DATA PROFILING
USER GUIDE

STEP Trailblazer 8.1

Table of Contents

Table of Contents 2

Data Profiling 4

Object Completeness 5

View Completeness Scores 5

User Privileges 7

Define Completeness Scores 7

Setting Completeness Scores for Data Containers 8

Export Completeness Values 10

Setting Up Completeness Metrics 10

Creating and Editing Completeness Metrics 12

Create a Completeness Metric 12

Set a Default Completeness Metric 15

Migrating Completeness Scores to Description Attributes 15

Migrate Completeness Scores 15

Object Completeness in Web UI 18

Adding the Completeness Meter Component 18

Configuring the Completeness Meter Component 18

Data Profiles 20

View Data Profiles 20

Enable Profiling on an Object 21

Generating, Updating, and Scheduling a Data Profile 22

Generating / Updating a Data Profile 22

Scheduling Data Profiling 23

Data Profile Dashboard 29

Accessing the Dashboard Configuration 30

Navigating the Dashboard Configuration 30

Configuring Dashboard Widgets 31

Saving the Configuration 32

Available Widgets 32

Value Details View 39

Accessing and Navigating Value Details View 39

Attribute Table Details 42

Value Details Tabs 43

Reference Details View 47

Accessing and Navigating Reference Details View 47

Reference Table Details 48

Reference Details Tabs 49

Widgets on the Overview tab. 50

Profile Configuration 51

Accessing the Profile Configuration 51

Select Profile Filter Business Condition .. 51

Include / Exclude Data in Attribute Groups 53

Select Business Conditions to Test During Profiling 54

Value Profile Configuration 56

Select Completeness Metrics to Profile ... 57

Global Dashboard 59

Displaying the Global Dashboard 59

Workflow Profile Widgets 61

Limiting Memory Usage During Profiling 63
Attribute Value Profiles 64
View an Attribute Value Profile 65
Config Property 65
Attribute Value Profile Tabs 66
Actions 66
The Overview Tab 66
The Frequent Values Tab 67
The Rare Values Tab 67
The Frequent Patterns Tab 67
The Rare Patterns Tab 68
The Attribute Usage Tab 68

Data Profiling

STEP offers a set of tools dedicated to measuring and improving the quality of data in the system:

- **Data Profiles** - The data profile functionality provides detailed overviews of data in specific branches of the hierarchy in Tree. Each profile contains information about the profiled data and provides easy access to correcting data errors.
- **Object Completeness** - Object completeness indicators provide information about how complete an object is based on scores users set for attributes, reference types, and product to classification link types.
- **Attribute Value Profiles** - Attribute value profiles enable users to see where and how an attribute is used. The profile provides an overview of how attribute values are distributed across the current context and workspace, and enables users to easily clean up data.

Some features are privilege controlled. For more information see **Action Sets** in the System Setup documentation.

Object Completeness

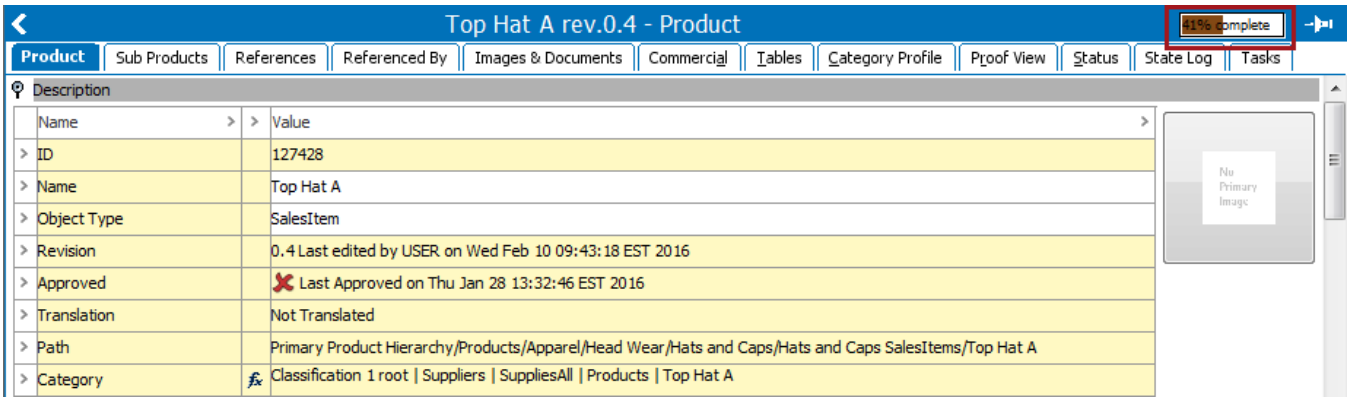
Object completeness indicators provide information about how complete an object is based on scores that you set for attributes, reference types, data containers, and product to classification link types. While it is possible to calculate a completeness percentage without settings scores for these individual data objects, it is recommended to do so, as you can weigh them based on relevance and outright omit those that should not apply to the overall completeness percentage.

Note: The object completeness functionality is not available on product overrides.

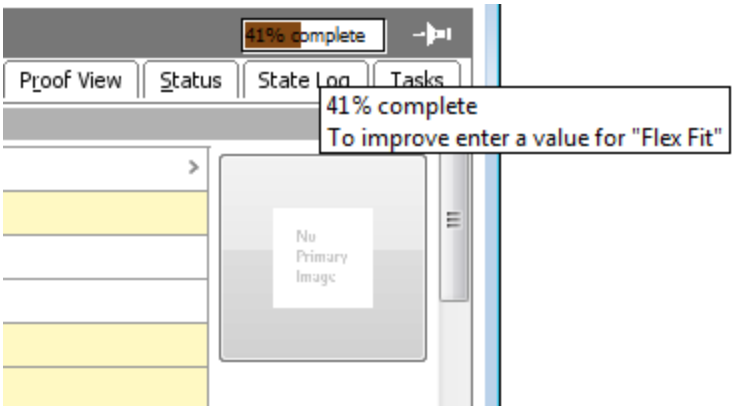
View Completeness Scores

1. In the **Tree**, expand the applicable Hierarchy, and then locate the relevant object.
2. Click the main object tab or the **References** tab. The **Completeness Meter** in the upper right corner displays the current status of the object expressed as a percentage.

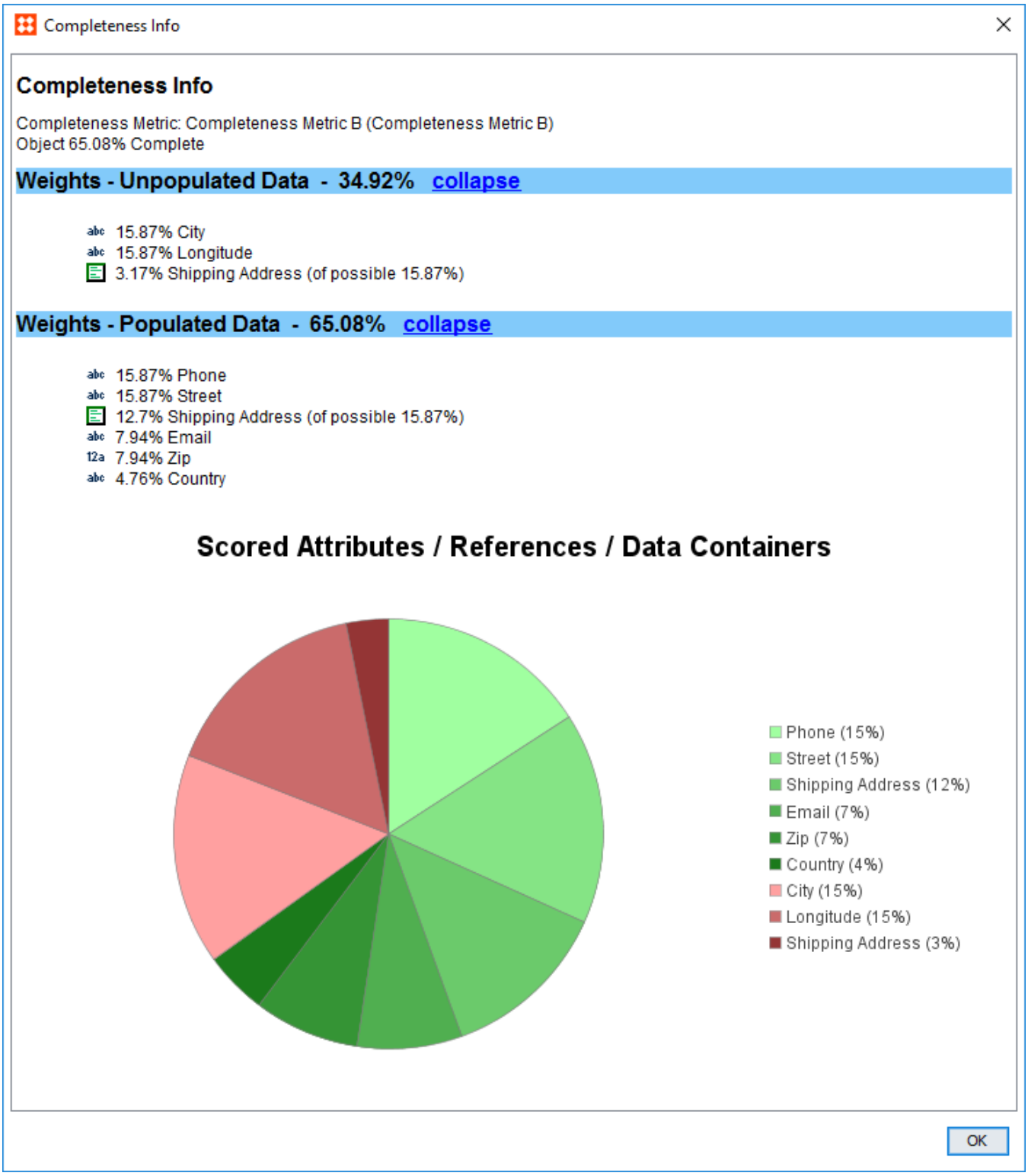
Note: If no indicator appears it is likely that no completeness scores have been set for the data valid for the object.



3. Hover the mouse over the **Completeness Meter** to view a text describing what you can do to improve the completeness percentage.



4. To get a more detailed view of how the completeness percentage for the selected object is calculated, double-click the **Completeness Meter**. A **Completeness Info** window appears, which displays all attributes, data containers, and referenced objects that contribute to the completeness percentage and how they are weighted.



Object completeness indicators are also available in Web UI. See the **Object Completeness in Web UI** section of the **Data Profiling** documentation.

User Privileges

The completeness of an object is always calculated and displayed as if the user had privileges to view and edit all relevant data. However, hints are only provided for data that the user is allowed to edit.

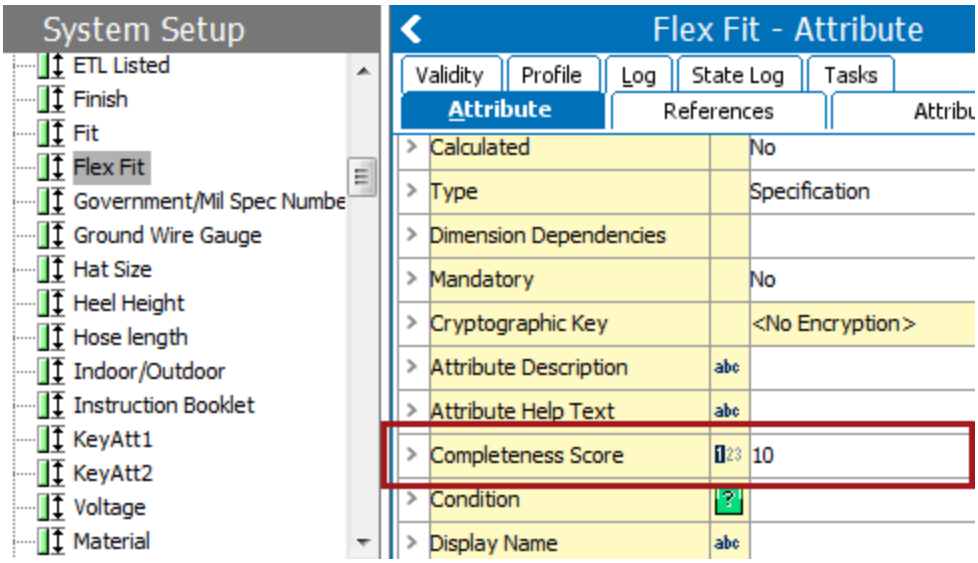
Define Completeness Scores

For a typical setup, in order to view completeness information, completeness scores must be configured on all relevant attributes, references, data containers, and product to classification link types. This score determines how much to weigh an element of the object when factoring the object's completeness percentage.

Note: For product reference types, images and document reference types, and product to classification link types, the scores are defined on the individual type's main editor.

---

1. In **System Setup** expand **Reference Types / Attribute Groups**, and then locate the relevant reference type / attribute.
2. In the **Completeness Score** field, under the **Description** section, enter the preferred score. Scores must be integers.



Note: You cannot specify scores for calculated attributes.

Alternatively, when setting the score on the link between a product or classification and a specification attribute, the completeness score of an attribute can be specified on the actual object in the **Tree**.

1. In the **Tree**, expand the applicable hierarchy, and then locate the desired object.
2. On the **References** tab, go to **Linked Attributes** and then click the flipper icon to expand the table.
3. In the **Completeness Score** column, enter the preferred score(s) for the relevant attribute links.

Linked Attributes from Product Hierarchy

	DisplaySequence	ID	Name	Completeness Score
		DescriptionWeb	Description, Web	
	2	DescSource	Description Source	
	2	DescTarg	Description Target	
		Disc Dimensions	Disc Dimensions	10

Setting Completeness Scores for Data Containers

Unlike other data objects that contribute to the object's completeness percentage, a data container can be incomplete itself, making it possible to contribute only a fraction of its maximum score. The weight of the total possible score of the data container is set on the data container type (just like for attributes and reference types), and any values absent from the individual attributes that comprise the data container are reflected in its score. Additionally, the attributes within a data container can have their own varying scores. In a sense, a data container has its own completeness percentage, and contributes either 100% of its score to the object's completeness percentage, or, if one or more attributes within are missing values, it contributes a fraction of its total possible score.

Completeness Info

Completeness Metric: Completeness Metric B (Completeness Metric B)
Object 11.11% Complete

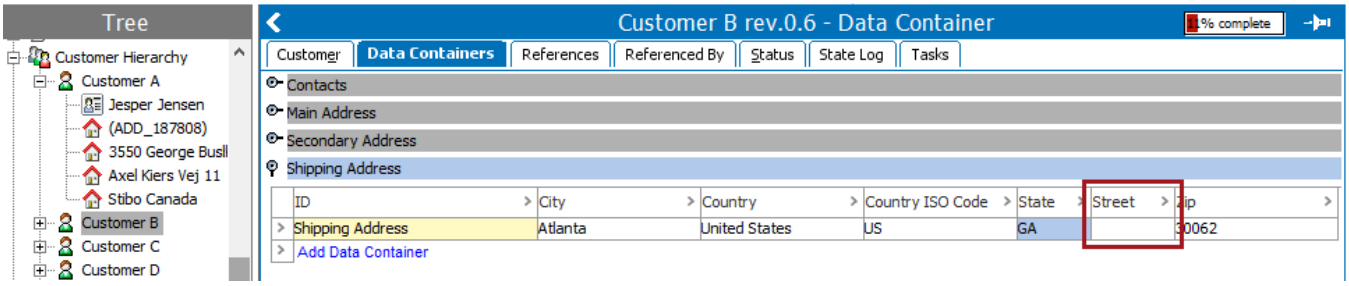
Weights - Unpopulated Data - 88.89% collapse

- abc 15.87% City
- abc 15.87% Longitude
- abc 15.87% Phone
- abc 15.87% Street
- abc 7.94% Email
- 12a 7.94% Zip
- 4.76% Shipping Address (of possible 15.87%)
- abc 4.76% Country

Weights - Populated Data - 11.11% collapse

- 11.11% Shipping Address (of possible 15.87%)

In the above example, the data container 'Shipping Address' can contribute a total of 15.87% completeness to the object's overall completeness percentage. However, because it has attributes with missing values, it is only contributing 11.11% completeness; roughly 70% of its maximum score.

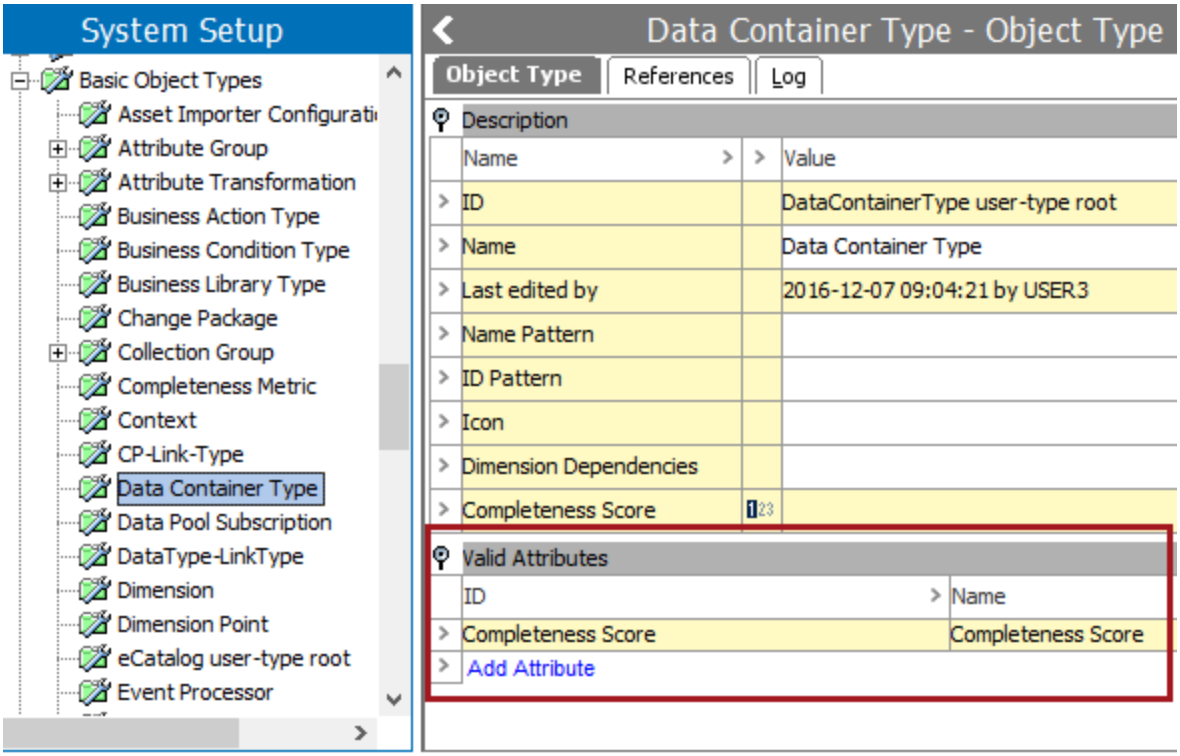


By viewing the data container itself (pictured above), you can see that the attribute 'Street' is missing a value, which explains why the data container is not contributing 100% of its score. Had all six attributes been weighted with a neutral score, each would contribute roughly 17% to the data container's score. However, based on the completeness breakdown, we know that the attribute 'State' is responsible for roughly 30% of the score. This is because 'State' has a higher weighted score than some of the other attributes.

Note: If a second 'Shipping Address' instance were added, its attributes would be factored in with those of the existing 'Shipping Address' when calculating the score, as they are both held within the same data container.

As with all attributes, the scores for attributes that comprise a data container are defined on the Completeness Score metadata field found on the attributes. The maximum score for the data container itself is defined by the same field on the data container type ('Shipping Address', in this example).

Before completeness scores can be applied to data containers, however, ensure that the Completeness Score attribute is valid for the Data Container Type object type.



Export Completeness Values

Completeness values can be exported to Excel and CSV.

1. Create a calculated attribute and enter the function "completeness()" in the Value template.

   For more information, see the **Calculated Attributes** section of the **System Setup / STEP Super User Guide** documentation.

2. From the **File** menu, hover over **Export** and then click **Data...**.
3. In **Step 4 - Map Data** of the Export Manager wizard, add **Select Attribute** to columns. The **Select Attribute or Attribute Group** dialog is displayed. Search or browse for the desired calculated attribute.
4. Select the calculated attribute, check **Force calculation**, and then continue with the export.

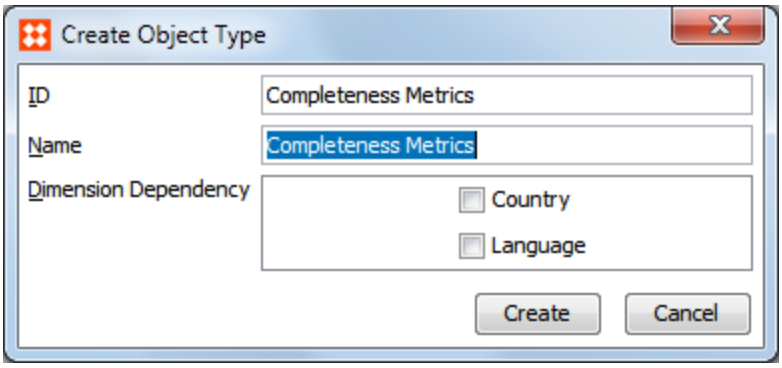
   For more information, see the **Exporting Data** section of the **Export Manager** documentation.

Setting Up Completeness Metrics

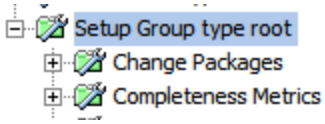
Completeness Metrics define what data is considered when evaluating the completeness of an object. Before completeness metrics can be created however, a Setup Group must be configured to hold them. Additionally, a Setup Group is also required if the user wants to migrate completeness scores to a Description attribute.

For more information, see the **Migrating Completeness Scores to Description Attributes** section of the **Data Profiling** documentation.

1. In **System Setup**, expand **Object Types & Structures**.
2. Right-click **Setup Group type root**, and then choose **New Object Type**.
3. Enter an **ID**, and a **Name** such as Completeness Metric, and then click **Create**.

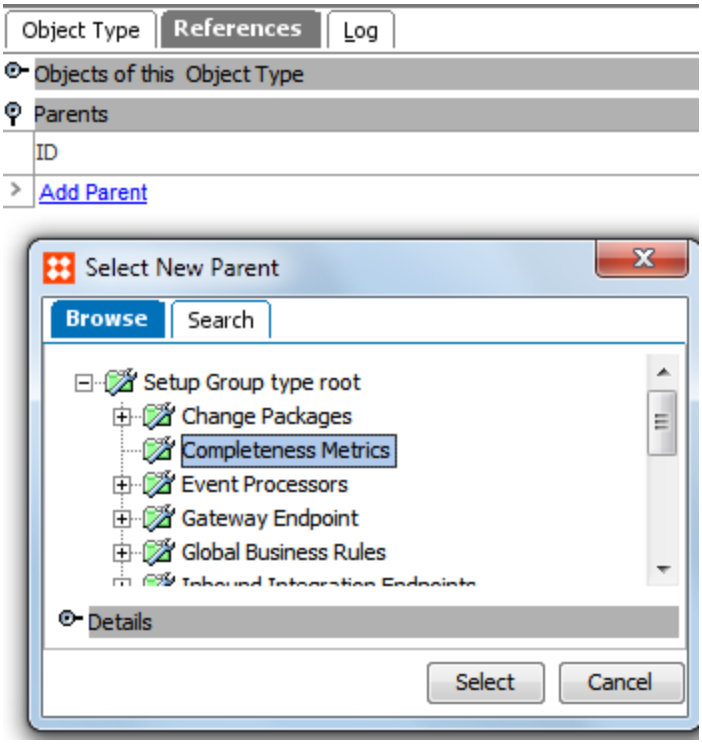


The **Setup Group** appears as a child below **Setup Group type root**.



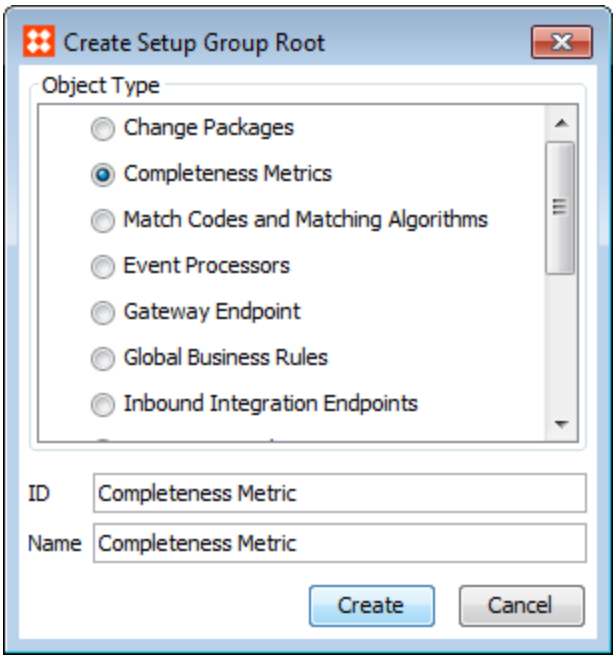
Next, you have to link the Completeness Metric object type to a Setup Group.

4. In **Object Types & Structures**, expand **Basic Object Types**, then select **Completeness Metric**.
5. On the **References** tab, click **Add Parent**.
6. In the **Select New Parent** dialog, select the Setup Group you just created, then click **Select** to make it a valid parent.



Next, you have to create an instance of the Completeness Metric object.

7. On the **System Setup** tab, click **Maintain**, mouse over **Insert**, and then select **Setup Group Root**.
8. In the **Create Setup Group Root** dialog, select the object type you just created.



