

**Oracle® Enterprise Data Quality for Product Data**

AutoBuild Reference Guide

Release 11g R1 (11.1.1.6)

**E29129-02**

February 2013

Oracle Enterprise Data Quality for Product Data AutoBuild Reference Guide, Release 11g R1 (11.1.1.6)

E29129-02

Copyright © 2001, 2013, Oracle and/or its affiliates. All rights reserved.

Primary Author: Lorna Vallad

This software and related documentation are provided under a license agreement containing restrictions on use and disclosure and are protected by intellectual property laws. Except as expressly permitted in your license agreement or allowed by law, you may not use, copy, reproduce, translate, broadcast, modify, license, transmit, distribute, exhibit, perform, publish, or display any part, in any form, or by any means. Reverse engineering, disassembly, or decompilation of this software, unless required by law for interoperability, is prohibited.

The information contained herein is subject to change without notice and is not warranted to be error-free. If you find any errors, please report them to us in writing.

If this is software or related documentation that is delivered to the U.S. Government or anyone licensing it on behalf of the U.S. Government, the following notice is applicable:

U.S. GOVERNMENT RIGHTS Programs, software, databases, and related documentation and technical data delivered to U.S. Government customers are "commercial computer software" or "commercial technical data" pursuant to the applicable Federal Acquisition Regulation and agency-specific supplemental regulations. As such, the use, duplication, disclosure, modification, and adaptation shall be subject to the restrictions and license terms set forth in the applicable Government contract, and, to the extent applicable by the terms of the Government contract, the additional rights set forth in FAR 52.227-19, Commercial Computer Software License (December 2007). Oracle America, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

This software or hardware is developed for general use in a variety of information management applications. It is not developed or intended for use in any inherently dangerous applications, including applications that may create a risk of personal injury. If you use this software or hardware in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure its safe use. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software or hardware in dangerous applications.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group.

This software or hardware and documentation may provide access to or information on content, products, and services from third parties. Oracle Corporation and its affiliates are not responsible for and expressly disclaim all warranties of any kind with respect to third-party content, products, and services. Oracle Corporation and its affiliates will not be responsible for any loss, costs, or damages incurred due to your access to or use of third-party content, products, or services.

---

---

# Contents

<b>Preface</b> .....	v
Audience .....	v
Documentation Accessibility .....	v
Related Documents .....	vi
Conventions .....	vi
<b>1 Overview</b>	
<b>Quantity of Structured Information</b> .....	1-2
<b>AutoBuild and Smart Glossaries</b> .....	1-2
Default Smart Glossary .....	1-2
Smart Glossary Foundation for Generated Data Lens .....	1-3
Replacing Item Definition Model Phrase structures with Smart Glossary Phrase Structures	
1-3	
<b>Prerequisites to Using AutoBuild</b> .....	1-3
PIM Metadata Prerequisites .....	1-3
Non-PIM Metadata Prerequisites .....	1-4
Identifying Structured Item Information .....	1-4
Sources of Structured Item Content .....	1-5
Supported Structured item Information Formats .....	1-5
Supported Category Formats .....	1-5
Supported Attribute Formats .....	1-6
<b>2 AutoBuild User Interface</b>	
<b>The AutoBuild User Interface</b> .....	2-1
Step 1: Identify Category Information .....	2-3
Category Source Information Section .....	2-3
Category / Attribute Metadata Table .....	2-6
Category Details .....	2-6
Step 2: Identify Attribute Information .....	2-8
Attribute Information Section .....	2-8
Required Attributes Box .....	2-9
Scoring Attributes Box .....	2-13
Optional Attributes Box .....	2-16
Category / Attribute Metadata Table .....	2-17
Attribute Details .....	2-17

Attribute Values Box .....	2-18
Standardizations Box.....	2-19
Step 3: Identify Sample Data Information .....	2-19
Select Sample Data Source Section .....	2-19
Category/Attribute Metadata Table .....	2-21
Sample Data Table .....	2-21
Step 4: Define Data Lens Output .....	2-22
Outputs Box .....	2-22
AutoBuild Item Definition Explorer .....	2-23
Output Details .....	2-24
<b>AutoBuild Report</b> .....	<b>2-25</b>

### 3 The AutoBuild Process

<b>Using AutoBuild with PIM Metadata</b> .....	<b>3-1</b>
AutoBuild with Fusion PIM Metadata.....	3-1
AutoBuild with R12 PIM Metadata .....	3-1
<b>Using AutoBuild with Non-PIM Metadata</b> .....	<b>3-1</b>
Selecting the Data Lens Type .....	3-3
Identifying the Category Information.....	3-3
Selecting the List of Category Columns .....	3-3
Selecting the Category Column Layout.....	3-3
Identifying the Attribute Information.....	3-4
Selecting the Required Attributes Information .....	3-4
Selecting the Scoring Attributes Information .....	3-5
Identifying the Sample Data Information.....	3-6
Selecting the Sample Data Source.....	3-6
Defining the Data Lens Output.....	3-7
<b>Finishing Your New Data Lens</b> .....	<b>3-10</b>
<b>Data Pairs Data Lens Example</b> .....	<b>3-11</b>
<b>Further Considerations</b> .....	<b>3-13</b>

---

---

# Preface

This reference guide is intended to explain the capabilities of the Oracle Enterprise Data Quality for Product Data AutoBuild Application. The document is organized as follows:

- Chapters 1 and 2 describe the basic application features.
- Chapter 3 describes the AutoBuild process using a metadata from various sources.

Review the following Oracle Enterprise Data Quality for Product Data documentation prior to the use of this manual is recommended:

- *Oracle Enterprise Data Quality for Product Data Knowledge Studio Reference Guide*

To understand all of the features presented, you must use this reference guide in conjunction with the Oracle Enterprise Data Quality for Product Data documents listed in "[Related Documents](#)" on page 2-vi.

## Audience

You should have a basic understanding of the DataLens Technology. This includes the Knowledge Studio (recognizing and standardizing data) and use of the Smart Glossaries.

A thorough understanding of the material in this guide is required for the following customer personnel:

- Business Analysts trained in either the Oracle Enterprise Data Quality for Product Data Knowledge Studio or Application Studio.
- Subject matter experts (SMEs), including Product Specialists and Category Managers, trained in the Oracle Enterprise Data Quality for Product Data Knowledge Studio.
- IT Workflow Analysts trained in the use of the Oracle Enterprise Data Quality for Product Data Application Studio
- Application and Solution Owners

## Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at

<http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

## Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=info> or visit <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=trs> if you are hearing impaired.

## Related Documents

For more information, see the following documents in the documentation set:

- The *Oracle Enterprise Data Quality for Product Data Getting Started Guide* provides information about how to get started with EDQP and use the Application Launch Pad to start client applications.
- The *Oracle Enterprise Data Quality for Product Data Application Studio Reference Guide* provides information about creating and maintaining Data Service Applications (DSAs).
- The *Oracle Enterprise Data Quality for Product Data Knowledge Studio Reference Guide* provides information about creating and maintaining data lenses.
- The *Oracle Enterprise Data Quality for Product Data Services for Excel Reference Guide* provides information about creating a DSA based on data contained in a Microsoft Excel worksheet.

See the latest version of this and all documents in the Oracle Enterprise Data Quality for Product Data Documentation Web site at

[http://docs.oracle.com/cd/E35636\\_01/index.htm](http://docs.oracle.com/cd/E35636_01/index.htm)

## Conventions

The following text conventions are used in this document:

Convention	Meaning
<b>boldface</b>	Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.
<i>italic</i>	Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.
monospace	Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, text that you enter, or a file, directory, or path name.
<b>monospace</b>	Boldface, monospace type indicates commands or text that you enter.

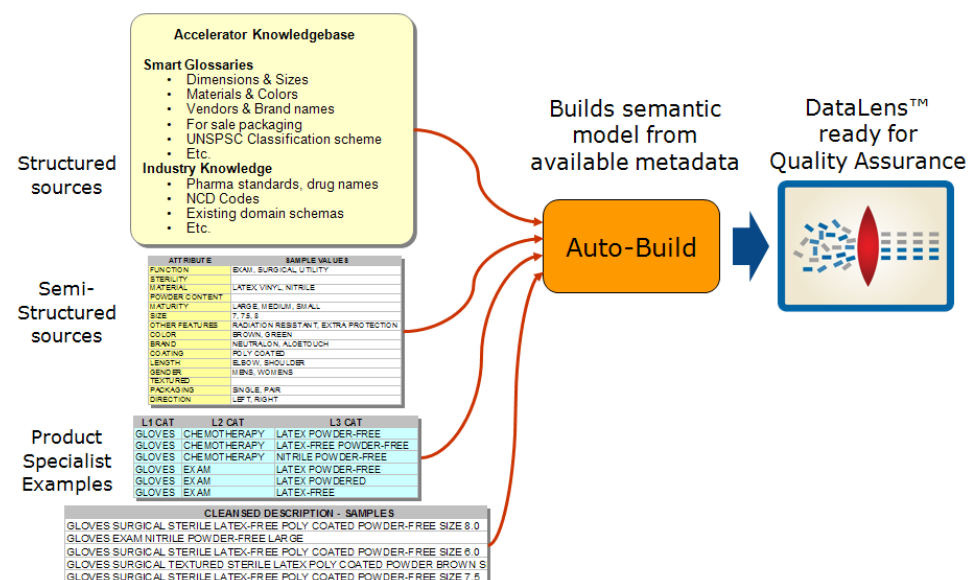
# Overview

The AutoBuild application is included in the Services for Excel additionally installed (add-in) product, which provides a custom toolbar that is added to Excel. For more information about this product, see *Oracle Enterprise Data Quality for Product Data Services for Excel Reference Guide*.

Autobuild can rapidly leverage your existing product information to create initial data lenses specific to your enterprise content. The use of Oracle Enterprise Data Quality for Product Data (EDQP) Smart Glossaries allows cross-domain knowledge to automatically be recognized in your data lenses. For example, a company with writing instruments inventory may have structured content containing information about each product category. This data knowledge can be transformed from an Excel worksheet into a data lens by the AutoBuild Application saving a significant amount of effort and cost.

Autobuild constructs the initial data lens using the structured category and attribute information. Given sufficient information, the AutoBuild application can accomplish the following:

- Construct a full Item Definition hierarchy, complete with required, scoring, and optional attributes.
- Construct rich term and phrase recognition rules.
- Provide an initial set of classification rules.



The AutoBuild application significantly reduces the level of effort required to:

- Create an initial data lens.
- Align initial data lenses to any business domain.
- Leverage the data lenses to automatically meet enterprise data requirements across a variety of internal and external data sources.

AutoBuild offers a familiar, easy-to-use graphical wizard interface to step you through the process from start to finish.

## Quantity of Structured Information

AutoBuild can use as much or as little structured information as you have available. If you only have a list of categories, AutoBuild creates a base data lens that contains a set of conforming Item Definitions from the provided categories. If you have attribute names associated with the categories, then AutoBuild can add attributes to the data lens Item Definitions and create the associated phrase rules. If you have attribute values associated with the attributes, then AutoBuild creates terminology (term) rules to recognize the attribute information and automatically associate the phrases with the correct attributes in the Item Definitions. AutoBuild makes the best use of any structured product information you provide.

AutoBuild works best when you provide a few examples of structured items for each category. These example items should represent the product categories and attributes you want to capture in your enterprise data.

## AutoBuild and Smart Glossaries

Enterprise DQ for Product Smart Glossaries are data lenses designed to be applied to a broad range of data domains. The Smart Glossaries delivered with the product recognize materials, colors, and units of measure for example. Each Smart Glossary is designed to be imported into an existing data lens, or can be used as the basis for creating a new data lens. Individual Smart Glossaries can be combined together to create a template that is used in the Autobuild process. Smart Glossaries can be created and edited using the Knowledge Studio. For more information about Smart Glossaries, see *Oracle Enterprise Data Quality for Product Data Knowledge Studio Reference Guide*.

When AutoBuild creates a new data lens from your structured item content, a combination of Smart Glossaries is used as the basis for the generated data lens.

## Default Smart Glossary

The default Smart Glossary used by AutoBuild is named `DLS_Import_Template`. This Smart Glossary is installed along with the other pre-packaged Smart Glossaries on the server. This Smart Glossary is a composite of glossaries from units of measure, counts, colors, materials and finishes, and product packaging.

Each time you create a data lens using Autobuild, you can select a different Smart Glossary to be used in the autobuild process. You can also create and configure your own combination of Smart Glossaries that are most applicable to your domain data for use by AutoBuild. Although only one Smart Glossary can be applied when creating a data lens with AutoBuild, additional Smart Glossaries can be imported after the data lens has been autobuilt and opened.

The Smart Glossary used by AutoBuild to generate a data lens is also known as a data lens template.



---



---

**Note:** The default Smart Glossary, `DLS_Import_Template`, is automatically available. You must check out any other Smart Glossary you want to use as a template to generate your data lens, from your Oracle DataLens Server. For information on Smart Glossaries, see *Oracle Enterprise Data Quality for Product Data Knowledge Studio Reference Guide*.

---



---

## Smart Glossary Foundation for Generated Data Lens

The Smart Glossary used by AutoBuild provides all of the initial settings for the generated data lens. Standard data lens options found in the Knowledge Studio options are copied from the selected Smart Glossary to the new data lens. Transformation types are copied from the Smart Glossary to the generated data lens. This includes the classification types, standardization types, and match types. Unit Conversion rules are also copied to the generated data lens.

AutoBuild combines your structured metadata and the selected Smart Glossary to create a new data lens that contains the generated item definitions, attributes, and associated attributes and values. If there is overlap between the Smart Glossary phrases and terms and the phrases and terms in the metadata, AutoBuild may use the Smart Glossary rules and structure instead of those found in the metadata. The primary exception to this rule is if the metadata attribute values are members of valuesets. This avoids creating ambiguous terms and phrases in the data lens. Additionally, AutoBuild prunes unnecessary rules so that there are no unintended collisions between rules.

### Replacing Item Definition Model Phrase structures with Smart Glossary Phrase Structures

AutoBuild can drill down into the structure of the generated phrases and replace any units of measure and count type phrase productions discovered in the structured item information with the EDQP standard phrase structures from the Smart Glossary. One benefit of using the Smart Glossary version of these phrase structures is that the Smart Glossary phrase structures have already been standardized. Except when valid values and units of conversion have been specified in the metadata, the AutoBuild data lens will inherit the unit conversion rules from the Smart Glossary.

## Prerequisites to Using AutoBuild

To use the AutoBuild application successfully, the following prerequisites must be met:

### PIM Metadata Prerequisites

There are specific PIM metadata prerequisites you must observe when using AutoBuild with Oracle PIM systems. Use one of the following guides to identify the PIM metadata prerequisites for the PIM system you are using:

*Oracle Enterprise Data Quality for Product Data Fusion PIM Integration Implementation and User's Guide* at

[http://docs.oracle.com/cd/E35636\\_01/doc.11116/e29148/toc.htm](http://docs.oracle.com/cd/E35636_01/doc.11116/e29148/toc.htm)

*Oracle Enterprise Data Quality for Product Data R12 PIM Connector User's Guide* at

[http://docs.oracle.com/cd/E35636\\_01/doc.11116/e29140/toc.htm](http://docs.oracle.com/cd/E35636_01/doc.11116/e29140/toc.htm)

The remaining sections in this chapter should be used as referential material only.

