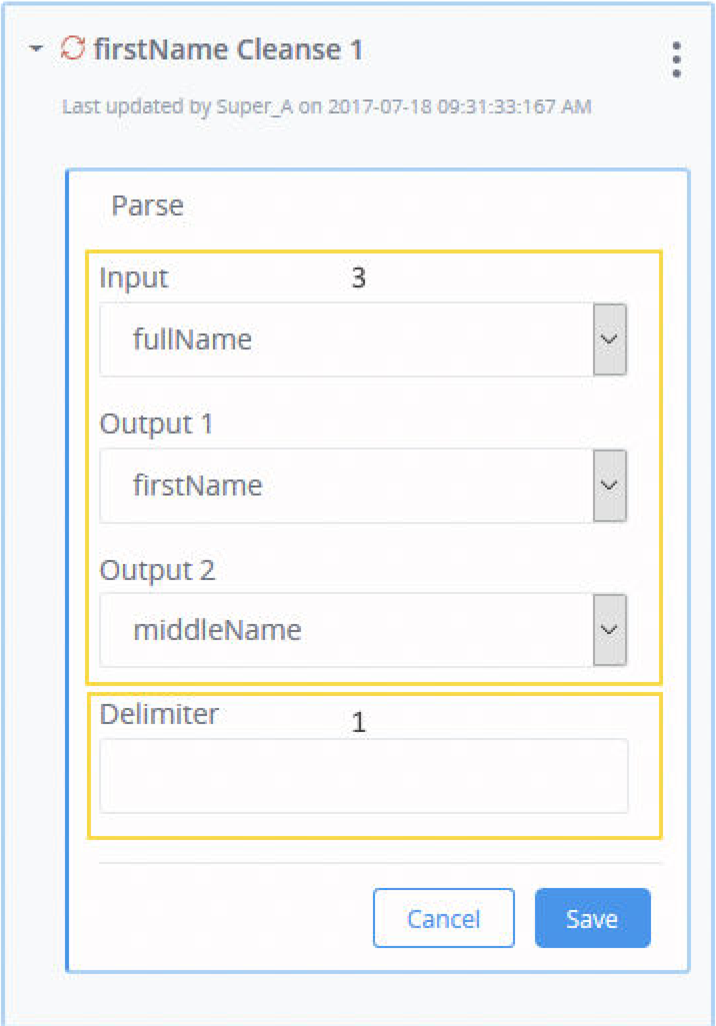
Operations can be edited from every involved attribute.

Only one multicolumn operation can be added to one rule.

Multicolumn operations cannot be added to the rules with standard operations, as shown in the following image. Operations can contain up to three sections:

1. **Behaviour parameters.** What the operation has to execute.
2. **Tags.** Annotate positive or negative aspects of your data.

3. **Mapping.**



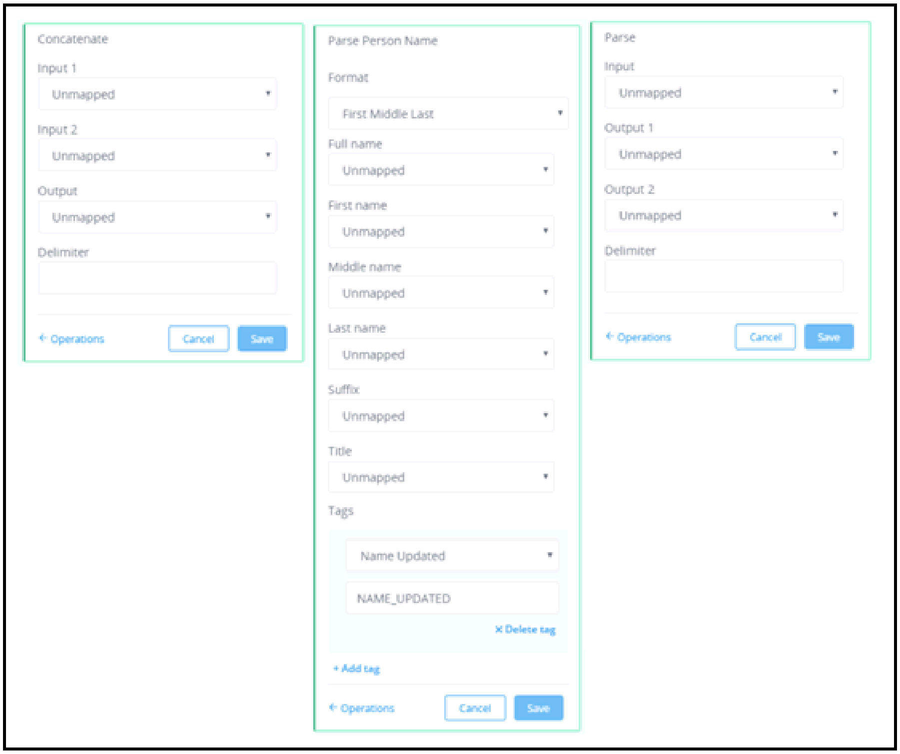
Note: By default, all input fields are unmapped.

Standard labels can be mapped for the input attributes operation with particular examples for the data model (for example, First name = Name1, Middle name = Name2, Last name = Name3). You can add fields only with the same data type selected. Any field can stay unmapped.

You can add as many simple operations as needed, but only one multi-column operation can be added.

The following image shows the three multi-column operations available in Omni-Gen™ Personal Edition:

- Concatenate
- Parse Person Name
- Parse



Appendix A

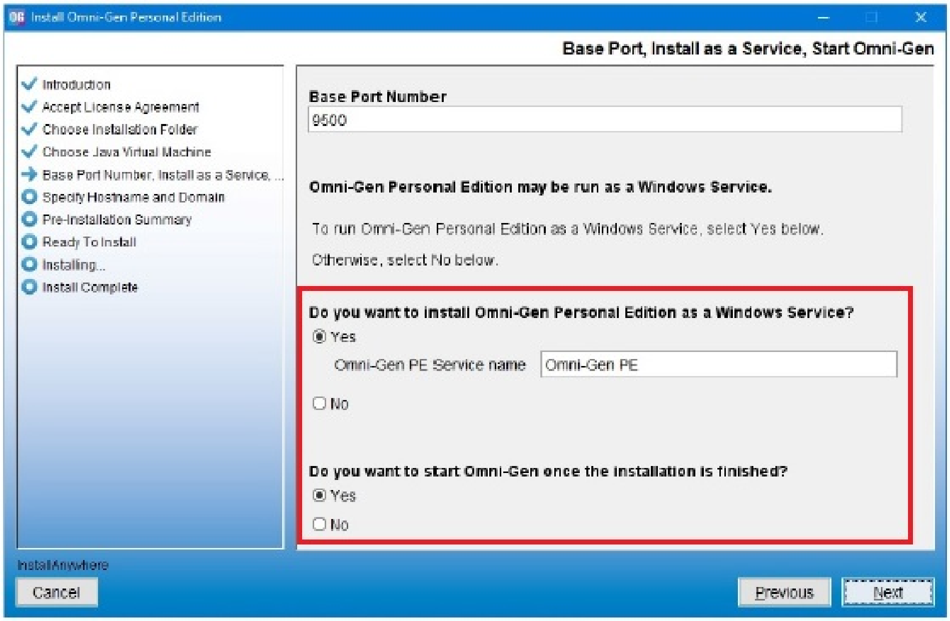
Managing the Windows Service

This section describes how to manage the Windows service that is installed with Omni-Gen™ Personal Edition.

In this appendix:

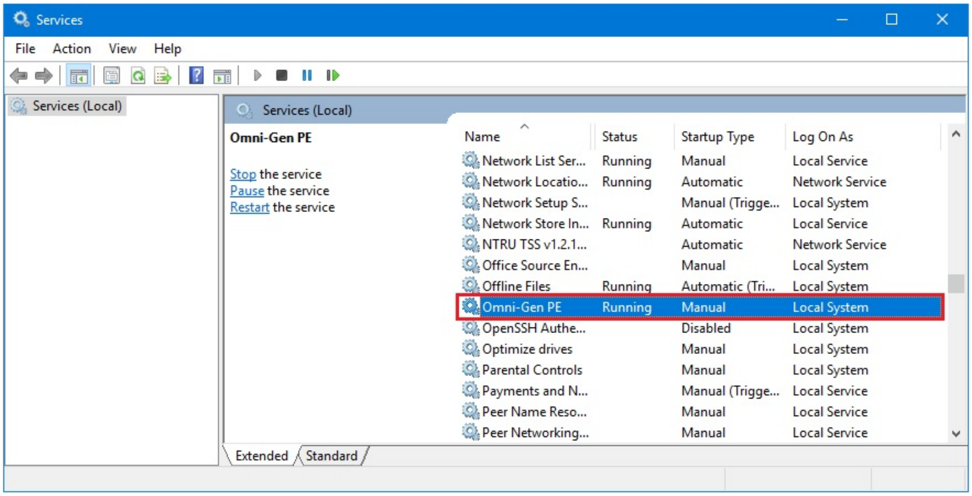
Stopping and Starting the Windows Service

During the installation of Omni-Gen™ Personal Edition, you are prompted to install the product as a Windows service, as shown in the following image.