9. Enter an **ID** and a **Name**, then click **Create**.

A Setup Group is created as a node in the System Setup hierarchy, and you can start creating completeness metrics.

For more information about creating completeness metrics and how to work with them, see the **Creating and Editing Completeness Metrics** section of the **Data Profiling** documentation.

Creating and Editing Completeness Metrics

STEP allows the use of multiple completeness metrics simultaneously. This is useful, for example, if multiple channel specific completeness measurements are required. An attribute can be created to hold completeness scores for a specific completeness metric, thus offering a means to display different completeness measurements for the same object. Additionally, multiple metrics can use the same completeness score attribute.

Note: A completeness metric does not have to be an attribute with associated completeness scores. In these cases, all data objects (attributes and references) are part of the completeness calculation and have an equal score.

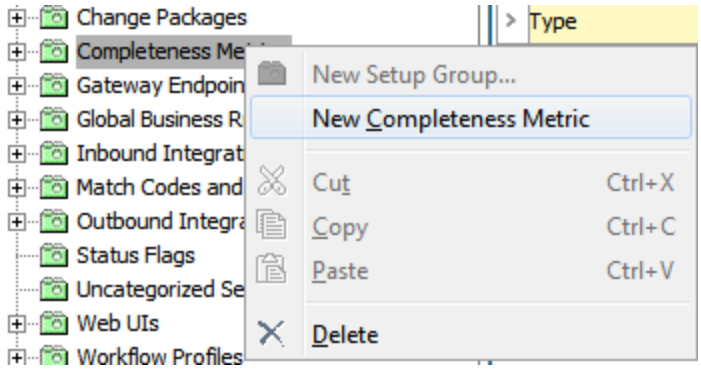
Create a Completeness Metric

1. Optionally, create an attribute that can hold the completeness score values.

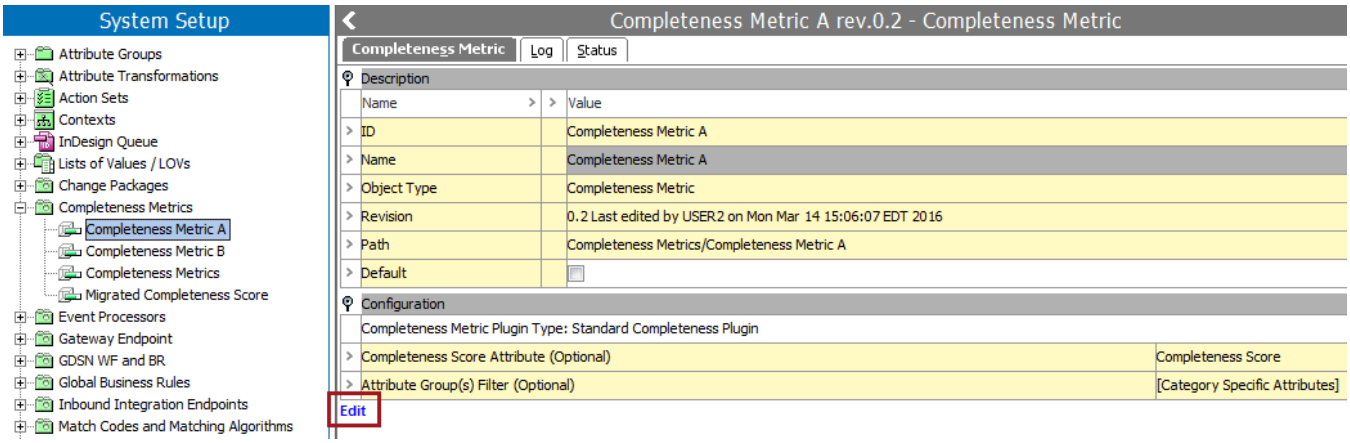
   Note that if you want to migrate legacy completeness values stored in a special system attribute to a Description attribute, the attribute must meet a series of requirements. For more information, see the **Migrating Completeness Scores to Description Attributes** section of the **Data Profiling** documentation.

   If you create an attribute that is not going to be used for migrating completeness scores, it does not have to meet these requirements.

2. In **System Setup**, right-click the **Setup Group Root** that holds completeness metrics, and then click **New Completeness Metric**.



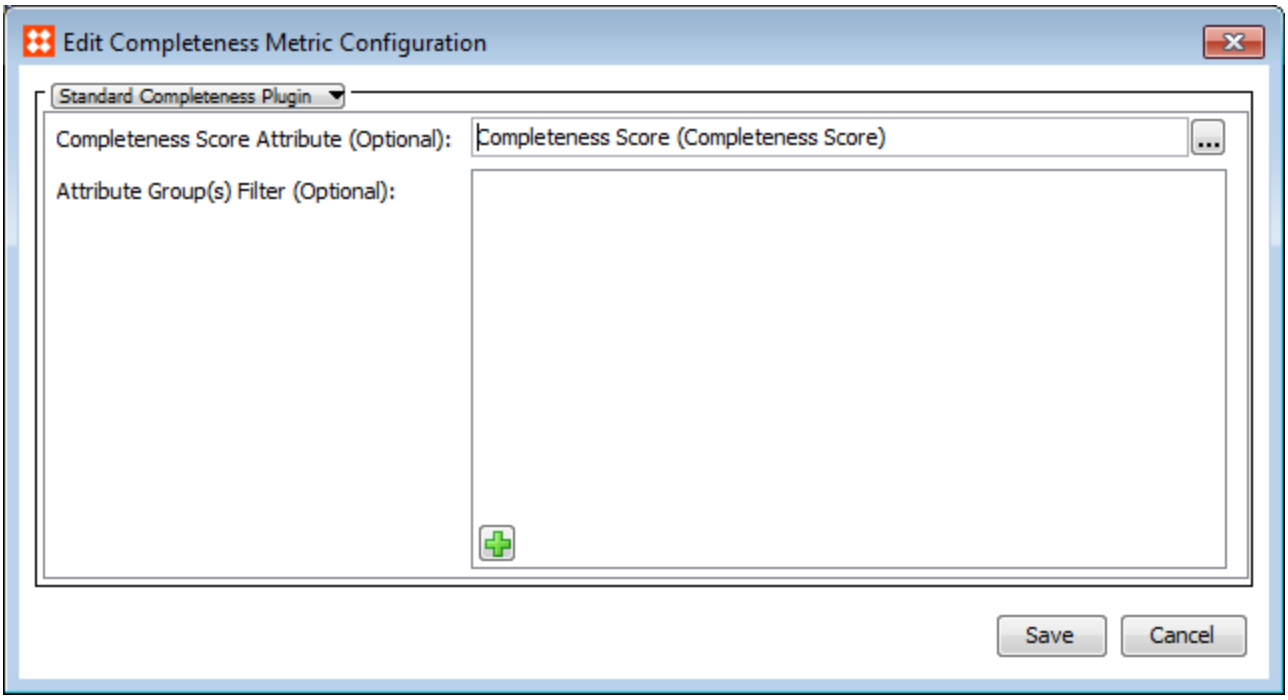
3. Enter an **ID** and a **Name** for the metric, and then click create. The completeness metric is added as a child below the selected **Setup Group Root**.
4. Select the metric that you just created, and then, on the **Completeness Metric** tab, click **Edit**.



5. In the **Edit Completeness Metric Configuration** editor, select the relevant completeness plug-in. The number of available plug-ins is system specific. A Standard Completeness Plugin is provided by Stibo Systems.

   The Completeness Score Attribute field allows you to select which attribute will hold the completeness scores for the completeness metric, if applicable. If nothing is specified here all elements of the object that could contribute to the overall object completeness score will, and they are all weighed evenly.

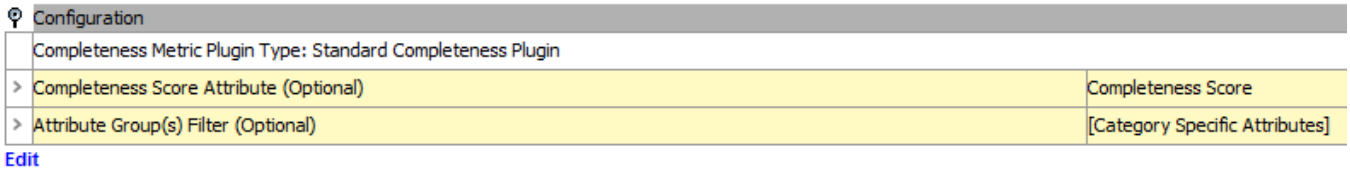
   The Attribute Group(s) Filter allows you to specify which attributes and references / links should be considered in the completeness calculation.



The Completeness Plugin provides the following options:

Selection	Description
No attribute or attribute groups is selected	All data objects (attributes and references) are part of the completeness calculation, and they have an equal score.
An attribute is selected	Only completeness scores from the selected attribute are used for the completeness calculation.
One or more attribute groups are selected	Only completeness scores from data objects in the selected attribute group(s) are used for the completeness calculation, and they have an equal score.
An attribute and one or more attribute groups are selected	Only the completeness score from the selected attribute is used for the completeness calculation.

On the **Completeness Metric** tab, in the **Configuration** area, you can view the configuration settings.



Note: If data containers should be considered for an entity's completeness score, the completeness metric must use default settings.

Set a Default Completeness Metric

When a completeness metric is set as the default metric, the completeness meter shown on the product and entity editor tabs in the Workbench display the value from the default completeness metric.

1. In **System Setup**, expand the **Setup Group Root** that holds completeness metrics, and then select the completeness metric you want to apply as default.
2. On the **Completeness** tab, in the **Description** area, select **Default**.

Migrating Completeness Scores to Description Attributes

Completeness scores are stored in a standard Description attribute. This means that all values can be stored as STEP XML and that completeness score setups can be moved easily between systems.

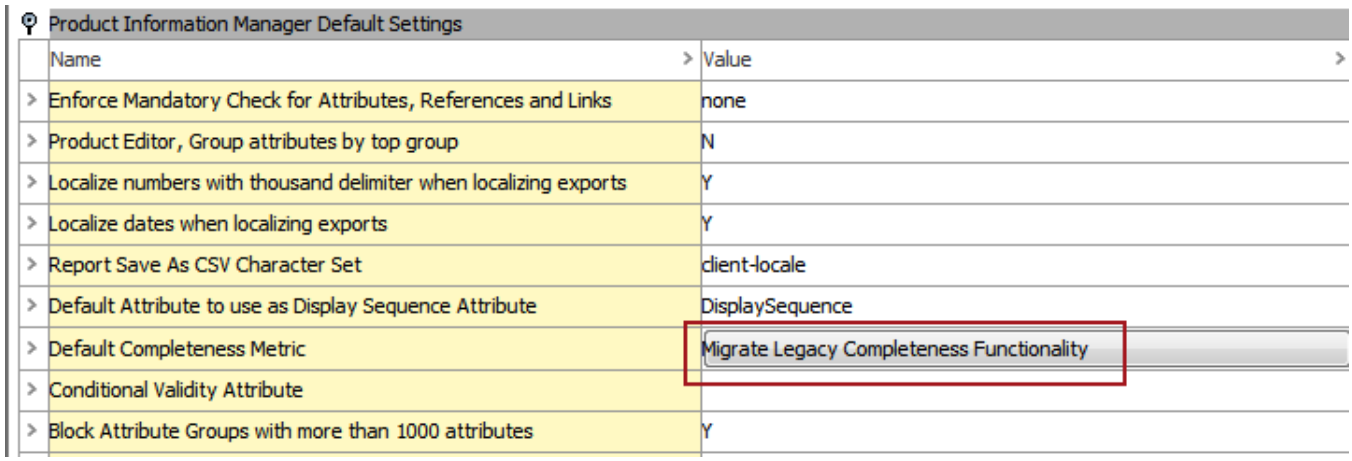
Before STEP Trailblazer 7.4, completeness score values were stored in a special system attribute. To migrate values from the system attribute to a Description attribute, use the **Migrate legacy completeness functionality** in **Users & Groups**.

Migrate Completeness Scores

The following describes the steps involved in migrating completeness scores to a Description attribute:

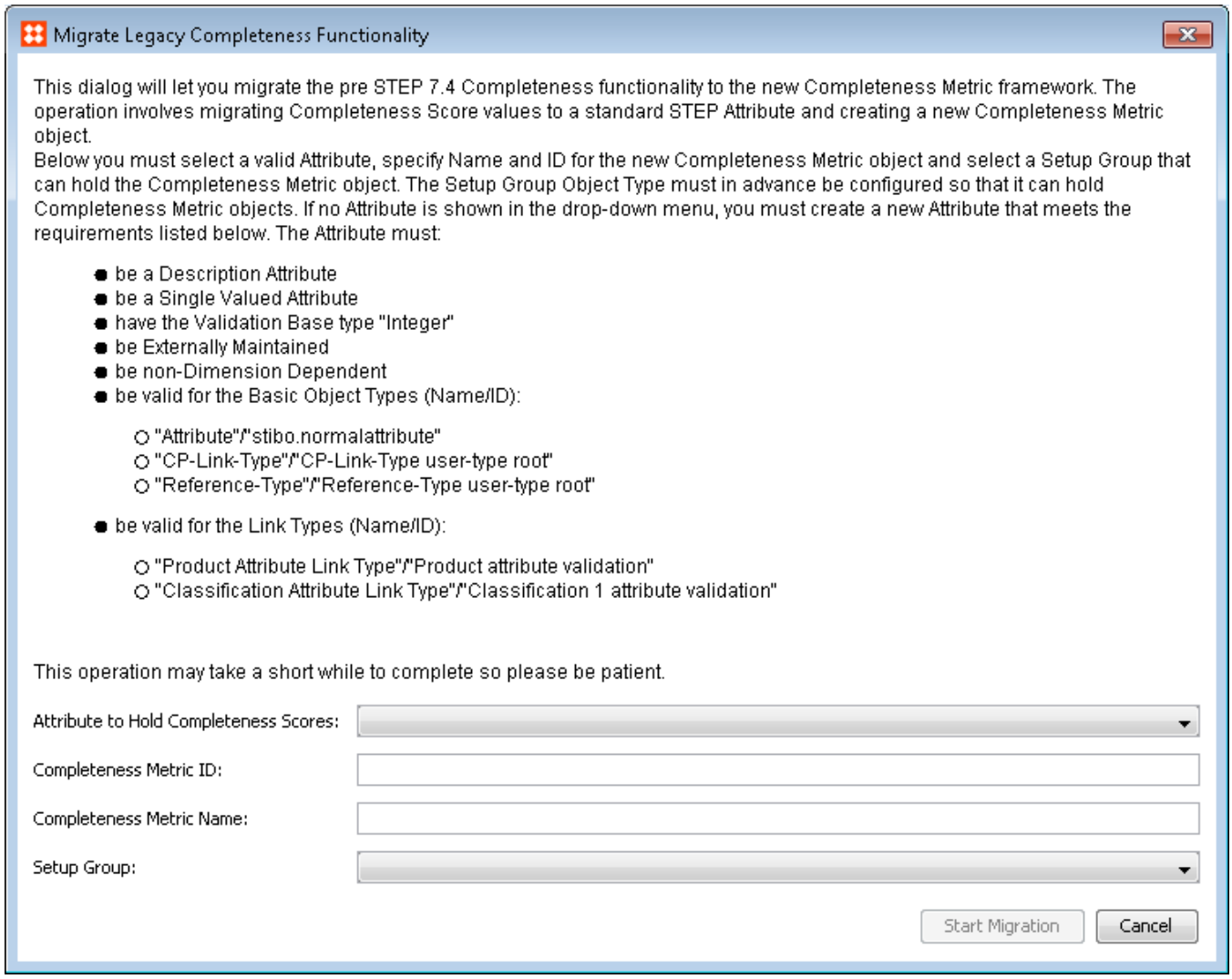
1. Create an attribute that can hold the completeness score values. The attribute must meet the following requirements:
   - It must be a Description attribute.
   - It must be a Single valued attribute.
   - It must have the Validation base type Integer.
   - It must be externally maintained.
   - It must not be dimension dependent.
   - It must be valid for the basic object type Attribute/stibo.normalattribute.
   - It must be valid for the basic object type CP-Link-Type/CP-Link-Type user-type root.
   - It must be valid for the basic object type Reference-Type/Reference-Type user-type root.
   - It must be valid for the link type Product Attribute Link Type/Product attribute validation.
   - It must be valid for the link type Classification Attribute Link Type/Classification 1 attribute validation.

   For information on how to set validity on an attribute, see the **Setting Validity of Description Attributes** section of the **System Setup / STEP Super User Guide** documentation.
2. In **System Setup**, click **Users & Groups**.
3. In the **Product Information Manager Default Settings** area, click the **Migrate Completeness Functionality** button.



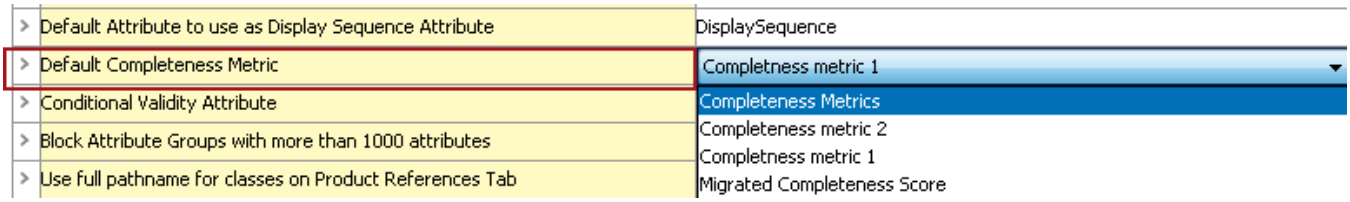
4. In the **Migrate Legacy Completeness Functionality** dialog, select the attribute you created in step 1 to hold the completeness scores. If you did not create an attribute, you have to do this to continue with the migration process.

   During the migration process, a Completeness Metric object is created, and you must specify an **ID** and a **Name** for the object. You also have to select a **Setup Group** to hold the Completeness Metric.



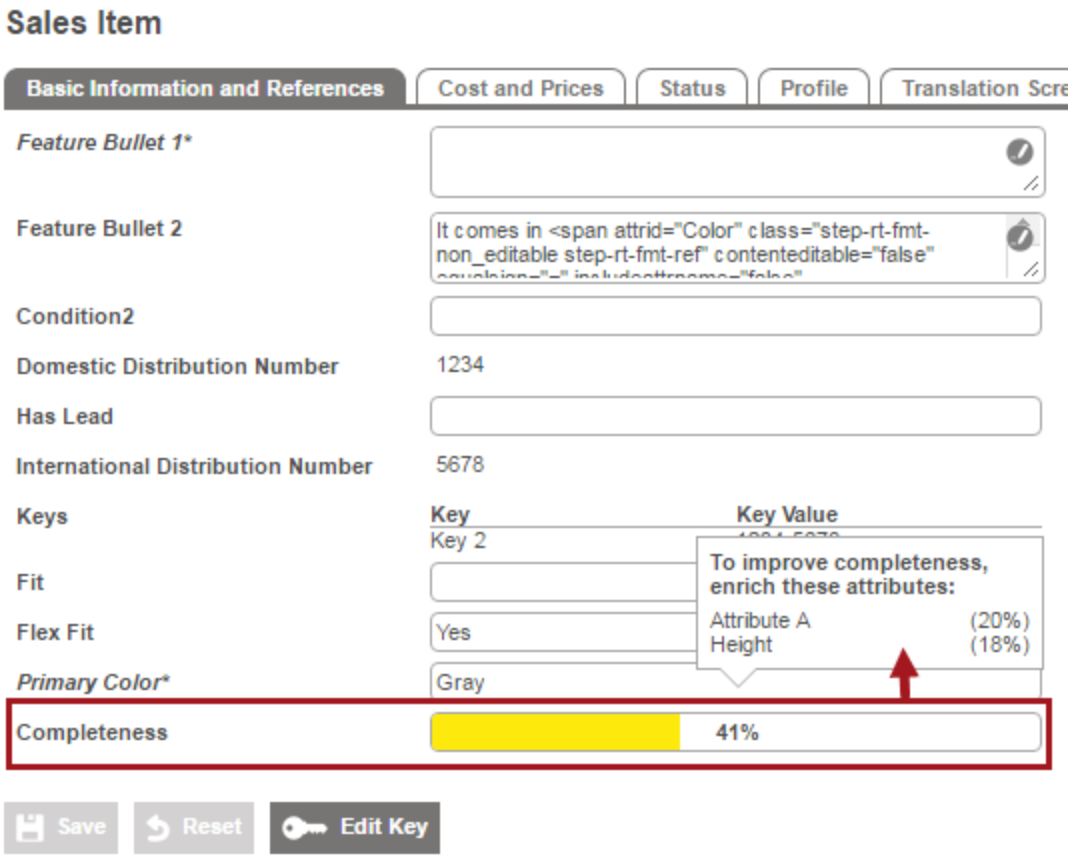
The migration is not performed as a background process, so the dialog stays open until the migration has completed.

The migration is a one-time process and the **Migrate Legacy Completeness Score** button is not available after the migration has been completed. Instead, you can select the default completeness metric from a drop-down list.



Object Completeness in Web UI

Object completeness indicators are available on Node Details component pages, allowing users to view the completeness of the currently selected node in Web UI. Like in workbench, object completeness is represented by a completeness meter. Additionally, mousing over the meter will provide text describing what you can do to improve the completeness percentage.

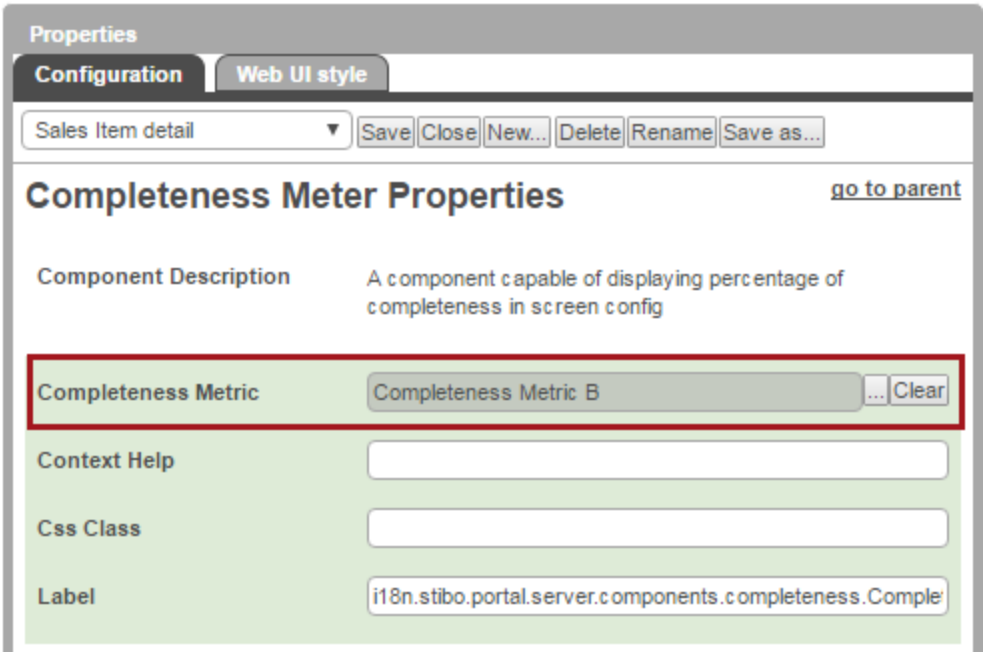


Adding the Completeness Meter Component

1. In design mode, navigate to Node Editor properties.
2. Under Child Components, click **Add**.
3. In the 'Add component' window the appears, select 'Completeness Meter' and click **Add**.
4. Click **Save**.

Configuring the Completeness Meter Component

1. Navigate to the Completeness Meter properties.
2. In the Completeness Metric field, click the ellipsis button **(…)**and select the desired completeness metric from the node selector. This selection will determine which completeness metric to use when evaluating the data on the object.


Properties
Configuration
Web UI style
Sales Item detail
Save
Close
New...
Delete
Rename
Save as...
Completeness Meter Properties
go to parent
Component Description
A component capable of displaying percentage of completeness in screen config
Completeness Metric
Completeness Metric B
...
Clear
Context Help
Css Class
Label
i18n.stibo.portal.server.components.completeness.Comple

3. Click **Save**.

Data Profiles

The Category / Data Profile functionality provides a detailed overview of data in a specific branch of the hierarchy in Tree. Each profile contains information about the profiled data and provides easy access to correcting data errors.

Note: For entities, classifications, and collections, these profiles are called 'Data Profiles'. For products, they are called 'Category Profiles'.

