Oracle® Product Data Quality

AutoBuild Reference Guide

Version 5.5

February 2010

ORACLE

Oracle Product Data Quality AutoBuild Reference Guide, Version 5.5

Copyright © 2001, 2010, Oracle and/or its affiliates. All rights reserved.

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

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 software or related documentation 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 USA, Inc., 500 Oracle Parkway, Redwood City, CA 94065.

This software 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 which may create a risk of personal injury. If you use this software in dangerous applications, then you shall be responsible to take all appropriate fail-safe, backup, redundancy, and other measures to ensure the safe use of this software. Oracle Corporation and its affiliates disclaim any liability for any damages caused by use of this software in dangerous applications.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

This software 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.

This product is currently English only.

Contents

Preface iv
About this Book iv
Intended Audience ..... *v*
Conventions Used in This Book ..... *v*
Related Information ..... *vi*
Chapter 1 1
Introduction 1
Prerequisites to Using AutoBuild 4
Data Prerequisites *4*
Identifying Structured Item Information *4*
Sources of Structured Item Content 6
Quantity of Structured Information 6
Supported Structured item Information Formats 6
Supported Category Formats *6*
Supported Attribute Formats *7*
AutoBuild and Smart Glossaries 8
Default Smart Glossary *8*
Smart Glossary Foundation for Generated Data Lens *8*
Item Definition Model Merged Into Smart Glossary 8
Replacing Item Definition Model Phrase structures with Smart Glossary Phrase structures 9
Chapter 2 10
AutoBuild User Interface 10
The AutoBuild User Interface 12
Step 1: Identify Category Information *14*
Category Source Information Section 14
Category/Attribute Data Table 16
Category Details 17

Step 2: Identify Attribute Information *19*
Attribute Information Section 20
Required Attributes Box *20*
Scoring Attributes Box *25*
Optional Attributes Box *28*
Category/Attribute Data Table 29
Attribute Details 30
Scoring Attributes in Separate File(s) Box *30*
Attribute Values Box *32*
Standardizations Box *33*
Step 3: Identify Sample Data Information *34*
Select Sample Data Source Section 34
Category/Attribute Data Table 36
Sample Data Table *37*
Sample Details 38
Step 4: Define Data Lens Output *38*
Outputs Section 39
AutoBuild Item Definition Explorer 39
Output Details 41
AutoBuild Reports 43
AutoLearning Metrics Report 44
Semantic Knowledge Structure 45
AutoBuild Details Report 46
Chapter 3 47
The AutoBuild Process 47
Using the AutoBuild Process 48
Identifying the Category Information *49*
Selecting the List of Category Columns 49
Selecting the Category Column Layout 50
Selecting the User 50
Identifying the Attribute Information *51*
Selecting the Required Attributes Information 52
Selecting the Scoring Attributes Information 53
Identifying the Sample Data Information *54*

Selecting the Sample Data Source 54
Defining the Data Lens Output *56*
Finishing Your New Data Lens 60
Data Pairs Data Lens Example 62
Further Considerations 66
Appendix A 67
Installing the Software 67
Installing the Software 68
Accepting the Security Warning *68*

Preface

About this Book

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

- Chapters 1 and 2 describe the basic application features.
- Chapter 3 describes the AutoBuild process using an example.

To understand all of the features presented, you must use this reference guide in conjunction with the Oracle Product Data Quality documents listed in Related Information.

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

- *Oracle Product Data Quality Knowledge Studio Reference Guide*

You must have the Oracle Product Data Quality client software installed on your computer. For more information, see Installing the Software on page 67. Additionally, you must have the Oracle Product Data Quality AutoBuild product installed on your system. For more information and installation instructions, see The *Oracle Product Data Quality Services for Excel Reference Guide.*

Intended Audience

You should have a basic understanding of the DataLens Technology. Including the functionality of the Oracle Product Data Quality Knowledge Studio, how it recognizes and standardizes data, and the Oracle Product Data Quality Governance Studio applications.