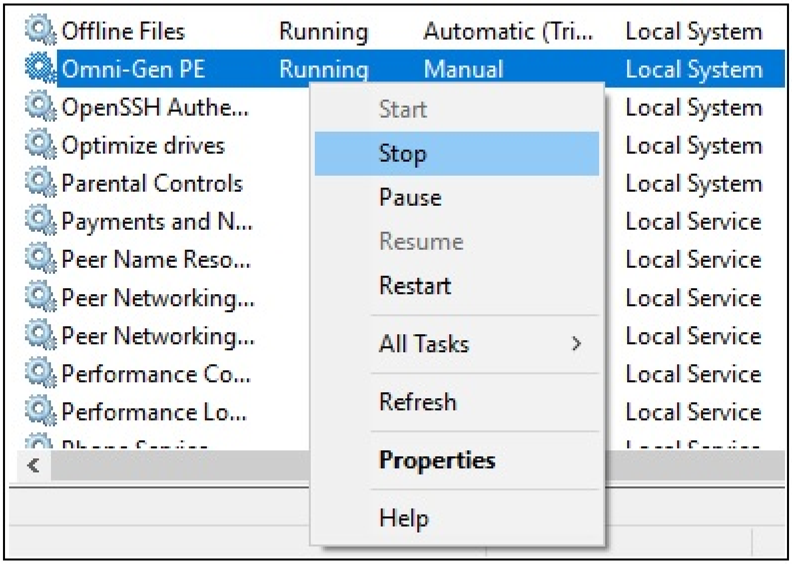


By default, the Windows service is labeled *Omni-Gen PE*, which is also set to start once the installation has completed.

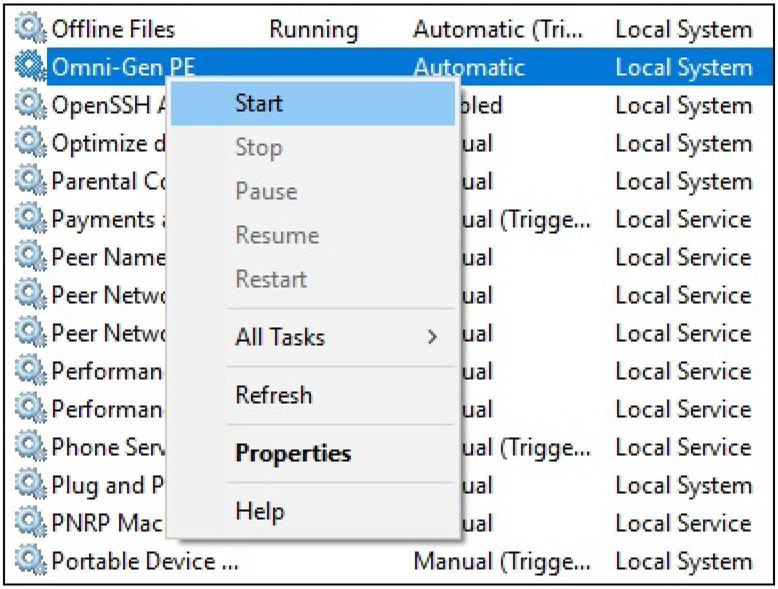
You can manage this service from the Windows Services utility (Administration Tools), which is shown in the following image.



To stop the *Omni-Gen PE* service at any point, right-click the service and select *Stop* from the context menu, as shown in the following image.

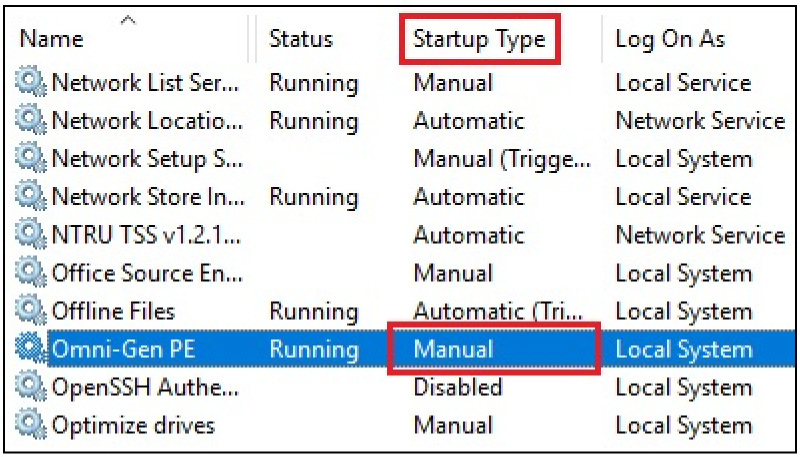


To start the *Omni-Gen PE* service, right-click the service and select *Start* from the context menu, as shown in the following image.

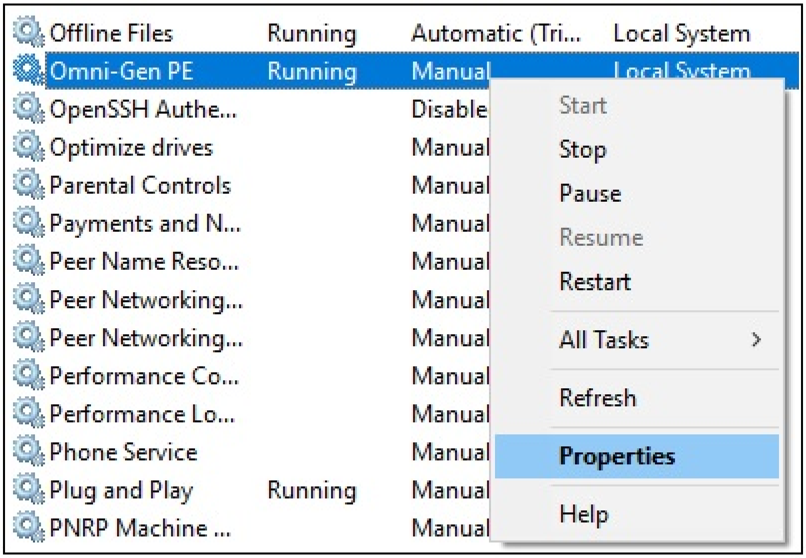


Changing the Startup Type to Automatic

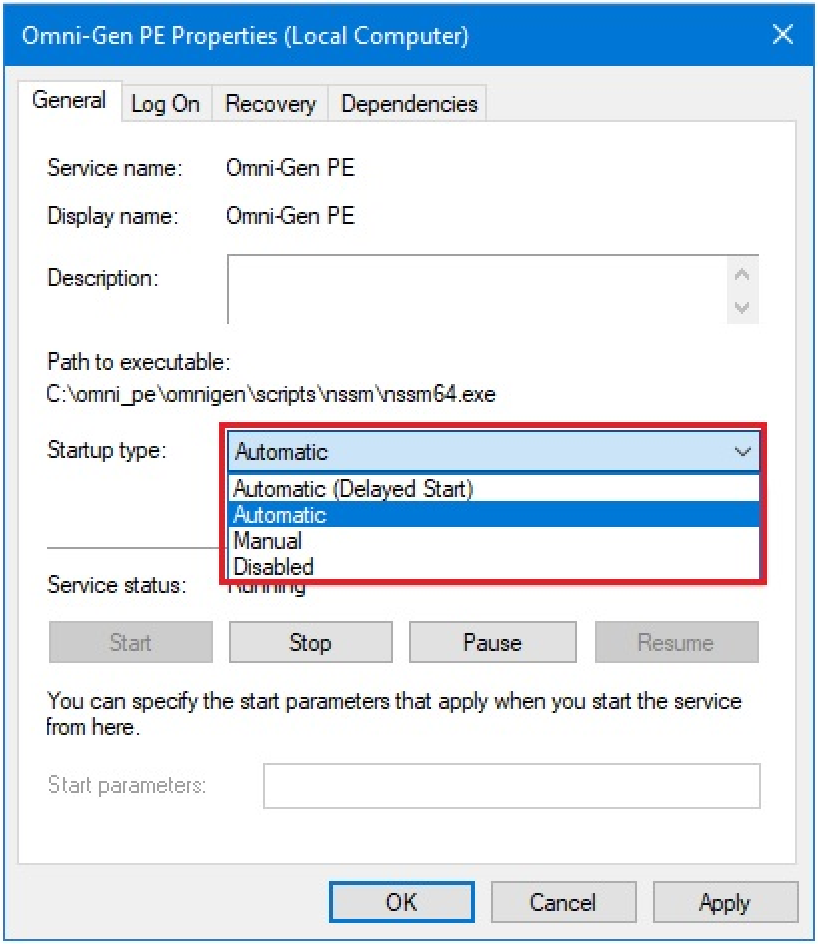
The startup type for the *Omni-Gen PE* Windows service is set to *Manual* by default, as shown in the following image.



This means that if you were to restart (reboot) your system, Omni-Gen™ Personal Edition will not be started automatically. As a result, it is recommended to set the startup type for the *Omni-Gen PE* service to *Automatic*. To do so, right-click the *Omni-Gen PE* service and select *Properties* from the context menu, as shown in the following image.