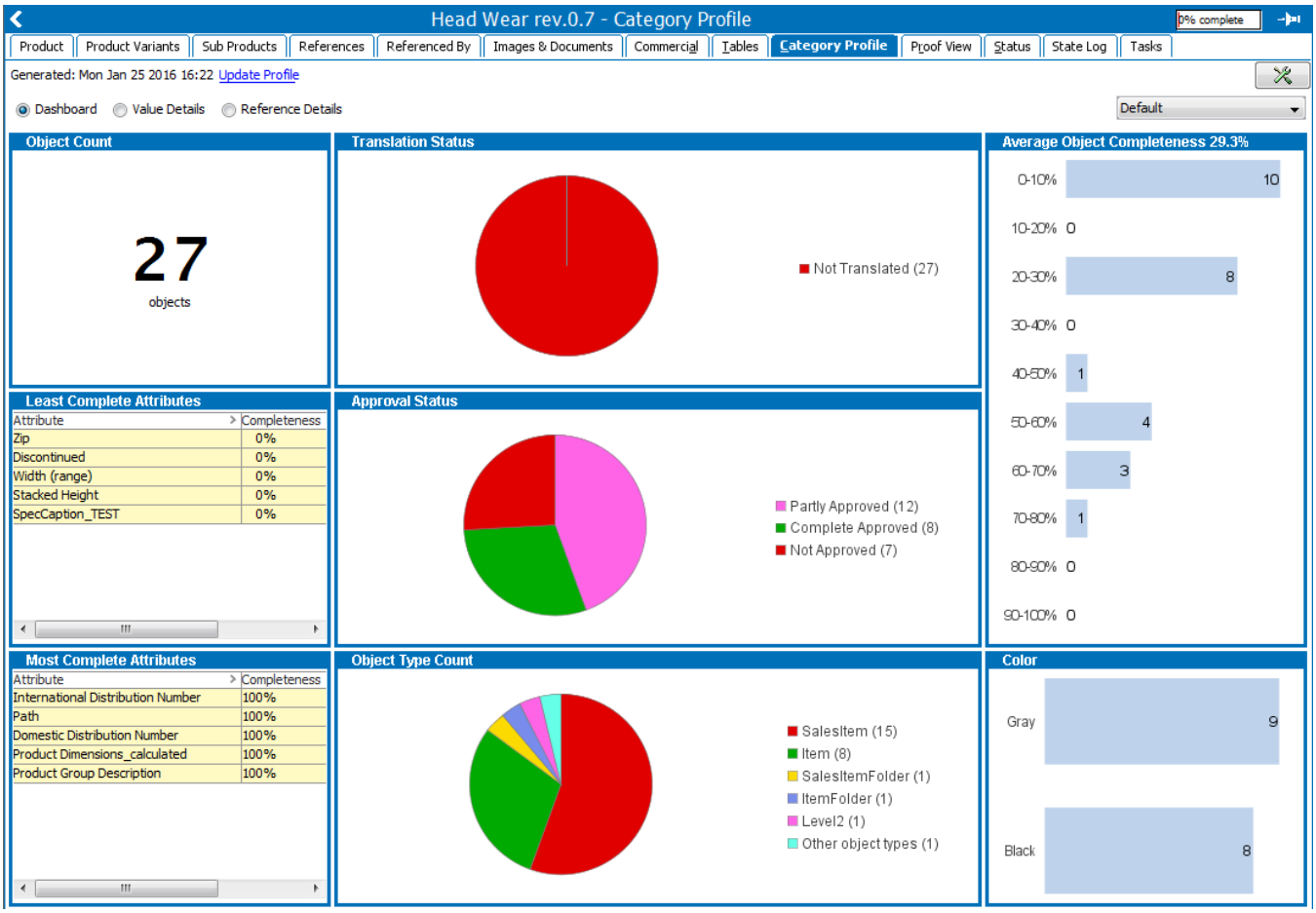
With a data profile, users can view information about the profiled data such as:

- The degree to which the attributes used in the hierarchy are populated.
- The most / least frequently used values for a given attribute, value ranges, used characters, and more.
- The approval and translation status of the profiled objects.
- The degree to which reference types are populated.
- The average product completeness (the degree to which important attributes, references, and product to classification links exist and have values.)

These profiles are generated on demand, and are context and workspace specific.

View Data Profiles

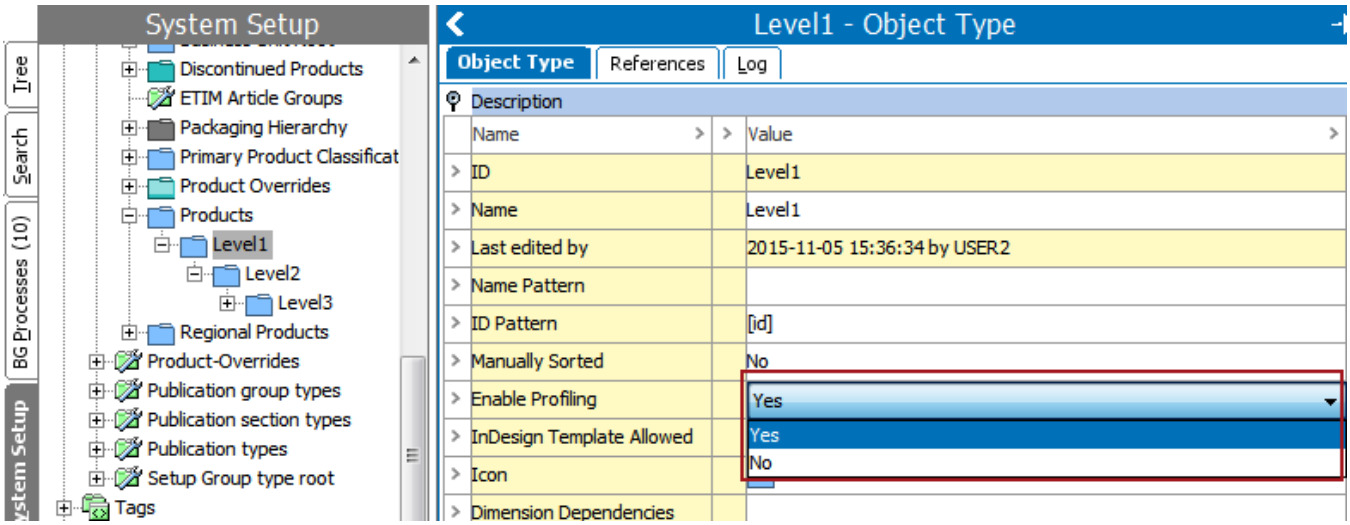
1. In the **Tree**, select the relevant hierarchy, and then click the **Category Profile** / **Data Profile** tab.
2. If no profile exists, click **Generate Profile**, otherwise click **Update Profile**



Enable Profiling on an Object

Objects can be flagged for profiling via object type configuration.

1. In **System Setup**, expand **Object Types & Structures**.
2. Select the relevant object type, then click the **Object Type** tab. Next, in the **Enable Profiling** field, select **Yes**. Note that you can only make changes here if you have the proper privileges. For more information see the **Action Sets** section of the **STEP Super User Guide / System Setup** documentation.



Because this configuration is made via object type, data / category profiles can be made on any level in the hierarchy.

Note: Information about product override objects is not included in category profiles.

Generating, Updating, and Scheduling a Data Profile

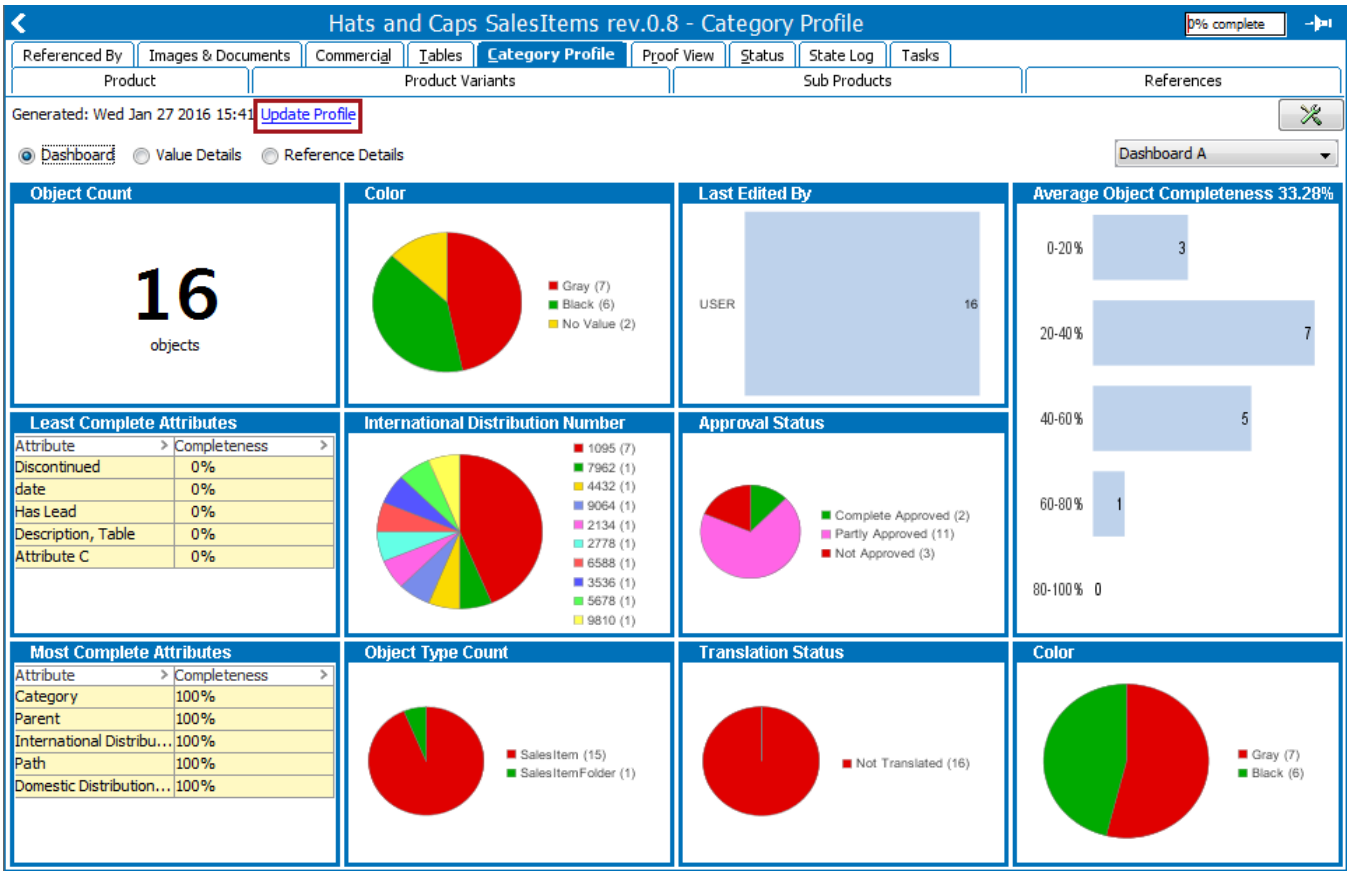
Before the data in a category / data profile can be viewed a profile must be generated. Generating a profile simply takes a snapshot of the data and does not actively monitor the profile for changes. Because of this, profiles must be manually updated whenever changes to the data are made (or a scheduled process must be configured).

Note: Only users with the proper privileges are able to generate or update a category profile.

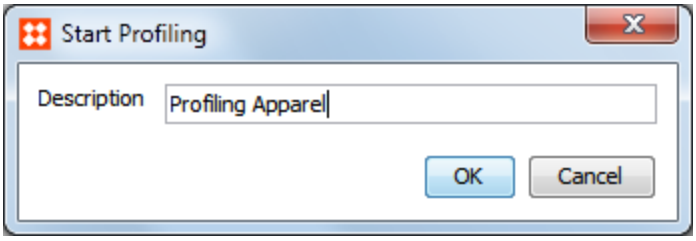
Generating / Updating a Data Profile

1. In the **Tree**, expand the applicable hierarchy, and then select the relevant object group or object.
2. On the **Category / Data Profile** tab, click the **Generate Profile** link. The **Start Profiling** dialog appears.

   If updating an existing profile the link is called **Update Profile**.



3. Click **OK**. A background process starts. Once finished, the profile is displayed.



Important: The background process runs using super user privileges, and the profiles are identical regardless of who started the process. Also, if a user has profile viewing privileges, all data is displayed in the profile no matter which view limitations otherwise apply to the user. By configuring the privileges, it is possible to control who can view, configure, and update profiles.

Scheduling Data Profiling

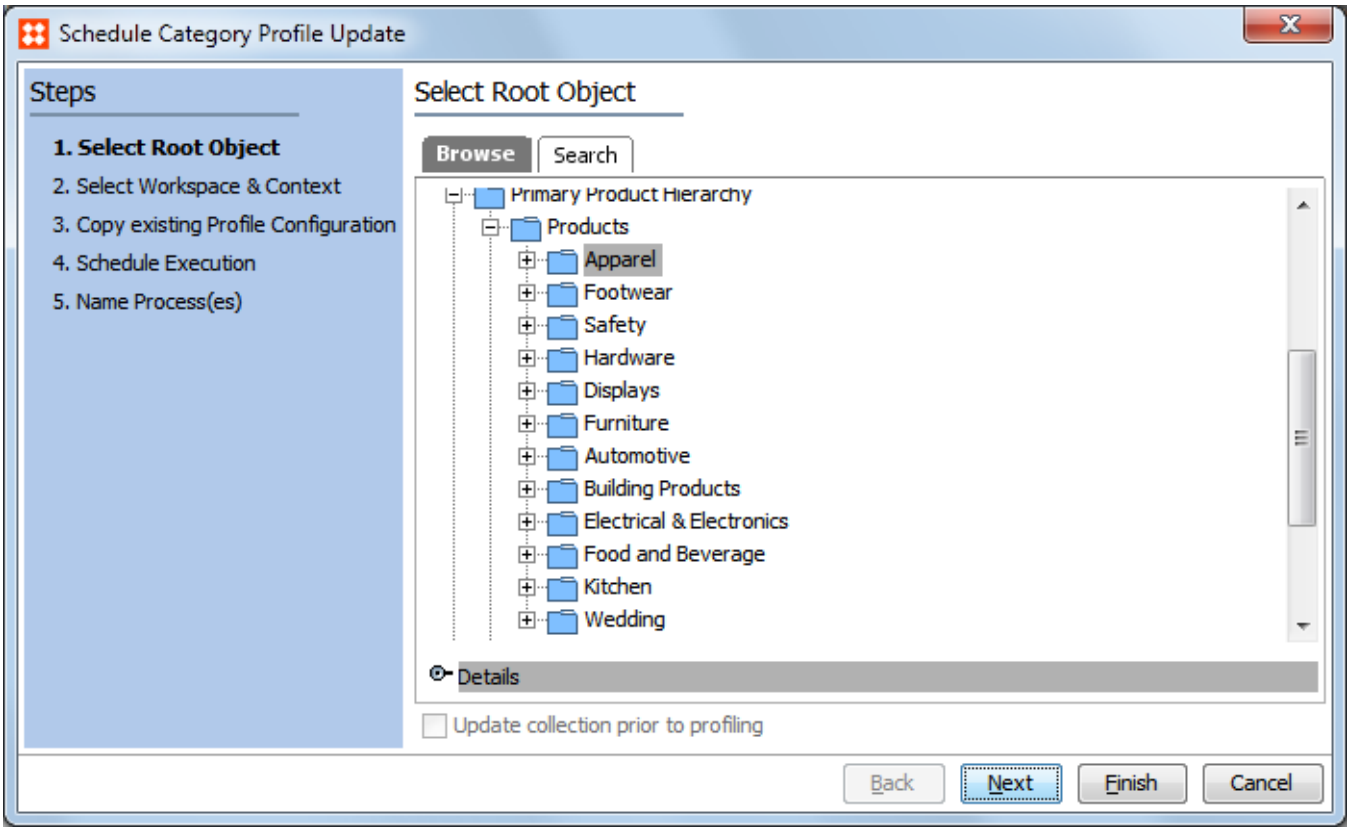
Category / Data profiles can be scheduled to update automatically using the **Schedule Category Profile Update** wizard.

In the **File** menu, select **Schedule Category Profiling** to open up the wizard:

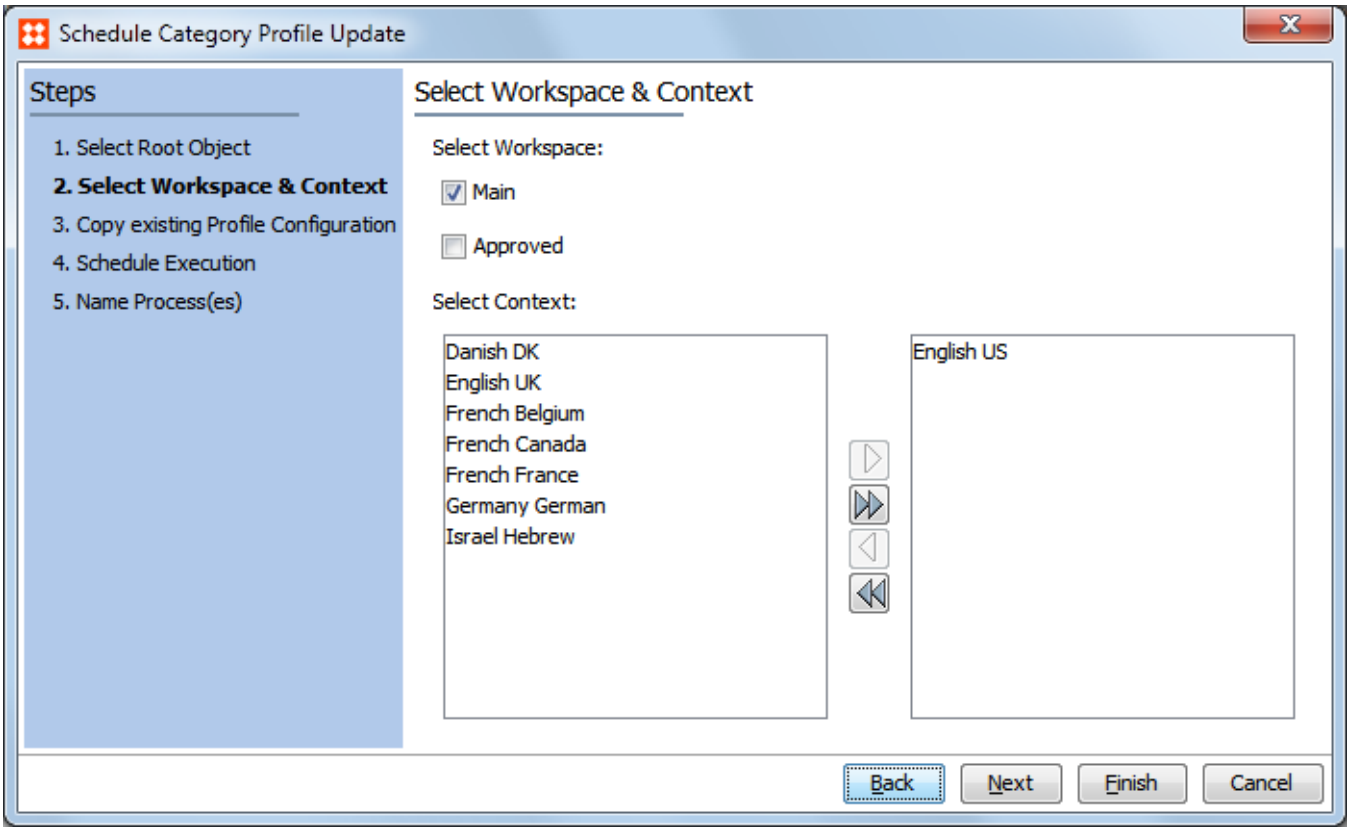
1. **Step 1 Select Root Object** - Browse or search for the preferred root object, and then click **Next**. Note that all category / data profiles underneath the selected node will also update.

   Valid selections include products, entities, classifications, and collections. If a collection is selected, profiles are generated or updated for all objects in the collection for which profiling is enabled.

   If the collection is search based, and you want the collection to be refreshed as part of the processing, select the **Update collection prior to profiling** check box.

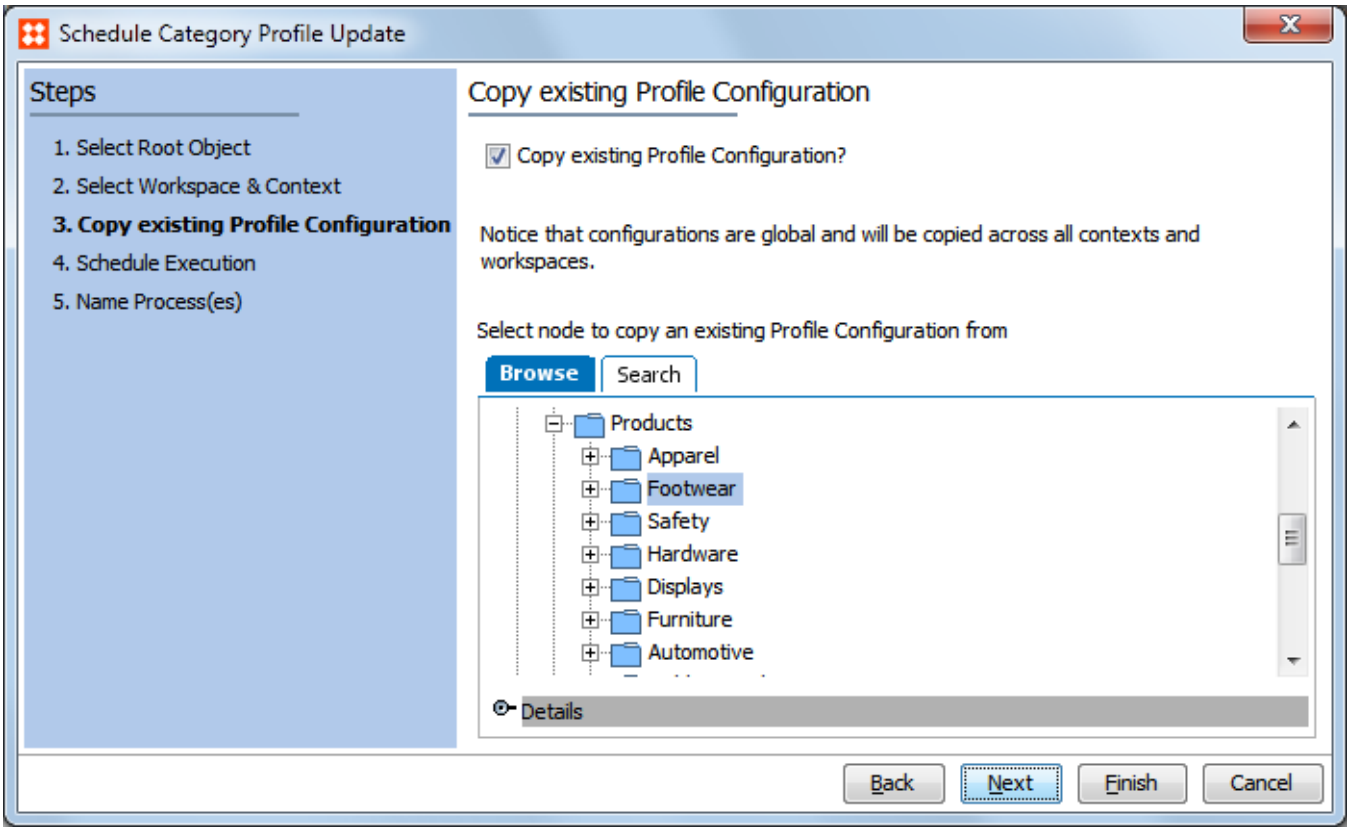


2. **Step 2 Select Workspace & Context** - Select the workspaces and contexts that the profile applies to. This allows the scheduled update to apply to all relevant context and workspace combinations.

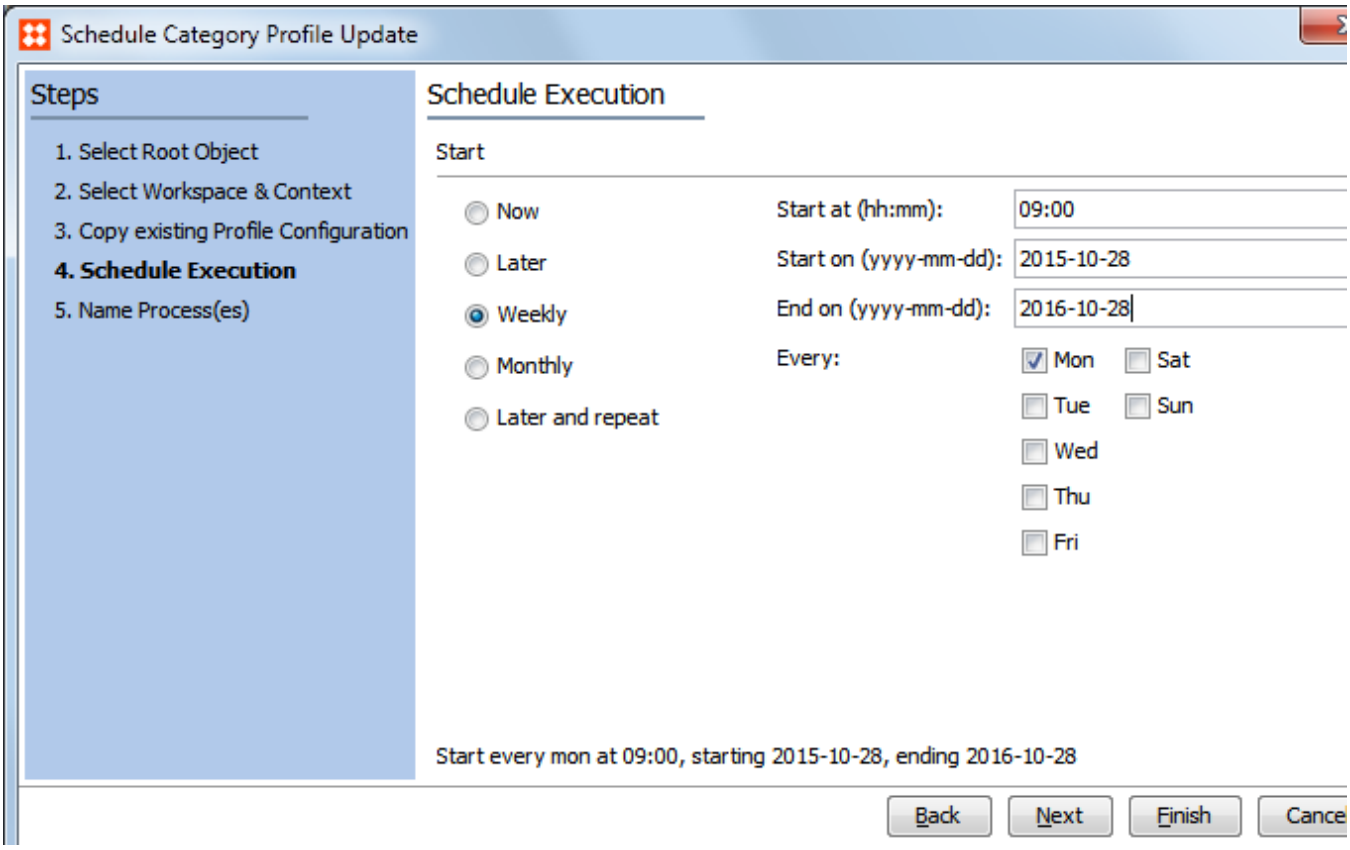


3. **Step 3 Copy existing Profile Configuration** - (Optional) Specify an existing configuration to use for the new one. This is useful, for example, if you need to profile many categories and want to reuse configurations.