This document is intended for all users of the DataLens Technology, including:

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

Conventions Used in This Book

The following typographical conventions that are used in this book:

file, directory, or path name

Used for the names of files, directories, or path names.

<server>

Used to indicate text that is to be replaced by user-supplied values.

bold

Used for new terms, new concepts, graphical user interface elements, or keyboard keys.

italics

Shows a book or cross-reference to related material or for emphasis.

Ctrl+x

Used to indicate a key sequence. A sequence such as **Ctrl-x** indicates that you must hold down the key-labeled Ctrl while you press another key or button.

Tip: Indicates useful hints.

Note: Indicates additional or supplemental information.

Important: Indicates essential information.

Warning: Indicates essential information to follow to avoid data loss, data corruption, or damage to hardware or software.

Related Information

The following documents and resources contain useful information:

- The *Oracle Product Data Quality Application Studio Reference Guide* provides information about creating and maintaining Data Service Applications (DSAs).
- The *Oracle Product Data Quality Knowledge Studio Reference Guide* provides information about creating and maintaining data lenses.
- The *Oracle Product Data Quality Glossary* provides definitions to commonly used Oracle Product Data Quality technology terms.
- The *Oracle Product Data Quality Governance Studio Reference Guide* provides information about creating and maintaining Data Service Applications (DSAs).
- The *Oracle Product Data Quality Services for Excel Reference Guide* provides information about creating a DSA based on data contained in a Microsoft (MS) Excel spreadsheet.
- The *Oracle Product Data Quality Task Manager Reference Guide* provides information about managing tasks created with the Task Manager or Governance Studio applications.
- The *Oracle Product Data Quality Oracle DataLens Installation Guide* provides detailed Oracle Product Data Quality Oracle DataLens Server installation instructions.
- The *Oracle Product Data Quality Oracle DataLens Server Administration Guide* provides information about installing and managing an Oracle DataLens Server.
- The *Oracle Product Data Quality Connector Implementation Guid*e provides information about installing and configuring Oracle Product Data Quality.
- The *Oracle Product Data Quality COM Interface Guide* provides information about installing and using the Oracle DataLens Server COM APIs.
- The *Oracle Product Data Quality Java Interface Guide* provides information about installing and using the Oracle DataLens Server Java APIs.
- The *Oracle Product Data Quality User Guide* provides information about how to use Oracle Product Data Quality.