## Non-PIM Metadata Prerequisites

In order to make efficient use of the AutoBuild application, you must work with your metadata in the form of Excel worksheets that contain some or all of the following information about your company's products:

### Category or Item Class

The categories that describe the item data you wish to process. The categories may be hierarchically grouped though this is not essential.

### Item Form, Fit, and Function Attributes

Examples of these attributes include color, weight, size, material, packaging, and so on.

### Valid Attributes

Examples of those attribute values that are valid for your data.

You can expect maximum benefit from the AutoBuild application when your structured item examples contain attribute values that are expressed in full, unabbreviated form. The Enterprise DQ for Product can leverage these *full-form* examples and automatically recognize a broad range of term abbreviations and variations in phrasing.

## Identifying Structured Item Information

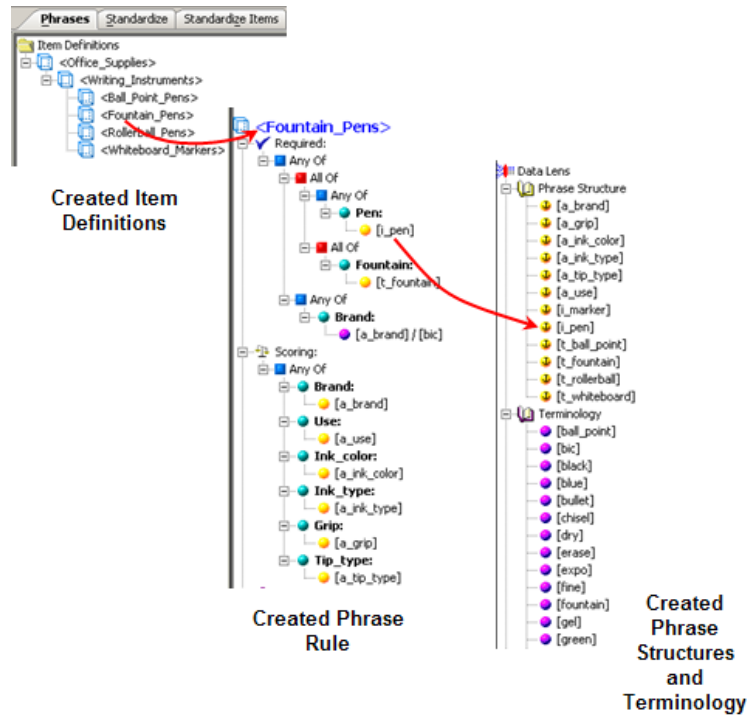
The main sources of structured item information that can be used with AutoBuild are the following:

- Item categories
- Item names and types
- Item brand information
- Item attributes
- Attribute values

The following figure provides an example of each of the above types of structured item information from a sample metadata file:

SKU	Category 1	Category 2	Category 3	Name	Type	Brand	Use	Ink Color	Ink Type	Grip	Tip Type
5671231	Office Supplies	Writing Instruments	Rollerball Pens	Pen	Rollerball	Monte Blanc	Use	Blue	Gel	Standard	Bullet
5671255	Office Supplies	Writing Instruments	Fountain Pens	Pen	Fountain	Bic	Use	Black		Rubber	Fine
5678220	Office Supplies	Writing Instruments	Ball Point Pens	Pen	Ball Point	Bic	Permanent	Red		Soft	Bullet
2259871	Office Supplies	Writing Instruments	Whiteboard Markers	Marker	Whiteboard Expo		Dry Erase	Green	Low Odor	Standard	Chisel
9383738	Office Supplies	Writing Instruments	Whiteboard Markers	Marker	Whiteboard Expo		Dry Erase	Orange		Wide	Medium

Given the structure of the information in the preceding example, AutoBuild can generate a data lens similar to the following:



### Sources of Structured Item Content

The source of structured item information could be any of the following:

- Existing electronic product catalog information
- Item master information
- Product information management (PIM) structures
- Category management or product marketing worksheets
- Existing eCommerce site information
- Structured examples used to define your current product categories

### Supported Structured item Information Formats

Product information exports can vary greatly among systems and database schemas. As a result, AutoBuild is designed to support a wide variety of product information export formats.

### Supported Category Formats

- Multiple column category names (each category column represents one level in a classification hierarchy).
- Multiple column category code/name pairs with the category columns grouped in pairs. The first column in the pair is a category code and the second column in the pair is a category name; each column pair represents one level in a classification hierarchy.
- Single column category names (single level classification hierarchy).
- Single column category names that are character separated (multiple level classification hierarchy with a character string separating each category name).
- Single column UNSPSC Category Codes.

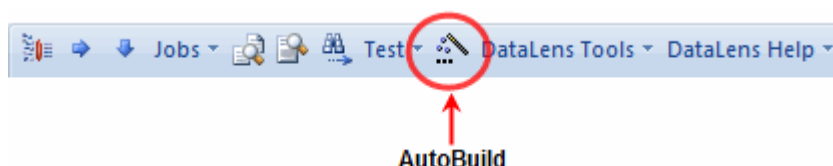
### Supported Attribute Formats

- Attribute names listed in the same row with the category information.
- Attribute name/value pairs listed in same row with the category information. A third column can be used to identify a valueset so that *only* the values present in the metadata worksheet are used to create term rules.
- Attribute name/value/unit of measure (UOM) triplets listed in same row with the category information.
- Attributes in the same file as the category information; attribute names listed as the column headers.
- Attributes in a separate file from the category information; attribute names listed as the column headers.
- Multiple categories in a single worksheet with each category grouped as a distinct set of rows separated by a blank line from another category group. Attributes are in same file as the category information, attribute names listed as column headers.
- Category and attribute information can repeat across multiple rows within the same worksheet. In addition, category and attribute information can be listed across multiple worksheets in a workbook.

---

## AutoBuild User Interface

The AutoBuild function is accessed using the **AutoBuild** button on the Services for Excel toolbar as shown in the following:



**Tip:** The tooltips appear when you rest your mouse pointer on a menu item, button, tab, icon, or similar content.

**Tip:** If you press and hold the **Shift** key and click the AutoBuild button, you are advanced directly to Step 4 of the AutoBuild User Interface Wizard. For more information, see "[Step 4: Define Data Lens Output](#)" on page 2-22.

If this toolbar does not appear in Excel, then Services for Excel may not be installed. For installation instructions and more information, see *Oracle Enterprise Data Quality for Product Data Services for Excel Reference Guide*.

When you click the **AutoBuild** button, the following message may be displayed:

```
No data lenses were found.  
Source DataLens DLS_Import_Template not found -- please select another lens or  
check out this lens before running AutoBuild
```

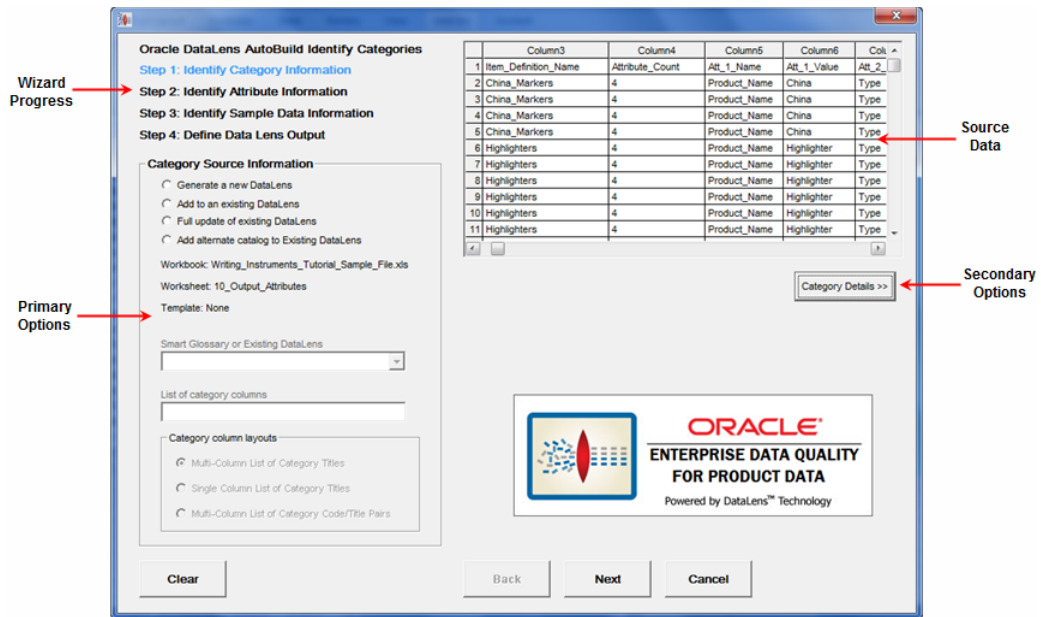
This means that the `DLS_Import_Template` Smart Glossary data lens has not been checked out of Knowledge Studio. Ensure that you open Knowledge Studio and check out the `DLS_Import_Template` data lens as read-only from your Oracle DataLens Server. For installation instructions, see *Oracle Enterprise Data Quality for Product Data Knowledge Studio Reference Guide*.

### The AutoBuild User Interface

The AutoBuild user interface is a four-step wizard that allows you to select how you want your initial data lens to be created. You start the wizard by clicking on the **AutoBuild** button on the AutoBuild toolbar.

**Tip:** It is good practice to save all open Excel workbooks before you start the AutoBuild wizard.

The first step of the AutoBuild wizard is **Identify Category Information**.



**Note:** The active Excel Workbook cannot be accessed while the AutoBuild wizard is in use.

Each step of the wizard contains the four following areas:

### Wizard Progress

The title of the active step of the wizard is colored blue.

### Source Metadata

The metadata from the active Excel worksheet is displayed. The columns can be resized as in Excel and you can use the arrows to scroll up and down to view more metadata. The columns are colored as you make selections.

### Primary Options

The controls in this area allow you to select the necessary (primary) options for the step.

### Secondary Options

The button in this area activates the additional (secondary) options that you can select for the step. The default secondary options are selected for you if no changes are made or the area is not activated.

The following buttons appear in each step and are active when appropriate:

#### Clear

Resets the wizard selections so that you can start a new data lens creation process with the active Excel workbook.

#### Back

Returns the wizard to the previous step to review or make changes in selections.

#### Next

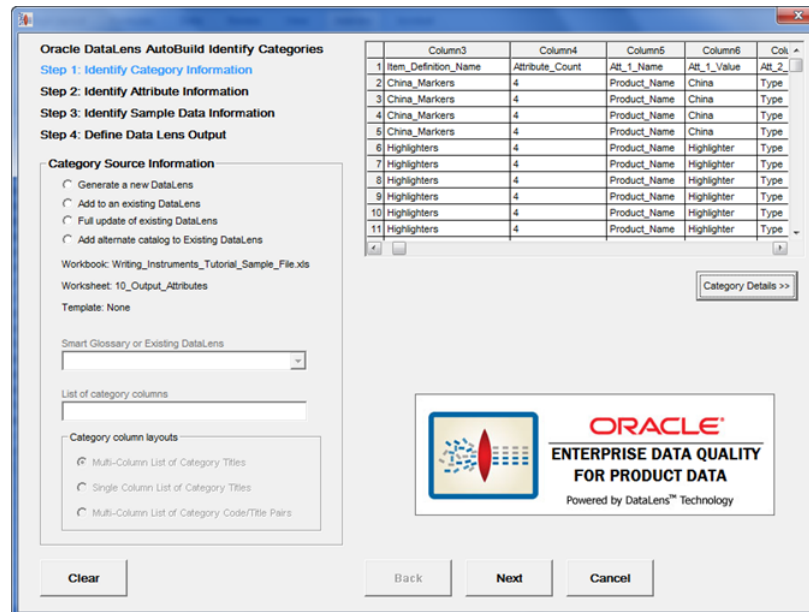
Advances the wizard to the next step in the process.

**Cancel**

Exits the AutoBuild wizard; does not save any selections.

**Step 1: Identify Category Information**

The first step of the AutoBuild wizard is to identify category information and its use is described in this section. The first time the AutoBuild wizard is opened from a worksheet, the active Excel worksheet is chosen by default as the data source.

**Category Source Information Section**

The controls in this section are used as follows:

**Data Lens Generation Options**

These options operate in conjunction with your selection from the **Smart Glossary or Existing DataLens** list and are active based on this list. Select how you want AutoBuild to generate your data lens from the following options:

- **Generate a new DataLens**

Creates a new data lens and is active when the `DLS_Import_Template` data lens is selected.

The following three options are visible when a data lens other than `DLS_Import_Template` is selected from the **Smart Glossary or Existing DataLens** list:

- **Add to an existing DataLens**

Adds new attributes and item definitions to an existing data lens.

- **Full update of existing DataLens**

Adds new and updates existing attributes and item definitions to an existing data lens. Classification schemas can also be updated using this option.

- **Add alternate catalog to Existing DataLens**

Adds a new user-defined classification schema to the selected existing data lens. This new alternate catalog may either be added to a data lens as simply a new

classification schema or it can be added and at the same time classification rules can be automatically created as follows:

**Adding an alternate catalog to a data lens (does not create classification rules) -**

You can add a classification schema to a data lens, the alternate catalog must have the following format and column names:

	A	B	C	D	E	F
1	catalog	parent_category_id	parent_category_desc	category_id	category_name	
2	Purchasing Catalog			2404	Electromechanical Components	
3	Purchasing Catalog	2404	Electromechanical Components	2405	Audio Devices	
4	Purchasing Catalog	2404	Electromechanical Components	2406	Circuit Breakers	
5	Purchasing Catalog	2425	IC's	2372	Memory	
6	Purchasing Catalog	2404	Electromechanical Components	2407	Fans	
7	Purchasing Catalog	2404	Electromechanical Components	2408	Industrial Controls & Sensors	
8	Purchasing Catalog	2404	Electromechanical Components	2409	Knobs & Dials	
9	Purchasing Catalog	2404	Electromechanical Components	2410	Power	
10	Purchasing Catalog	2410	Power	2411	Batteries	
11	Purchasing Catalog	2410	Power	2412	Battery Holders & Snaps	

After you have added this catalog to your data lens, you can manually create classification rules within the data lens (using Item Definitions and/or rules).

**Note:** This functionality can also be used more generally for creating classification schemas to add to data lenses. As long as the classification schema has the correct format, Autobuild will create a classification schema for use in a data lens.