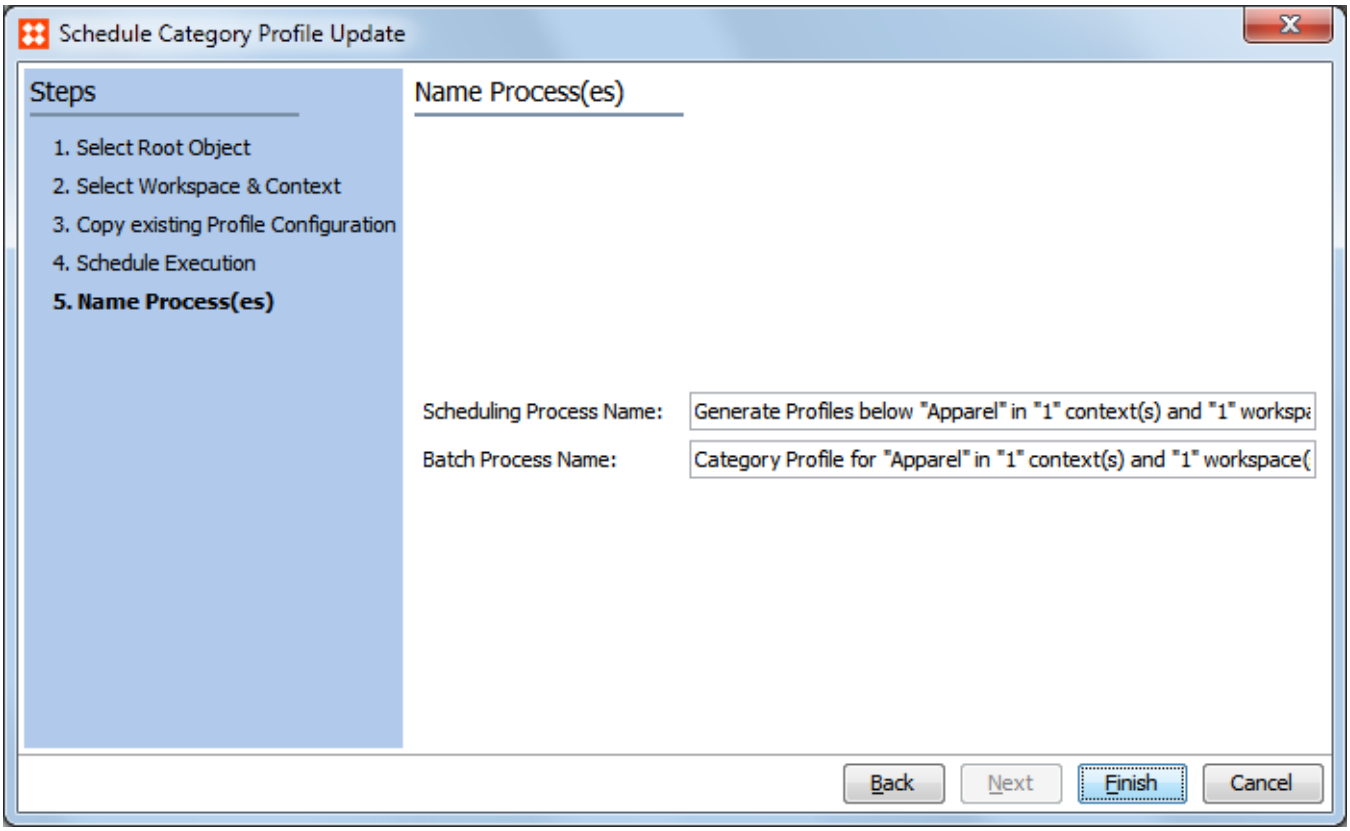
   Select the **Copy existing Profile Configuration** check box, and then search or browse for the object that contains the desired configuration. When the scheduled process is executed, the configuration is copied from the selected object to each of the objects for which profiles are generated or updated.



4. **Step 4 Schedule Execution** - Define the preferred schedule. You can specify whether you want the process to run immediately, at a later point in time, or as a recurring weekly or monthly process.



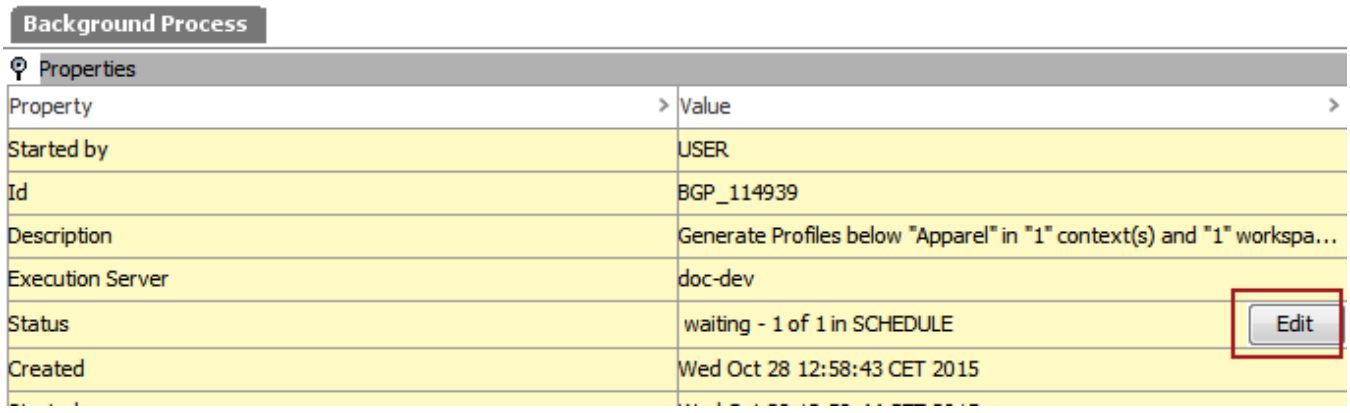
5. **Step 5 Name Process(es)** - Give the process a name and then click **Finish**. A background process is started.



Access the Schedule Category Profile Update Wizard from the Background Process

It is possible to edit the profiling schedule from the Background Process.

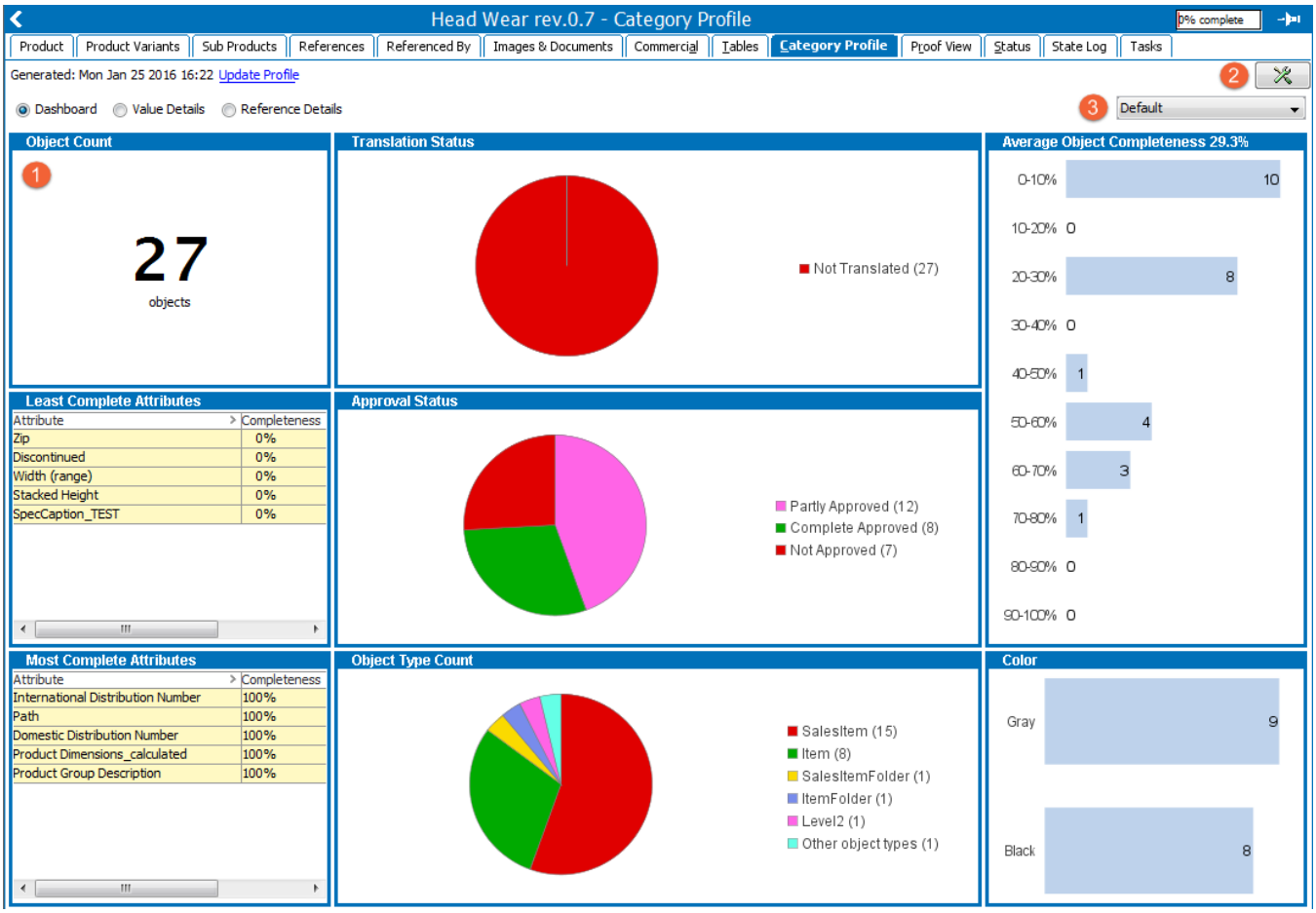
1. On the **Background Process** tab, locate and select the relevant background process.
2. In the **Status** property **Value** field, click **Edit**. The **Schedule Category Profile Update** wizard opens.



Data Profile Dashboard

The category / data profile dashboard is a highly customizable view that allows users to display profiled data using a collection of configurable widgets. Each profile can have its own set of dashboards, allowing users to group widgets together in a logical manner, and to create different views for specific user roles or tasks. Dashboard configurations are tied to the profile but are shared across all contexts and workspaces.

Pictured below is an example of a Category Profile Dashboard:



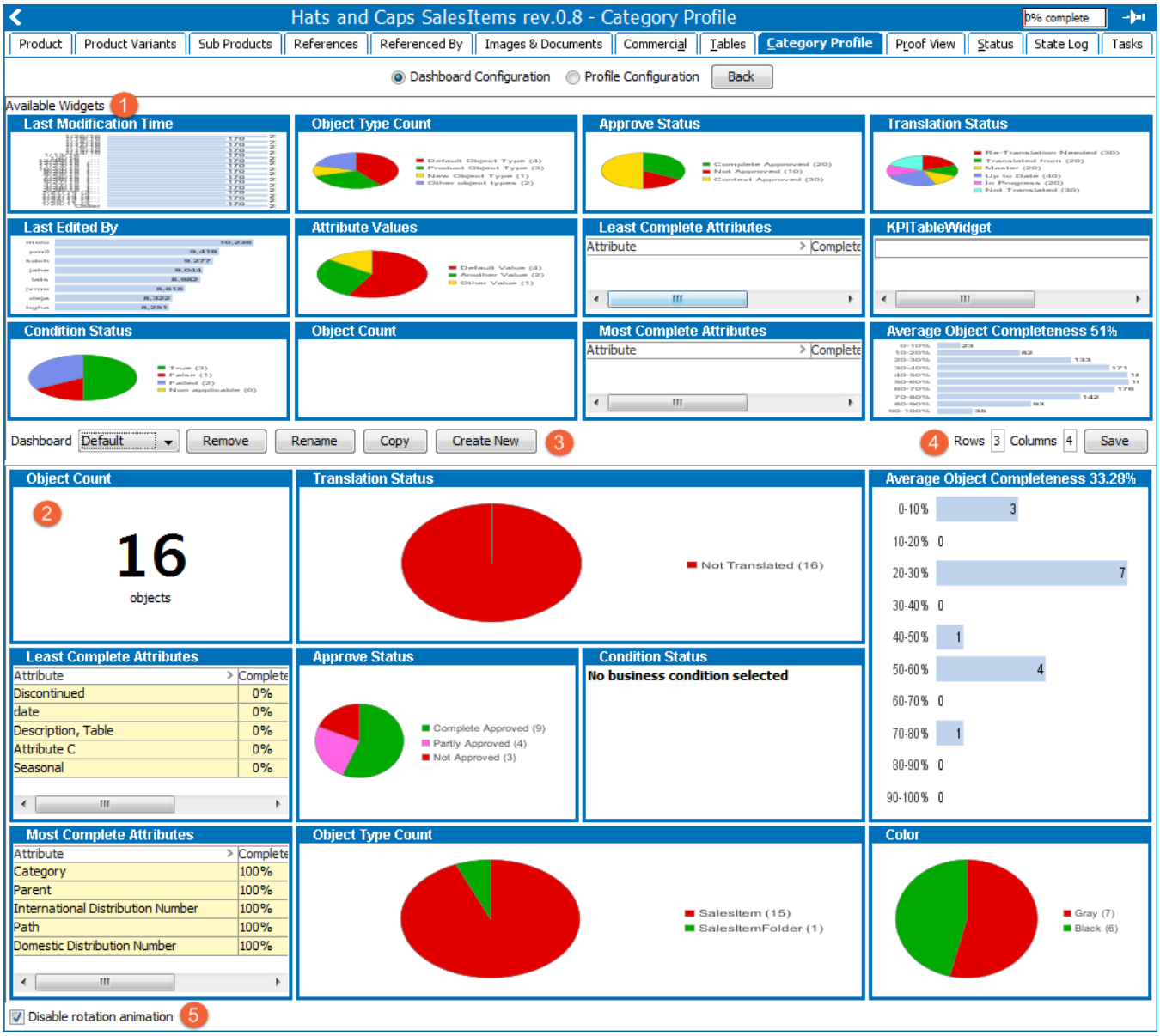
1. This area displays the currently selected dashboard.
2. Clicking the **Tools** button allows users to access the **Dashboard Configuration** and **Profile Configuration** views.
3. Users can toggle between dashboards using this dropdown.

Changes to a category / data profile dashboard are made via the **Dashboard Configuration** view, as detailed below.

Accessing the Dashboard Configuration

1. In the **Tree**, select the relevant object hierarchy or object.
2. On the **Category / Data Profile** tab, click the **Tools** icon in the upper right corner, and then select **Dashboard Configuration**.

Navigating the Dashboard Configuration



1. The top area of this view contains all **Available Widgets** for the dashboard. These widgets can be added to the currently selected dashboard by left-clicking the title bar and dragging the widget to an empty space in the dashboard preview area (the lower half of the screen).

2. The bottom area of this view contains the **Dashboard Preview**. Once a widget is added to this area, they can be configured by right-clicking the title bar and selecting the desired configuration option. Further details on configuring the widgets are detailed below.

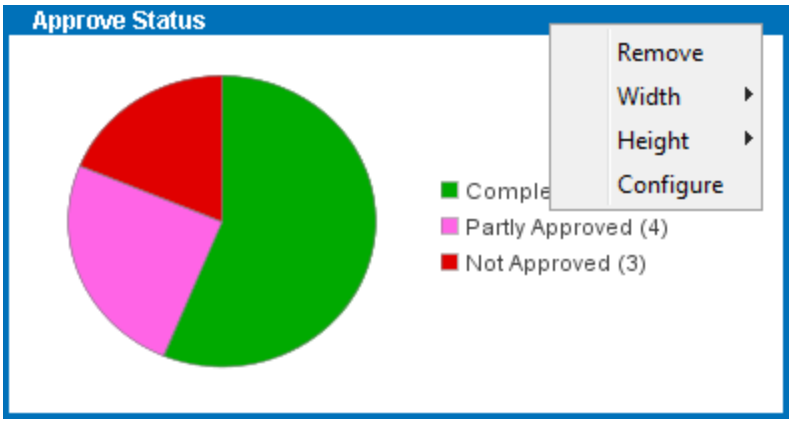
Note: The data shown in the **Available Widgets** area is sample data. The **Dashboard Preview**, however, displays correct data if the category has been profiled.

---

3. Different dashboards can be configured using the buttons found in-between the **Available Widgets** and **Dashboard Preview** areas on the left:
    - The **Dashboard** dropdown list allows users to select whichever dashboard they want to display / edit. Note that this dropdown selector is also available in the Category / Data Profile Dashboard view, underneath the **Tools** icon, and is used to toggle between dashboard displays.
    - The **Remove** button allows users to delete the currently selected dashboard.
    - The **Rename** button allows users to rename the currently selected dashboard.
    - The **Copy** button allows users to copy the currently select dashboard, including all of its widget configurations. If clicked, a naming prompt will appear. Once the copied dashboard is named, it will automatically appear as an option in the **Dashboard** dropdown list.
    - The **Create New** button allows users to create a new dashboard for the currently selected profile.
4. In the **Rows** and **Columns** fields, found in-between the **Available Widgets** and the **Dashboard Preview** on the right-side of the view, the user can specify how many rows and columns the currently selected dashboard should have.
5. Check the **Disable rotation animation** checkbox to disable the animation that occurs when toggling between the Category / Data Profile Dashboard and the Dashboard Configuration views.

Configuring Dashboard Widgets

Once added to the Dashboard Preview, users can configure the size and shape of their widgets, as well as filter what data appears. To access these configuration options, right-click the title bar of the desired widget and a list of options will appear.



Clicking **Remove** will delete the widget from the currently selected dashboard. Clicking **Width** or **Height** will display a list of values - the higher the value selected the taller / wider the widget becomes. Note that increasing the size of a widget may push other widgets out of view if there is not enough space. As mentioned above, the **Rows** and **Columns** fields dictate how large the dashboard can be.

Clicking **Configure** displays a configuration window with additional options. These options vary between widgets, but common configuration options include: widget title, what context to display data from, what workspace to display data from, filtering data via object type, sorting order, and chart type. For more details on which options are available to each individual widget, see the **Available Widgets** section in the **Data Profiling** documentation.

Saving the Configuration

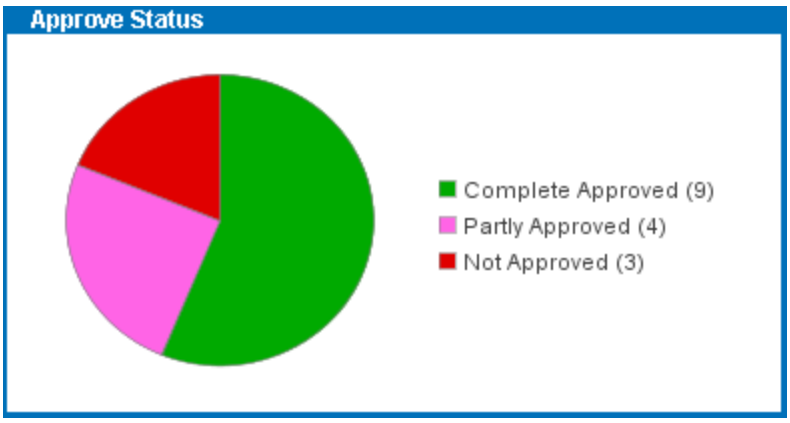
1. Once the dashboard has been configured, click **Save**, located next to the **Rows** and **Columns** fields, then click the **Back** button at the top of the screen to return to the category / data profile.
2. Click **Update Profile** for the changes to take effect.

Available Widgets

All available category / data profile dashboard widgets are detailed below. These widgets are also available in Web UI unless noted otherwise. For more information about configuring dashboard widgets in Web UI, see the **Data Profile Widgets in Web UI** section of the **Web User Interfaces** documentation.

Approve Status

This widget displays the approval status of objects in the data profile.

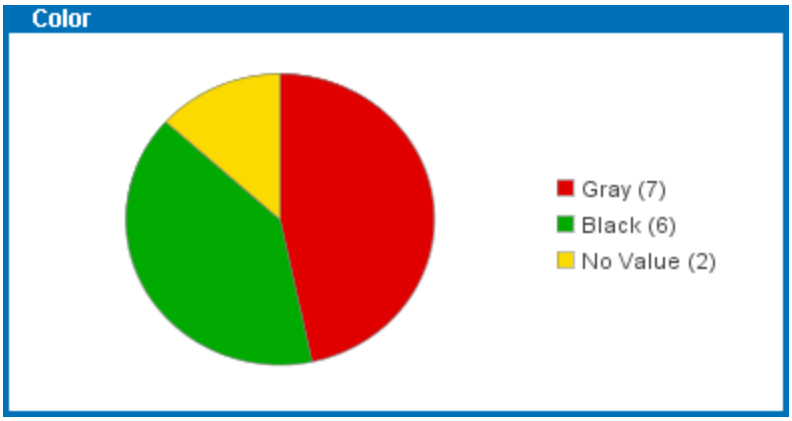


Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Attribute Values

This widget shows the distribution of values for the specified attribute across objects in the data profile.



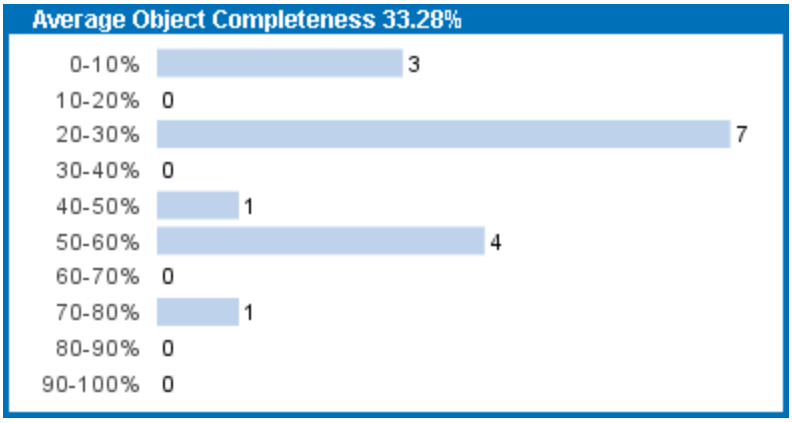
Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.
- **Attribute ID** - Specifies which attribute to pull values from. Defining this configuration is mandatory.
- **Chart / Dashboard Type** - Determines whether to display the data as a pie chart or bar chart.
- **Max Number of Values / Max Values** - Specifies the maximum number of distinct values to display. By default this field is set to -1, which means 'all values'. If the specified maximum is exceeded, the remaining (smaller) values will be placed in the "Other values" group.
- **Show No Values** - Displays the number of products that have no value for the specified attribute when the box is checked.

Important: Any attribute that is full text indexed, multi-valued, calculated, or has values with inline references is not compatible with this widget.

Average Object Completeness

This widget shows the average object completeness in the data profile.



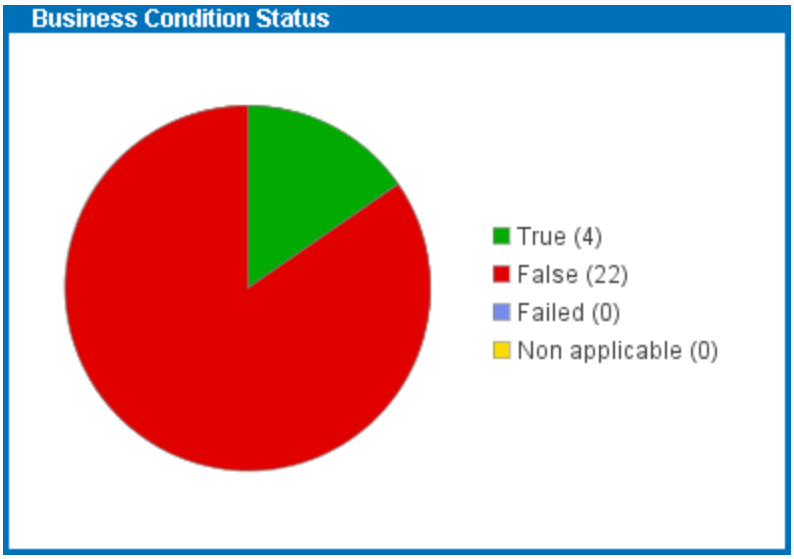
Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.
- **Completeness Metric** - Determines which completeness metric to evaluate the data with. Note that the completeness metric must be specified in the Profile Configuration and data must have been profiled using that configuration before the widget can show any data. For more information, see the **Profile Configuration** section of the **Data Profiling** documentation.
- **Intervals** - Displays the bars of the graph in intervals of 2, 4, 5, 10, or 20.

Business Condition Status

This widget displays the result of testing a business condition during profiling.

Note: If the test of the business condition results in an error, the test is reported as failed.