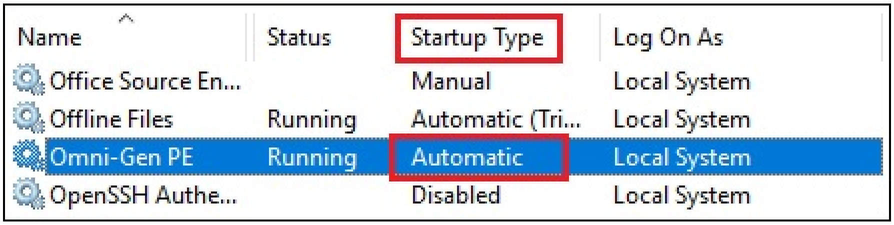


The Omni-Gen PE Properties dialog opens, as shown in the following image.



Select *Automatic* from the Startup type drop-down list, click *Apply*, and then *OK*.

The startup type for the *Omni-Gen PE* service is now set to *Automatic*, as shown in the following image.



Now, if you were to restart (reboot) your system, Omni-Gen™ Personal Edition will also be started automatically.

Appendix B

Additional Functionality

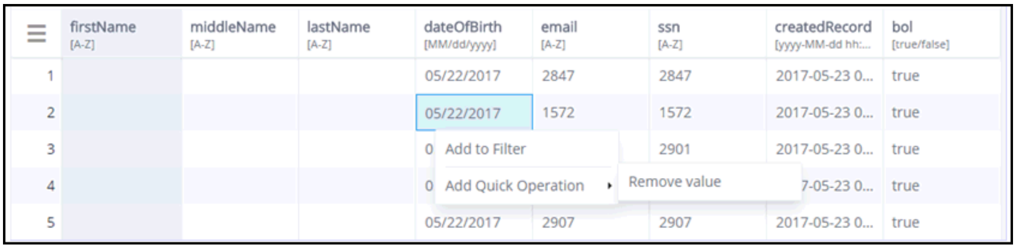
This appendix describes the additional functions of Omni-Gen™ Personal Edition Getting Started.

In this appendix:

Adding Quick Operations from the Data Grid

This section describes how to add quick operations from the data grid.

You can add basic operations from the data grid value using the context menu, as shown in the following image.



The behavior of adding operations is the same as for the Rules panel.

After selecting *Add Quick Operation*, the option Rules panel expands, and the operation is added. The following list describes several ways of how quick operations are added to a new or existing rule.

- **The rule is not present in the Rules panel or there are only external rules present.** A new operation will be added to a new rule. In addition to an empty description, the system automatically generates a default name, for example:

  ```
  <Column name> Cleanse [number representing the last existing rule with this name + 1]
  ```

The rule will inherit the Global Filter of the data set that you used at the moment of adding an operation, and the rule is automatically selected and highlighted.

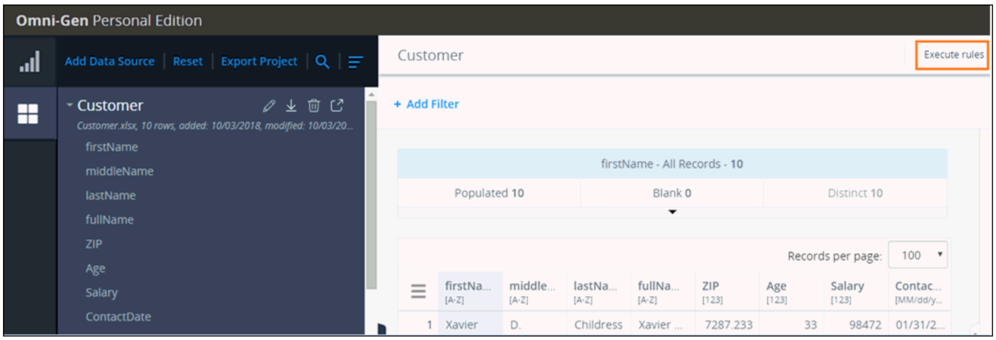
- **The rule is present in the Rules panel.**
  - **If the rule is selected,** and if the rule's filter matches the global filter, then the operation will be added to the latest and selected rule. However, if it doesn't match the global filter, then the latest rule in the rules panel will be verified and the operation will be added to the new rule.
  - **If the rule is not selected,** then the latest rule in the rules panel will be verified. If the latest rule's filter matches the global filter, then the operation will be added to the latest rule. If the latest rule's filter doesn't match the global filter, then the operation will be added to the new rule.

You can add as many quick operations as you wish. Added operations and rules are marked as uncommitted.

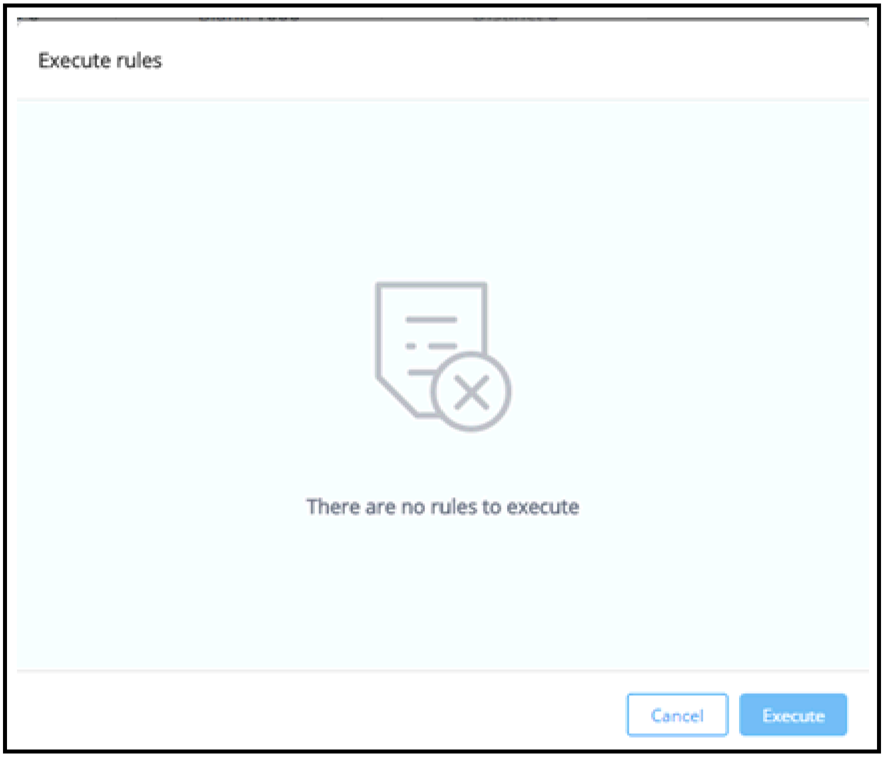
Configuring and Executing Rules

If the added rules are not executed, then data will not be changed.

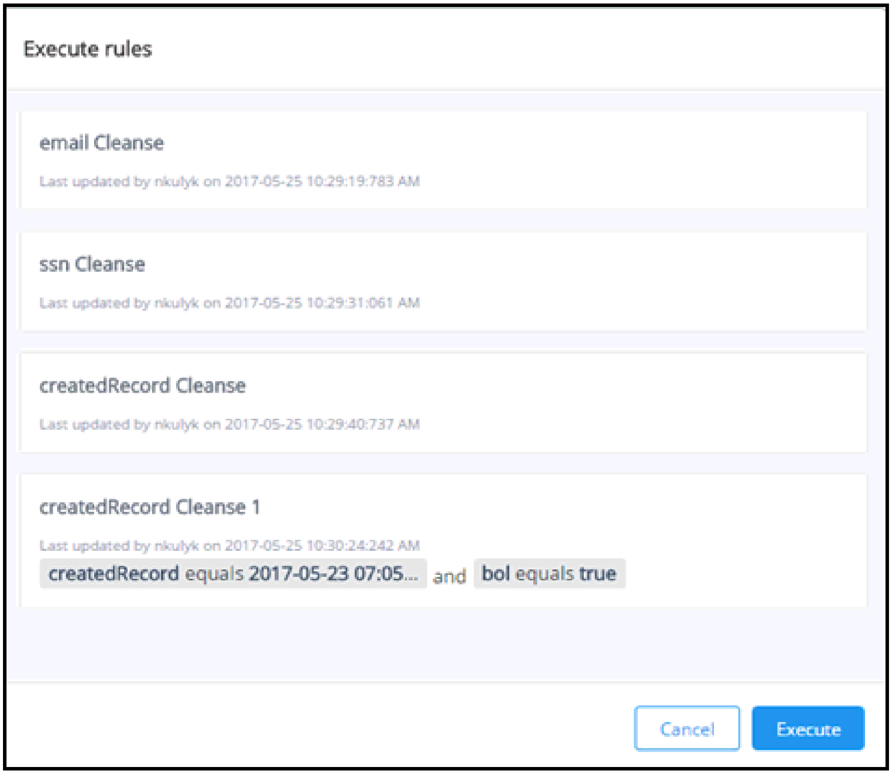
Click *Execute rules* to apply all changes. It is located at the top of every Personal Edition page, as shown in the following image.



A dialog appears with all rules that can be executed. If there are no rules to execute, the following image appears.