**Adding an alternate catalog to a data lens and create classification rules -**

You can create classification rules at the same time that you add an alternate catalog. When this function is performed, you are creating classification rules that map Item Definitions in your existing data lens to the categories in your new alternate catalog (classification schema). To do this, you must meet the following requirements:

- The alternate catalog spreadsheet must contain the columns described in "Adding an alternate catalog to a data lens (does not create classification rules) -" on page 2-4.
- The Item Definitions in your existing data lens must have the Description fields populated.
- The alternate catalog spreadsheet must contain the column, mapped\_primary \_ categories.
- If you have a mapping for a alternate catalog to your Item Definition, the mapped\_primary \_categories column must contain the Description of the Item Definition that should be classified to that category.

The following is an example of these requirements:

	A	B	C	D	E	F
1	catalog	parent_category_id	parent_category_desc	category_id	category_name	mapped_primary_categories
2	Purchasing Catalog			2404	Electromechanical Components	
3	Purchasing Catalog	2404	Electromechanical Components	2405	Audio Devices	Digital Audio Devices
4	Purchasing Catalog	2404	Electromechanical Components	2406	Circuit Breakers	Breaker Panel
5	Purchasing Catalog	2425	IC's	2372	Memory	IC-Memory
6	Purchasing Catalog	2404	Electromechanical Components	2407	Fans	
7	Purchasing Catalog	2404	Electromechanical Components	2408	Industrial Controls & Sensors	
8	Purchasing Catalog	2404	Electromechanical Components	2409	Knobs & Dials	



---



---

**Note:** Since the category to Item Definition mapping is performed using the Item Definition description, the Item Definition description field (in the data lens) must be populated, and values in the `mapped_primary_catgories` column *must* reference the Item Definition descriptions (not the Item Definition names).

---



---

### Excel Information

AutoBuild displays the current Excel workbook, worksheet, and template information.

### Smart Glossary or Existing Data Lens List

Select a Smart Glossary or data lens to be used for data lens generation. You can create and use your own combination of Smart Glossaries that is most applicable to your domain data.

---



---

**Note:** The "DLS\_" prefix to a data lens name is reserved to indicate a Smart Glossary thus the selection of any data lens using this naming convention is assumed to be a Smart Glossary. For example, the default Smart Glossary is `DLS_Import_Template`.

---



---



---



---

**Note:** If the data lens names do not display properly (as is sometimes the case with double-byte characters), ensure that your Windows region and language options are set properly. For more information, see Windows Help.

---



---

### List of category columns Field

Use this field to specify the columns to use for the item definition hierarchy in the data lens. For example, the item definitions in the data lens will correspond to the item catalog categories in your item master. The column order is assumed to be from left to right with the parent categories to the left and the child categories to the right. You can use numbers, dashes, and commas to specify a range of columns, or a combination of range(s) and single column(s).

You must confirm your category column list entry by pressing the **Tab** key or selecting another user interface option. Once the column list is confirmed, the source metadata table is updated colorizing your category column selections in green.

### Category column layout Options

The three possible category column layouts to choose from are as follows:

- **Multi-Column List of Category Titles**

This column layout assumes that there are multiple levels in the category hierarchy and that each column represents one level in the hierarchy. AutoBuild assumes that the category hierarchy is expressed left to right with the first column starting from the left as the highest level in the category hierarchy, and the last column on the right is the lowest level in the hierarchy. This option is active and selected by default when the **List of category columns** field contains more than one column.

- **Single Column List of Category Titles**

This layout is active and selected for you automatically when you list only one category column. It is the default when the AutoBuild wizard first opens; it is the

only option that is active when one column is selected in the **List of category columns** field.

- **Multi-Column List of Category Code/Title Pairs**

This layout is similar to the **Multi-Column List of Category Titles** with the addition that each category is expressed as a pair of columns. The first column contains a category code and the second column contains the category name. This option is active when the column list contains more than one column.

### Category/Attribute Metadata Table

The Category / Attribute Metadata table section is displayed as follows:

	Column1	Column2	Column3
1	Id	Description	Item_Definition
2	WI12	China Mrkr Green Ink 12 / Box	China_Markers
3	WI13	China Markers red ink	China_Markers
4	WI14	China Markers White ink	China_Markers
5	WI15	China Mrk yellow ink	China_Markers
6	WI16	Chisel Point Highlighter Turquoise ink	Highlighters
7	WI31	Highlighter Chisel Tip Fluorescent Purple	Highlighters
8	WI32	Highlighter Chisel Tip Fluorescent Yellow	Highlighters
9	WI33	Highlighter Chisel Tip Orange	Highlighters
10	WI34	Highlighter Chisel Tip Pink	Highlighters
11	WI35	Highlighter Chisel Tip Yellow	Highlighters

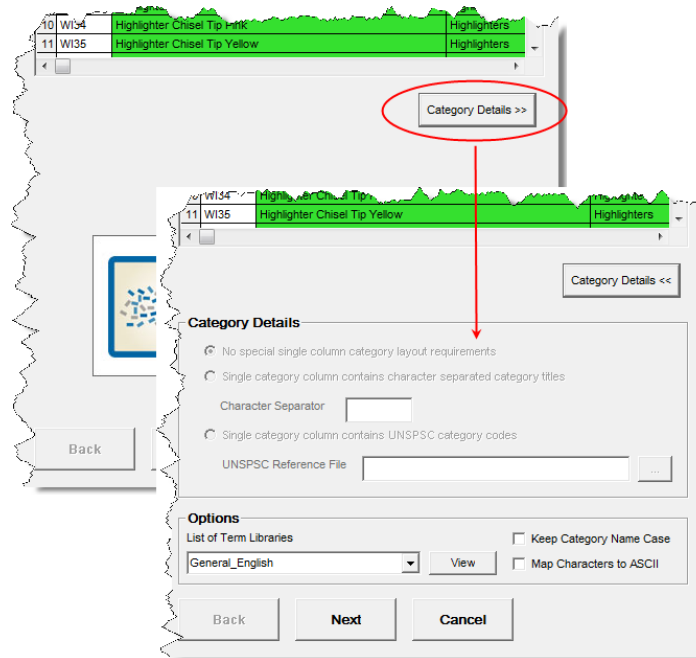
Category Details <<

### Metadata Table

The metadata table displays the first 30 rows and 80 columns of the structured Excel worksheet metadata source. As the columns are identified in the **List of category columns** field, they are colored green as in the preceding example.

### Category Details

The **Category Details** section is activated by clicking on the button of the same name and is used as follows:



None of the **Category Details** options are active unless you select only *one* column in the **List of category columns** field. These options are used as follows:

#### **No special single column category layout requirements option**

This option is automatically selected for you when a single category column is identified. It indicates that either the column contents are not delimited or that they contain category codes.

#### **Single category column contains character separated category titles Option**

This option indicates that there are multiple character separated category names in the defined category column. Selecting this option activates the following:

- **Character Separator Field**

When the previous option is selected, use this field to inform the system of the character used to separate the category titles.

#### **Single category column contains UNSPSC category codes Option**

This option indicates that the single identified category column contains UNSPSC category codes. AutoBuild is UNSPSC aware and can automatically build the complete UNSPSC hierarchy for you from a UNSPSC reference file that you supply with the following:

---

**Note:** The Enterprise DQ for Product ships with several versions of the UNSPSC classification system.

---

- **UNSPSC Reference File Browser Button**

Use to locate and select the UNSPSC reference file that you want to use. The file browser automatically opens a local directory. You can select a file of your own or a UNSPSC file installed with the product and is found in:

C:\Users\user1\AppData\Roaming\DataLens\system\schemas

You must select this option to output the UNSPSC schema as a classification type in Step 4 of the wizard. For more information, see "Output Details" on page 2-24.

**List of Term Libraries List**

Select the Term Library that you want to apply to the data lens being created. If there is a match between a metadata attribute value and the defined full-form in the term library, the term variants from the term library are added to the generated data lens.

**View Button**

Starts Internet Explorer and displays the available list of Term Libraries based on your selection in the **List of Term Libraries** list.

**Keep Category Name Case**

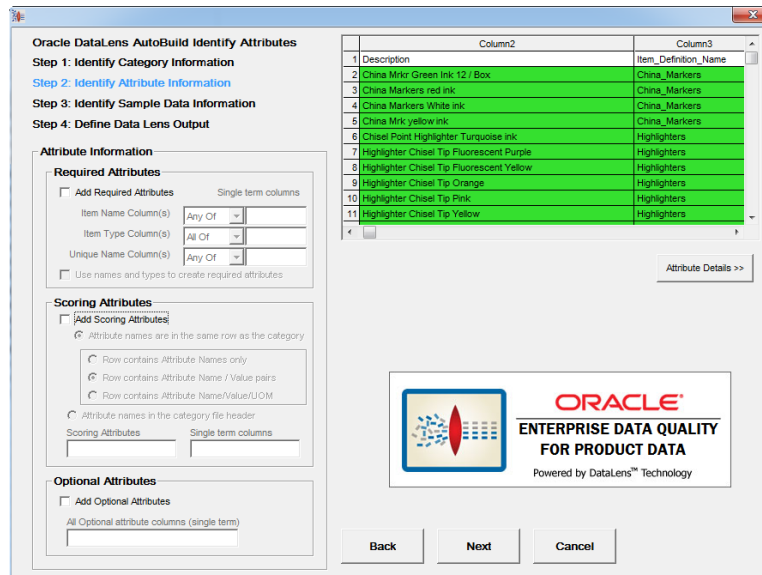
This option ensures that the category name case in your worksheet is used to create the corresponding Item Definition rather than the case rules in the selected Smart Glossary or data lens. Selecting this option ensures that AutoBuild applies proper case to the category name with the use of underscores. For example, 'Ink Pen' is changed to 'Ink\_Pen' rather than 'Ink\_pen'.

**Map Characters to ASCII Check Box**

This option will map the metadata source diacritical characters (for example, à and ñ) to their ASCII equivalents. It is useful when the input data is a romance language containing diacritical characters and you want to create a data lens with the ASCII version of categories and rules.

**Step 2: Identify Attribute Information**

The second step of the AutoBuild wizard is to identify attribute information and its use is described in this section.



**Attribute Information Section**

The following section describes each option in the **Attribute Information** section.

### Required Attributes Box

The purpose of these options is to identify attributes that can be used to identify each item uniquely. The controls in this section are used as follows:

#### Add Required Attributes

This check box activates the controls in this box so that you can select which columns of the source metadata will be the required attributes in the Item Definition.

- **Item Name, Item Type, and Unique Name Columns Fields**

Use these fields to specify the columns that are to be used to define the required logic name attributes, which are the item name, item type, and a unique name respectively.

To specify the list of attributes, you can use numbers, dashes, and commas to specify a single column, a range of columns, or a combination of ranges and single columns. You confirm your entry by pressing the **Tab** key or selecting another control in the step. Once the column list is confirmed, the metadata table is updated by coloring your item name columns in violet.

- **Use names and types to create required attributes Check box**

When this option is selected, the required logic attributes are generated by AutoBuild. The attribute names are used for the Required Attribute names, rather than using the default `Item_Name` and `Item_Type` as attribute names. Required attribute logic attempts to add attributes to the set of required items automatically by determining the uniqueness of the attributes. Some scoring or form/fit/function attributes are unique to a category or unique to a set of sibling categories (categories that are at the same level and report to the same parent category).

Use of enhanced attribute logic works best when the structured source contains *a broad range of categories* and most categories contain attribute definitions. This allows the system to analyze the uniqueness of the attribute information across the category hierarchy. If your structured source metadata only contains a couple of categories, or if your attribute information is sparse across your categories, then you should not select this option.

Unique names are considered to be uniquely associated with a single category of product so the first category associated with a specific brand name will be recorded as the category identified by that brand name.

---



---

**Note:** A Unique Name as defined for required logic *must* be a unique identifier. If your brand names are not unique across categories then you should use your unique names just as item names.

---



---

- **Logic generated if this option is selected is as follows:**

The default "Any Of" and "All Of" nodes described in the following examples are generated by AutoBuild when the **Use names and types to create required attributes** check box is selected and attribute selections are made using the controls in the **Attribute Information** box:

Example	Required logic generated for each Item Definition
No item name, item type, or brand name columns are defined and <b>Add Required Attributes</b> is selected	<b>Required</b> <b>All Of:</b> Item_Name: Item_Type:
Only item name columns are defined	<b>Required</b> <b>All Of:</b> <b>Any Of:</b> Name_Attr1: Name_Attr2: Attributes that are unique to my sibling categories if any ...
Only item type columns are defined	<b>Required</b> <b>All Of:</b> <b>All Of:</b> Type_Attr1: Type_Attr2: Attributes that are unique to this category alone if any ...

Example	Required logic generated for each Item Definition
One or more columns of both item names and item types are define	<p><b>Required</b></p> <p><b>All Of:</b></p> <p><b>Any Of:</b></p> <p>Name_Attr1:</p> <p>Name_Attr2:</p> <p>Attributes that are unique to my sibling categories if any</p> <p>...</p> <p><b>All Of:</b></p> <p>Type_Attr1:</p> <p>Type_Attr2:</p> <p>Attributes that are unique to this category alone if any</p> <p>...</p>
One or more brand name columns are defined	<p><b>Required</b></p> <p><b>Any Of:</b></p> <p><b>All Of:</b></p> <p>Item_Name:</p> <p>Item_Type:</p> <p><b>Any Of:</b></p> <p>Brand:</p> <p>[a_brand]\[some_brand_name1]</p> <p>[a_brand]\[some_brand_name2]</p> <p>...</p>

Example	Required logic generated for each Item Definition
One or more brand name columns and item name and item type columns are defined	<p><b>Required</b></p> <p><b>Any Of:</b></p> <p><b>All Of:</b></p> <p><b>Any Of:</b></p> <p>Name_Attr1:</p> <p>Name_Attr2:</p> <p>Attributes that are unique to my sibling categories if any</p> <p>...</p> <p><b>All Of:</b></p> <p>Type_Attr1:</p> <p>Type_Attr2:</p> <p>Attributes that are unique to this category alone if any</p> <p>...</p> <p><b>Any Of:</b></p> <p>Brand:</p> <p>[a_brand]\[some_brand_name1]</p> <p>[a_brand]\[some_brand_name2]</p> <p>...</p>

- Logic generated if this option is *not* selected**

The following examples describe the required logic attributes that are generated by AutoBuild when the **Use names and types to create required attributes** check box is *not* selected:

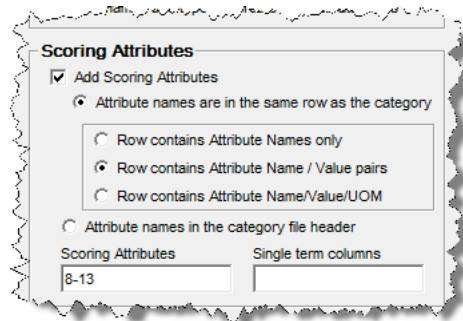
Example	Required logic generated for each item definition
If no item name or item type columns are defined	<p><b>Required</b></p> <p><b>All Of:</b></p> <p>Item_Name:</p> <p>Item_Type:</p>
If only item name columns are defined	<p><b>Required</b></p> <p><b>All Of:</b></p> <p>Item_Name:</p> <p>[i_column1]</p> <p>[i_column2]</p> <p>...</p> <p>Item_Type:</p>



Example	Required logic generated for each item definition
If only item type columns are defined	<p><b>Required</b></p> <p><b>All Of:</b></p> <p>Item_Name:</p> <p>Item_Type:</p> <p>[t_column1]</p> <p>[t_column2]</p> <p>...</p>
If one or more columns of both item names and item types are defined	<p><b>Required</b></p> <p><b>All Of:</b></p> <p>Item_Name:</p> <p>[i_column1]</p> <p>[i_column2]</p> <p>...</p> <p>Item_Type:</p> <p>[t_column1]</p> <p>[t_column2]</p> <p>...</p>
If one or more brand name columns are defined	<p><b>Required</b></p> <p><b>Any Of:</b></p> <p><b>All Of:</b></p> <p><i>(item name and item type logic as previously described)</i></p> <p>Brand:</p> <p>[a_brand]\[some_brand_name1]</p> <p>[a_brand]\[some_brand_name2]</p> <p>...</p>

### Scoring Attributes Box

The following section describes each option found in the **Scoring Attributes** box.