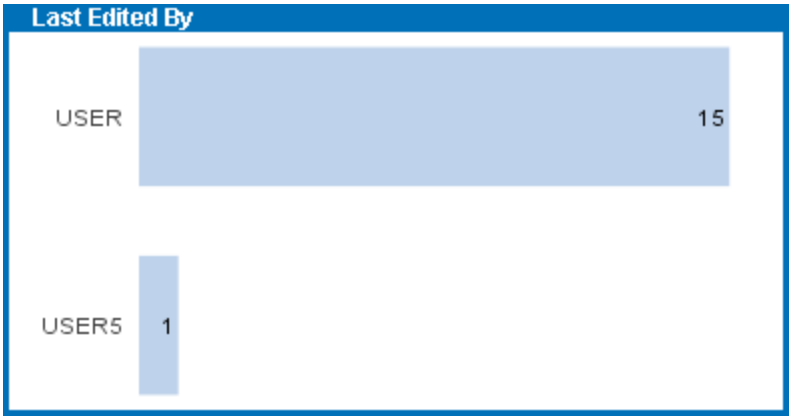
Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.
- **Condition** - Determines which business condition to test when profiling. Note that the condition must be specified in the Profile Configuration and data must have been profiled using that configuration before the widget can show any data. For more information, see the **Profile Configuration** section of the **Data Profiling** documentation.

Last Edited By

This widget shows which users last edited the object in the data profile.

Note: Not available in Web UI.



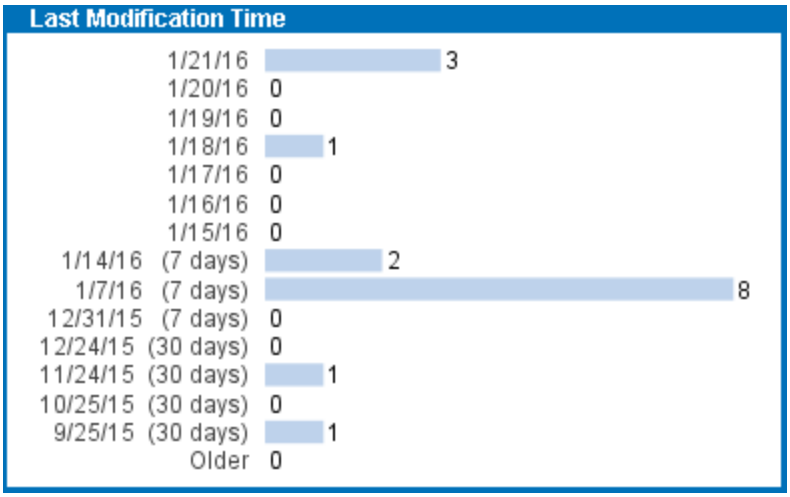
Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.
- **Sort Order** - Sorts the data in descending order by users with the most frequent edits or by users with the rarest.
- **Chart Type** - Determines whether to display the data as a pie chart or bar chart.
- **Max Number of Values** - Specifies the maximum number of distinct values to display. By default this field is set to -1, which means 'all values'. If the specified maximum is exceeded, the remaining (smaller) values will be placed in the "Other values" group.

Last Modified Time

This widget shows the last time each object in the profile was edited.

Note: Not available in Web UI.



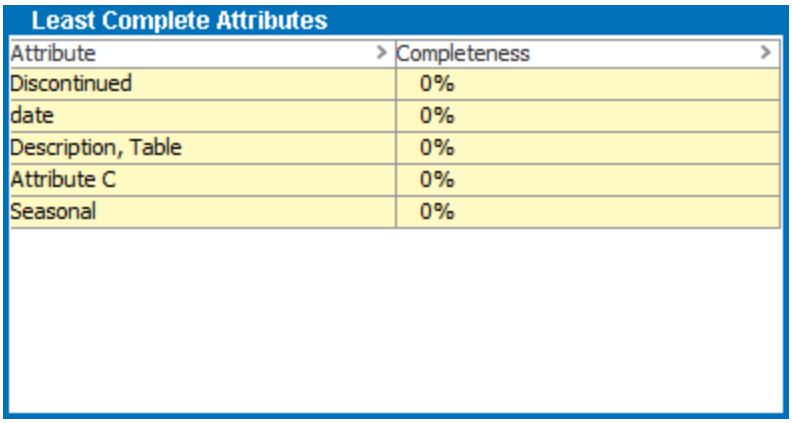
Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Least Complete Attributes

This widget displays the five least complete attributes in the data profile. Completeness is rounded to the nearest whole percentage and attributes with the same completeness are presented in the order in which they are retrieved from the database. Inherited values are included in the calculation.

Note: This widget has no relation to the Completeness Metric defined in the Profile Configuration and will evaluate all attributes in the data profile.



Configuration options include:

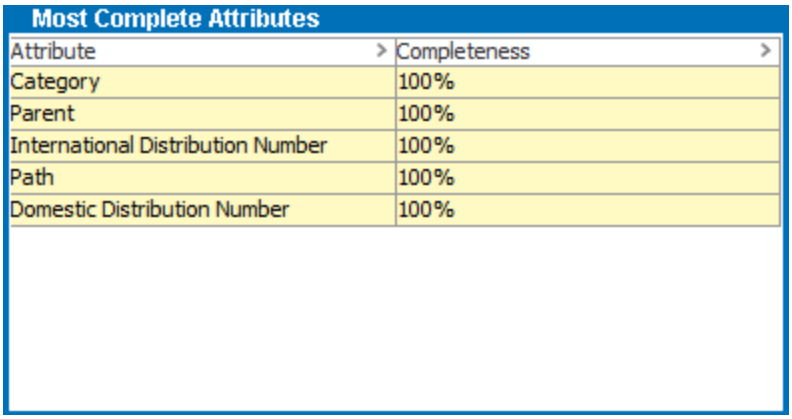
- **Title** - Changes the name that appears on the widget's title bar.

- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Most Complete Attributes

This widget lists the five most complete attributes in the data profile. Completeness is rounded to the nearest whole percentage and attributes with the same completeness are presented in the order in which they are retrieved from the database. Inherited values are included in the calculation.

Note: This widget has no relation to the Completeness Metric defined in the Profile Configuration and will evaluate all attributes in the data profile.



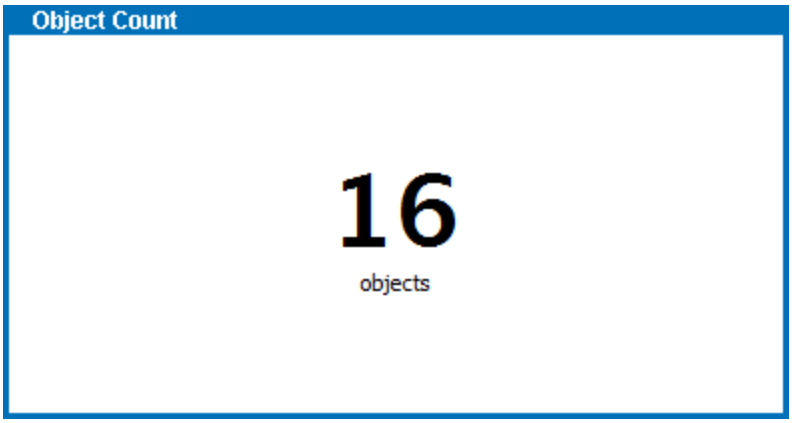
Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Object Type** - Filters data via object type.

Object Count

This widget displays the number of objects in the data profile.

Note: If any objects being profiled are linked multiple times in the sub-tree, they are also counted multiple times. This only applies to sub-trees under classifications and entities, as products can only link to other products once in a parent / child relationship.



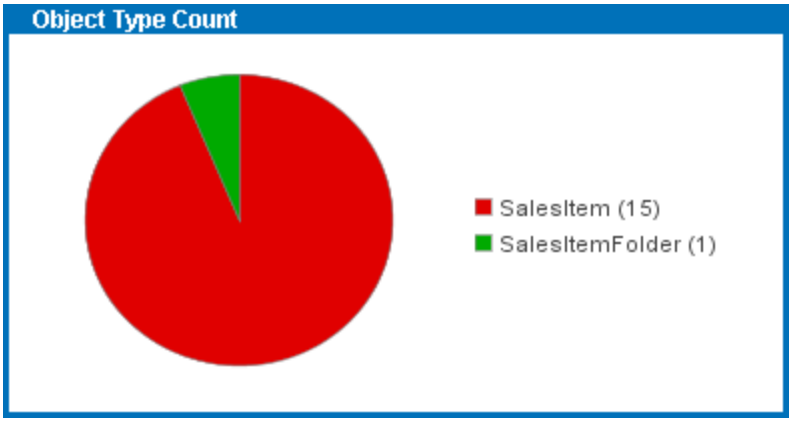
Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.

Object Type Count

This widget displays the distribution of object types in the data profile.

Note: Only the five most common types are displayed as separate entries. The remaining types are grouped in "Other object types".

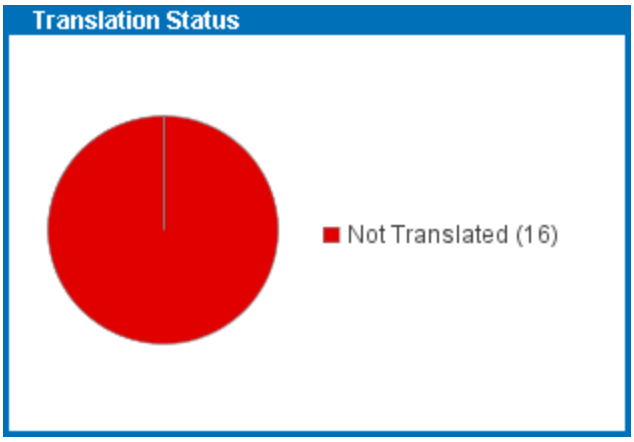


Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Chart Type** - Determines whether to display the data as a pie chart or list.

Translation Status

This widget displays the translation status of the objects in the data profile.



Configuration options include:

- **Title** - Changes the name that appears on the widget's title bar.
- **Context** - Pulls data from the specified context.
- **Workspace** - Pulls data from the specified workspace.
- **Chart Type** - Determines whether to display the data as a pie chart or list.

Note: The KPI Table Widget has been deprecated.

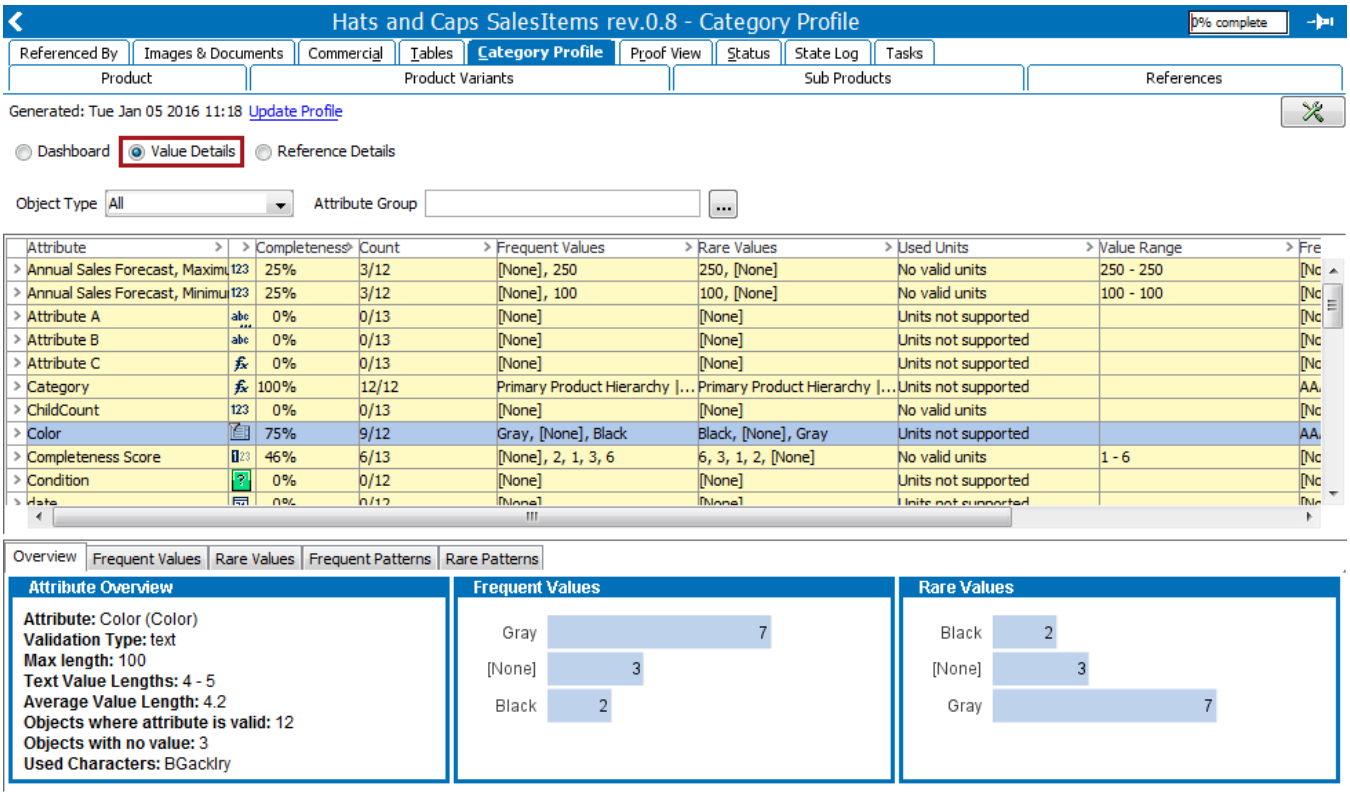
Value Details View

The **Value Details** view provides information about all attributes found within a category / data profile. As described below, this view provides users with multiple ways to display attribute data, including tables, bar graphs, and lists.

Accessing and Navigating Value Details View

1. In the **Tree**, select the relevant object hierarchy or object.
2. On the **Category Profile / Data Profile** tab, click **Generate Profile** or **Update Profile**, and then select **Value Details**.

   The **Value Details** page is divided into two areas, though at first only the upper area displays. The upper area lists all the attributes that are valid for the objects in the data profile. When selecting a row, detailed information about the selected attribute appears in the lower area.



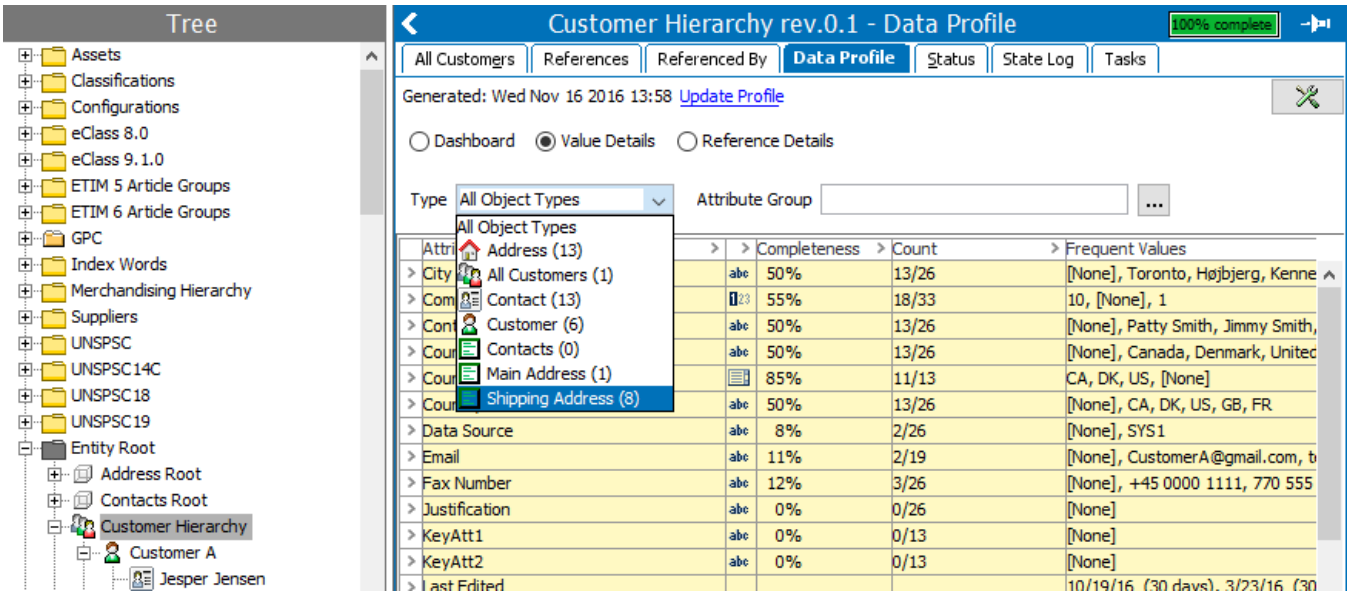
3. From the **Type** list, select a single object type or all object types. The number in parenthesis after the object type name shows how many instances of each object type are in the profile.
4. The user may filter their results so that only those attributes that belong to the specified attribute group will appear on the list. This can be done by entering the name of the specific attribute group into the **Attribute Group** field. Alternatively, you may also click the ellipsis button **(…)** and select the attribute group from the list that appears.

Note: Valid attributes are attributes that are valid on the object types used in the data profile. This includes specification attributes that are linked into the hierarchy in or above the profiled nodes. Specification attributes that are only made valid through the classification hierarchy are not profiled. Attributes and values on references and links, as well as orphan attributes and values, are not profiled.

For information on the tabs in the lower area of this view, see the Value Details Tabs section of the Data Profiling documentation.

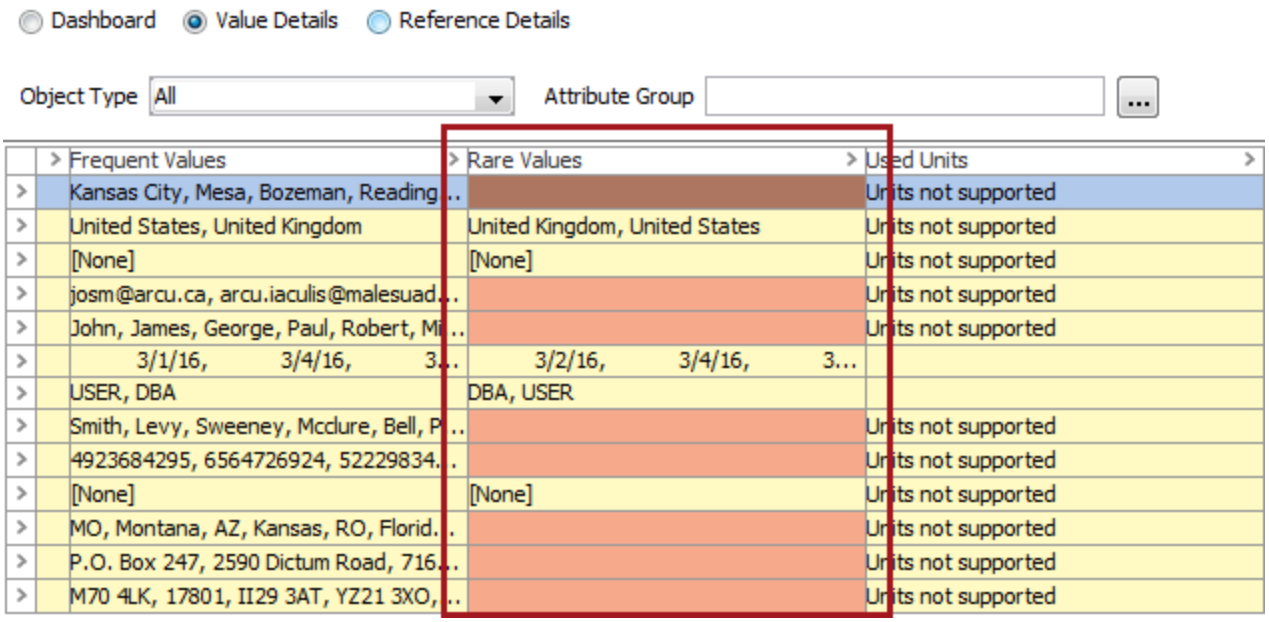
Profiling Entities with Data Containers

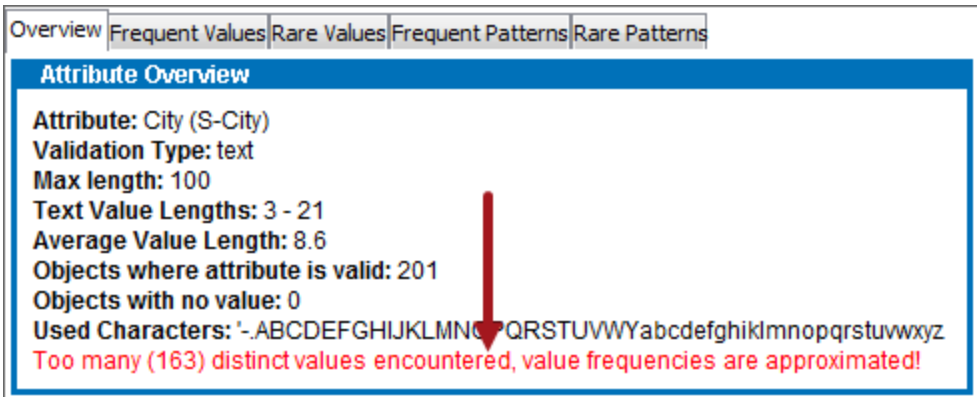
When profiling entities, it is possible to filter the attribute values based on Data Container types.



Maximum Distinct Values

It is possible for an attribute to have too many values to display. In these cases, the applicable rows will turn red and a message will appear in the Attribute Overview section.





By default, the system limits the amount of values the functionality will consider to 100. This limit can be changed by changing these properties via the sharedconfig.properties file:

- DataProfile.MaxDistinctAttributeValuesConsideredDuringProfileGeneration

Additionally, altering this property changes the maximum number of different reference targets that will be considered:

- DataProfile.MaxDistinctTargetsConsideredDuringProfileGeneration

Attribute Table Details

The following list details the columns found in the attribute list table. All columns can be filtered and sorted.

Column	Description
Attribute	Displays the name of the attribute. When clicking the name, the attribute definition in System Setup appears.
Column 2	Shows the Attribute Validation Base Type icon.
Completeness	Shows the degree to which the attribute is populated in the profile rounded to the nearest whole percent. Inherited values are included in the calculation.
Count	Lists the number of objects in the profile where the attribute has a value (including inherited values) and the number of objects where the attribute could possibly have a value.
Frequent Values	Displays a comma-separated list of the most frequently appearing values for the attribute in the profile. The list is sorted in ascending order based on the frequency of that value's appearance. The list includes inherited values. Tags are stripped.

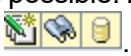
Column	Description
Rare Values	Comma-separated list of the least frequently appearing values for the Attribute in the profile. The list is sorted descending on frequency. Includes inherited values. Tags are stripped.
Used Units	Displays which units are used for the attribute. Contains the following information: • List of units used in values for attributes in the profile. • "Units not supported" - This text is displayed for attributes of a validation base type that does not support units. • "No valid units" - This text is displayed for attributes of a validation base type that supports units but where no units are configured. • "No units used" - This text is displayed for attributes of a validation base type that supports units but where no attribute values in the profile have a unit.
Value Range	Shows the value range for attributes with the validation base type "Integer" or "Number". If units that share the same base validation are used, the value ranges are shown with Units. For example: 123 cm - 2 m".
Frequent Patterns	Displays the most frequent patterns of the values of a given attribute in addition to the number values that match the pattern. More information can be found in the **Value Details Tabs** section of the **Data Profiling** documentation.
Rare Patterns	Displays the most uncommon patterns of the values of a given attribute in addition to the number values that match the pattern. More information can be found in the **Value Details Tabs** section of the **Data Profiling** documentation.

Value Details Tabs

The lower area of Value Details is divided into a number of tabs that contain detailed attribute information.

- Overview Tab
- Frequent Values Tab
- Rare Values Tab
- Frequent Patterns Tab
- Rare Patterns Tab

Actions

With the exception of the Overview tab, all tabs contain a table where three different actions are possible. From left to right, the icons represent the following actions: Bulk Update, Search, and Save as Collection  .

- **Bulk Update** When the Bulk Update icon is clicked, a search is run within the profile for objects that have the selected attribute value. The result is used as input data set for the Bulk Update wizard.

The Bulk Update wizard opens on the **Operations** step with the attribute selected in the **Set Value** operation.

This functionality can be used to fix data errors. Since the input data set is the result of a search, you will not accidentally update the wrong data due to running the operation from an outdated profile.

If viewing value details for a specific object type, the search is filtered so that only objects of the selected type are used as input data set for the Bulk Update operation. Likewise, if you have selected "Only show values entered as local values", the search only includes local values.

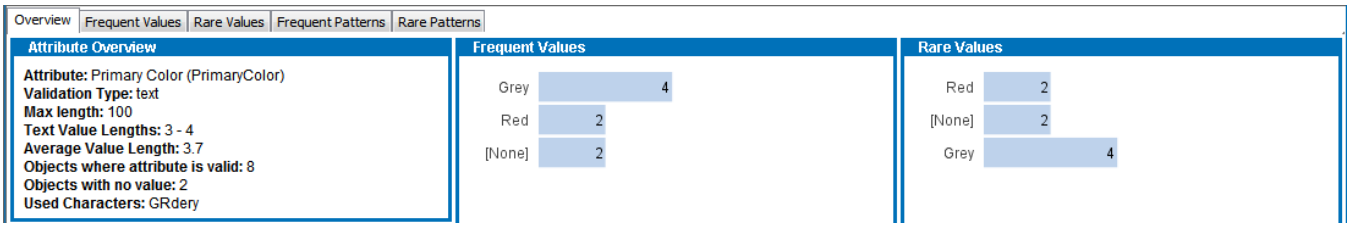
For more information about the Bulk Update wizard, see the **Running Bulk Updates** section of the **Getting Started / STEP User Guide** documentation.

Important: If "Only show values entered as local values" is not checked before you run the bulk update, you run the risk of creating local instances of data from before the operation was inherited through the hierarchy.