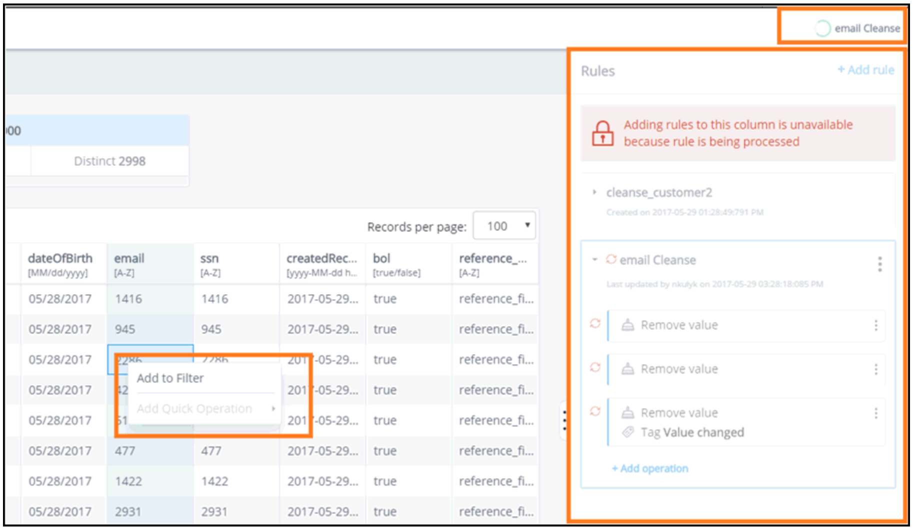
After executing a rule, all created rules will be processed, but only those which were modified will be displayed in the Execute rules dialog, as shown in the following image.



All added rules including previously created and processed rules will appear in the Execute rules dialog. The Execute rule dialog informs you of when an executed rules affect data.

Click *Execute* to start the execution process.

Once processing has started, the rule panel for selected columns or for columns involved in multi-column operations are blocked until the rule has finished processing. You will also not be able to add quick operations for selected columns, as shown in the following image.

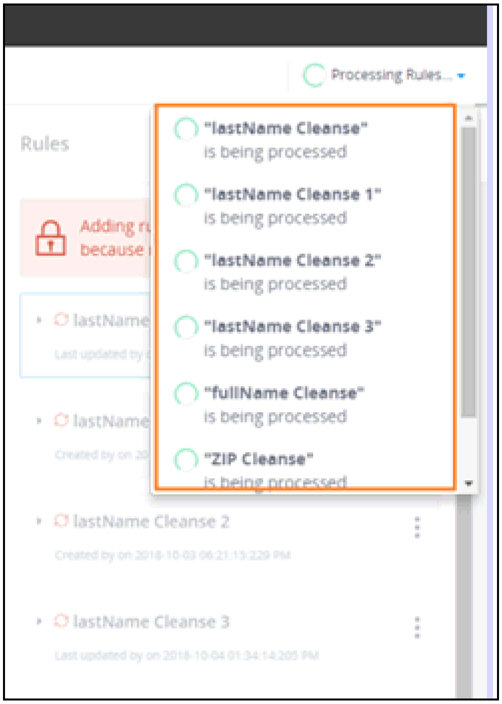


A notification will appear if a panel is blocked. You can work with other columns and rules during the execution process, and you can create rules for columns that are not involved. However, new rules cannot be execute until processing of the previous rules are completed.

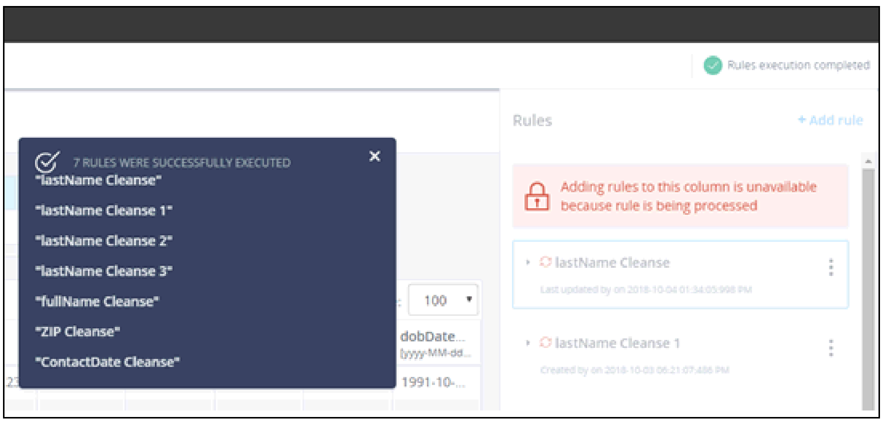
The Rule processing indicator appears in the bread crumbs area during the execution process.

If only one rule is processed, then its name will appear instead.

There are no limitations for the quantity of rules that can run in the background at the same time. If many rules are processed simultaneously, then the Rules processing indicator will contain all rules, which are collapsed by default. You can expand it to see the detailed information about which rules are processed, as shown in the following image.



When the process is complete, a notification appears that a change was completed in the notification area, and the Rules panel becomes available, as shown in the following image.

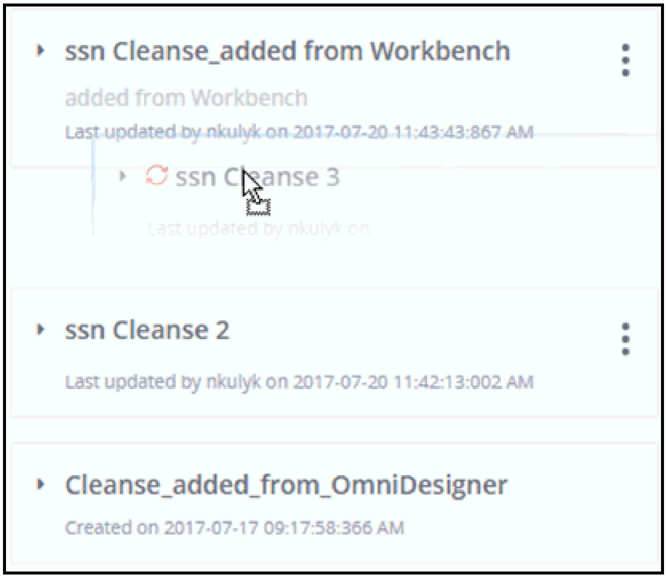


You cannot cancel or revert rules during processing.

Reordering Rules

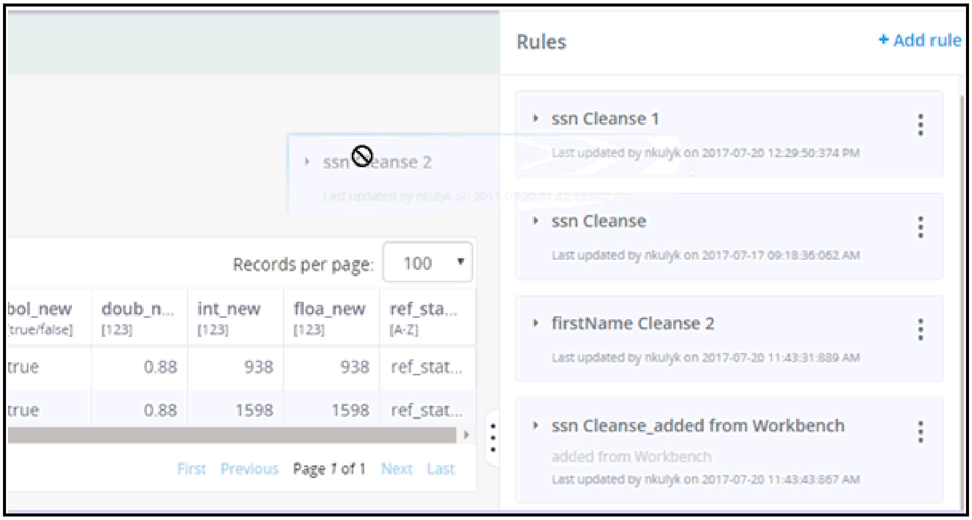
Reordering rules or operations allows you to change their positions in order to execute them in another sequence.

To move the rule to another section, drag and drop the rule to the section you wish. An icon appears for the rule that you are dragging, as shown in the following image.



Rules can be moved to any location of the Rules panel except for external rules where they are always the first on top.

Rules cannot be dropped outside of the Rule panel, but can be expanded and collapsed, as shown in the following image.