#### Add Scoring Attributes Check Box

Select this option to add scoring attributes to your new data lens Item Definitions.

**Attribute names are in the same row as the category Option**

Select this option when your attribute information is located in the same row as the associated categories. In this case, one or more columns contain the different category levels and in the same worksheet, there are adjacent columns that contain the attribute names and possibly example attribute values. The headers for the attribute columns are typically generic, like "Attribute Name 1" or "Attribute Value 1", and so on. The attribute value information found in each row is specific to the category defined by that row as in the following example:

Category 1	Category 2	Category 3	Category 4	AttrName1	AttrValue1	AttrName2	AttrValue2	AttrName3	AttrValue3
Office Supplies	Writing Instruments	Pens	Rollerball	Ink Color	Blue	Grip	Standard	Point	Fine
Office Supplies	Writing Instruments	PENS	Fountain	Ink Color	Black	Grip	Rubber	Point	Fine
Office Supplies	Writing Instruments	Pens	Ball Point	Ink Color	Red	Grip	Soft	Point	Fine

Category column

Category specific attribute names and values

Generic attribute column

The following three options are active only when this option is selected.

■ **Row contains Attribute names only Option**

Select this option to instruct AutoBuild that attribute names, but not attribute values, are available in the structured metadata source as in the following example:

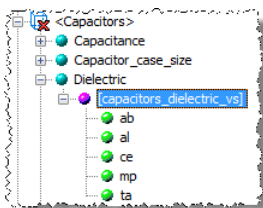
Category 1	Category 2	Category 3	Category 4	AttrName1	AttrName2	AttrName3
Office Supplies	Writing Instruments	Pens	Rollerball	Ink Color	Grip	Point
Office Supplies	Writing Instruments	PENS	Fountain	Ink Color	Grip	Point
Office Supplies	Writing Instruments	Pens	Ball Point	Ink Color	Grip	Point
Office Supplies	Writing Instruments	Markers	Whiteboard	Ink Color	Grip	Point

■ **Row contains attribute name / value columns Option**

Select this option to instruct AutoBuild that the associated attribute columns contain both attribute names and attribute values as paired columns as in the following example:

C	D	E	F	G
catalog_group_id	catalog_group	attr_id	attr_name	attr_value
571	Capacitors	Capacitor_Specs Dielectric	Dielectric	AB
571	Capacitors	Capacitor_Specs Dielectric	Dielectric	AL
571	Capacitors	Capacitor_Specs Dielectric	Dielectric	CE
571	Capacitors	Capacitor_Specs Dielectric	Dielectric	MP
571	Capacitors	Capacitor_Specs Dielectric	Dielectric	TA
571	Capacitors	Capacitor_Specs CAPACITANCE	CAPACITANCE	000??

If your metadata contains valuesets, the names of terms built using valuesets are appended with '\_vs' and Smart Glossary terms that may exist are ignored. If the attributes in the preceding example were valuesets, the [capacitors\_dielectric\_vs] term is built and contains the ab, al, ce, mp, and ta valueset as in the following:



If the valueset is 'Yes' and 'No' (or equivalent variations such as Y and N), the rules are created to recognize both the presence and absence of the attribute (for example, [shielded] and [not][shielded]). Additionally the valueset values are used to create the standardization rules for these attributes.

---

**Note:** Your product description data may or may not conform to the Y/N rules created by AutoBuild. Thus, when your metadata contains Y/N valueset values, it is recommended that you closely review the grammar and standardization rules to ensure that they will work for your particular product data.

---

- **Row contains attribute name / value / UOM columns Option**

Select this option to instruct AutoBuild that the associated attribute columns contain attribute names, attribute values, and units of measure grouped together as three columns sets as in the following example:

Category 1	Category 2	Category 3	Category 4	AttrName1	AttrValue1	AttrUOM1	AttrName2	AttrValue2	AttrUOM2
Hardware	Fasteners	Screws	Wood	Material	Stainless Steel		Point Size		8 points
Hardware	Fasteners	Bolts	Lag	Material	Stainless Steel		Shear Strength		2500 lbs
Hardware	Fasteners	Bolts	Toggle	Material	Stainless Steel		Shear Strength		1500 lbs
Hardware	Fasteners	Bolts	Carriage	Material	Stainless Steel		Shear Strength		3500 lbs

When your source metadata includes a UOM column named, `attr_uom`, this option is automatically selected and the selection of the **Single term columns Scoring Attribute** option is forced.

- **Attribute names in the category file header Option**

Select this option when the attribute names reside in the header row (row 1) of the structured metadata source. This worksheet structure is often used when the worksheet contains either a single category or a set of closely related categories that all share the same attribute information. An example would be many different types of pens, specifically different types of pens where each pen type has the same attributes.

The following is an example of a worksheet with one table three related categories:

Category 1	Category 2	Category 3	Use	Ink Color	Ink Type	Grip	Tip Type
Office Supplies	Writing Instruments	Rollerball Pens		Blue	Gel	Standard	Bullet
Office Supplies	Writing Instruments	Fountain Pens		Black		Rubber	Fine
Office Supplies	Writing Instruments	Ball Point Pens	Permanent	Red		Soft	Bullet
Office Supplies	Writing Instruments	Whiteboard Markers	Dry Erase	Green	Low Odor	Standard	Chisel
Office Supplies	Writing Instruments	Whiteboard Markers	Dry Erase	Orange		Wide	Medium

You can process several tables containing different source metadata when the attribute names in the header row of each table are the same (for at least the first five columns) as in the following example:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
18	W104	Pen Rll Med Bk 10/Bx	Writing_InstrumentsPens_and_Pen_Refills	Pen_Refills	Pen_Refills	Rll	Rolleball	Ink Black																			
19	W193	Refill Rollerball Ink Black	Writing_InstrumentsPens_and_Pen_Refills	Pen_Refills	Pen_Refills	Refill	Rollball	Ink Blue																			
20	W194	Refill Rollerball Ink Blue	Writing_InstrumentsPens_and_Pen_Refills	Pen_Refills	Pen_Refills	Refill	Roll Ball	Ink Blue																			
21	W1103	Roller Ball Pen Refill Medium Point Blue Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen_Refills	Pen_Refills	Refill	Roller Ball	Ink Blue																			
22																											
24	W16	Balpoint Pen Medium Point Black Barrel black ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Balpoint	Balpoint	Ink Black																			
25	W117	Comfort Grip Retractable Ballpoint Pen medium Point Blue ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Blue																			
26	W118	Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Red																			
27	W127	Retractable Ballpoint Pen Non-Refillable Black Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Black																			
28	W128	Erasable Ink Ballpoint Pen Medi Point Blue Barrel/Blue Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Balpoint	Balpoint	Ink Blue																			
29	W129	Felt Tip Pen Extra-Fine Point Translucent blue Barrel Purple Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Felt Tip	Felt Tip	Ink Purple																			
30	W130	Felt Tip Pen Extra-Fine Point Translucent black Barrel Red Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Felt Tip	Felt Tip	Ink Red																			
31	W179	Pen Extra-Fine Point Blue Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Pen	Pen	Ink Blue																			
32	W180	Pen Extra-Fine Point Green Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Pen	Pen	Ink Green																			
33	W188	Porous Pen Extra-Fine Point Red Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Porous	Porous	Ink Red																			
34	W191	Porous Pen Extra-Fine Tip White Barrel/Black Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Porous	Porous	Ink Black																			
35	W195	Balpoint Pen Retractable Fine Point Black Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Balpoint	Balpoint	Ink Black																			
36	W196	Retractable Ballpoint Pen Fine Point Clear Barrel/Black Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Black																			
37	W197	Retractable Ballpoint Pen Medium Point Black Barrel/ black ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Black																			
38	W198	Retractable Ballpoint Pen Refillable Medium Point Black Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Black																			
39	W199	Balpoint Pen Rubber Grip Fine Point Retractable Black Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Black																			
40	W1100	Retractable Ballpoint Pen Rubber Grip Fine Point Red Ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Retractable Ballpoint	Retractable Ballpoint	Ink Red																			
41	W1101	Roller Ball Pen Gel Ink Refillable Medium Blue	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Roller Ball	Roller Ball	Ink Blue																			
42	W1102	Roller Ball Pen Refillable Medium Red gel ink	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Roller Ball	Roller Ball	Ink Red																			
43	W1107	Pen, Ballpoint, Blue Ink, Medium Point	Writing_InstrumentsPens_and_Pen_Refills	Pen	Pen	Balpoint	Balpoint	Ink Blue																			
44																											
45	W112	China Marker Green Ink 12 / Box	Writing_InstrumentsMarkers	China_Markers	China_Markers	China	Makers	Ink Green																			
46	W113	China Markers red ink	Writing_InstrumentsMarkers	China_Markers	China_Markers	China	Makers	Ink Red																			
48	W114	China Markers White ink	Writing_InstrumentsMarkers	China_Markers	China_Markers	China	Makers	Ink White																			
49	W115	China Mkr yellow ink	Writing_InstrumentsMarkers	China_Markers	China_Markers	China	Mkr	Ink Yellow																			
50																											
51	W16	Chisel Point Highlighter Turquoise ink	Writing_InstrumentsMarkers	Highlighters	Highlighters	Highlighter	Chisel Point	Ink Turquoise																			
52	W31	Highlighter Chisel Tip Fluorescent Purple	Writing_InstrumentsMarkers	Highlighters	Highlighters	Highlighter	Chisel Tip	Ink Fluorescent Purple																			

**Scoring Attributes Field**

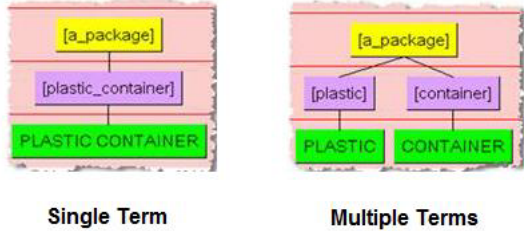
Use this field to specify scoring attribute columns. The order of addition of your attributes is from the lowest column to the highest column number. To specify the list of attributes, you may use numbers, dashes, and commas to specify a single column, a range of columns, or a combination of range(s) and single column(s). You confirm your attribute column list entry by pressing the **Tab** key or selecting another user interface option. Once the column list is confirmed, the **Category/Attribute Data** table is updated and colorizes your selected scoring attribute columns blue.

**Single term columns Field**

Use this option to force the creation of a single term for an attribute value containing more than one word. When AutoBuild creates the data lens phrases and terms for identified attributes values, and the attribute values are identified as single term columns, then AutoBuild creates a single term to recognize the attribute value no matter how many words (tokens) are in the attribute value. For example, when the attribute value is "pens" setting this option has no effect. However, if the attribute value is "rollerball pens" then both words are used to create a single term, "[rollerball\_pens]".

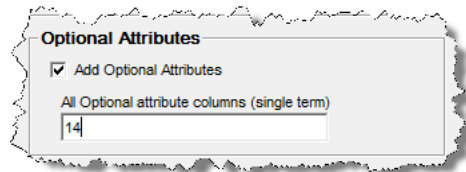
**Note:** The Predict and AutoSuggest functions operate best when terms are single-termed. Typically, columns with text are single termed while UOM columns are not.

The following example shows how the terms are built when the attribute value is two words 'PLASTIC' and 'CONTAINER' for either single term or multiple terms:



**Optional Attributes Box**

The following section describes each option found in the **Optional Attributes** box.



### Add Optional Attributes Check Box

Select this option to add optional attributes and activate the **Add Optional Attributes columns (single term)** check box.

### Add Optional Attributes columns (single term) Check Box

Select this option to add data lens primary business attributes common to all items like 'Manufacturer', 'Brand\_or\_Model\_Name', 'Model\_or\_Part\_Number', 'Units\_per\_Package', 'Additional\_Information', and 'URL'. Optional attributes are created as for any attributes that you provide in the field.

### Category/Attribute Metadata Table

The **Category/Attribute Metadata** table section is displayed as follows:

	Column3	Column4	Column5	Column6	Column7	Column8
1	Item_Definition_Na	Attribute_	Att_1_Name	Att_1_Value	Att_2_Name	Att_2_Value
2	China_Markers	4	Product_Name	China	Type	Mrkr
3	China_Markers	4	Product_Name	China	Type	Markers
4	China_Markers	4	Product_Name	China	Type	Markers
5	China_Markers	4	Product_Name	China	Type	Mrk
6	Highlighters	4	Product_Name	Highlighter	Type	Chisel Poin
7	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip
8	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip
9	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip
10	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip
11	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip

### Metadata Table

The metadata table displays the first 30 rows and 80 columns of the structured Excel worksheet metadata source. As you identify category and attribute information, the appropriate columns are colorized as in the preceding example. The category columns are colorized as follows:

Color	Meaning
Green	category
Violet	item name, item type, and brand names
Blue	scoring attributes

### Attribute Details

The **Attribute Details** section is activated by clicking on the button of the same name and is used as follows:

### Attribute Values Box

The following section describes each option in the **Attribute Values** box.

#### Phrase Separator Field

Use this field to list the characters used to delimit multiple attribute values in a single attribute value column. The list must be a space-separated list of punctuation characters. AutoBuild creates a separate attribute value for each value that is separated by one of the defined character separators. By default, the phrase separator character list is set to a semi-colon (;).

#### Maximum #Terms in a Phrase Field

Use this field to control the maximum number of terms in an attribute value that AutoBuild will attempt to add to the associated phrase. If the number of attribute value terms is greater than the number found in this field, AutoBuild ignores the extra attribute values. This option must be set to a number between 1 and 50. The default value is set to five.

#### Map values directly to each attribute Check Box

Use this check box when your attribute values are well defined and structured (considered to be *gold* data) and you want *only* these values assigned. The generated or updated data lens will contain full phrase and term context associations to the scoring attributes of Item Definitions.