- **Search** - When the Search icon is clicked, a search is run within the profile for objects that have the selected attribute value. When the search is complete, the result is displayed on the **Search** navigation tab. Searches are filtered on the object type selection in the profile and inherited values are included or excluded depending on whether you select "Only show values entered as local values".
- **Save as Collection** When the Save as Collection icon is clicked, the Save as Collection in Background Process dialog is displayed. Select a location and enter a Collection Name and a Collection ID, and then click **OK**. A search is run within the profile for objects that have the selected attribute value, and the result is saved in the collection.

Overview Tab

The **Overview** tab contains a number of widgets that each present different information about the selected attribute. Users cannot customize this view.



Widgets on the Overview tab.

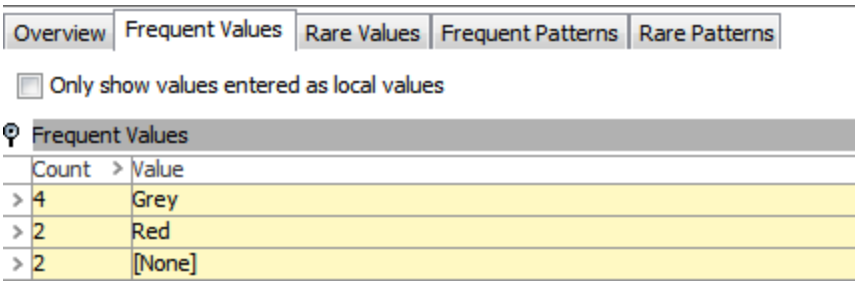
Widget	Description
Attribute Overview	Displays information retrieved from the attribute definition in System Setup and information generated in the profiling. You can, for example, see which characters are used, the value length span, and the numerical value span of number and integer attributes.

Widget	Description
Frequent Values	Bar chart that displays the ten most frequently appearing values or words. Includes inherited values. Tags are stripped. Values with an equal count are presented in the order they are retrieved from the database.
Rare Values	Bar chart that displays the ten least frequently appearing values or words. Includes inherited values. Tags are stripped. Values with an equal count are presented in the order they are retrieved from the database.
Used Units	This widget is only displayed for attributes of validation base types that support units. The widget can contain the following content: • Pie chart of units used in values for attributes in the profile. • "No valid units" - This text is displayed for attributes of a validation base type that supports units but where no units are configured. • "No units used" - This text is displayed for attributes of a validation base type that supports units but where no attribute values in the profile have a unit. Inherited values are included in the pie chart calculation.

Frequent Values Tab

On this tab, the 100 most common values in the profile of the selected attribute are listed. Tags are resolved.

The default view includes inherited values, however, if you check "Only show values entered as local values", inherited values are excluded. Values inherited from other contexts are still displayed. Values with an equal count are presented in the order they are retrieved from the database.



Rare Values Tab

On this tab, the 100 least common values in the profile of the selected attribute are listed. Tags are resolved.

The default view includes inherited values, however, if you check "Only show values entered as local values", inherited values are excluded. Values inherited from other contexts are still displayed. Values with an equal count are presented in the order they are retrieved from the database.

Used Unites Tab

This tab lists all units used in the values of the selected attribute, such as length in centimeters (cm), weight in pounds (lbs), price in dollars ($), etc. This tab only appears for those attributes with values containing units.

Frequent Patterns Tab

This tab lists the most frequent patterns of the values of a given attribute and the number values that match the pattern.

A value pattern is a way of describing the structure of a value in terms of characters, numbers, and symbols. When the patterns of the values in a set are compared, certain types of errors can be detected by finding values whose patterns are different from the rest. For example, phone numbers with missing digits or numbers that are mixed with letters, and so on.

A pattern for a given value is created by replacing any letter with A, any number with 9, and any unrecognized character with X. Other symbols are left as they are.

The following is an example of a value pattern:

Value	Pattern
123-ABB	999-AAA
234.23	999.99
mail@mail.com	AAAA@AAAA.AA

The pattern [None] represents an empty value.

Rare Patterns Tab

This tab resembles the Frequent Patterns but lists the least frequently used patterns rather than the most frequent. This information is useful when searching for faulty values.

Special Attribute and Value Handling

Full text indexed attributes and multi-valued attributes are handled specially.

- **Full text indexed attributes**

For attributes configured as full text indexable, the most frequent or rare words are listed with the exception of words on Oracle's English Default Stoplist. However, if attribute values have units, the attribute is not profiled as full text indexable.

- **Multivalued attributes**

For multivalued attributes, the entries in each value are split out. In the following example, only three products in the profile have a value for the multivalued attribute color:

- Product A: Red, Green, Blue
- Product B: Yellow, Red
- Product C: Blue, Red

The order of the frequent values information is:

Red, Blue, Green, Yellow

- **Values with inline references**

If a value consists of an inline reference only, and the reference does not resolve to any value, the value appears as an empty string. In overview charts, however, the text "EMPTY" is displayed.

- **Empty values and references**

Empty attribute values and empty reference values are handled as ordinary values. Empty values show up in detail views and graphs as [None].

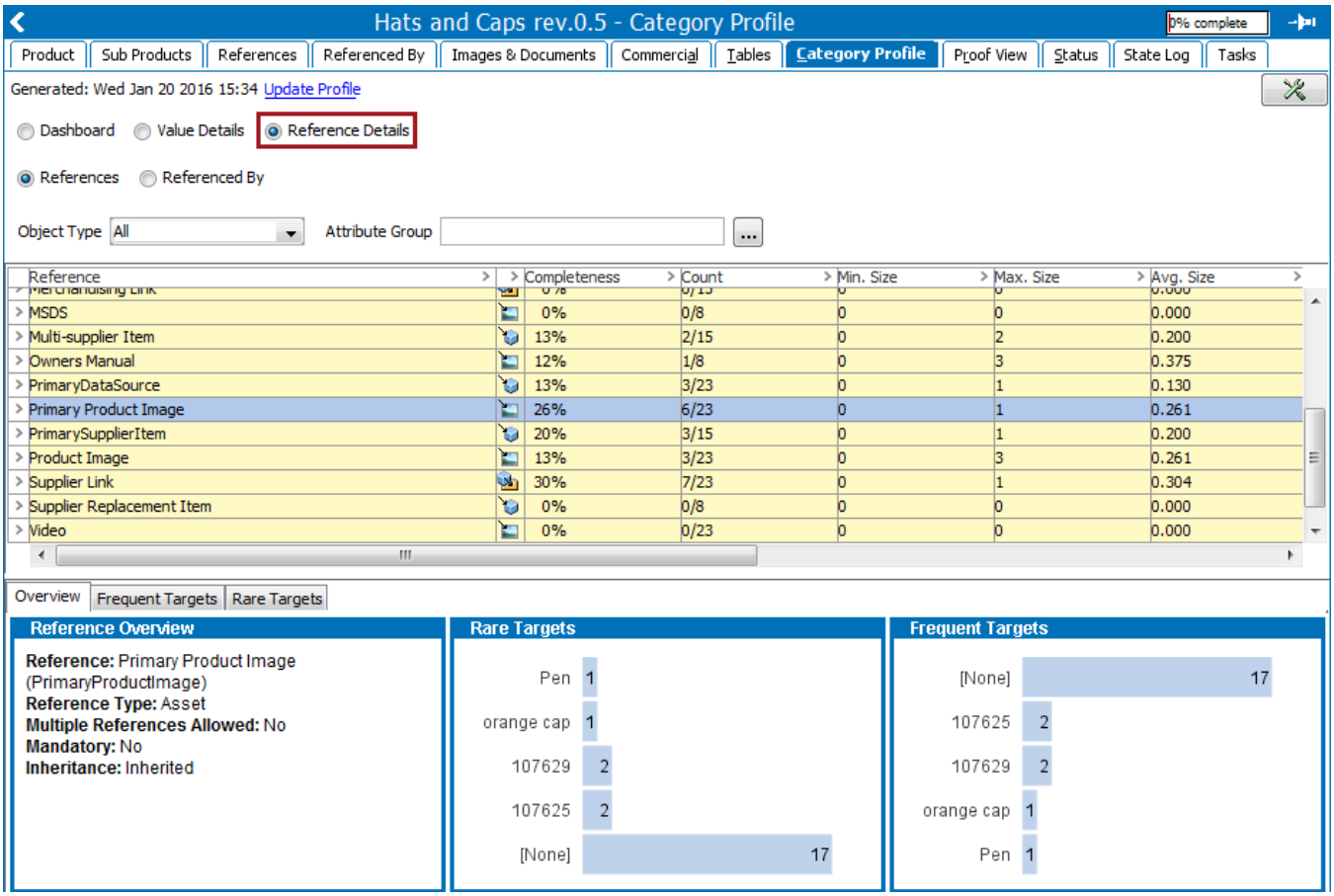
Reference Details View

The **Reference Details** view provides information about all references, image and document references, and product to classification links for which objects within the selected profile are valid sources.

Accessing and Navigating Reference Details View

1. In the **Tree**, select the relevant hierarchy or object.
2. On the **Category / Data Profile** tab, click **Generate Profile** or **Update Profile**, and then select **Reference Details**.

   The **Reference Details** page is divided into two areas, though at first only the upper area displays. The upper area lists all the references and links that are valid for objects in the profile. When you select a row, detailed information about the selected reference or link will then display in the lower area.



3. From the **Object Type** list, select a single object type or all object types. The number in parenthesis after the object type name shows how many instances of that object type are in the category.
4. To view references available to the objects, select **Referenced By**.
5. The user may filter their results so that only those attributes that belong to the specified attribute group will appear on the list. This can be done by entering the name of the specific attribute group into the **Attribute Group** field. Alternatively, you may also click the ellipsis button **(…)** and select the attribute group from the list that appears.

For information on the tabs in the lower area of this view, see the Reference Details Tabs section of the Data Profiling documentation.

Reference Table Details

The following list details the columns found in the reference list table. All columns can be filtered and sorted.

Column	Description
Reference	Displays the reference and link type titles. When clicking the name, the reference or link type

Column	Description
	definition in System Setup appears.
Column 2	Shows the Attribute Validation Base Type icon.
Completeness	Shows the degree to which the reference or link type is populated in the profile rounded to the nearest whole percent. Inherited references and links are included in the calculation. If one reference or link of the specified type exists on an object, it is counted as populated regardless of whether or not the type is configured to allow multiple references or links.
Count	Displays two numbers as a ratio. The first is the number of objects in the profile where the reference or link type is populated (including inherited references and links), and the second is the number of objects in the profile that are valid sources for the reference or link type.
Min. Size	Minimum number of references of this type on an object in the profiled data set.
Max. Size	Maximum number of references of this type on an object in the profiled data set.
Avg. Size	Average number of references of this type on objects in profiled data set.
Frequent Targets	Displays a comma-separated list of the most frequently referenced or linked targets for the reference or link type in the profile. The list is sorted in ascending order by frequency and includes inherited references and links.
Rare Targets	Displays a comma-separated list of the least frequently referenced or linked targets for the reference or link type in the profile. The list is sorted in ascending order by frequency and includes inherited references and links.

Reference Details Tabs

The lower area of Reference Details is divided into a number of tabs that contain detailed attribute information.

- Overview Tab
- Frequent Targets Tab
- Rare Targets Tab

The Frequent Targets and the Rare Targets tabs contain a table where three different actions are possible. From left to right, the icons represent the following actions: Bulk Update, Search, and Save as Collection .

- **Bulk Update** When the Bulk Update icon is clicked, a search is run within the profile for the selected reference or link type. The result is used as input data set for the Bulk Update wizard.

The Bulk Update wizard opens on the **Operations** step with the reference or link type selected in the **Add Reference** operation. Choose a target that is valid for the selected type.

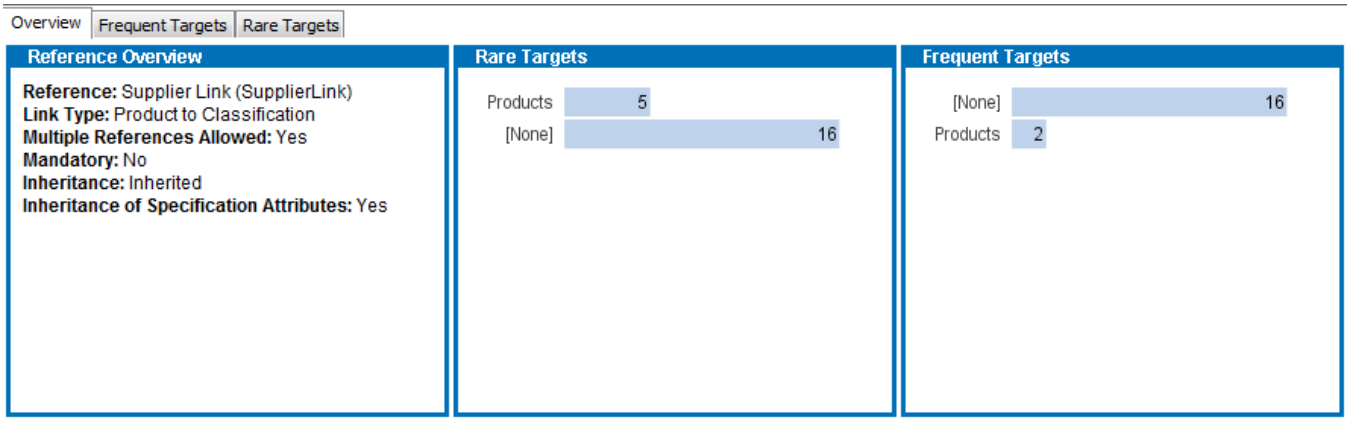
This functionality can be used to fix data errors. Since the input data set is the result of a search, you will not accidentally update the wrong data due to running the operation from an outdated profile.

For more information about the Bulk Update wizard, see the **Running Bulk Updates** section of the **Getting Started / STEP User Guide** documentation.

- **Search** - When the Search icon is clicked, a search is run within the profile for objects that have the selected reference or link type. When the search is complete, the result is displayed on the **Search** navigation tab. Searches are filtered on the object type selection in the profile.
- **Save as Collection** When the Save as Collection icon is clicked, the Save as Collection in Background Process dialog is displayed. Select a location and enter a Collection Name and a Collection ID, and then click **OK**. A search is run within the profile for objects that have the selected reference or link type, and the result is saved in the collection.

Overview Tab

The **Overview** tab contains a number of widgets that each present different information about the selected reference or link type. You cannot customize this view.



Widgets on the Overview tab.

Widget	Description
Overview	Displays information retrieved from the reference or link type definition.
Frequent Targets	Bar chart that displays the ten most frequently referenced or linked targets for the reference or link type in the profile. Includes inherited references and links. Targets with an equal count are presented in the order they are retrieved from the database.

Widget	Description
Rare Targets	Bar chart that displays the ten least frequently referenced or linked targets for the reference or link type in the profile. Includes inherited references and links. Targets with an equal count are presented in the order they are retrieved from the database.

The Frequent Targets Tab

On this tab, the 100 most common referenced / linked targets for the Reference / Link Type in the Category are listed. Inherited References/Links are included in the count. Note that targets with an equal count are presented in the order they are retrieved from the database.

The Rare Targets Tab

On this tab, the 100 least common referenced / linked targets for the Reference / Link Type in the Category are listed. Functionality-wise, the tab works exactly like the detail **Frequent Targets** tab described above.

Profile Configuration

Category Profiles can be configured with strict profiling parameters via the **Profile Configuration** screen. From this view, users can:

- Use a business condition to specify which objects to profile (allows you to exclude discontinued products, for example)
- Include or exclude data from specified attribute groups when profiling the data
- Designate business conditions to test when a profile is run
- Split up the value of a specified attribute so that the individual words in the attribute's value are profiled as separate entries
- Define which Completeness Metrics to use when profiling the data

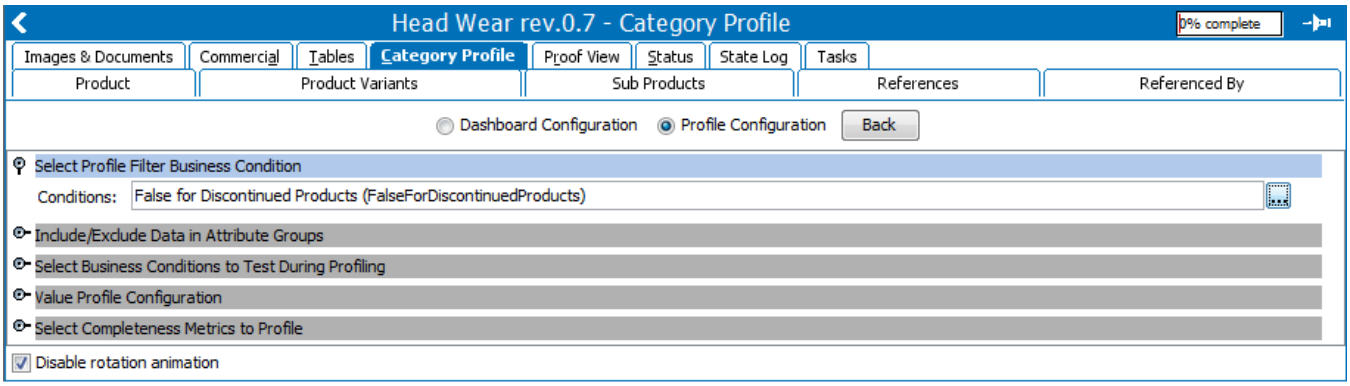
Note: The profile must be updated manually every time changes are made to any of the parameters.

Accessing the Profile Configuration

1. In the **Tree**, select the relevant hierarchy or object.
2. On the **Category Profile** tab, click the **Tools** icon in the upper right corner, and then select **Profile Configuration**.

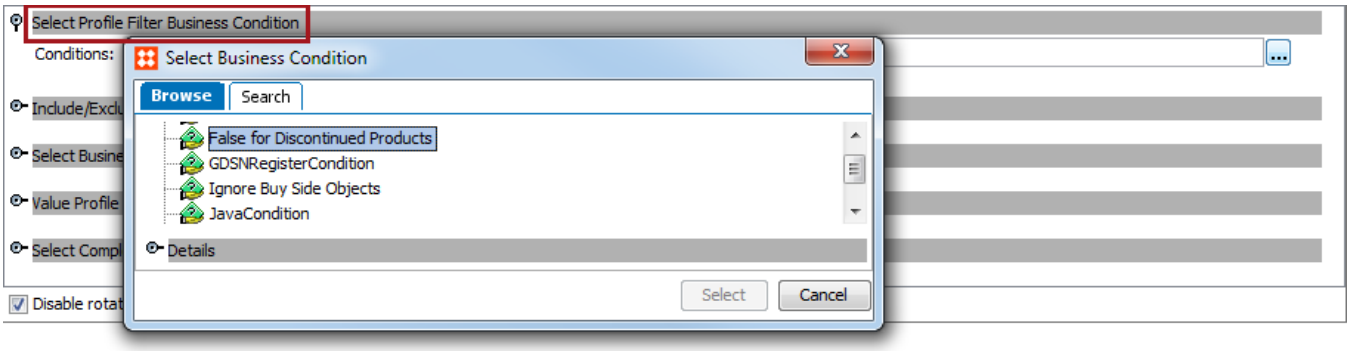
Select Profile Filter Business Condition

The **Select Profile Filter Business Condition** parameter allows users to filter the data being profiled with the specified business rule. This can be particularly useful if only objects of a specific type should be profiled or objects with specific attribute values should be left out.

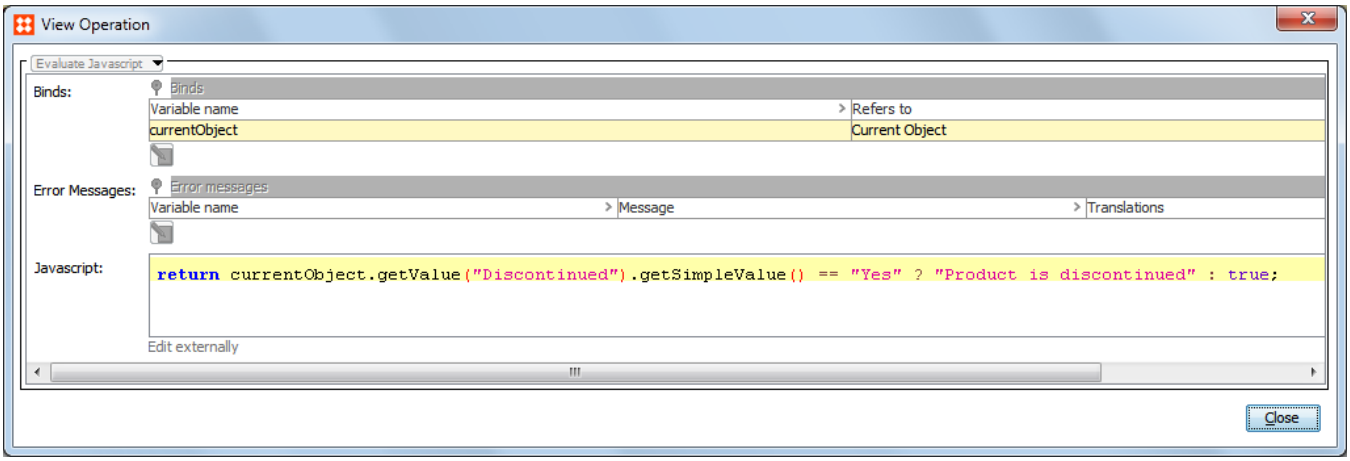


1. In the **Select Profile Filter Business Condition** section, click the ellipsis button, then browse or search for the relevant business condition.

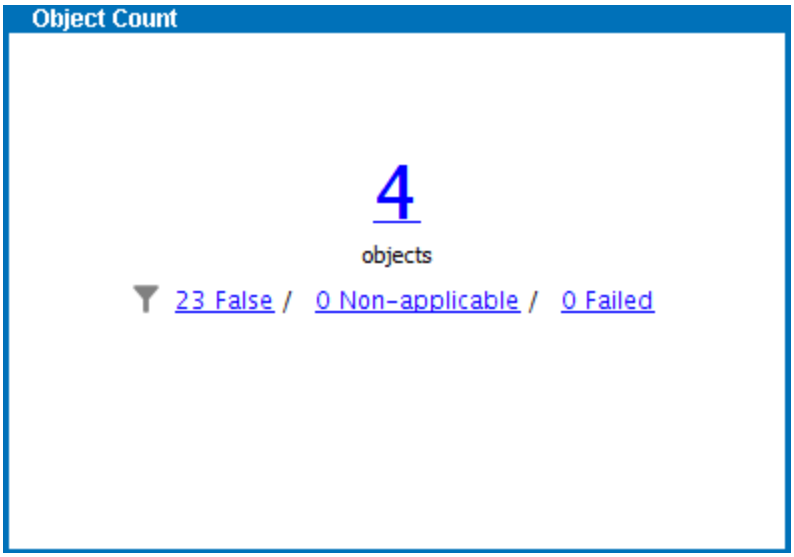
   As an example, in the following screenshot a business condition that excludes discontinued products is selected:



   The selected condition specifies that if the attribute **Discontinued** is set to **Yes**, the object is excluded from the profiling. The selected business condition is pictured below:



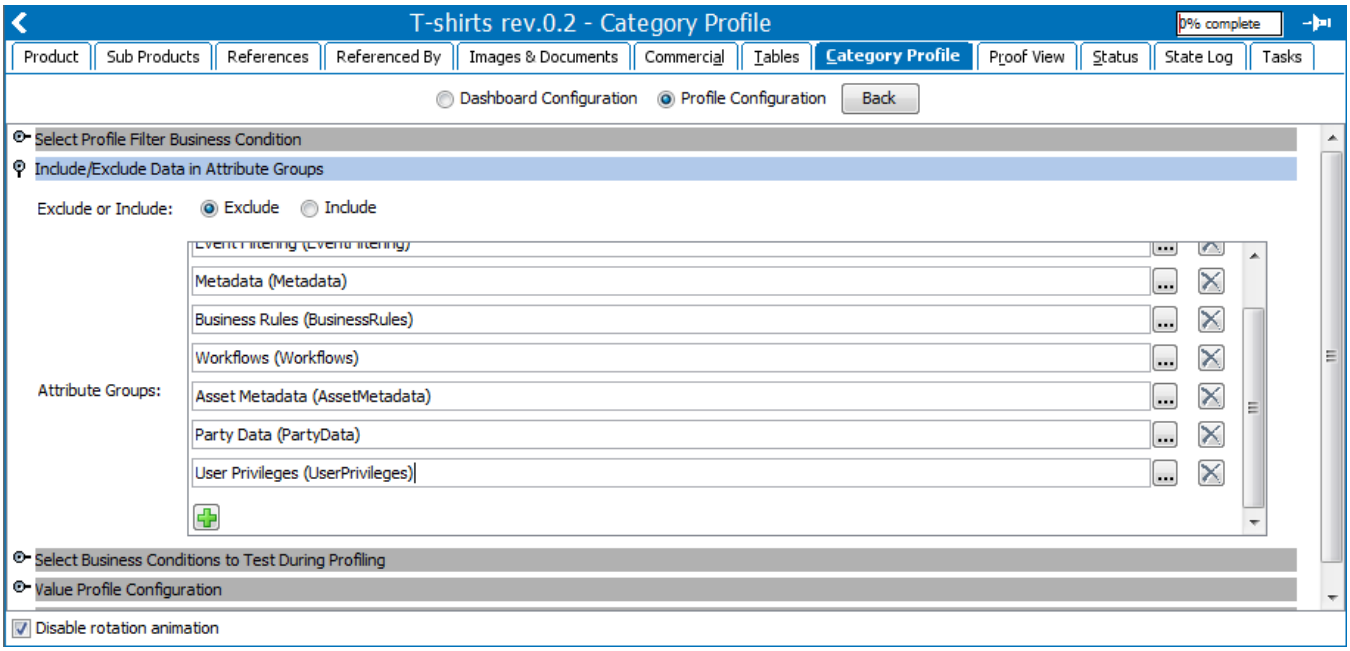
2. Click **Select** to apply the business condition to the profile.
3. Click the **Back** button to return to the **Category Profile Dashboard**, then click the **Update / Generate profile** link. If an **Object Count** widget is configured on the dashboard it will display the number of objects that were filtered out by the condition. In the example below, twenty-three discontinued objects were excluded as these were deemed False during profiling:



For more information on how to configure the Object Count widget, see the **Category Profile Dashboard** section of the **Data Profiling** documentation.