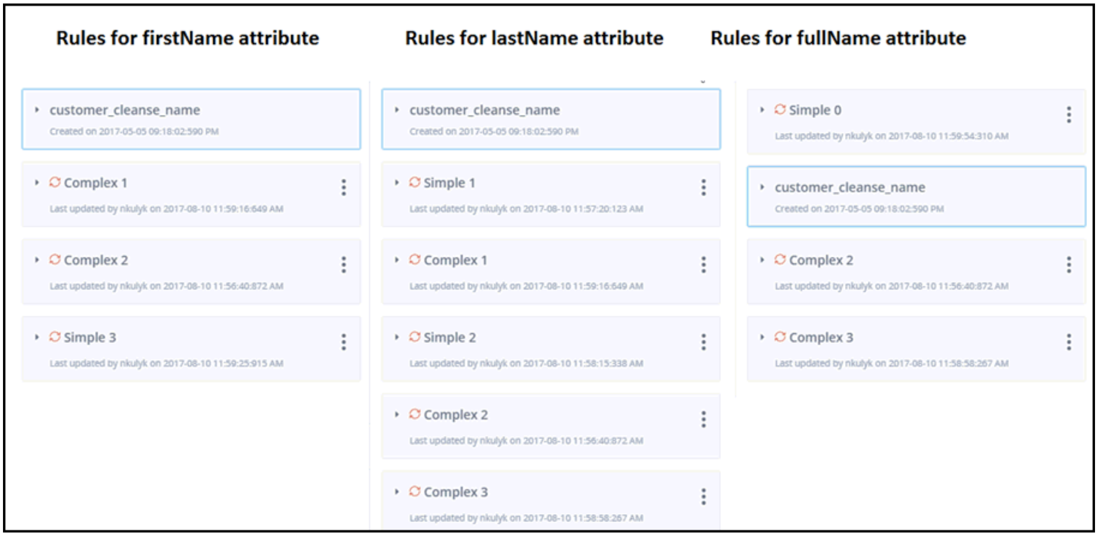


Reordering rules with simple operations will not affect the Rules list for other columns since simple operations are related to single attributes. However, when rules contain multicolumn operations that are related to several attributes, reordering multicolumn rules will impact the order of rules of those attributes.

When reordering multicolumn rules containing multi-column operations, you must remember the following principles:

- Rules containing multicolumn operations are moved below the rule with simple operations. It will not affect the position of the rule in other columns.
- Rules containing multicolumn operations are moved below the rule with complex operations and will affect the position of the rule in other columns.
- Rules containing multicolumn operations are moved below the rule with multicolumn operations and will affect the position of the rule in other columns.

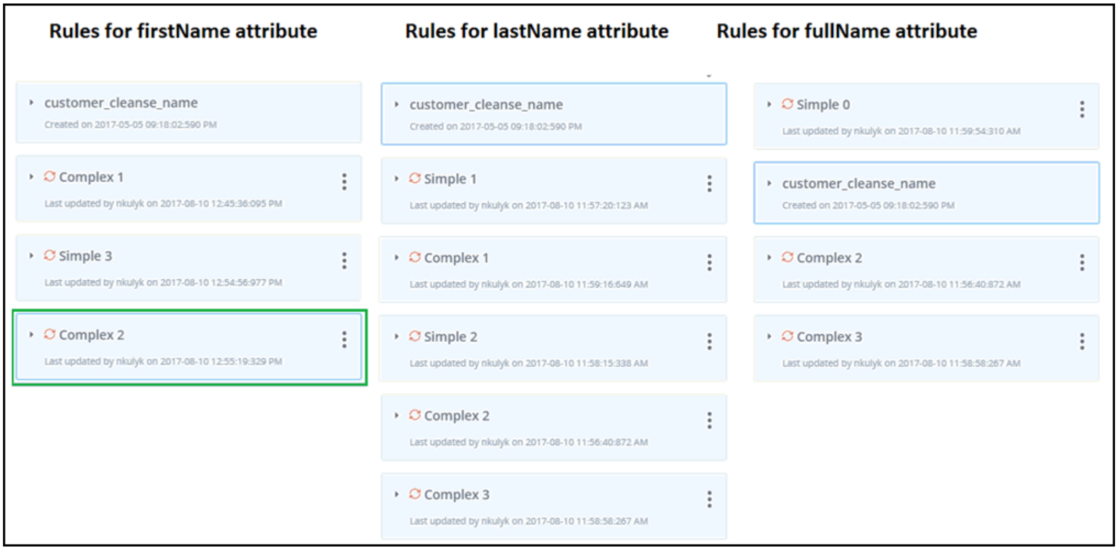
The following image shows three attributes prior to being reordered.



Example 1

In this example, in the firstName attribute column, the Complex 2 rule (containing a complex operation) is dragged to the bottom, below Simple 3 (containing a simple operation), as shown in the following image.

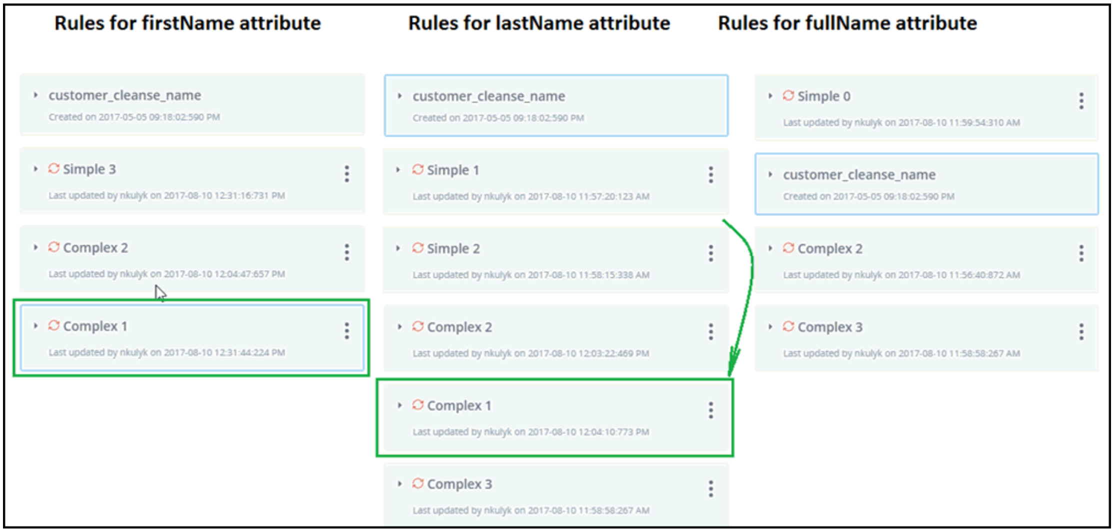
As a result, only the firstName column is affected, as shown in the following image.



Example 2

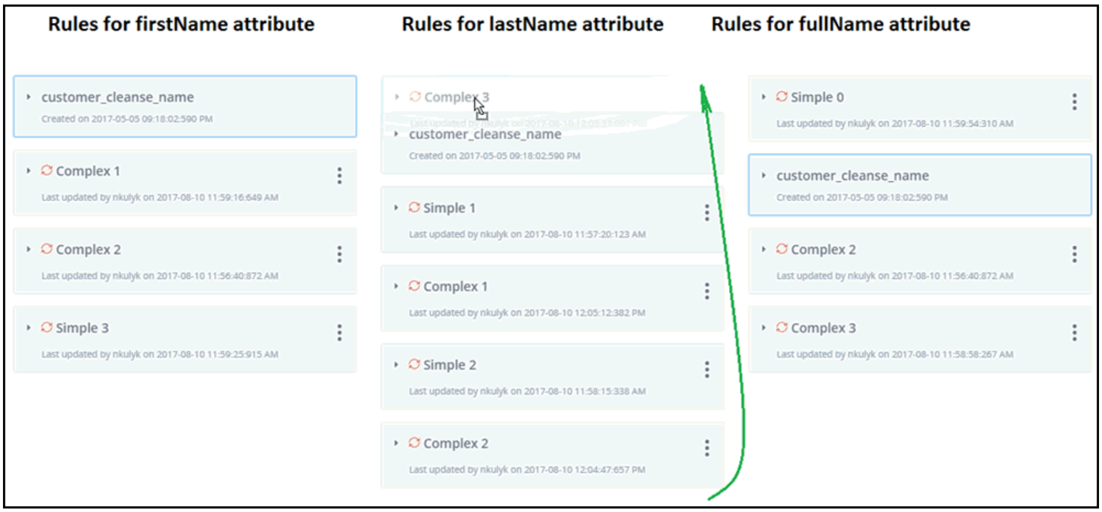
In this example, in the firstName attribute column, the multicolumn operation Complex 1 is dragged to the bottom, below a Simple 3 (containing a simple operation) and Complex 2 (containing a multicolumn operation), as shown in the following image.

As a result, the firstName and lastName attribute columns are affected, as shown in the following image.

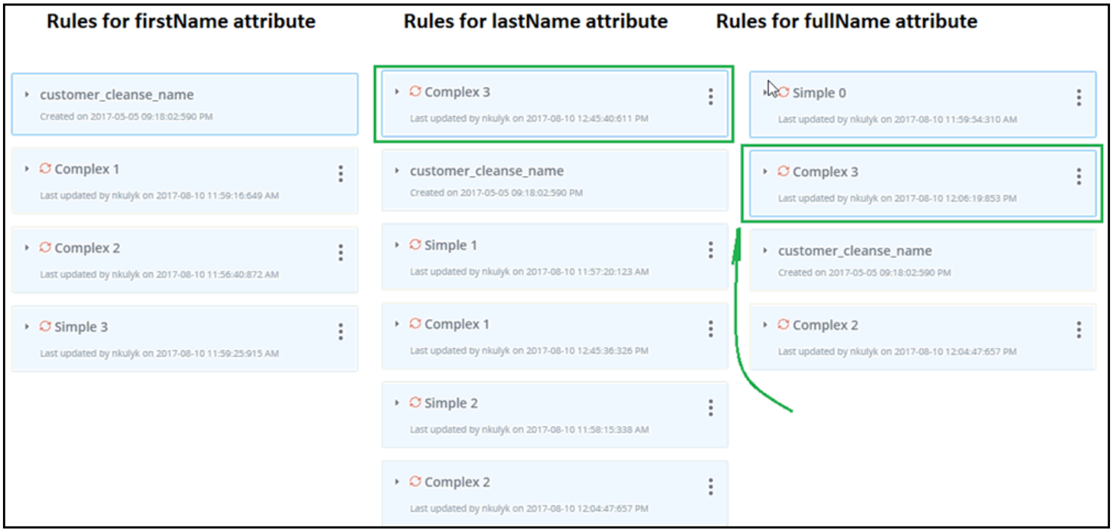


Example 3

In this example, in the lastName attribute column, the multicolumn operation Complex 3 is dragged to the bottom below all other rules containing simple and multicolumn operations, as shown in the following image.