#### Attribute values are Part Number Codes Check Box

Use this check box to indicate that your attribute values are part numbers.

#### Include attribute name in phrase for context Options

Select one of the three options to control the context that is added to attribute values by AutoBuild to create unique attribute value recognition. AutoBuild adds the attribute name to the attribute value; this combination of terms becomes a production in the phrase rule. This eliminates the possible ambiguities, allowing attribute values to be uniquely recognized.

- **Append as needed**

Select this option to append the attribute name to the end of the attribute value only when the attribute value is not unique, or when the attribute value is a number or a unit of measure.

- **Append always**

Select this option to append the attribute name to the end of the attribute value for all attribute values.

- **Insert always**

Select this option to insert (prepend) the attribute name at the beginning of the attribute value for all attribute values.

- **Never**

Select this option so that attribute names are never added to the attribute values.

### Standardizations Box

The following section describes each option in the **Standardizations** box.

#### #Required/Scoring/Optional Standardization Columns Lists

If your metadata worksheet contains target standardization forms for the attribute values, you can use this information to automatically create standardization rules in the datalens being autobuilt. These options allow you to select of the number of standardization columns that appear to the right of each associated attribute column that contains the corresponding standardized attribute values. For each identified standardization column, the column header is used to create a unique standardization type in the new lens.

#### Create Replace All Rules Check Box

Creates standardization rules for all attributes in the default standardization. These rules are all of the type 'Replace all'.

## Step 3: Identify Sample Data Information

The third step of the AutoBuild wizard is to identify the sample data information and its use is described in this section.

### Select Sample Data Source Section

The following section describes each option in the **Select Sample Data Source** section.

#### Use DataLens Knowledge Studio to collect samples Option

Select this option when you want to collect your sample data for data lens maintenance using the Knowledge Studio. Use this selection when your sample text is not immediately available during the AutoBuild process.



**Create Samples from Attributes Option**

Select this option when you want to create your sample data directly from the attribute values identified through the AutoBuild wizard. Use this selection when sample text is not available and you feel that the existing attribute values are representative of the metadata that you need to process via the Oracle DataLens Server. When this option is selected, the **Sample Data Rows and Columns** box and the following check box is active:

- **Prefix Sample Data Columns (listed below) Check Box**

Select this option when you want to prefix the sample data created from your attribute values with names, types, brands, or short descriptions found in other columns of your example structured metadata. When this check box is checked, you must use the **List of Sample Data Columns** field to define the columns to be prefixed with the attribute values.

**Sample Data File Box**

This box contains the options that you use to identify the file or worksheet containing your sample data. As you make your file selections, the example sample data file display is updated with colorized columns to provide feedback on the options you have selected. The file you select can be the file that you are currently using as the source of example-structured data.

- **Use current file Option**

Use this option to activate the **Sample Data Worksheet** list to select the worksheet you want to use from the open Excel workbook.

- **Use separate Option**

Use this option to activate the **Sample Data File** field to enter the full file path to your sample data file or click the button adjacent to this field to locate the file.

- **Sample Data File Field**

This field displays the full file path to the currently selected Excel workbook (or Excel compatible file). Use it to locate and select, or enter the attribute directory (folder) path name that you want to use.

- **Sample Data Worksheet List**

Use this list to select the worksheet you want to use in the selected Excel workbook.

**Sample Data Rows and Columns Box**

This box contains options for you to use in specifying how you wish to create sample data files using the available columns of your sample data file.

- **Id Column Field**

Use this option to identify the unique id column found in your sample data source. A unique id is required when creating sample data from attribute values. This value must be a single column number.

- **Group By Column Field**

Use this option to identify a column to be used to create sample files that are grouped by the unique values found in the specified column. By default, AutoBuild creates the data lens standard 100 record, randomly selected sample files from your identified source file. When this option is specified, the *group by* files are also created. The benefit of using this option is to allow your category experts to perform lens maintenance on a single related set of items at one time. This option is not available when creating samples from the attribute values;



AutoBuild automatically creates the *group by* samples from the supplied category information when creating samples from the attribute values.

- **List of Sample Data Columns Field**

Use this option to specify the columns that are to be concatenated together to create your data samples. The data in the specified columns are concatenated together with a space between each column value. Your columns will be concatenated in lowest column number to highest column number order. To specify the list of columns, you may use any combination of numbers, dashes, and commas to specify a single column, a range of columns, or a combination of range(s) and single column(s). For example, 5-8 or 5,6,8,9, 5-8,10. You confirm your column list entry by selecting the **Tab** key or selecting another user interface option.

### Category/Attribute Metadata Table

The following describes the **Category/Attribute Metadata** table on the Identify Attribute Information dialog.

	Column3	Column4	Column5	Column6
1				
2	Level 1	Level 2	Level 3	Product Name
3	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
4	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
5	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
6	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
7	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
8	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
9	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
10	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
11	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill

### Metadata Table

The metadata table displays the first 30 rows and 80 columns of the structured Excel worksheet metadata source. As you identify category and attribute information, the appropriate columns are colorized as in the preceding example. The category columns are colorized as follows:

Color	Meaning
Green	category
Violet	item name, item type, and brand names
Blue	scoring attributes

### Sample Data Table

The following describes the Sample Data Table that is displayed when you identify the location of your sample data. This table operates like the **Data** table.

Sample Data Table

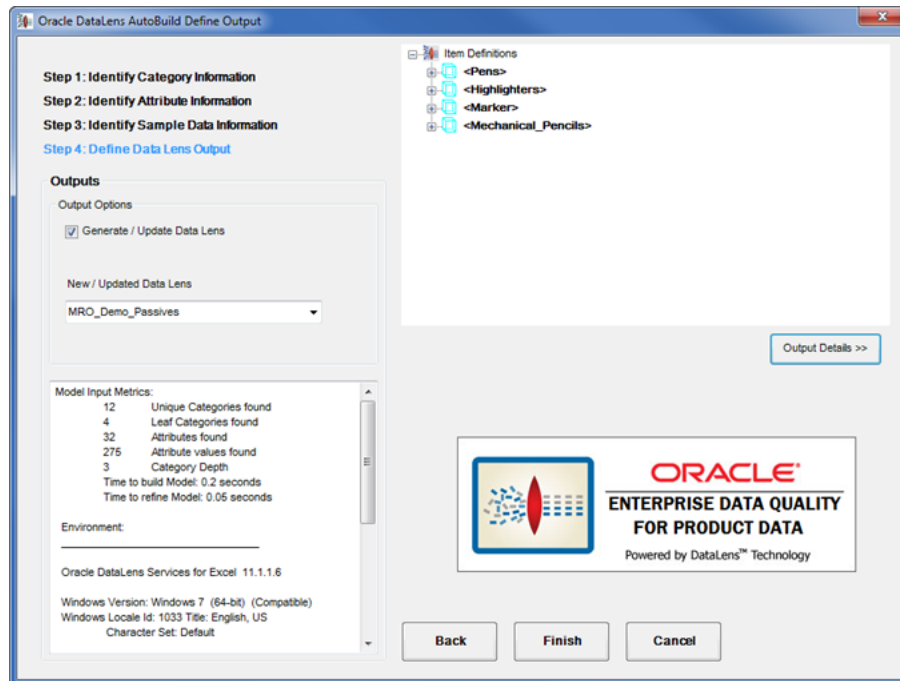
Column3	Column4	Column5	Column6	Col
1				
2	Level 1	Level 2	Level 3	Product_Name Type
3	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill Ballpoint
4	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill Ballpoint
6	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill Ballpoint
7	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill Ballpoint
8	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill Ballpoint
9	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
10	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
11	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill Rollerball

Column1	Column2	Column
2	Description	Level 1
3	Ballpoint Pen Refill Fine Point Blue Ink 2 / Pack	Writing_Instrun
4	Ballpoint Pen Refill Med Pt Black Ink 2 / Pk	Writing_Instrun
5	Ballpoint Pen Refill Medi Point black ink 2 / Pack	Writing_Instrun

### Step 4: Define Data Lens Output

The final step of the AutoBuild wizard is defining the data lens output and its use is described in this section.



### Outputs Box

The controls in this section are used as follows:

#### Generate / Update Data Lens

Default selection to create or update a data lens.

#### New / Updated Data Lens Field

Use this field to name your new data lens or select an existing data lens to add to update. Your name must be unique when creating a new data lens and you can select an existing data lens to use as the basis for a new data lens name.

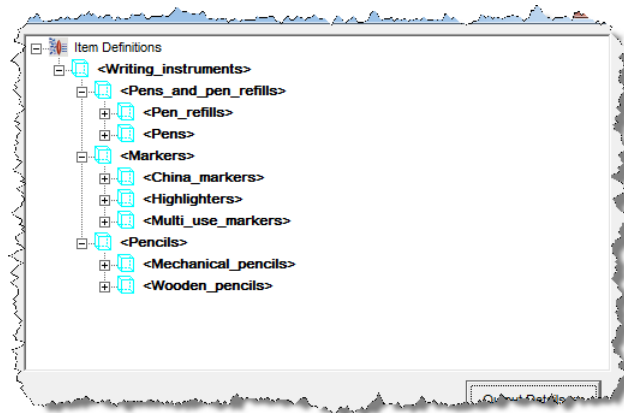
## Report Field

This read-only field contains a report that provides metrics on the model generated for your new data lens. AutoBuild reports all warning messages concerning the model generation process in this field. For example, the presence of duplicate category names results in a warning message.

If there are warning or error messages reported you should take time to review the messages to determine whether action is needed to correct the structured metadata input before proceeding. AutoBuild generates a data lens when the **Finish** button is selected whether or not there are warning messages. Warning messages indicate that AutoBuild may not interpret your structured metadata correctly.

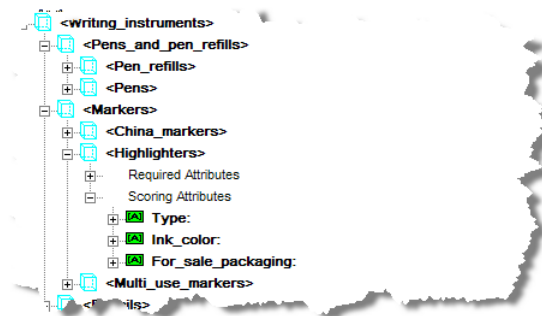
## AutoBuild Item Definition Explorer

The Item Definition Explorer provides a way to review the lens model created by AutoBuild. The generated model reflects the structured metadata provided and your option selections.

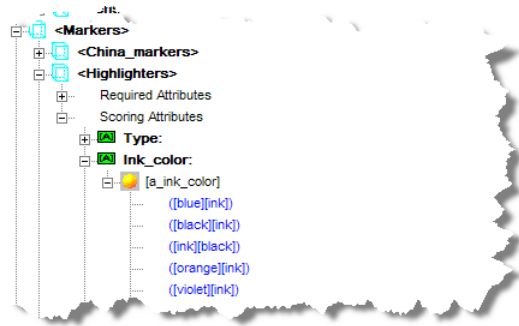


The **AutoBuild Item Definition Explorer** provides a way to review the result of your selections prior to generating the data lens. If you find inconsistencies in the model, you can return to previous AutoBuild wizard steps and make the necessary adjustments or corrections to ensure that your data lens is created properly.

The **AutoBuild Item Definition Explorer** displays the hierarchy of item definitions as a set of indented nodes; each node with a blue cube icon indicates an item definition or a level in the item definition hierarchy. For any node in the explorer, you can view more details by clicking on the plus icon, or fewer details by clicking on the minus icon.

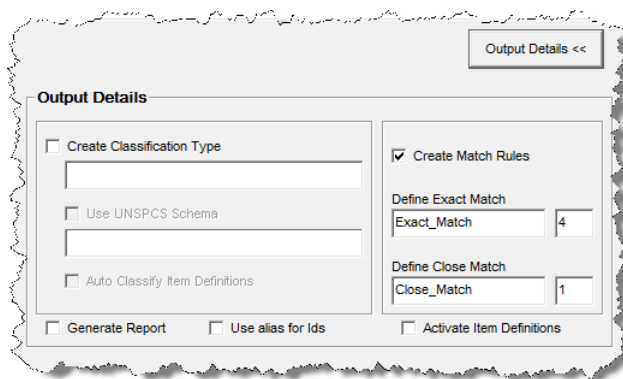


Finally, you can drill down deeper into each phrase and view the phrase productions as a set of one or more grouped term names. For each attribute, you can view the assigned phrase(s) if any. A phrase generally has an associated lens phrase yellow ball icon as in the following example.



## Output Details

The **Output Details** section is used activated by clicking on the button of the same name and is used as follows:



### Create Classification Type Check Box

Select this check box to add a classification to your new lens. When this check box is selected, the field below it and the **Auto Classify Item Definitions** check box are activated.

- **Name Field**

Use this field to name your new classification type.

- **Use UNSPSC Schema Check Box**

This check box is active when the **Single category column contains UNSPSC category codes** option is set in "[Step 1: Identify Category Information](#)" on page 2-3.

When your categories are defined by UNSPSC category codes, you can use this selection to add the associated UNSPSC reference schema as the schema file for the defined classification type. If this check box is selected, the read-only field below it displays the name of the selected UNSPSC reference schema file. For more information about configuring a UNSPSC schema, see "[Category Details](#)" on page 2-6.

- **Auto Classify Item Definitions Check Box**

Use this field to ensure that AutoBuild creates a classification type (the same as your Item Definition structure) and creates the appropriate classification rules.

### Generate Report Check Box

Use this check box to generate AutoBuild metrics reports that contain all of the information found in the model metrics report described in "[Step 3: Identify Sample Data Information](#)" on page 2-19, as well as details on the generated data lens, and a

complete list of the option selections used during the AutoBuild session. For more information, see ["AutoBuild Report"](#) on page 2-25.

#### **Use alias for Ids**

Select this check box if you want your item definition and attribute alias fields to be populated with their respective identifiers. Your metadata may have specific identifiers for categories and attributes. If your metadata does not have specific identifiers, checking this box results in populating the item definition alias with the category name, exactly as it appears in the metadata worksheet.

If your metadata does contain category and attribute identifiers, these values are also stored in the `system_id` and `group_id` internal fields and can be viewed in the Item Definition Alias field.

#### **Activate Item Definitions**

Select this check box to activate the Item Definitions created in your data lens from your source metadata. If an Item Definition is activated, it can be used for attribute extraction and standardization on the Production or QA Transform Servers. If an Item Definition is inactive, than an item recognized by that Item Definition may be processed only by the Development Transform Server.

The Item Definitions created by AutoBuild may be incomplete, so the Item Definitions must be closely reviewed in Knowledge Studio before the data lens is checked in for production use. If a data lens is checked in for production use without being reviewed, one possible result may be an unexpectedly high number of returned records upon running a match application. This is applicable to Production and QA Oracle DataLens Servers only; Development servers ignore the inactive flag.