Include / Exclude Data in Attribute Groups

The **Include / Exclude Data in Attribute Groups** parameter allows users to filter the data being profiled via attribute group.



1. Navigate to the **Include / Exclude Data in Attribute Groups** section and click either the **Exclude** or **Include** radio button. If **Include** is selected, the data profile will only pull data from the attribute groups specified by the user in the Attribute Groups field. If **Exclude** is selected, the data profile will pull data from all attribute groups except those specified in the Attribute Groups field. Typically, the choice between **Include** and **Exclude** will largely depend on which requires the shorter list.

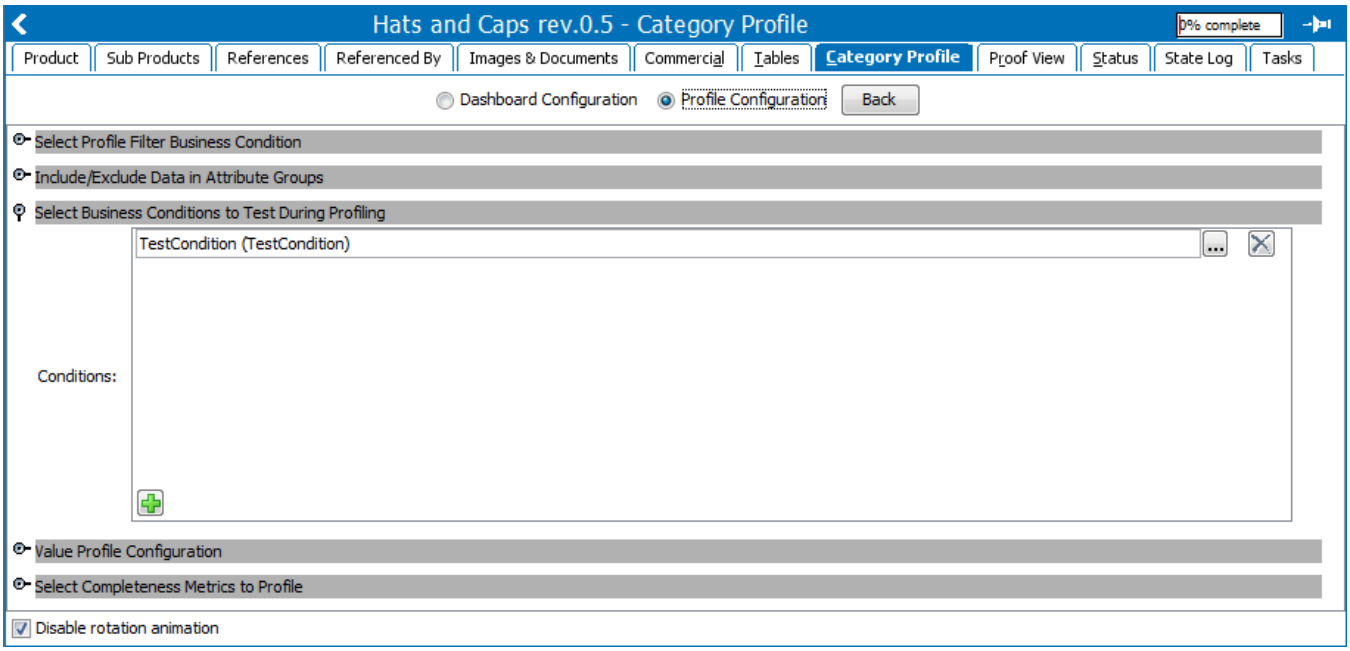
Note: To include all attribute groups in the system, the user can simply select **Exclude** and leave the Attribute Groups field blank.

---

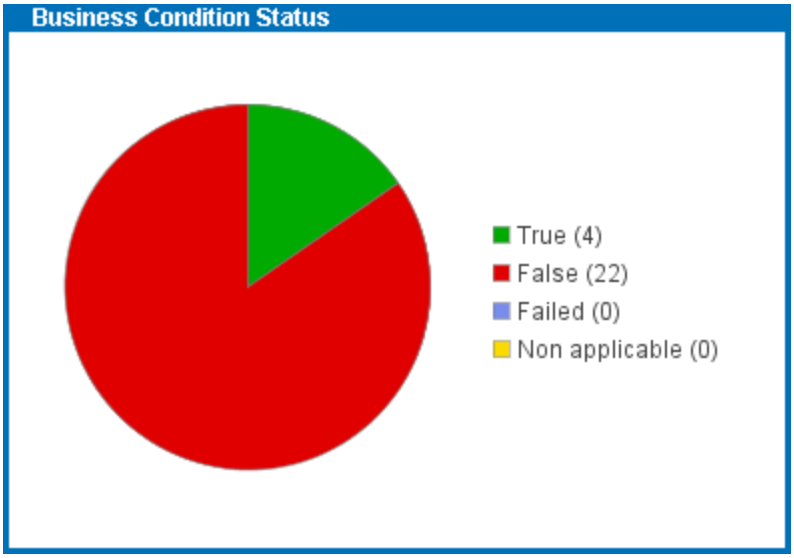
2. Click the green plus sign to add a new attribute group selector field.
3. Click the ellipsis button, then browse or search for the relevant attribute group.
4. Click **Select**. If no other attribute groups need to be specified, click the **Back** button to return to dashboard.

Select Business Conditions to Test During Profiling

The **Select Business Conditions to Test During Profiling** parameter allows users to test business conditions for all products, entities, classifications, and assets when profiling a selected hierarchy. The result of the tests can be displayed in Business Condition Status widgets on the dashboard.



1. Navigate to the **Select Business Conditions to Test During Profiling** section and click the green plus sign to add a new business condition to the selector field.
2. Click the ellipsis button, then browse or search for the relevant business condition.
3. Click **Select**. If no other business conditions need to be specified, click the **Back** button to return to dashboard. If the Business Condition Status widget is configured on the dashboard it will display the results of the test.

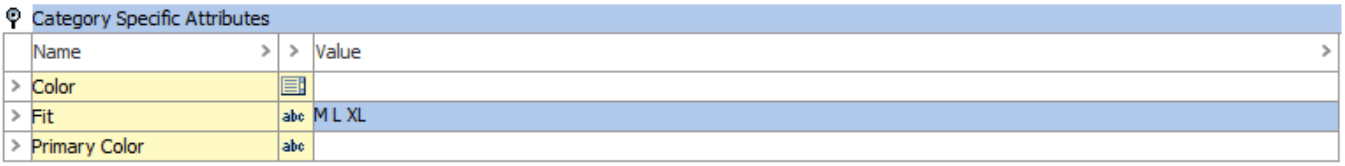


For more information on how to configure the Business Condition Status widget, see the **Category Profile Dashboard** section of the **Data Profiling** documentation.

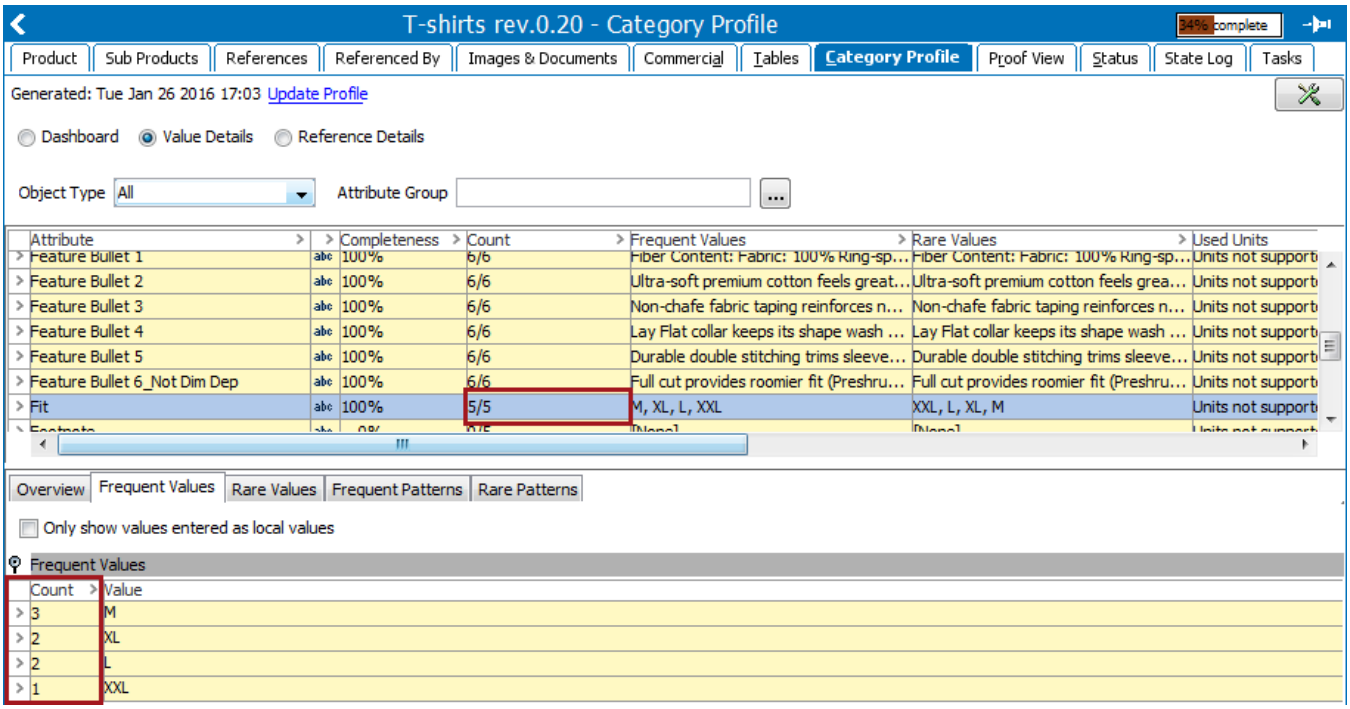
Value Profile Configuration

The **Value Profile Configuration** parameter allows users to profile the individual words of an attribute's value. As an example, 'Fit', an attribute with validation base type 'Text', lists all applicable shirt sizes for a product. These individual sizes can be profiled separately so that the attribute value 'M L XL' is profiled as three different values, 'M', 'L', and 'XL'.

Attribute value as seen on the product node:



Profiled values as seen in the Value Detail view:

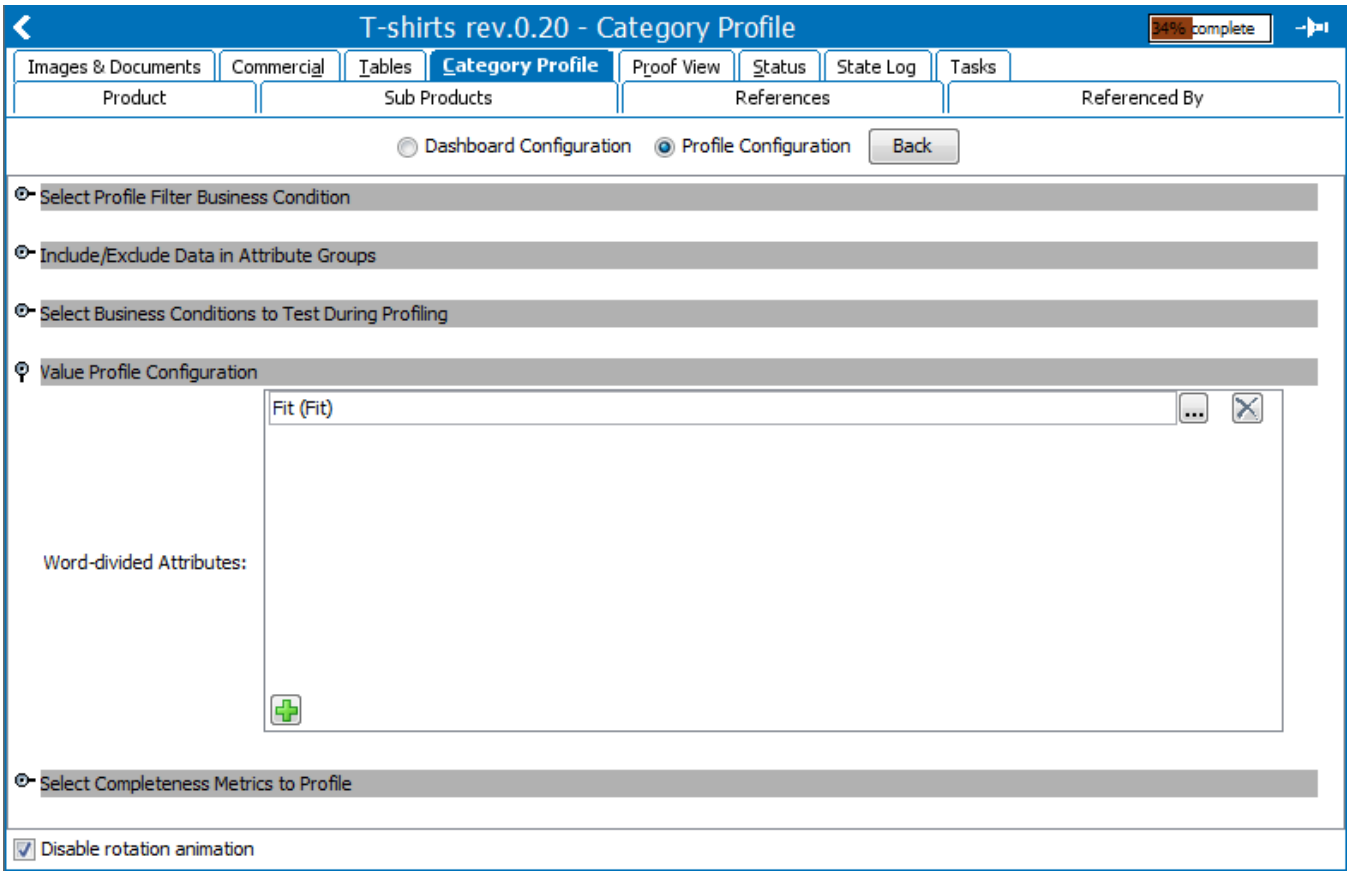


In addition to the **Value Details** view, this separation of values is also reflected on all applicable dashboard widgets.

For more information on the Value Details view, see the **Value Details** section of the **Category Profile** documentation.

Configuration Steps

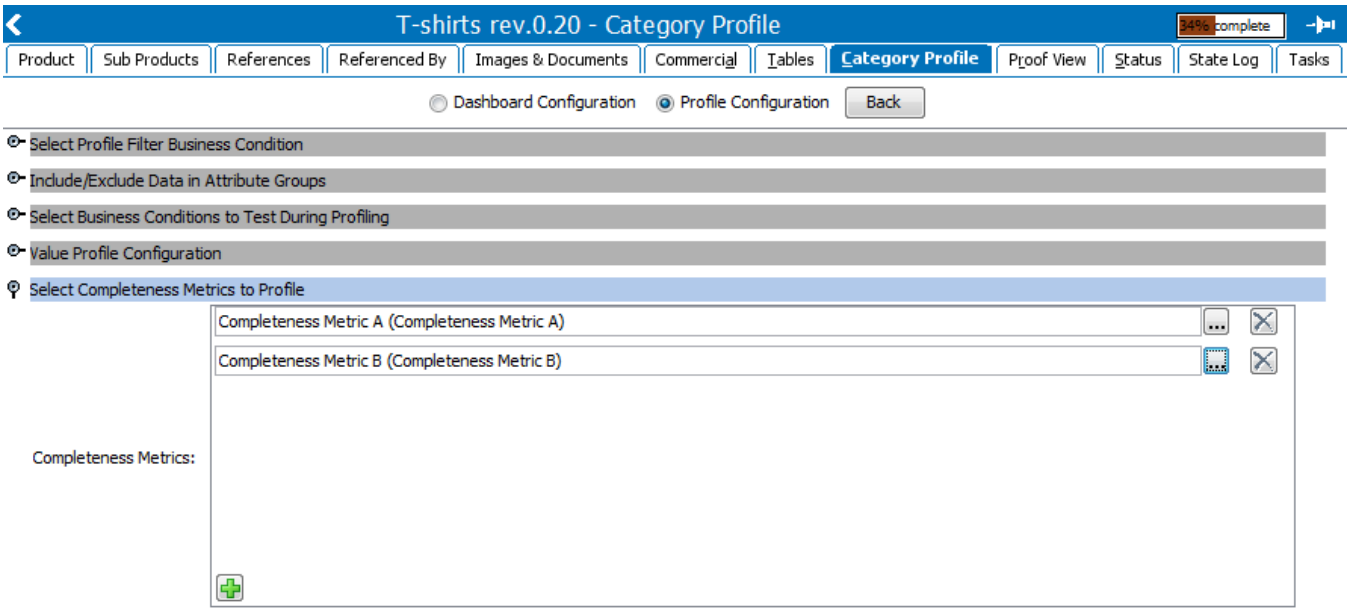
Select the attributes whose values should be profiled separately.



1. Navigate to the **Value Profile Configuration** section and click the green plus sign to add a new word-divided attribute to the selector field.
2. Click the ellipsis button, then browse or search for the relevant attribute.
3. Click **Select**. If no other attributes need to be specified, click the **Back** button to return to dashboard.

Select Completeness Metrics to Profile

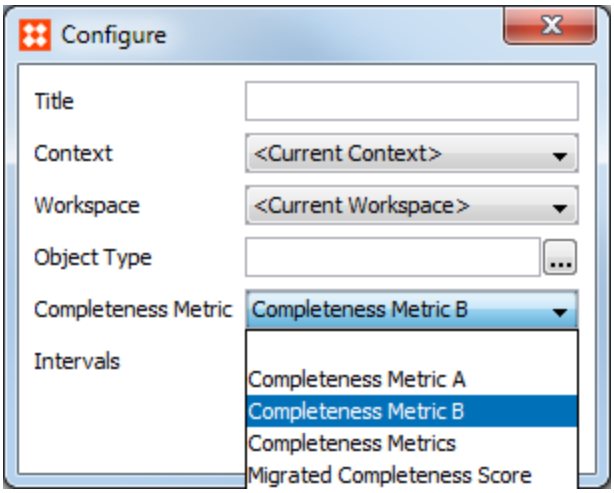
The **Select Completeness Metrics to Profile** parameter allows users to specify which completeness metrics to use when profiling the data.



1. Navigate to the **Select Completeness Metrics to Profile** section and click the green plus sign to add a new completeness metric to the selector field.
2. Click the ellipsis button, then browse or search for the relevant completeness metric.
3. Click **Select**. If no other completeness metrics need to be specified, click the **Back** button to return to dashboard.

If using the Average Object Completeness widget, ensure that this field includes whichever completeness metric is configured on the widget. If they do not match, a 'no data available' message will appear in place of the data on the widget.

Important: Even if the completeness metric specified by the widget is the same as the system default, it will still display the 'no data available' message. To use this widget the proper completeness metric must be specified in the Profile Configuration.



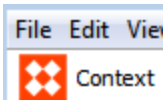
For more information see the **Available Widgets** section of the **Data Profiling** documentation.

Global Dashboard

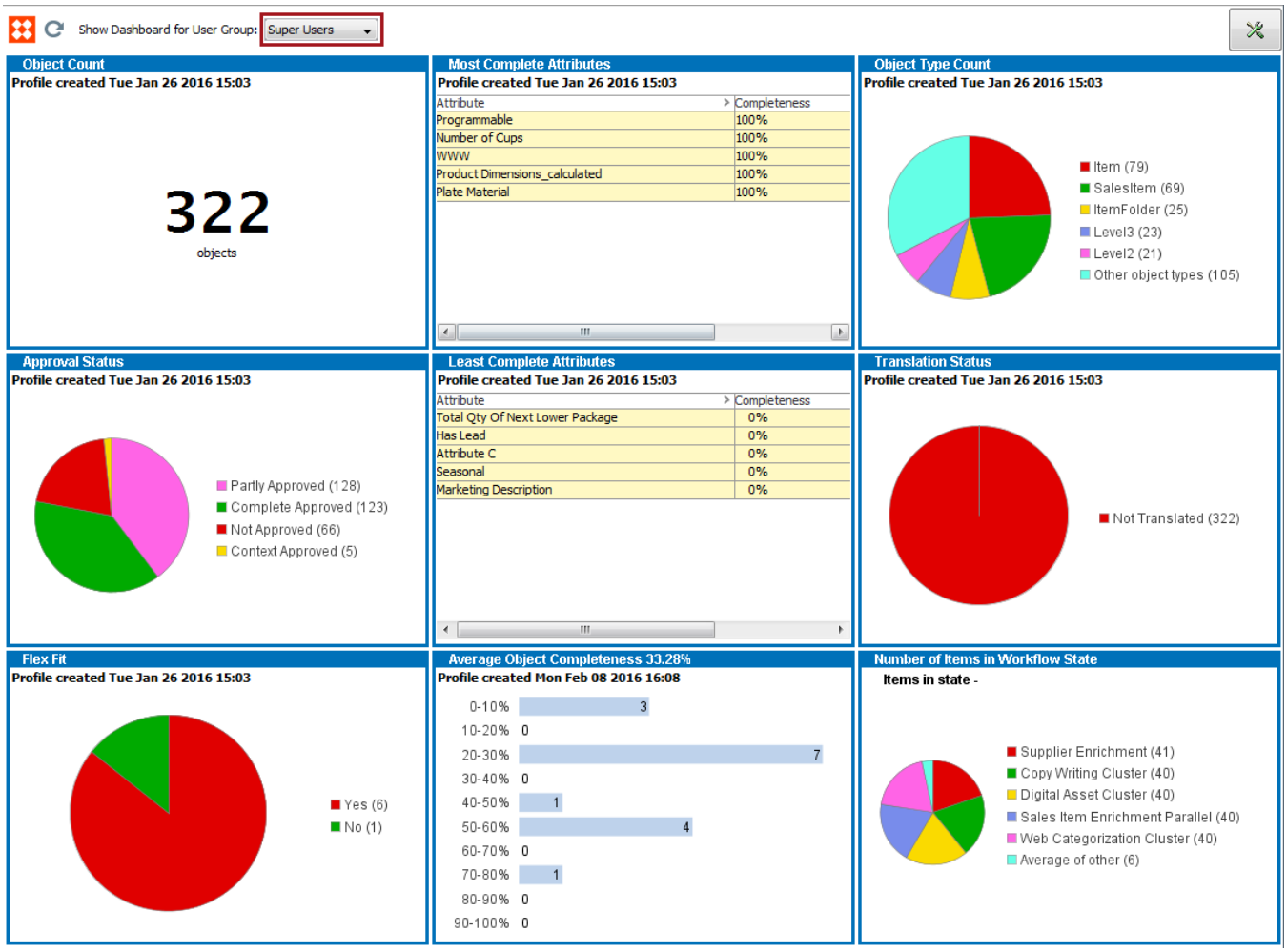
The global dashboard is a highly customizable view that allows users to display profiled data using a collection of configurable widgets. Multiple dashboards can be configured to display the most relevant data for specific user groups. Additionally, widgets can be configured to display data from specific category / data profiles. This allows users to view profile information without having to navigate through the **Tree** to locate the relevant profile.