Chapter 1

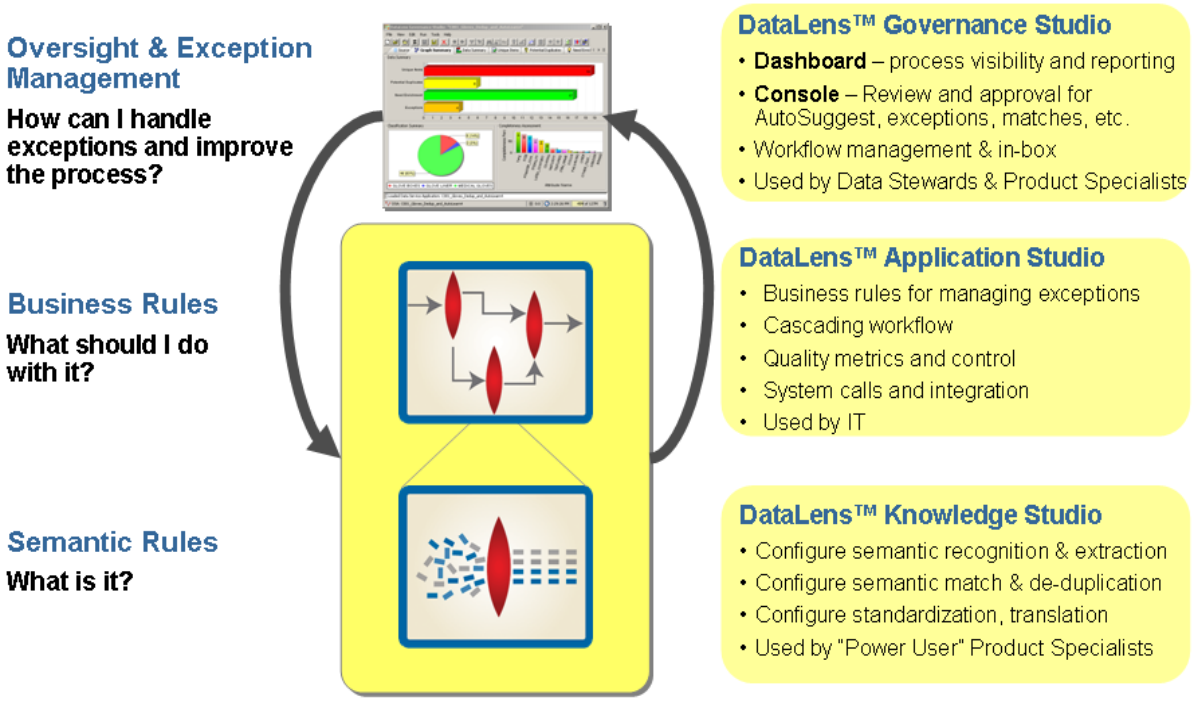
Introduction

In this chapter

- Prerequisites to Using AutoBuild 4
- AutoBuild and Smart Glossaries 8

Oracle Product Data Quality is built on industry-leading DataLens™ Technology to standardize, match, enrich, and correct product data from different sources and systems. The core DataLens Technology uses patented semantic technology designed from the ground up to tackle the extreme variability typical of product data.

Oracle Product Data Quality uses three core DataLens Technology modules: Governance Studio, Knowledge Studio, and Application Studio. The following figure illustrates the process flow of these modules.

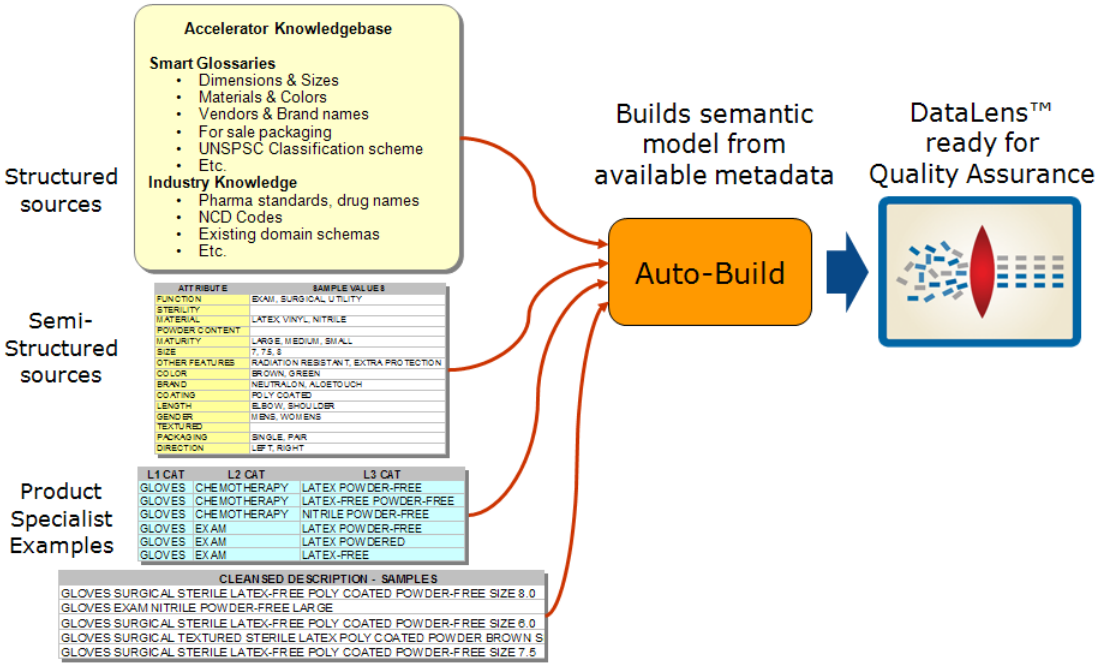


The AutoBuild Application is available from the Services for Excel additional installed (add-in) product. For more information about this product, see *Oracle Product Data Quality Services for Excel Reference Guide*.