By default, Item Definitions created by AutoBuild are set to inactive so this check box is not selected.

#### **Create Match Rules Check Box**

Select this check box to create default matching rules. A default attribute order for each item definition, containing only scoring attributes in the order specified in the worksheet columns, is automatically created. The match types 'Exact\_Match' and 'Close\_Match' are created based on the standardization type 'Match\_Attributes'.

When this check box is selected, the **Define Exact Match** and **Define Close Match** controls are activated and used as follows:

- **Define Exact Match Controls**  
Defines the number of required attributes for the Exact\_Match rule per item definition.
- **Define Close Match Controls**  
Defines the number of required attributes for the Close\_Match rule per item definition.

## **AutoBuild Report**

The AutoBuild application can automatically generate a Semantic Knowledge Structure report about your data lens. This report is created after your data lens is constructed if you have selected the Generate Report option. For more information about setting this option, see ["Output Details"](#) on page 2-24.

The report is added to your Excel workbook as a new worksheet so that you can easily review and save the information. It is updated in your Excel workbook each time you

run AutoBuild to generate a data lens. This is applicable whether there are one or multiple worksheets in your workbook.

**Tip:** You can keep prevent the report from being overwritten and keep it for future use by renaming the worksheet.

The Semantic Knowledge Structure report details all action taken related to item definitions, phrases, terminology, and alternate catalogs in the creation or update of your data lens.

In an incremental update of a data lens, additions are made to an existing data lens, but deletions must be made manually.

This report describes what items should be manually deleted from your data lens, based on the differences between the original data lens and the metadata structure used in the incremental AutoBuild update. Use this report as a guide for making deletions from your updated data lens.

	A	B	C	D	E
1	AutoBuild Wizard				
2	Semantic Knowledge Structure				
3	DataLens: test_report				
4	Windows User: user				
5	Date / Time: 1/11/2013 11:00:45 AM				
6	Full update of existing DataLens				
7					
8					
9					
10	<a href="#">Item Definitions</a>				
11	<a href="#">Item Definitions with Attributes</a>				
12	<a href="#">Phrases</a>				
13	<a href="#">Terms</a>				
14	<a href="#">Alternate Catalog</a>				
15					
16					
17					
18					
19					
20	<b>Item Definitions</b>				
21				<a href="#">New Meta-Data</a>	
22				<a href="#">Renamed Meta-Data</a>	
23				<a href="#">Existing Meta-Data</a>	
24				<a href="#">Not Present in Meta-Data</a>	
25					
26	<b>Updates</b>	<b>Level: 1</b>	<b>Level: 2</b>	<b>Level: 3</b>	
27	New	<a href="#">Level_1</a>			
28	New		<a href="#">Level_2</a>		
29	New			<a href="#">Level_3</a>	
30	New	<a href="#">Writing_Instruments</a>			
31	New		<a href="#">Pens_and_Pen_Refills</a>		
32	New			<a href="#">Pen_Refills</a>	
33	New			<a href="#">Pens</a>	
34	New		<a href="#">Markers</a>		
35	New			<a href="#">China_Markers</a>	
36	New			<a href="#">Highlighters</a>	
37	New			<a href="#">Multi_Use_Markers</a>	

You can use the links at the beginning of the report to navigate to the specific data indicated.

You can delete this report by selecting the **DataLens Tools** menu and selecting **Remove AutoBuild Reports....**

A deletion confirmation prompt is displayed and selecting **Yes** deletes the AutoBuild report from the active workbook.

---

---

## The AutoBuild Process

This chapter provides an end-to-end example of how to use AutoBuild to read structured metadata containing item information and generate a working data lens for use in the Knowledge Studio. It uses a sample structured data set that contains several records of standard writing instruments like pens and pencils.

### Using AutoBuild with PIM Metadata

The following sections describe where to find the information to use AutoBuild with Oracle PIM metadata as your source.

#### AutoBuild with Fusion PIM Metadata

The EDQP solution with the Fusion PIM Data Hub is fully integrated and the process of using them together with AutoBuild is detailed in *Oracle Enterprise Data Quality for Product Data Fusion PIM Integration Implementation and User's Guide* at

[http://docs.oracle.com/cd/E35636\\_01/doc.11116/e29148/toc.htm](http://docs.oracle.com/cd/E35636_01/doc.11116/e29148/toc.htm)

#### AutoBuild with R12 PIM Metadata

The EDQP R12 PIM Connector is integrated with most Oracle R12 PIM systems and the process of using them together with AutoBuild is detailed in *Oracle Enterprise Data Quality for Product Data R12 PIM Connector User's Guide* at

[http://docs.oracle.com/cd/E35636\\_01/doc.11116/e29140/toc.htm](http://docs.oracle.com/cd/E35636_01/doc.11116/e29140/toc.htm)

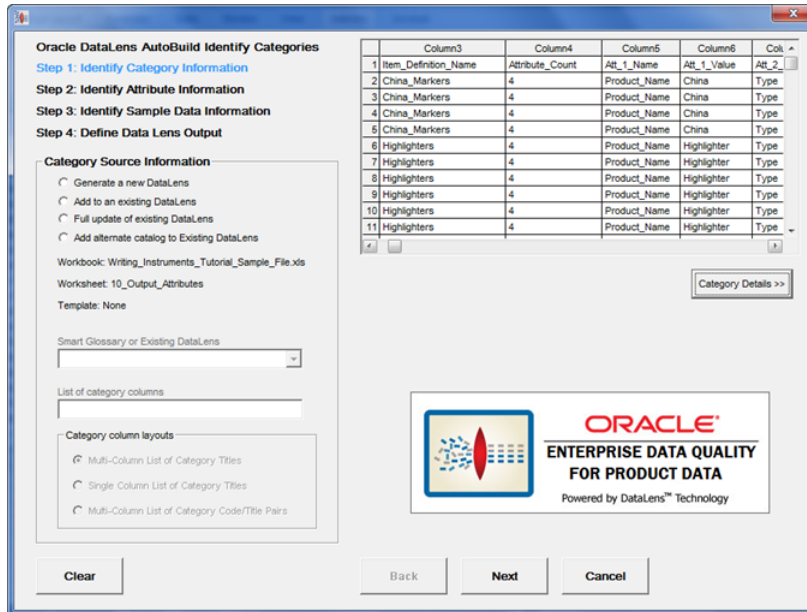
### Using AutoBuild with Non-PIM Metadata

You start the AutoBuild process by opening the Excel workbook file that contains your structured metadata. Once the workbook is open, locate the Services for Excel Add-In toolbar. For more information about installation or the AutoBuild Wizard, see "[The AutoBuild User Interface](#)" on page 2-1.

Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color
2	Balpoint Pen Refill Fine Point Blue Ink 2 / Pack	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Blue Ink
3	Balpoint Pen Refill Med Pt Black Ink 2 / Pk	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Black Ink
4	Balpoint Pen Refill Med Pt Black Ink 2 / Pk	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Black Ink
5	Balpoint Pen Refill Point Mdm Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Black Ink
6	Balpoint Pen Refill Point Mdm Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Black Ink
7	Balpoint Pen Refill mdm point .7mm ink Black	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Black Ink
8	Counterpoint pen refill Medium Point black ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Black Ink
9	W022 Security Pen Refills Medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Blue Ink
10	W023 Counter Pen Refills Medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Blue Ink
11	W027 Gel Grip Rollerball Pen Refill 1MM Red	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rollerball	Red Ink
12	W038 Gel Ink Roller Ball Pen Refills 7MM Black Ink 2 / Pack	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller Ball	Black Ink
13	W042 Gel Ink Roller Pen Refill Fine Point Orange Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller	Orange Ink
14	W043 Roller Pen Refill Fine Point Violet Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller	Violet Ink
15	W081 Pen Refill Fine point Blue 2 / Pack	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Blue Ink
16	W082 Pen Refill Ink Black 2	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Balpoint	Black Ink
17	W093 Pen Rfl Bpnt Med S / Bx	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Rfl	Bpnt	Black Ink
18	W084 Pen Rfl Med Bk 10/Bx	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Rfl	Bpnt	Black Ink
19	W093 Refill Rollerball Ink Black	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rollerball	Black Ink
20	W094 Refill Rollerball Ink Blue	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rollerball	Blue Ink
21	W103 Roller Ball Pen Refill Medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller Ball	Blue Ink
22							
23							
24	W08 Balpoint Pen Medium Point Black Barrel Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Balpoint	Black Ink
25	W107 Comfort Grip Retractable Ballpoint Pen medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Blue Ink
26	W116 Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Red Ink
27	W027 Retractable Ballpoint Pen Non-Refillable Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
28	W028 Erasable Ink Ballpoint Pen Medi Point Blue Barrel/Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Balpoint	Blue Ink
29	W029 Felt Tip Pen Extra-Fine Point Translucent blue Barrel/ purple Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Felt Tip	Purple Ink
30	W030 Felt Tip Pen Extra-Fine Point Translucent black Barrel Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Felt Tip	Red Ink
31	W079 Pen Extra-Fine Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Balpoint	Blue Ink
32	W080 Pen Extra-Fine Point Green Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Balpoint	Green Ink
33	W088 Porous Pen Extra-Fine Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Porous	Red Ink
34	W091 Porous Pen Extra-Fine Tip White Barrel/Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Porous	Black Ink
35	W095 Balpoint Pen Retractable Fine Point Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Balpoint	Black Ink
36	W096 Retractable Ballpoint Pen Fine Point Clear Barrel/Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
37	W097 Retractable Ballpoint Pen Medium Point Black Barrel/ black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
38	W098 Retractable Ballpoint Pen Refillable Medium Point Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
39	W099 Balpoint Pen Rubber Grip Fine Point Retractable Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
40	W100 Retractable Ballpoint Pen Rubber Grip Fine Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Red Ink
41	W101 Roller Ball Pen Gel Ink Refillable Medium Blue	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Roller Ball	Blue Ink
42	W102 Roller Ball Pen Refillable Medium Red gel ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Roller Ball	Red Ink
43	W107 Pen , Balpoint , Blue Ink, Medium Point	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Balpoint	Blue Ink
44							
45							
46	W12 China Mkrk Green Ink 12 / Box	Writing_Instruments	Markers	China_Markers	China	Mkrk	Green Ink
47	W13 China Markers red ink	Writing_Instruments	Markers	China_Markers	China	Makers	Red Ink
48	W14 China Markers White ink	Writing_Instruments	Markers	China_Markers	China	Makers	White Ink
49	W15 China Mkrk yellow ink	Writing_Instruments	Markers	China_Markers	China	Mkrk	Yellow Ink
50							
51							
52	W16 Chisel Point Highlighter Turquoise Ink	Writing_Instruments	Markers	Highlighters	Highlighter	Chisel Point	Turquoise Ink
53	W01 Highlighter Chisel Tip Fluorescent Purple	Writing_Instruments	Markers	Highlighters	Highlighter	Chisel Tip	Fluorescent Purple
54	W032 Highlighter Chisel Tip Fluorescent Yellow	Writing_Instruments	Markers	Highlighters	Highlighter	Chisel Tip	Fluorescent Yellow

When creating your structured data file using Excel, it is important to note Excel interprets certain data as Excel specific formats for formulas, dates, and so on. It is best to ensure that Excel interprets the data as strictly text by using a leading single quote character (') in every data cell. The leading single quote indicates to Excel that the information in the cell is to be treated as text. For example, if the cell contains a part number with slashes, dashes, leading spaces, or periods.

Click the **AutoBuild** button or select the **DataLens Tools** menu, and then click the **AutoBuild** button.



The first time the AutoBuild application is started and the wizard is displayed, the active Excel worksheet is chosen by default as the data source. The AutoBuild wizard remembers the settings chosen in all of the four wizard steps as long as Excel remains



open. The settings are saved permanently as long as the workbook is saved. At any time, to clear the current option selections, just return to the first step and click **Clear**.

## Selecting the Data Lens Type

Select the whether you want to generate a new data lens, add to an existing one, update (synchronize) an existing data lens, or add an alternate catalog to an existing data lens. In the example in the following sections, **Generate a new DataLens** is selected to show how to create a new data lens.

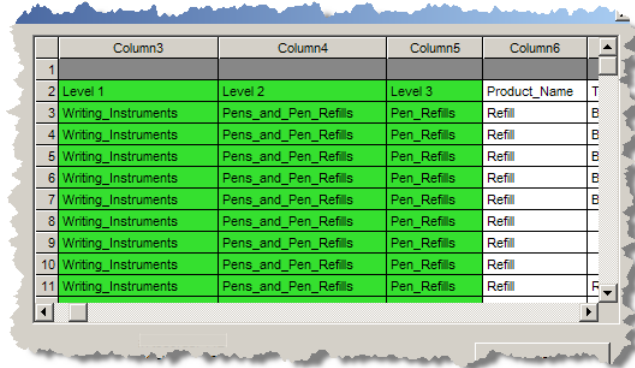
## Identifying the Category Information

You must identify which of the columns of your structured metadata contain the category information. These category columns are used by AutoBuild to create data lens Item Definitions, which automatically synchronizes the data lens knowledge to your category hierarchy. The Item Definition hierarchy is defined by either the layout of the category columns in the worksheet or by the hierarchy defined in a separate taxonomy reference file, such as the UNSPSC industry standard.

### Selecting the List of Category Columns

To identify the category columns in the selected worksheet, enter the list of columns in the **List of category columns** field. The category column(s) may contain a single column number, column ranges like "1-10", or a combination of column numbers and ranges that are comma separated like "1,5-10,16".

Once the list of category columns is entered, pressing the **Tab** key or clicking on another control causes the **Category/Attribute Data** table to redisplay the information. The selected columns are colored green.



	Column3	Column4	Column5	Column6	
1					
2	Level 1	Level 2	Level 3	Product_Name	T
3	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	B
4	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	B
5	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	B
6	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	B
7	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	B
8	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	
9	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	
10	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	
11	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	R

### Selecting the Category Column Layout

There are three possible category column layouts to choose. First and most commonly used, is the **Multi-Column List of Category Titles** option. This column layout assumes that there are multiple levels in the category hierarchy and that each column represents one level in the hierarchy. AutoBuild assumes that the category hierarchy is expressed left to right with the first column starting from the left as the highest level in the category hierarchy, and the last column on the right is the lowest level in the hierarchy. This layout is selected by default whenever the column list contains more than one column.

The second category layout is the **Single Column List of Category Titles** option. This layout is selected automatically when you list only one category column.

The third category layout is **Multi-Column List of Category Code/Title Pairs**. This layout is similar to the multi-column list previously described with the addition that each category is expressed as pair of columns. The first column contains a category code and the second column contains the category name.

This example uses the automatically selected **Multi-Column List of Category Titles** option.