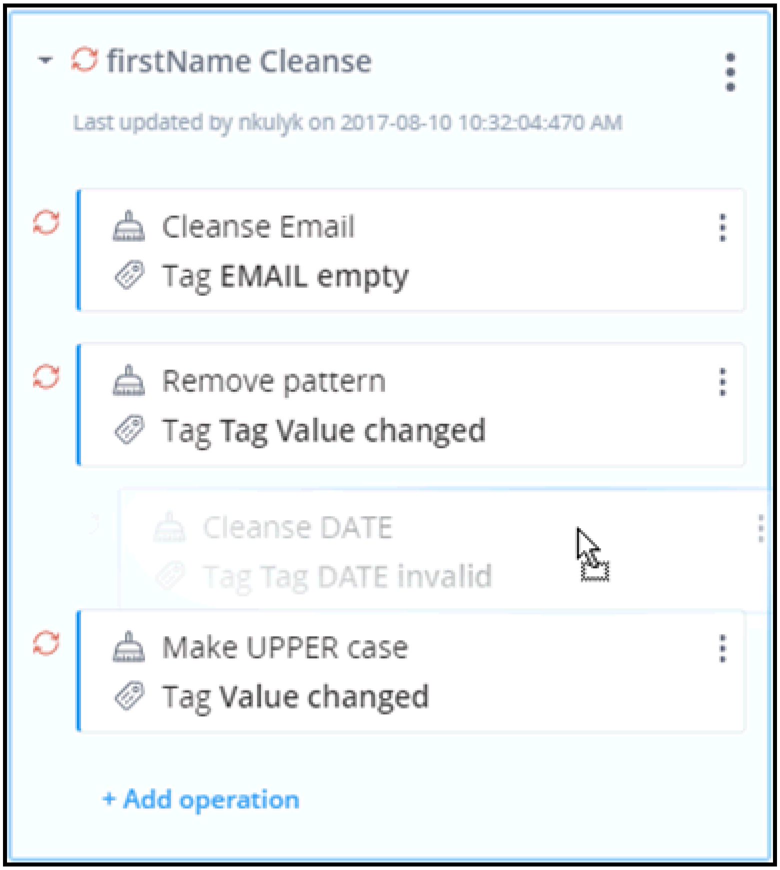
As a result, the lastName and fullName attribute columns are affected, as shown in the following image.



Reordering Operations

You can move an operation to a different location by dragging and dropping it to the location you wish.

An icon appears for the operation you are dragging, as shown in the following image.

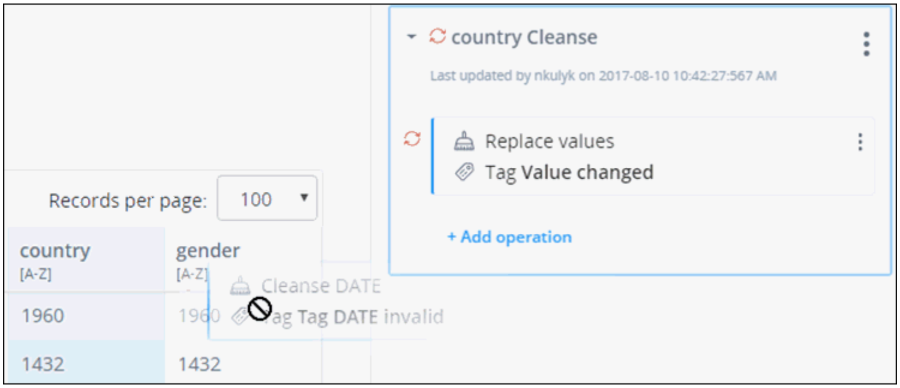


Only collapsed operations can be moved.

Operations can also be moved to any location in the Rules panel. Operations cannot be dropped outside of the Rules panel.

Operations cannot be moved if another operation inside the rule is opened in configuration mode.

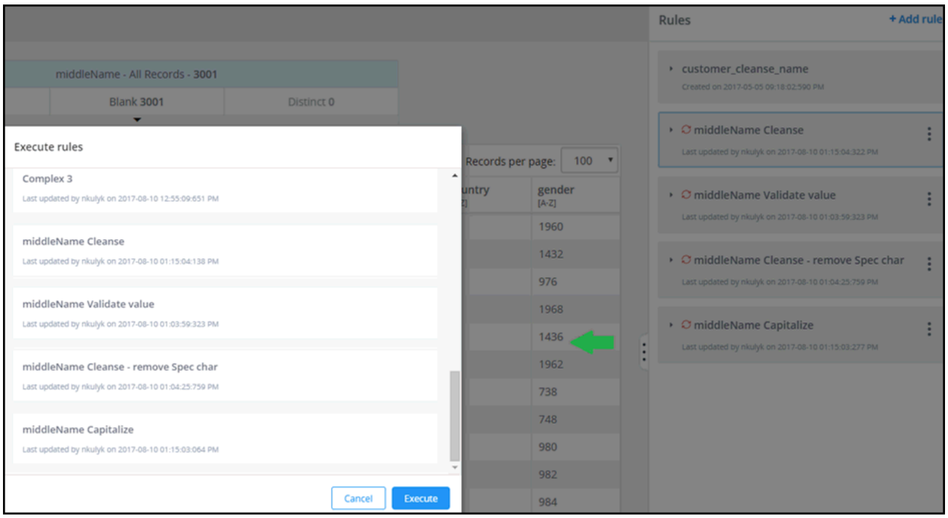
Multicolumn operations cannot be moved at all, as shown in the following image.



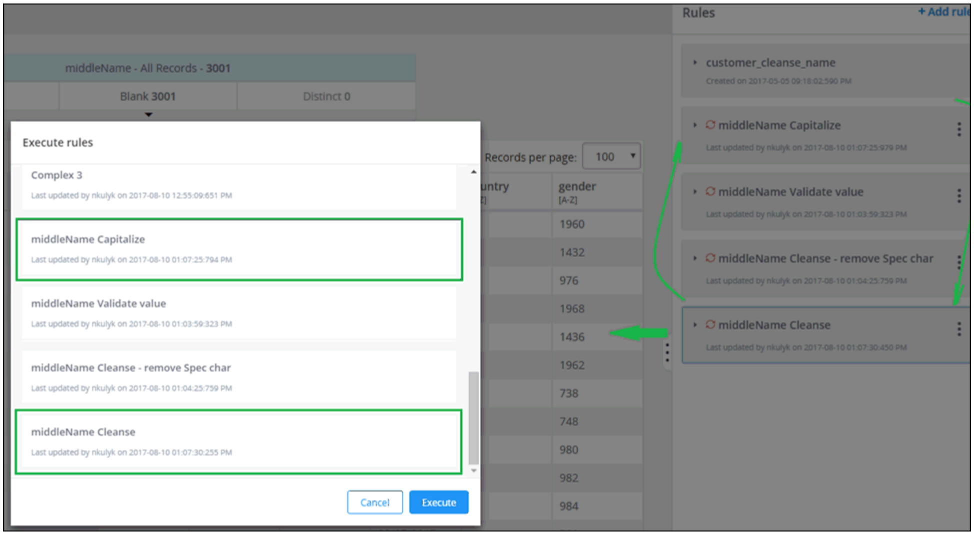
Executing Rules and Operations

After the rules or operations have been reordered in the Rules panel, their order will also be changed in the Execute rules dialog, and the rules will be executed in the new order and operation they appear in.

The following image shows the Execute rules dialog prior to reordering the rules.



The following image shows the new order of Rules in the Execute rules dialog after the rules have been reordered in the Rules panel.