Displaying the Global Dashboard

1. To view the dashboard, click the STEP logo icon in the upper left corner of the workbench.

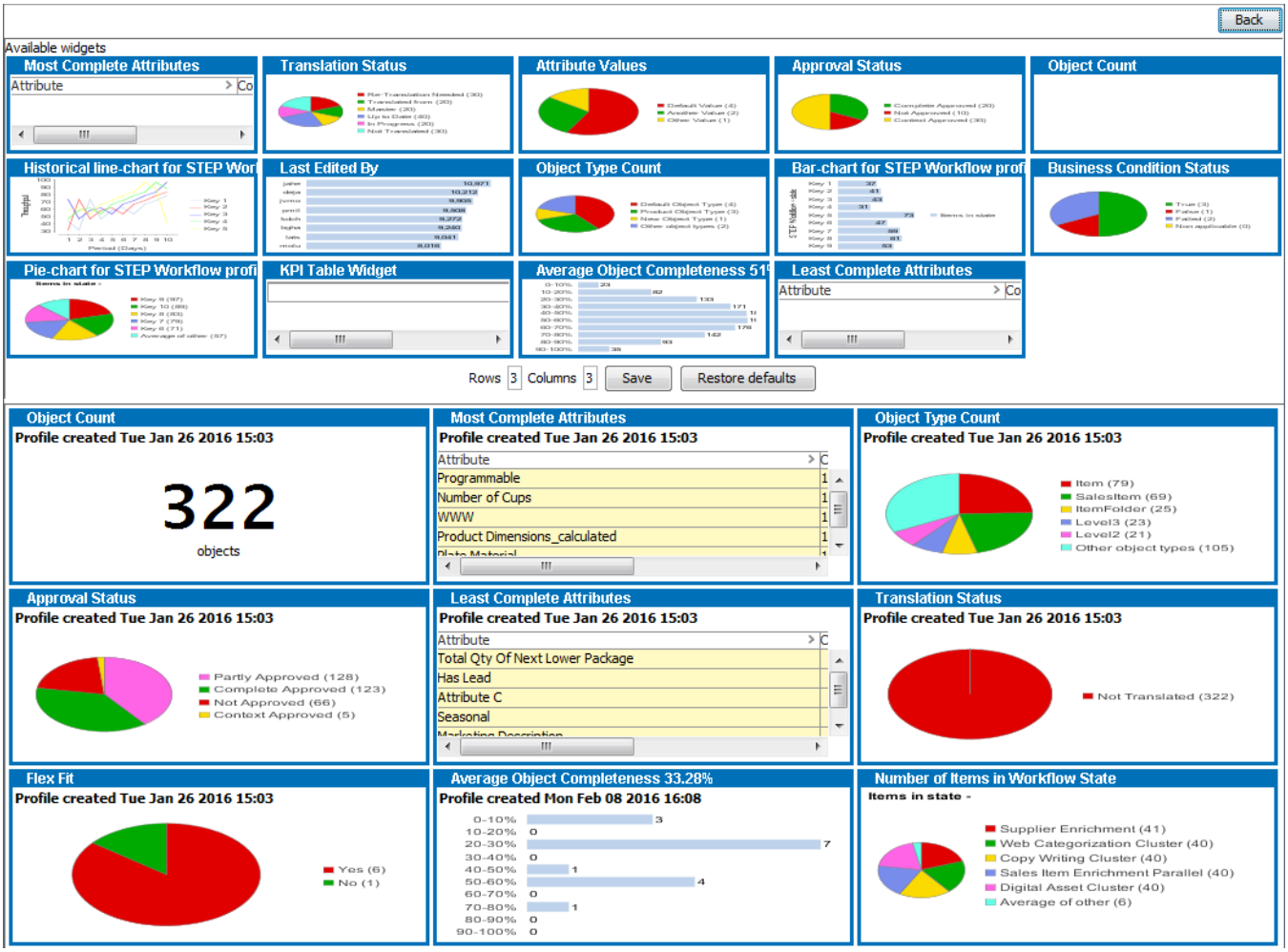


2. From the **Show Dashboard for User Group** list, select a user group you belong to view their dashboard.

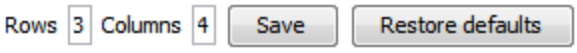


Customizing the Global Dashboard

1. On the **Global Dashboard**, select the preferred User Group, and then click the **Tools** icon . The configuration view of the dashboard is displayed. Note that the upper half of the view displays the available widgets, and the lower half is a preview of the dashboard. Adding these widgets to the dashboard is detailed in Step 3 below.



2. In the **Rows** and **Columns** fields, located between the available widgets and dashboard preview areas, specify how many rows and columns you want on the dashboard.



3. Configure the widgets for the currently selected dashboard:
   - To add a widget, left-click the header bar of the preferred widget and drag the widget to an empty area in the preview.

- To remove a widget, in the preview, right-click the header bar of the widget you want to remove, and then select **Remove**.
- To resize a widget, in the preview, right-click the header bar of the widget you want to re-size, mouse over **Width** or **Height**, and select the preferred value.
- To configure what data is displayed, in the preview, right-click the header bar of the applicable widget, and then select **Setup**.

For more information about configuring specific widgets, see the **Available Widgets** section of the **Data Profiling** documentation.

The main difference between configuring widgets for the global dashboard and for category /data profiles is that for each widget the user must specify which category / data profile to draw data from. The Workflow Profile widgets are specific to the global dashboard and are detailed below.

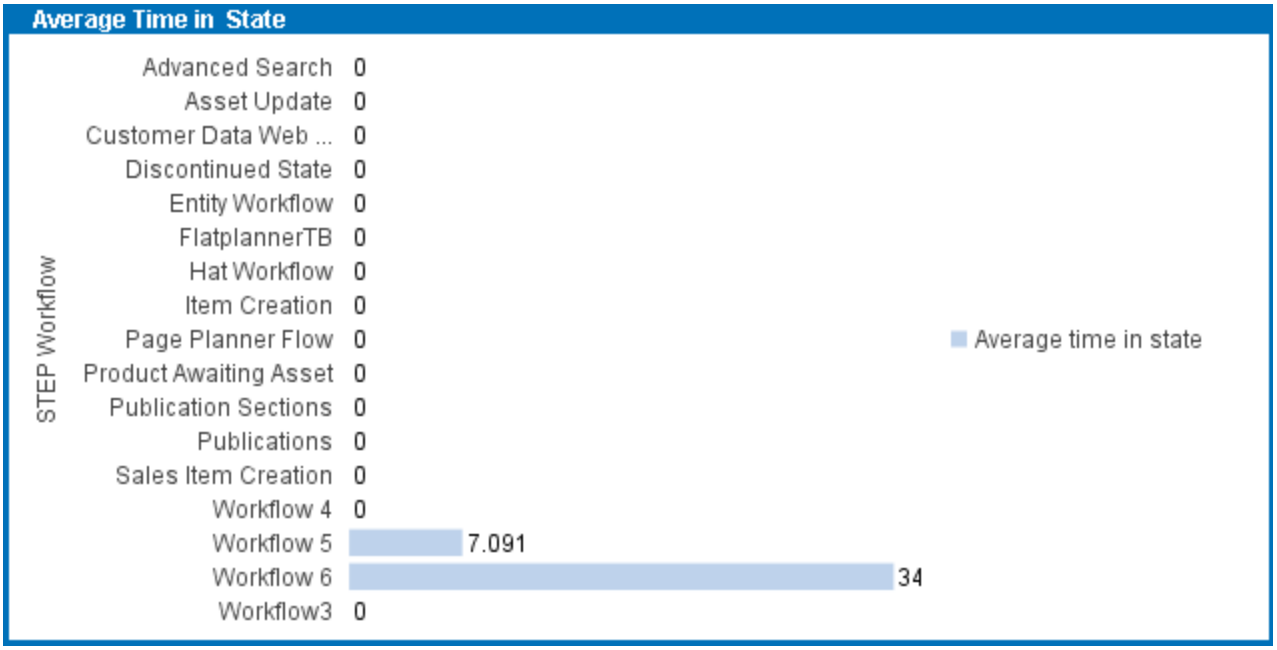
4. Once the dashboard has been configured, click **Save** and then click the **Back** button at the top to return to the main global dashboard view.
5. Click the **Refresh** icon for the changes to take effect.

Workflow Profile Widgets

Like the other widgets, the Workflow Profile widgets need to be configured to display the desired data.

Bar-chart for STEP Workflow Profile

This widget displays data for any available Workflow Profile in a bar graph.

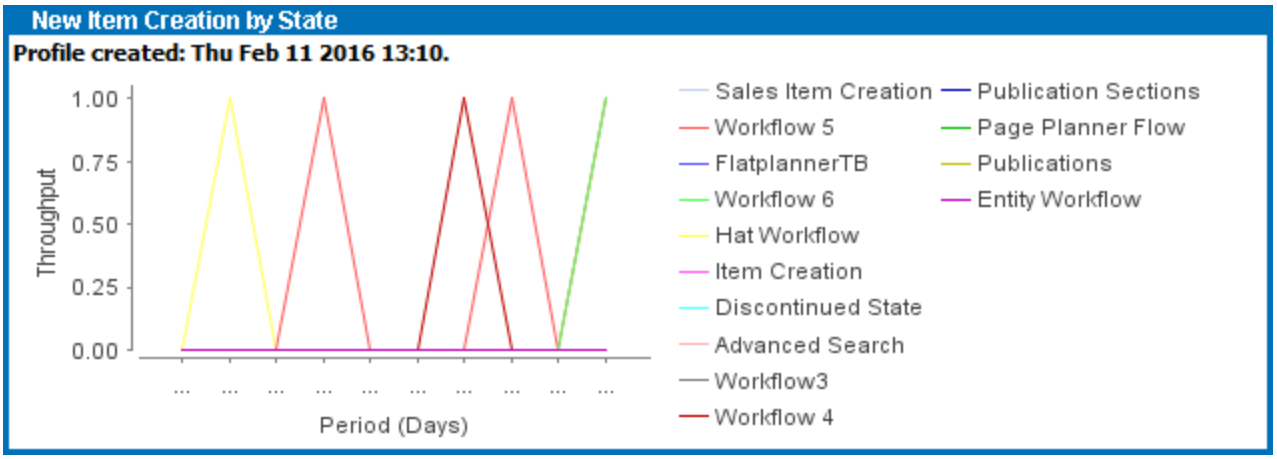


Configuration options include:

- **Chart Title** - Changes the name that appears on the widget's title bar.
- **Available Profiles** - Pulls data from the specified Workflow Profile.
- **Legend** - Determines which aspect of the Workflow Profile to model the data around (such as the workflows themselves or the states of the workflows).
- **Value** - Determines what values of the Workflow Profile to measure (such as 'Max Time in State', 'Items in State', etc.).
- **Orientation** - Determines if the bar graph displays vertically or horizontally.

Historical line-chart for STEP Workflow Profile

This widget displays historical data for any available Workflow Profile in a line graph. To ensure this widget is displaying up-to-date data navigate to the applicable STEP Workflow Profile node and click the **Update on-demand data** button.

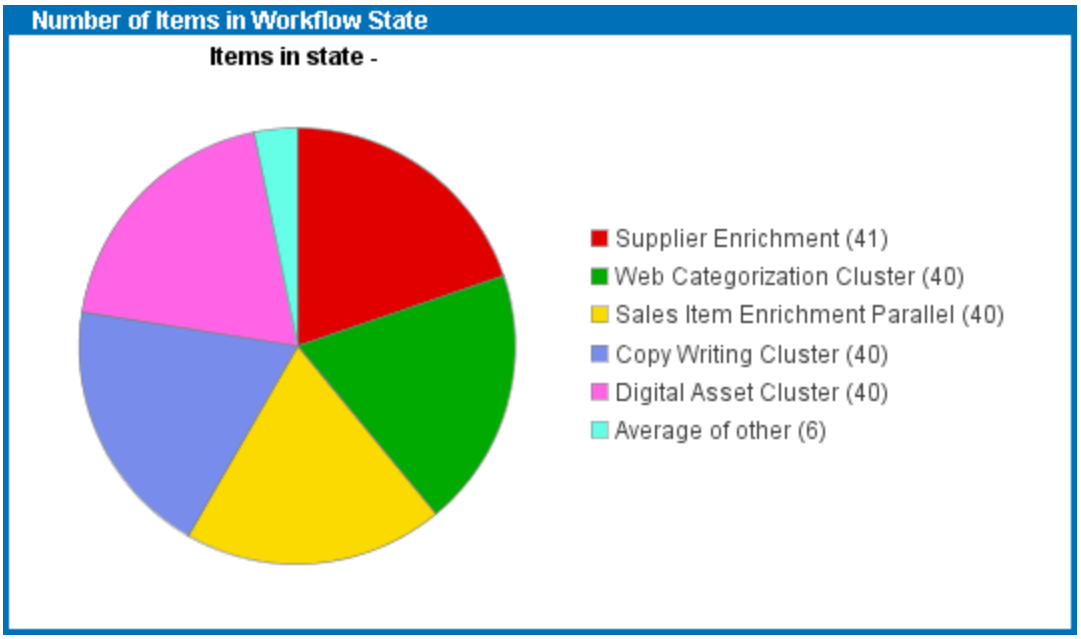


Configuration options include:

- **Chart Title** - Changes the name that appears on the widget's title bar.
- **Available Profiles** - Pulls data from the specified Workflow Profile.
- **Legend** - Determines which aspect of the Workflow Profile to model the data around (such as the workflows themselves or the states of the workflows).
- **Historical Column** - Determines which column of the Workflow Profile to display data from. The options may very based on how the individual profiles are configured in the system.

Pie-chart for STEP Workflow Profile

This widget displays data for any available Workflow Profile in a pie graph.



Configuration options include:

- **Chart Title** - Changes the name that appears on the widget's title bar.
- **Available Profiles** - Pulls data from the specified Workflow Profile.
- **Legend** - Determines which aspect of the Workflow Profile to model the data around (such as the workflows themselves or the states of the workflows).
- **Value** - Determines what values of the Workflow Profile to measure (such as 'Max Time in State', 'Items in State', etc.).
- **No. of pie-slices** - Determines the maximum number of pie slices that can be displayed in the chart.
- **Show average of the other** - Displays the average value for those workflows and/or states who's values were too low to be represented by themselves.

Limiting Memory Usage During Profiling

When large categories are profiled, the system uses a lot of memory. However, the following two config properties can be used to limit memory usage.

Note: Config properties can be set on the server by advanced users or by Stibo technical support.

Config properties

Use one of the following properties to limit memory usage.

`DataProfile.MaxDistinctAttributeValuesConsideredDuringProfileGeneration`

Sets the maximum number of distinct values to profile for attributes. The default setting is 100. When the limit is reached the following happens:

- Frequent value counts might become inaccurate. (STEP uses a counting implementation dedicated for counting in big data collections with a limited memory usage from Clearspring Analytics)
- The rare value count is disabled because only a frequent count can be maintained. In the profile, the frequent and rare values cells for attributes with too many distinct values are displayed with a light red background color. The attribute completeness and count and the value instance counts for profiled attributes are correct.

`DataProfile.MaxDistinctTargetsConsideredDuringProfileGeneration`

Sets the maximum number of distinct targets for the reference or link type that is profiled. The default setting is 100.

Attribute Value Profiles

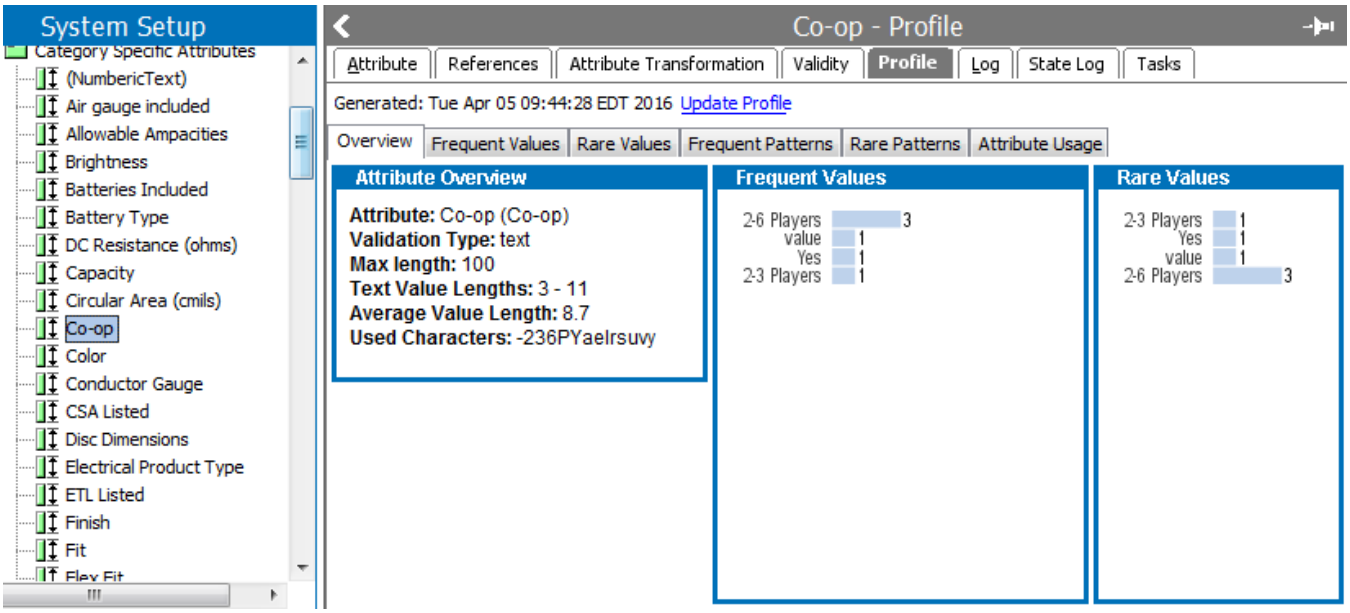
Attribute value profiles enable you to see where and how an attribute is used. The profile provides an overview of how attribute values are distributed across the current context and workspace, and enables users to easily clean up data.

Attribute profiling consists of two parts: Value Profiling and Attribute Usage.

- Value Profiling inspects all values for the selected attribute and collects information about these values.
- Attribute Usage uncovers how an attribute is used across business rules, calculated attributes, attribute transformations, integration endpoints, export and import configurations, key definitions, tables, and so on.

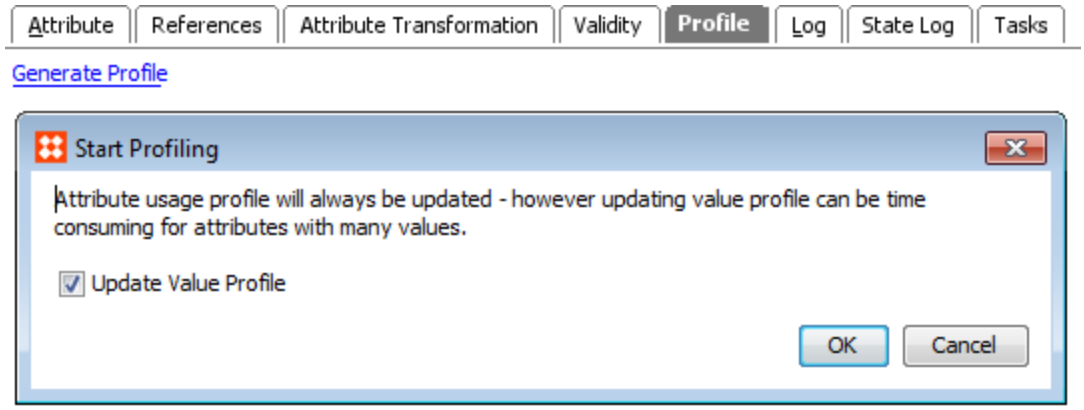
The value profile enables you to identify the most used values and the least used values. If the attribute is full text indexed, then all words are profiled as opposed to the entire value. Other aspects of the values are profiled as well.

Note: Value Profiles do not include inherited values, but do include values on objects in Recycle Bin.



View an Attribute Value Profile

1. In **System Setup**, expand **Attribute Groups**, and then select the preferred attribute.
2. On the **Profile** tab, click **Generate Profile** if no profile exists, otherwise click **Update Profile**. The **Start Profiling** message appears.



3. If you only need information about attribute usage, clear the **Update Value Profile** box, and then click **OK**.

Note: Collecting attribute values can be time-consuming depending on how prolific the attribute is in the system.

Pre-existing value profiles are not deleted if you choose only to update attribute usage. The profile is ready once the background process completes.

Config Property

To limit the system memory usage when profiling attributes, you can use the following config property:

```
AttributeProfile.MaxDistinctValuesConsiduredDuringAttributeProfileGeneration
```

The property limits the number of distinct values to be profiled. The property defaults to 1.000.

Attribute Value Profile Tabs

The information provided by the Attribute Value Profile is divided into a number of tabs:

- Overview
- Frequent Values
- Rare Values
- Frequent Patterns
- Rare Patterns
- Attribute Usage

Actions

All tabs, except the Overview and Usage tab, contain a table where different actions are possible. From left to right the icons represent the following functionality: Bulk Update, Search, Save as Collection, and Export Frequencies .

- **Bulk Update** When you click the Bulk Update icon, a search is run for objects that have the selected attribute value. The result is used as input data set for the Bulk Update wizard.

  The Bulk Update wizard opens on the **Operations** step with the attribute selected in the **Set Value** operation.

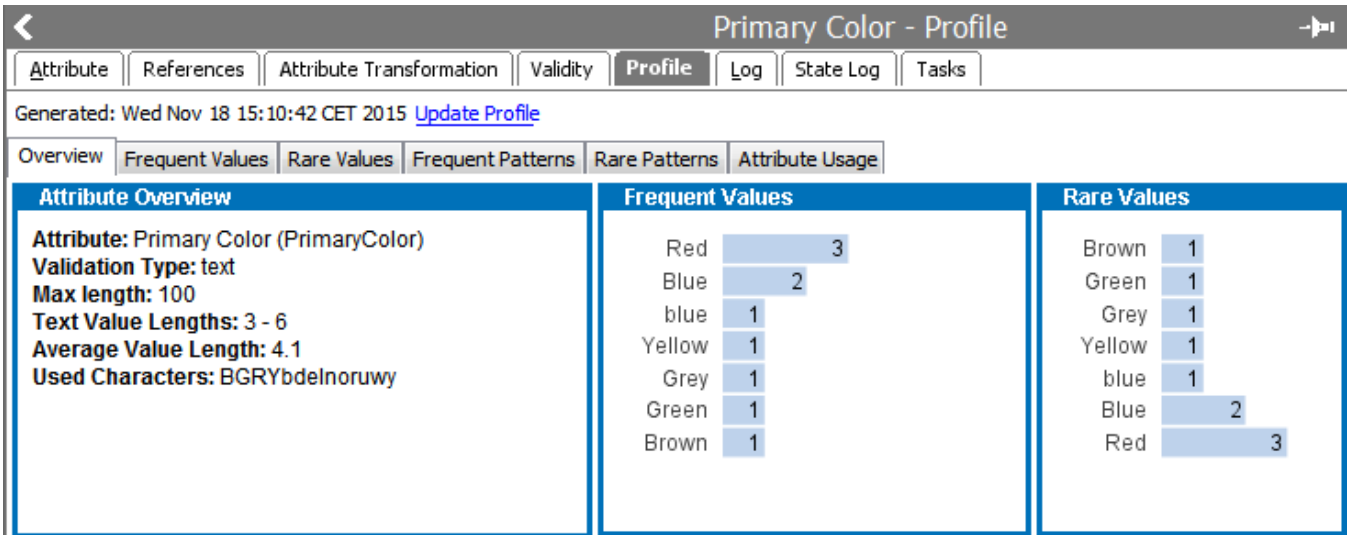
  This functionality makes it easy for you to correct data errors, and because the input data set is the result of a search, you can not accidentally update the wrong data.

  For more information about the Bulk Update wizard, see the **Running Bulk Updates** section of the **Getting Started / STEP User Guide** documentation.

- **Search** - When you click the Search icon, a search is run for products that have the selected attribute value. When the search is complete, the result is displayed on the **Search** navigation tab. Searches are filtered on the object type selection in the profile.
- **Save as Collection** When you click the Save as Collection icon, the Save as Collection in Background Process dialog is displayed. Select a location and enter a Collection Name and a Collection ID, and then click OK. A search is run within for products that have the selected attribute value, and the result is saved in the collection.
- **Export Frequencies** When you click the Export Frequencies button, the Export Frequencies dialog is displayed. Select output file and if you want absolute or relative frequency. A CSV file with semi colon separated frequencies, will be create on the local system.

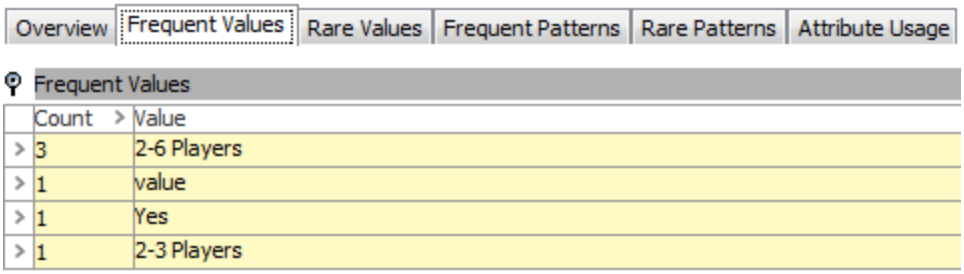
The Overview Tab

The **Overview** tab contains a number of widgets that each present different information about the selected attribute. Users cannot customize this view. Inherited values are not included.



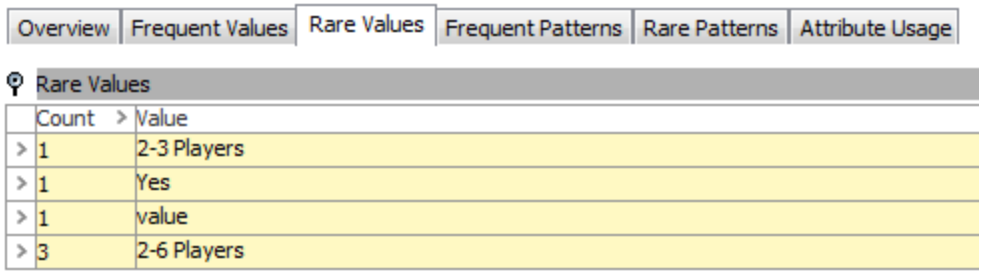
The Frequent Values Tab

On this tab, the 100 most frequently appearing values of the selected attribute are listed. Tags are resolved.



The Rare Values Tab

On this tab, the 100 least frequently appearing values of the selected attribute are listed.



The Frequent Patterns Tab

This tab lists the most frequent patterns of the values of a given attribute and the number values that match the pattern.

A value pattern is a way of describing the structure of a value in terms of characters, numbers, and symbols. When the patterns of the values in a set are compared, certain types of errors can be detected by finding values whose patterns are different from the rest. For example, phone numbers with missing digits or numbers that are mixed with letters, and so on.

A pattern for a given value is created by replacing any letter with A, any number with 9, and any unrecognized character with X. Other symbols are left as they are.

The following is an example of a value pattern:

Value	Pattern
123-ABB	999-AAA
234.23	999.99
mail@mail.com	AAAA@AAAA.AA

The pattern [None] represents an empty value.

The Rare Patterns Tab

This tab resembles the Frequent Patterns but lists the least frequently used patterns rather than the most frequent. This information is useful when searching for faulty values.

The Attribute Usage Tab

On this tab users can view all the places where the attribute is used. The attribute usage can be qualified as either Used or Possibly Used. Possibly Used means there is a good indication that the attribute is used but the system cannot detect it with 100% certainty. This is, for example, the case in calculated formulas if the ID of the attribute is found in the formula text.