**Oracle DataLens AutoBuild Identify Categories**

Step 1: Identify Category Information  
 Step 2: Identify Attribute Information  
 Step 3: Identify Sample Data Information  
 Step 4: Define Data Lens Output

**Category Source Information**

Generate a new DataLens  
 Add to an existing DataLens  
 Full update of existing DataLens  
 Add alternate catalog to Existing DataLens

Workbook: Writing\_Instruments\_Tutorial\_Sample\_File.xls  
 Worksheet: 10\_Output\_Attributes  
 Template: None

Smart Glossary or Existing DataLens  
 DLS\_import\_Template

List of category columns  
 3-5

**Category column layouts**

Multi-Column List of Category Titles  
 Single Column List of Category Titles  
 Multi-Column List of Category Code/Title Pairs

	Column3	Column4	Column5	Column6	Col
1	Item_Definition_Name	Attribute_Count	Att_1_Name	Att_1_Value	Att_2
2	China_Markers	4	Product_Name	China	Type
3	China_Markers	4	Product_Name	China	Type
4	China_Markers	4	Product_Name	China	Type
5	China_Markers	4	Product_Name	China	Type
6	Highlighters	4	Product_Name	Highlighter	Type
7	Highlighters	4	Product_Name	Highlighter	Type
8	Highlighters	4	Product_Name	Highlighter	Type
9	Highlighters	4	Product_Name	Highlighter	Type
10	Highlighters	4	Product_Name	Highlighter	Type
11	Highlighters	4	Product_Name	Highlighter	Type

Category Details >>

ORACLE  
 ENTERPRISE DATA QUALITY  
 FOR PRODUCT DATA  
 Powered by DataLens™ Technology

Clear Back Next Cancel

Click **Next** to advance to the next step.

## Identifying the Attribute Information

You must identify the location and layout of the attribute and attribute value information found in the structured data. This attribute information is used by AutoBuild to create the data lens Item Definitions attributes and automatically synchronize the data lens knowledge to your business domain schema. The attribute value information is used by AutoBuild to create data lens phrases and terminology for automatic text recognition. Additionally, AutoBuild automatically associates the proper data lens phrases to the correct Item Definition attributes thus providing automatic attribute extraction.

### Selecting the Required Attributes Information

First, you identify the attribute columns that are used for the Item Definition required attributes. These required attribute columns typically consist of item names, types, and brands.

In this example, column 6 contains the product name as in the following:

	Column3	Column4	Column5	Column6
1				
2	Level 1	Level 2	Level 3	Product_Name
3	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
4	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
5	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
6	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
7	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
8	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
9	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
10	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill
11	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill

As each required logic column specification is entered, the **Category/Attribute Data** table is updated, colorizing the selected columns in purple as in the following:

	Column4	Column5	Column6	Column7
1				
2	Level 2	Level 3	Product_Name	Type
3	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
4	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
5	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
6	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
7	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
8	Pens_and_Pen_Refills	Pen_Refills	Refill	
9	Pens_and_Pen_Refills	Pen_Refills	Refill	
10	Pens_and_Pen_Refills	Pen_Refills	Refill	
11	Pens_and_Pen_Refills	Pen_Refills	Refill	Rollerball

Sometimes structured metadata columns can be used for more than one purpose. For example, name and type information often overlaps with category information.

### Selecting the Scoring Attributes Information

The attribute information is laid out as a set of columns just to the right of the category information as in the following example.

	Column6	Column7	Column8	Column9	Column10	Column11	Column12
1							
2	Product_Name	Type	Ink_color	Size	Ink_Type	Barrel_Color	Grip_Ty
3	Refill	Ballpoint	Blue Ink				
4	Refill	Ballpoint	Black Ink				
5	Refill	Ballpoint	black, ink				
6	Refill	Ballpoint	Black Ink				
7	Refill	Ballpoint	ink Black	.7 mm			
8	Refill		black Ink				
9	Refill		Blue Ink				
10	Refill		Blue Ink				
11	Refill	Rollerball		1 MM			Gel Grip

In this example, columns 7 through 13 are used as the attribute scoring information.

To add attribute scoring, select the **Add Scoring Attributes** check box. In this example, the **Attribute names in the category file header** option must be selected because the names are in a header row.

Next, you must enter the attribute column list in the **Scoring Attribute** field. You can supply a comma-separated list of column numbers, as well as column ranges.

Once the list of attribute columns is accepted, the selected scoring attributes are colored blue as previously described.

	Column5	Column6	Column7	Column8	Column9
1	Level 3	Product_Name	Type	Ink_color	Size
2	Pen_Refills	Refill	Ballpoint	Blue Ink	
3	Pen_Refills	Refill	Ballpoint	Black Ink	
4	Pen_Refills	Refill	Ballpoint	black, ink	
5	Pen_Refills	Refill	Ballpoint	Black Ink	
6	Pen_Refills	Refill	Ballpoint	ink Black	.7 mm
7	Pen_Refills	Refill		black Ink	
8	Pen_Refills	Refill		Blue Ink	
9	Pen_Refills	Refill		Blue Ink	
10	Pen_Refills	Refill		Blue Ink	
11	Pen_Refills	Refill	Rollerball		1 MM

The example selections are as follows:

**Oracle DataLens AutoBuild Identify Attributes**

Step 1: Identify Category Information  
 Step 2: Identify Attribute Information  
 Step 3: Identify Sample Data Information  
 Step 4: Define Data Lens Output

**Attribute Information**

**Required Attributes**

Add Required Attributes      Single term columns

Item Name Column(s)    Any Of    6

Item Type Column(s)    All Of   

Unique Name Column(s)    Any Of   

Use names and types to create required attributes

**Scoring Attributes**

Add Scoring Attributes

Attribute names are in the same row as the category

Row contains Attribute Names only

Row contains Attribute Name / Value pairs

Row contains Attribute Name/Value/UOM

Attribute names in the category file header

Scoring Attributes      Single term columns

7-13

**Optional Attributes**

Add Optional Attributes

All Optional attribute columns (single term)

**Attribute Details**

Attribute Values

Phrase Separator:    Include attribute name in phrase for context:

Maximum Terms in a Phrase: 5     Append as needed

Map values directly to each attribute     Insert always

Attribute values are Part Number Codes     Append always

Never

Standardizations

#Required Standardization Columns: 0     Create Replace All Rules

#Scoring Standardization Columns: 0

#Optional Standardization Columns: 0

Back    Next    Cancel

Click **Next** to advance to the next step.

## Identifying the Sample Data Information

You must identify the location and layout of any available sample data that can be used by AutoBuild when creating a data lens. This sample data information is used by AutoBuild to add both category-based sample data files and 100 record random sample data files to the resulting data lens. The random sample files allow a data analyst to determine the convergence of the data lens knowledge. The category-based files allow a content expert to perform maintenance easily by focusing on specific item commodities. When no sample data is available, you can use the option to create a sample from the available attribute information.

### Selecting the Sample Data Source

First, select the data source for the sample data information. You have the options of using the Knowledge Studio to collect the sample data, create sample data from the attribute information identified in the previous wizard step, or by selecting sample data from a separate file.

This example creates the samples from attributes so the appropriate option is selected, as well as the **Prefix Sample Data Columns (listed below)** check box.

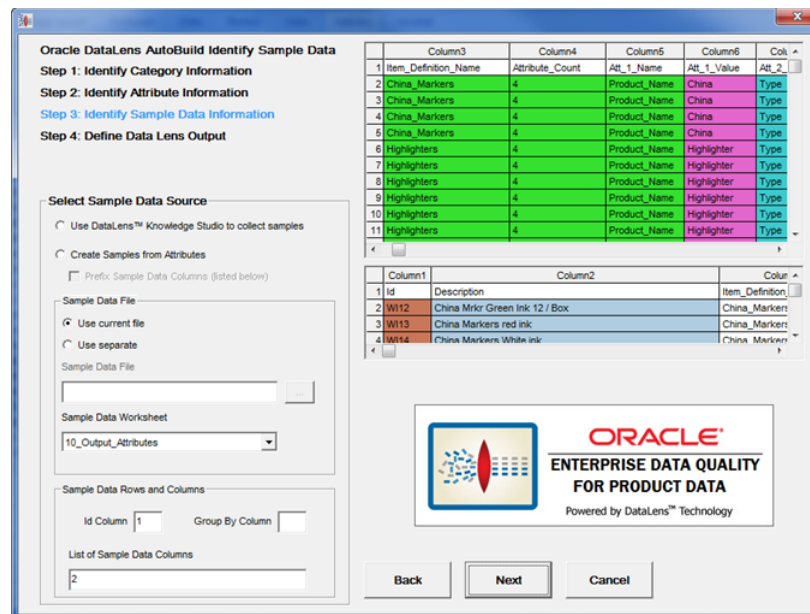
With these selections, the system automatically creates the sample data descriptions for use in the Knowledge Studio from the available attribute information. This can be very useful when example descriptions are not available. In addition, the generated descriptions are prefixed with other columns from our structured metadata source. In this example, the name, type, and brand name columns are included in the generated descriptions.

In order to generate sample descriptions effectively from the supplied attribute information, the attribute information must be linked together through a unique key. These columns are columns 5 through 7 and this information is entered in the **List of Sample Data Columns** field.

The source metadata in the 10\_Output\_Attributes worksheet of the open workbook is used in this example.

In this example, the Id number is used as the unique key and the column number is entered in the **Id Column** field.

The example selections are as follows:

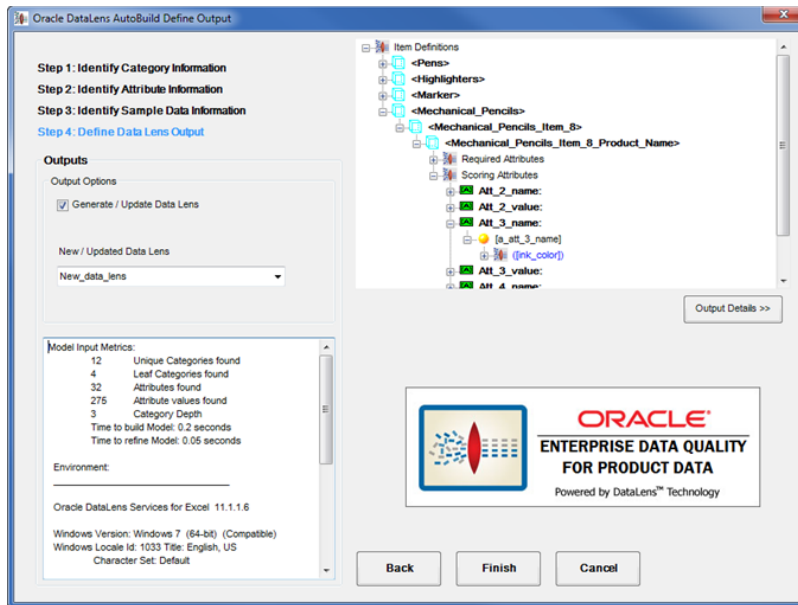


Click **Next** to advance to the final step.

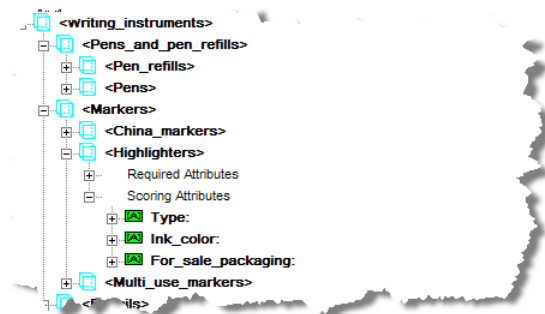
Before advancing, AutoBuild reads the category and attribute information that you have identified using the wizard and a model is generated containing the resulting Item Definitions, Item Definition attributes, attribute to phrase associations, and related phrases and terms.

## Defining the Data Lens Output

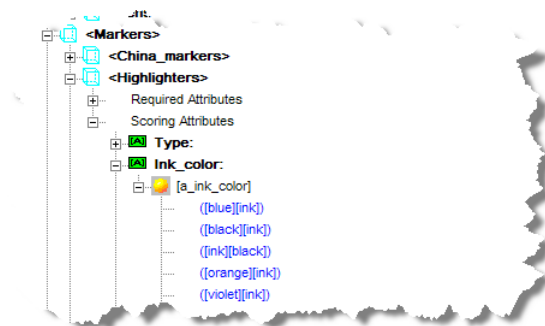
The final step of the wizard enables you to review the structure of the data lens that will be created, name it, and select any specific options before generating the data lens.



Review the Item Definitions and their associated attributes to ensure that the expected results will be built into the data lens. In this example, the 'Writing\_Instruments' Item Definition displays the entire set of *child* Item Definitions. Expanding 'Mechanical\_Pencils', and then the **Scoring Attributes** allows you to view these attributes as in the following:

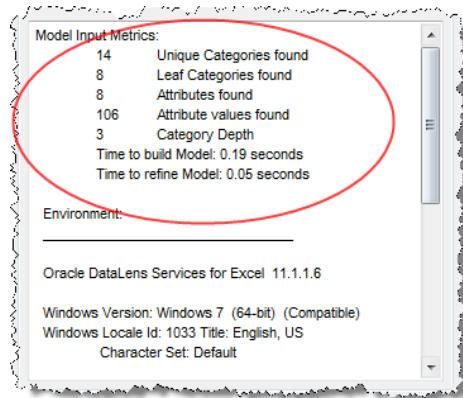


You can drill even further. Expanding the 'Ink\_color' attribute, and then the '[a\_ink\_color]' phrase allows you to view the entire phrase production as in the following:



If your item definition hierarchy did not turn out as expected, then select the **Back** button and return to the prior AutoBuild dialogs to refine your selections.

You should review the messages in the status pane to ensure that the metrics reported make sense and that there are no warning or error messages.



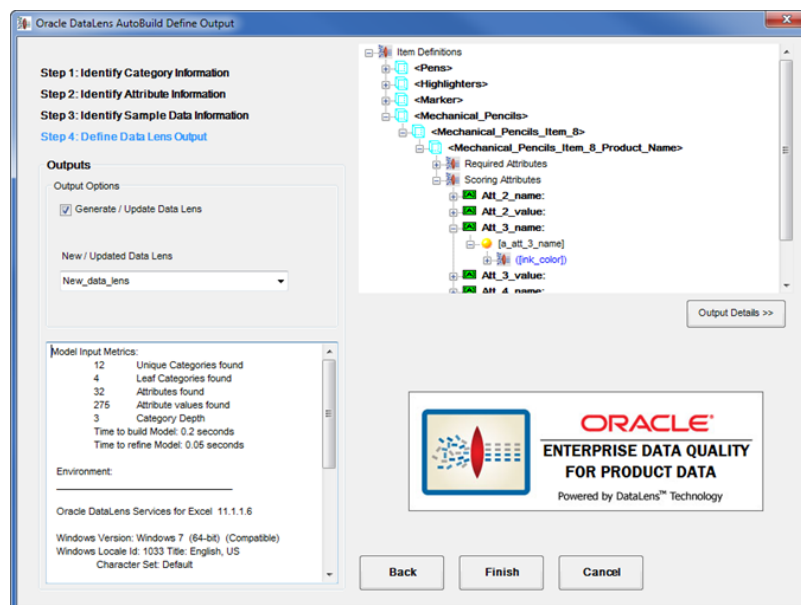
Typically, when warnings or errors are displayed AutoBuild allows you to continue to create a new data lens. Most often AutoBuild reports duplicate or conflicting category information that was found in your structured metadata examples. You should review the warnings to determine whether the structured input should be corrected.