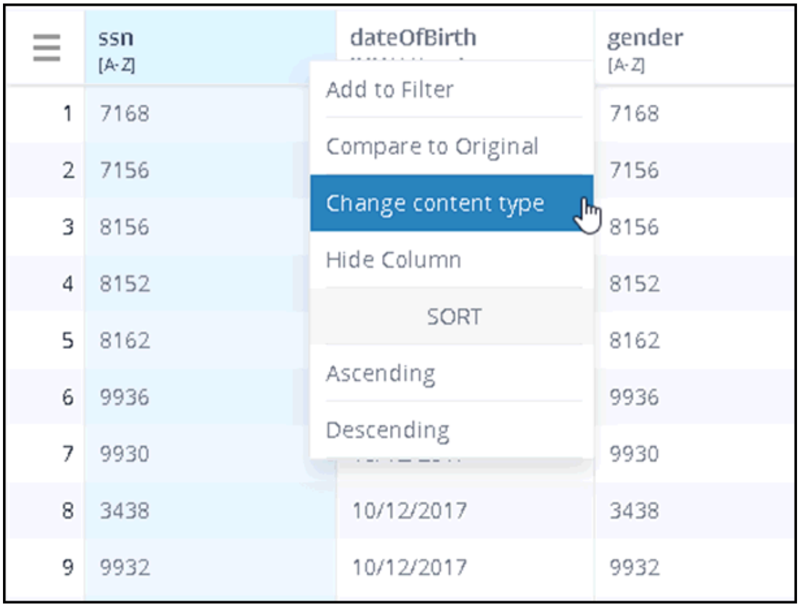


Changing Content Types

Changing content types allows you to create alternative types of data related to attributes. You can use custom data types to relate to an attribute, operation, and rule that will help you better understand your work and data involved.

For example, if you use the SSN attribute often, you can create corresponding content types that will be related to that attribute and as well as others.

You can add and change content types from the content menu by clicking on *Change content type*, as shown in the following image.

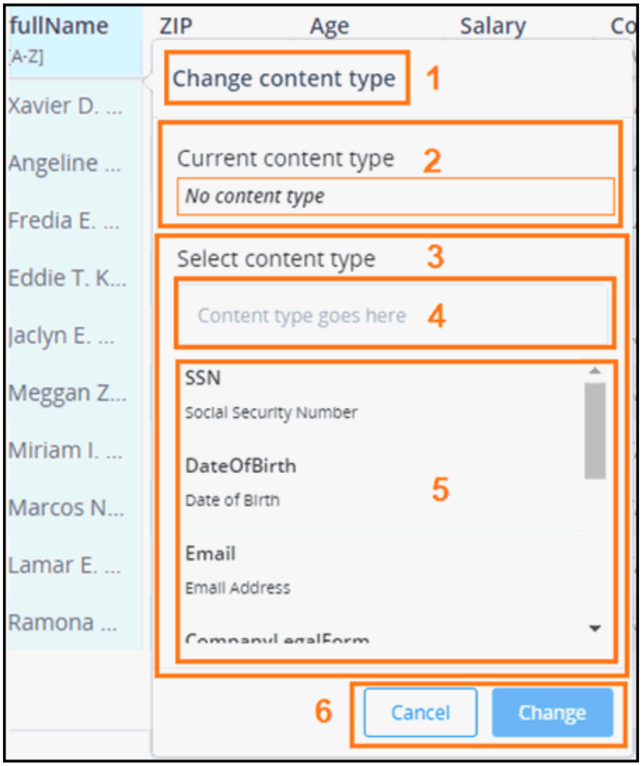


The Change content type dialog appears where you can select and change content types.

The following list corresponds to the image below.

1. Change content type header
2. Current content type section
3. Select content type section
4. Search field
5. Content type list

6. Cancel/Change buttons

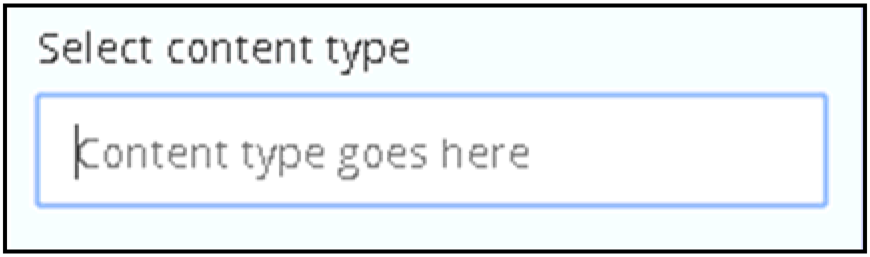


The Current content type section displays information regarding the current content type assigned to the attribute. If a content type is added, it will be shown in the grid header, as shown in the following image.

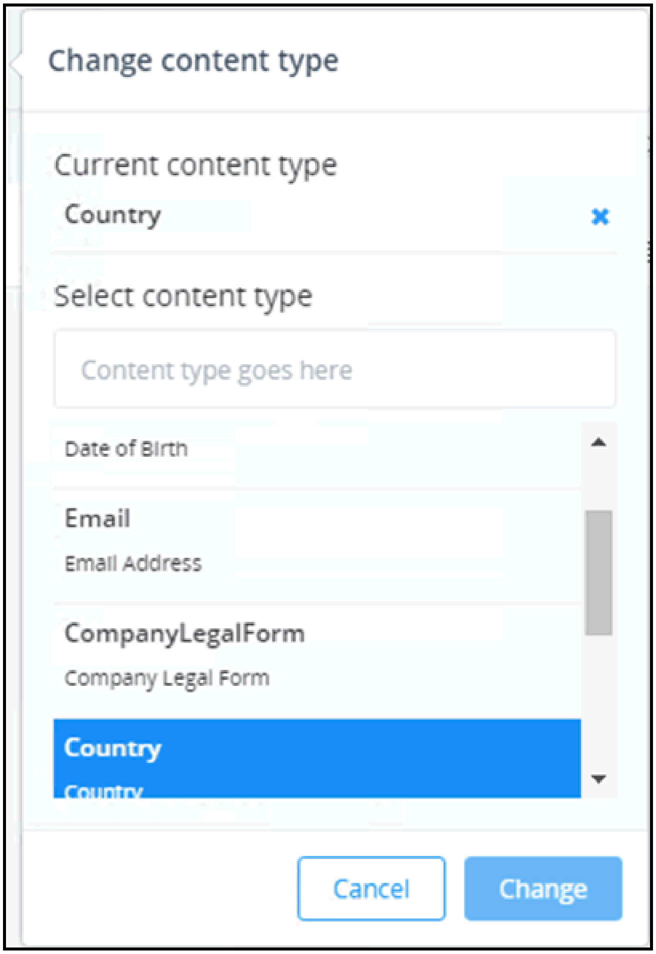
If there is no content type that is added, then there will be no content type in the grid header and the following message will appear in the Current content type section for the Change content type dialog:

```
No content type
```

The Select content type section contains information regarding all available content types, as swell as a search field for searching content types, as shown in the following image.

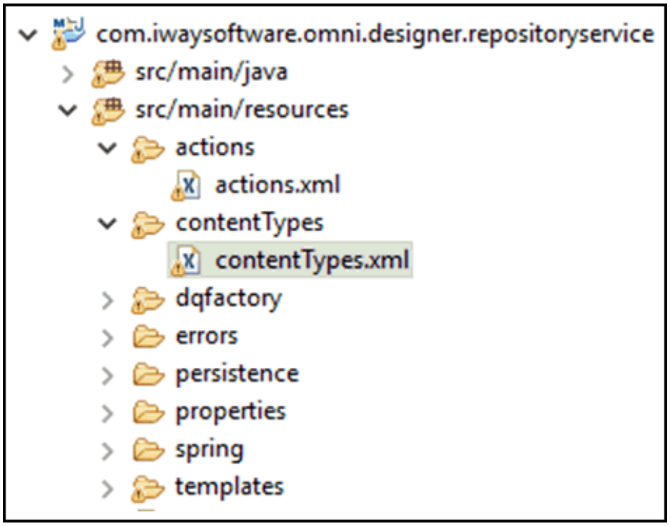


The available content types appear in the content type list, as shown in the following image.



The content type name and description appears for each item in the list.

The list of available content types is defined by the content_types.xml file located inside the Repository Service resources, as shown in the following image.



The following image shows a sample syntax structure of the XML file.

```
<?xml version="1.0" encoding="UTF-8"?>
<conetntTypes>
    <contentType name="SSN" description="A String Social Security Number" fieldType="String" />
    <contentType name="SSN" description="A Long Social Security Number" fieldType="Long" />
    <contentType name="Address" description="A common Address content type" />
</conetntTypes>
```

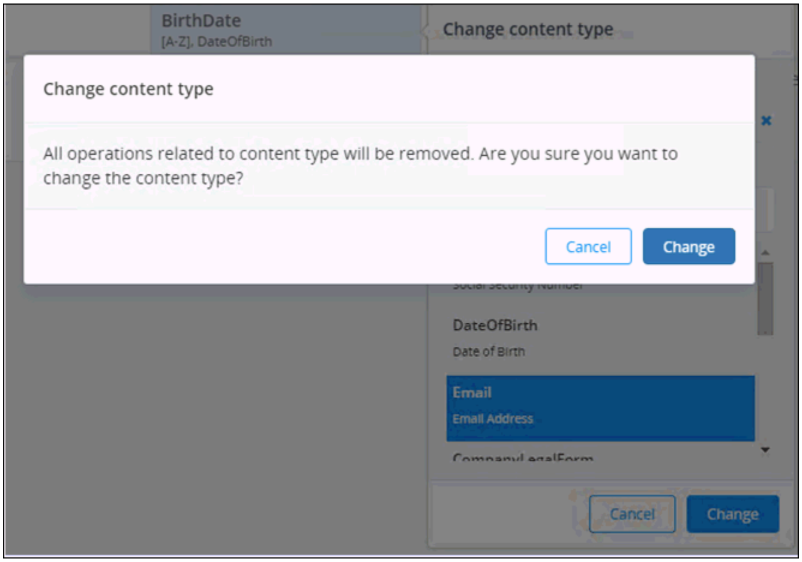
Notes:

- The *name* and *description* set in the content_type.xml file above will appear as the name of the content type in OmniDesigner and Workbench user interface.
- *fieldType* (if specified) is a Content type that will be available only for the given attribute type in the Personal Edition user interface. If it is not specified, then it is available for all attribute types.