---

**Note:** Do not use the "DLS\_" prefix to any data lens name as this is reserved to indicate that a data lens is a Smart Glossary.

---

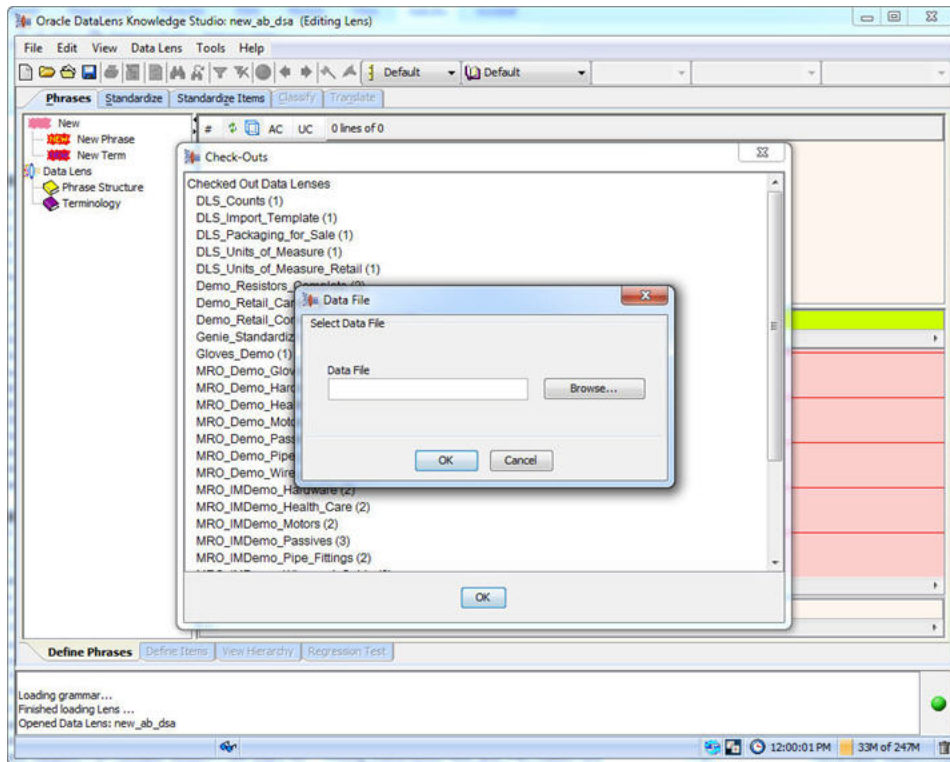
The example selections for the final wizard step are as follows:



To complete the AutoBuild process and create your new data lens, click **Finish**.

The new data lens is created from the category and attribute information provided in your structured data. When the AutoBuild process completes, you are returned to your worksheet. You can now start Knowledge Studio and open the data lens you just created as in the following:





You are now ready to modify and extend your data lens to meet your application requirements.

## Finishing Your New Data Lens

The data lens generated by AutoBuild contains the structure you need in order to complete your data lens development effort. The Item Definitions, attributes, phrases and terms generated by AutoBuild function as a guide for the finishing steps of the development process. You should become familiar with the generated data lens structure before proceeding.

During the completion of your data lens, you should reuse as much of the generated components as possible to remain consistent with the original intent of the structured item information.

The following is an overview of the steps you should take in completing your new data lens using the Knowledge Studio:

1. Complete the Item Definition recognition logic.
 

This is particularly important when no identifying item name or item type information is available. Once you have completed the Item Definition recognition logic many of the initial recognition ambiguities are resolved by the Knowledge Studio. This ensures that the remaining steps are successful.
2. Examine all of the phrase level ambiguities in your data lens.
 

You should review each of the standard random sample files and remove or resolve these multiple references.
3. Use standard sample convergence to complete the item recognition.
 

Apply the standard data lens development methodology for completing the data lens recognition using random sample files.



#### 4. Add standardization rules.

Create the standardization rules that are required by your data quality standards and/or the data application that you are developing. This effort can be minimized using a standard Smart Glossary in the AutoBuild process. The Smart Glossary can provide immediate standardization for all of your global terms and phrases. The Smart Glossary can also be used to provide all of your unit conversions up front. For the terms that are unique to this data lens, you can use the **Copy Full Form** function found on the **Standardize** tab to jump-start your standardization efforts.

Once you have completed your term and phrase standardization, you can finish your standardization efforts by selecting the desired attributes for output and setting up any required matching rules.

---

**Note:** If you selected **Create Replace all Rules** option, the preceding standardization occurs automatically.

---

#### 5. Create and validate data lens regression test sets.

Use the Knowledge Studio to create regression test set(s) in the data lens so that future maintenance efforts can be properly tested. You should also check the regression set carefully the first time through to make sure that the results are all as expected and make adjustments as needed.

#### 6. Check in, deploy to development, and test your new data lens against the application test cases.

Each data application created with the Application Studio has associated test cases that need to be executed whenever a data lens is added or updated.

#### 7. Improve the AutoBuild configuration.

Use the results from the preceding steps to update and improve the AutoBuild Smart Glossary that you use to create new data lenses. When possible, you should use the process results to improve the content of the structured item information. Refining the information used by AutoBuild provides clear improvements in the AutoBuild generated data lenses.

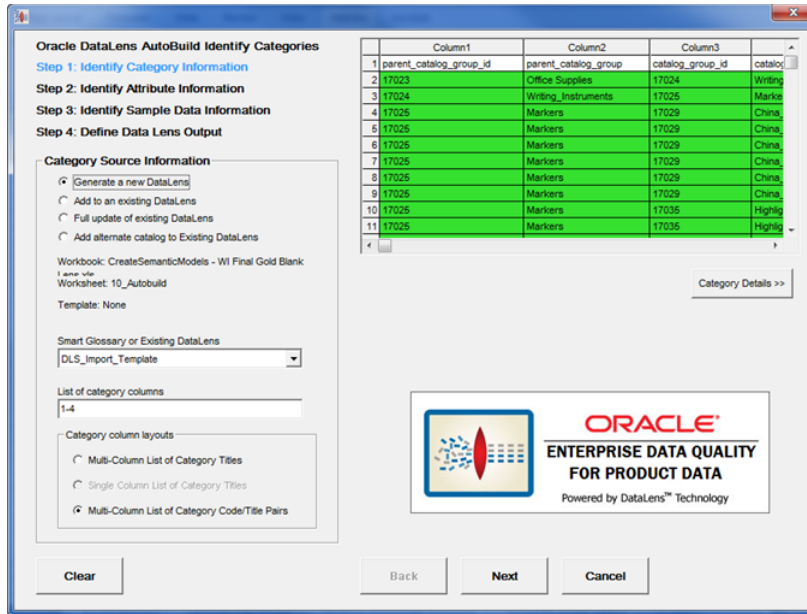
## Data Pairs Data Lens Example

This section provides an example of using data that contains both the attribute name and value in each record. This paired metadata is used in the AutoBuild process in the same manner as previously described though the Attribute Information options are different.

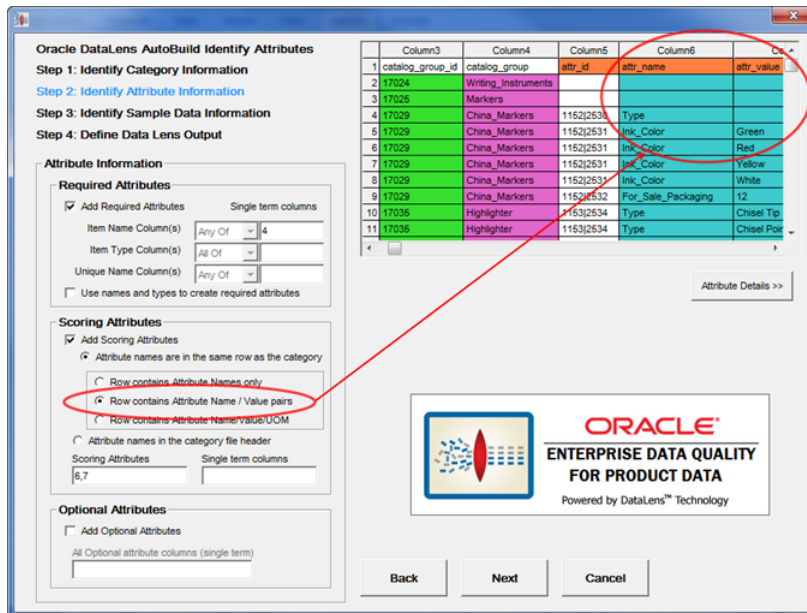
An example of a paired metadata Excel worksheet is as follows:

A	B	C	D	E	F	G
parent	parent_catalog_group	catalog_group_id	catalog_group	attr_id	attr_name	attr_value
17023	Office Supplies	17024	Writing_Instruments			
17024	Writing_Instruments	17025	Markers			
17025	Markers	17029	China_Markers	1152 2530	Type	
17025	Markers	17029	China_Markers	1152 2531	Ink_Color	Green
17025	Markers	17029	China_Markers	1152 2531	Ink_Color	Red
17025	Markers	17029	China_Markers	1152 2531	Ink_Color	Yellow
17025	Markers	17029	China_Markers	1152 2531	Ink_Color	White
17025	Markers	17029	China_Markers	1152 2532	For_Sale_Packaging	12
17025	Markers	17035	Highlighter	1153 2534	Type	Chisel Tip
17025	Markers	17035	Highlighter	1153 2534	Type	Chisel Point
17025	Markers	17035	Highlighter	1153 2535	Ink_color	Pink

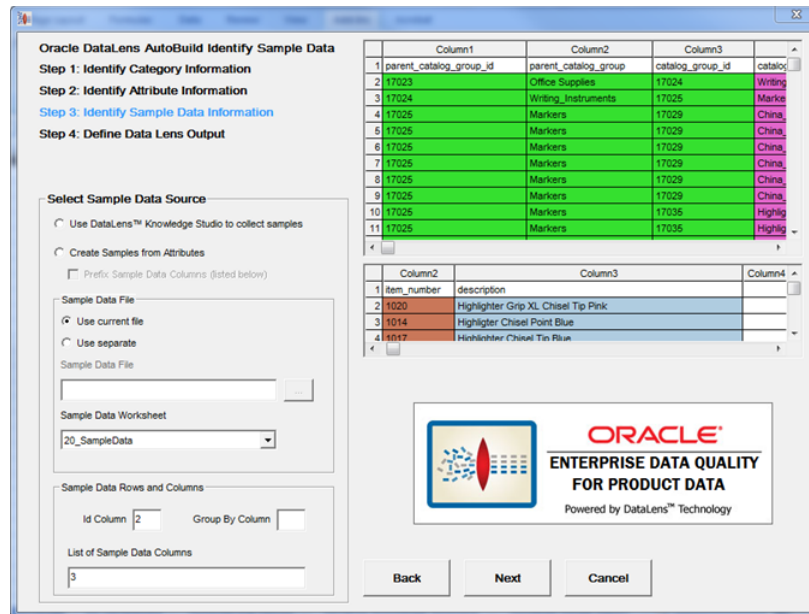
Selecting the category information could look like the following:



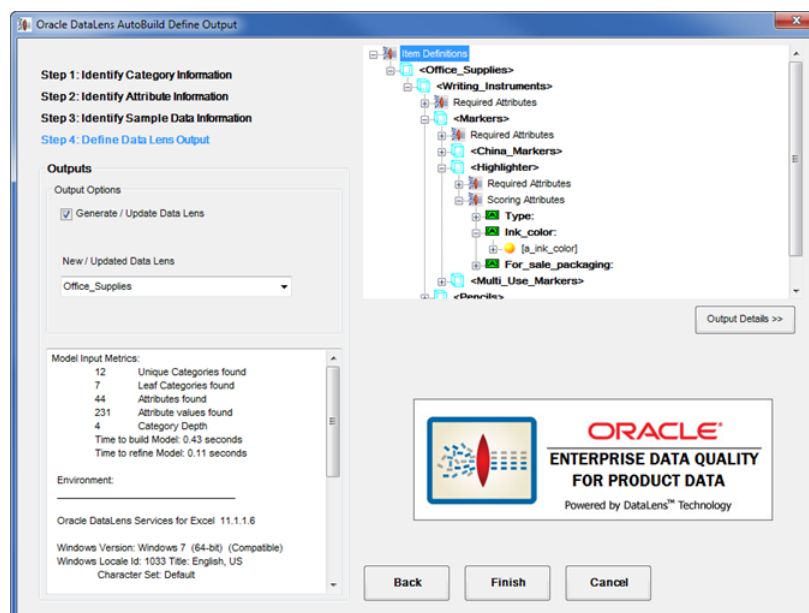
To indicate that the source metadata is paired occurs in Step 2 of the AutoBuild wizard as in the following:



In this example, the scoring attributes are columns 6 and 7, which is the attribute name and value pair.



In the final step of the wizard, the Item Definitions appear similar in structure to the example that did not have paired metadata in the Step 4: Define Data Lens Output section.



Your metadata can be organized into columns in a myriad of ways in an Excel worksheet. How you select the AutoBuild options defines the knowledge structure that is built in the resulting data lens.

## Further Considerations

You are encouraged to review the following to make the most of the use of the AutoBuild application.

### **Categories**

Categories can be used for sample descriptions, item names, and item types while item names and types can be used for categories.

### **Extending an Existing Data Lens**

Create a new data lens from your structured item information then refine your data lens as appropriate. Rerun AutoBuild using one of the add or update an existing data lens options in "[Step 1: Identify Category Information](#)" to modify your existing data lens.

### **Create Examples**

When there is no well-structured, *gold* data, creating just a few example records per category can greatly decrease data lens development time. Additionally, you should use the data lens terminology importing function to close the gaps in data recognition. For more information about importing phrases and terms, see *Oracle Enterprise Data Quality for Product Data Knowledge Studio Reference Guide*.

### **Overloaded Category Names**

In general, category names should only use item names and types; those that contain additional information are considered *overloaded*. For example, the overloaded category name "Surgical Gloves Latex-Free Powder-Free Nitrile Size 7" should be "Surgical Gloves" to avoid overloading.

You should standardize your category names or use the correct category level because this reduces effort and optimizes the data lens attribute centric matching.

### **Overloaded Attribute Names**

Similar to category names, attribute names should not be overloaded. For example, 'Heart Rate Monitor Operating Voltage' should be 'Voltage'

You should standardize attribute names before using AutoBuild to create a data lens because this reduces effort and optimizes the data lens attribute centric matching.