After selecting the content type from the list, it will be set in the Current content type section.

Click *Change* to assign it to a column and have it appear in the column header.

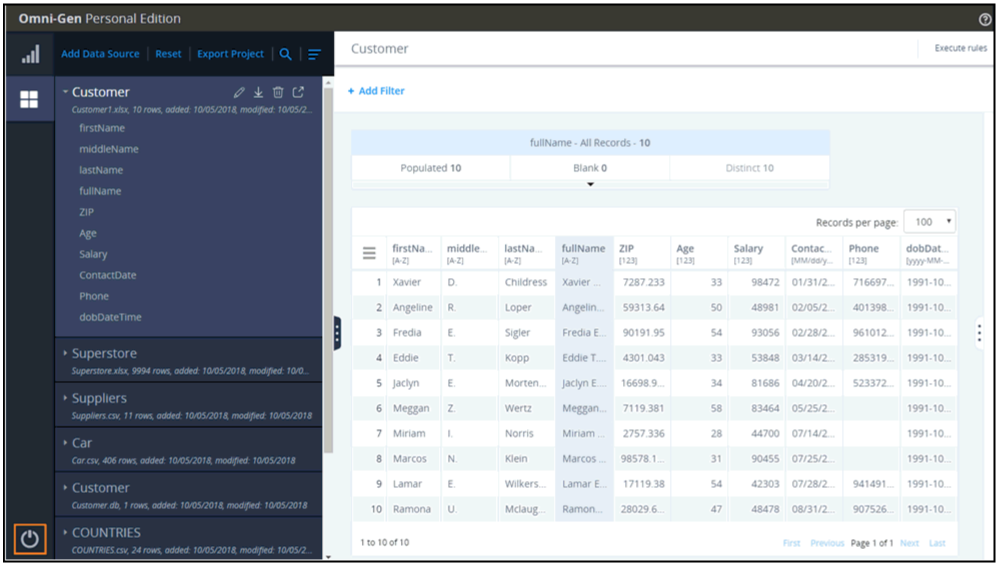
If you change one content type to another type in a specific column, a confirmation message appears, as shown in the following image.



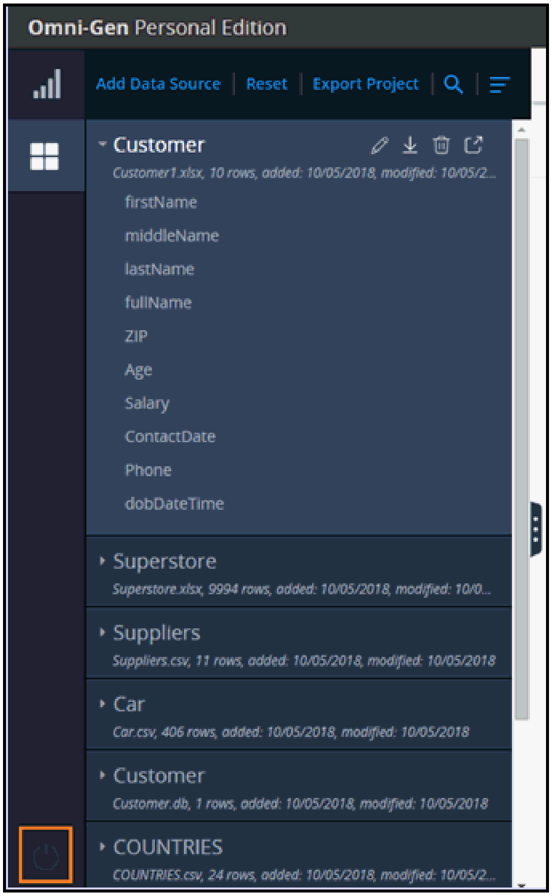
Resetting Your Session

The Reset Session button allows you to return the Personal Edition data to the default state without removing all the rules you have added.

The Reset Session button is located at the bottom left corner of the Profile and Data Grid pages, as shown in the following image.



If you first navigate to the workbench and no changes have been made, or if the Reset Session functionality was already performed, then the Reset Session button will be disabled, as shown in the following image.

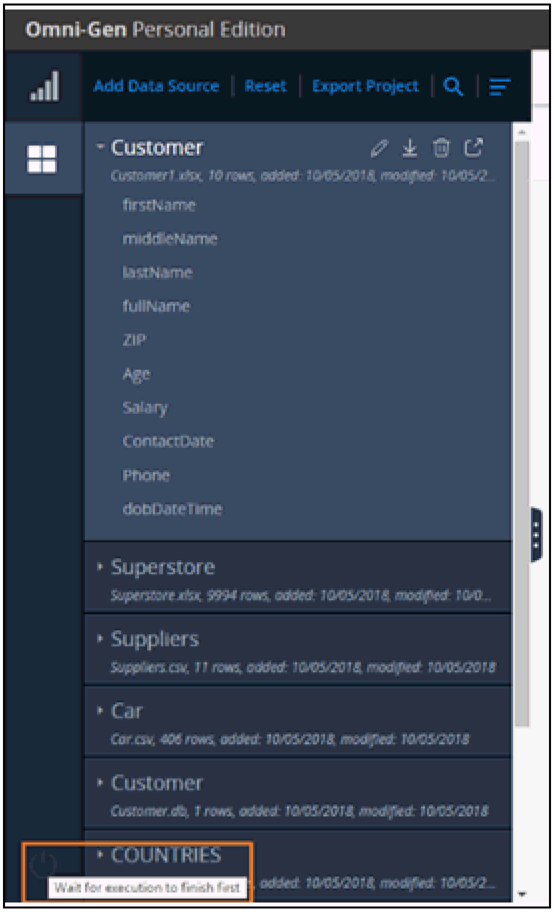


The Reset Session button will be available for use after adding a rule or filter.

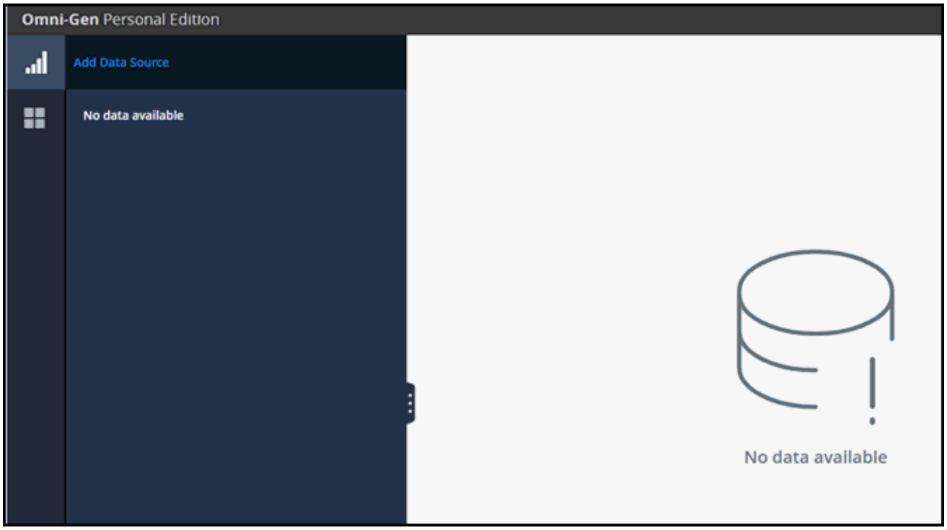
A confirmation dialog appears after clicking the Reset Session button, as shown in the following image.

After resetting a session, the option page will be reloaded and all the added filters and rules that were created or executed will be removed. Additionally, the Personal Edition data will return to its default state.

However, the ability to reset a session is not always available. You cannot reset a session while the rules are processing. If you hover the cursor over the Reset Session button during a rules execution, a tooltip appears informing you to wait for the execution to finish, as shown in the following image.



If no data is available, the Reset Session button will also be disabled, as shown in the following image.



Feedback

Customer success is our top priority. Connect with us today!

Information Builders Technical Content Management team is comprised of many talented individuals who work together to design and deliver quality technical documentation products. Your feedback supports our ongoing efforts!

You can also preview new innovations to get an early look at new content products and services. Your participation helps us create great experiences for every customer.

To send us feedback or make a connection, contact Sarah Buccellato, Technical Editor, Technical Content Management at *Sarah_Buccellato@ibi.com.*

To request permission to repurpose copyrighted material, please contact Frances Gambino, Vice President, Technical Content Management at *Frances_Gambino@ibi.com.*

Information Builders

iWay

Omni-Gen™ Personal Edition Getting Started Guide

Release 3.9

DN3502343.0719

Information Builders, Inc.
Two Penn Plaza
New York, NY 10121-2898