AutoBuild can rapidly leverage your existing product information using Oracle Product Data Quality Smart Glossaries to create initial data lenses specific to your enterprise content. For example, a company may already know that their inventory includes pens, pencils, magic markers, and highlighters and may already have example structured content that contain information about each of these products. This data knowledge can be translated from an Excel spreadsheet into a data lens by the AutoBuild Application, which can save a significant amount of effort and cost.

AutoBuild constructs the initial data lens by examining the product data examples. 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.

Prerequisites to Using AutoBuild

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

Data Prerequisites

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

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/Function Attributes

Examples of these attributes include color, weight, size, material, packaging, etc.

Valid Attributes

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

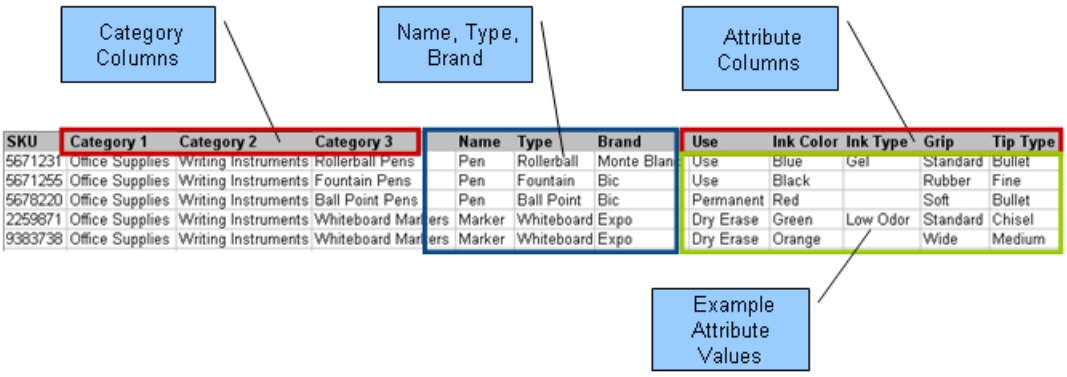
You can expect maximum benefit from the AutoBuild application when your structured item examples contain attribute values that are expressed in full, unabbreviated form. The Oracle Product Data Quality can leverage these *full-form* examples and automatically recognize a broad range of term abbreviations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

Identifying Structured Item Information

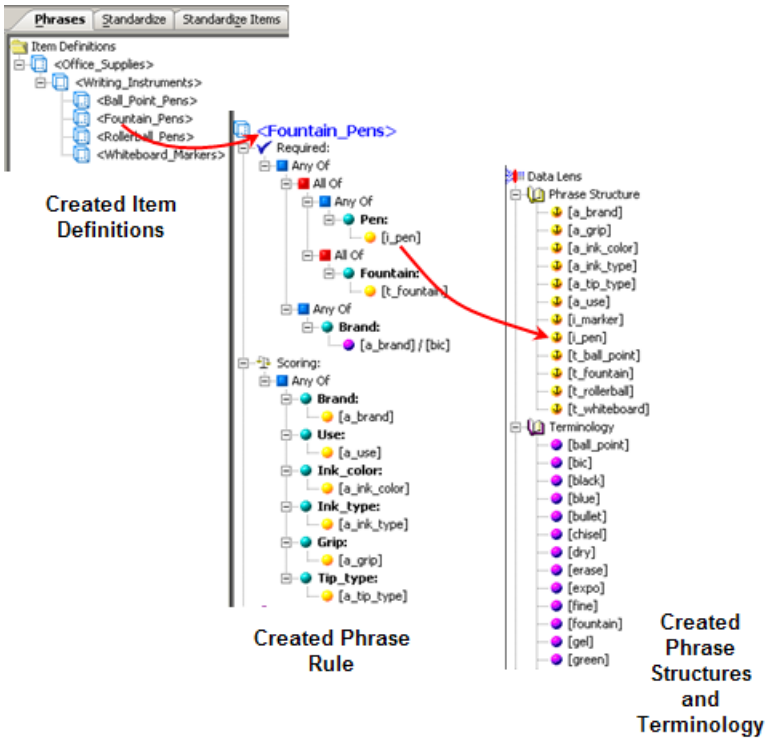
The main sources of structured item information that can be used with AutoBuild are the following:

- 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 data file:



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 spreadsheets
- Existing eCommerce site information
- Structured examples used to define your current product categories

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 Lens Item Definitions. If you have example attribute values associated with the attributes, then AutoBuild can create phrases and terminology (term) 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.

In general, AutoBuild works best when you provide a few example structured items per category. These example items should represent the product categories and attributes you want to capture in your enterprise systems.

Supported Structured item Information Formats

Product information exports can vary greatly between 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.
- 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 and Smart Glossaries

Oracle Product Data Quality Smart Glossaries are data lenses designed to be applied to a broad range of data domains. There are Smart Glossaries addressing generally applicable recognition of materials, colors, trade names, and units of measure for example. Each Smart Glossary is design to be importable into an existing lens, or can be used as the basis for creating a new lens. It is easy to combine Smart Glossaries to provide the combination of recognition for your domains of data.

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 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, and product packaging.

You can change the Smart Glossary AutoBuild uses for data lens generation each time you create a data lens. You can also create and configure AutoBuild to use your own combination of Smart Glossaries that is most applicable to you domain data.

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

> Note: You must check out the Smart Glossary you want to use as a template to generate your data lens, from your Oracle Product Data Quality as you would with any other Smart Glossary. For information on Smart Glossaries, see the *Oracle Product Data Quality 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 thus it is the foundation. Standard data lens options found in the Knowledge Studio options are copied from the selected Smart Glossary to the new data lens. All of the transformation types are copied from the Smart Glossary to the generated data lens as well. This includes the classification types, standardization types, and the translation types. Any Unit Conversion rules are copied to the generated data lens. The generated data lens also contains all of the phrases and terminology defined in the Smart Glossary.

Item Definition Model Merged Into Smart Glossary

AutoBuild merges all of the generated item definitions, attributes, and associated phrases and terms into the selected Smart Glossary. If there is overlap between the Smart Glossary phrases and terms and the phrases and terms generated from the supplied structured item information, AutoBuild keeps the Smart Glossary phrases and terms and ignores the generated phrases and terms.

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 Oracle Product Data Quality standard phrase structures from the Smart Glossary. The benefit of using the Smart Glossary version of these phrase structures is that the Smart Glossary phrase structures have already been standardized. In most cases, the unit conversion rules have already been applied, and the quality of the Smart Glossaries has been verified.

Chapter 2

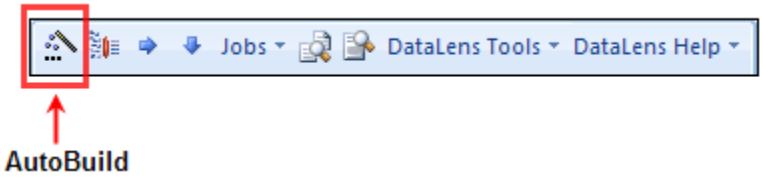
AutoBuild User Interface

In this chapter

- The AutoBuild User Interface 12
- AutoBuild Reports 43

The AutoBuild application is installed as an add-in program for Microsoft Excel, which provides a custom toolbar that is added to Excel. AutoBuild is installed as part the Services for Excel add-in product. For more information and installation instructions, see The *Oracle Product Data Quality Services for Excel Reference Guide.*

The AutoBuild function is accessed using the AutoBuild button on the AutoBuild toolbar as shown in the following:



Note: The **AutoBuild** button is only visible when the appropriate Preference option is set. For more information, see the *Oracle Product Data Quality Services for Excel Reference Guide*.

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 38.

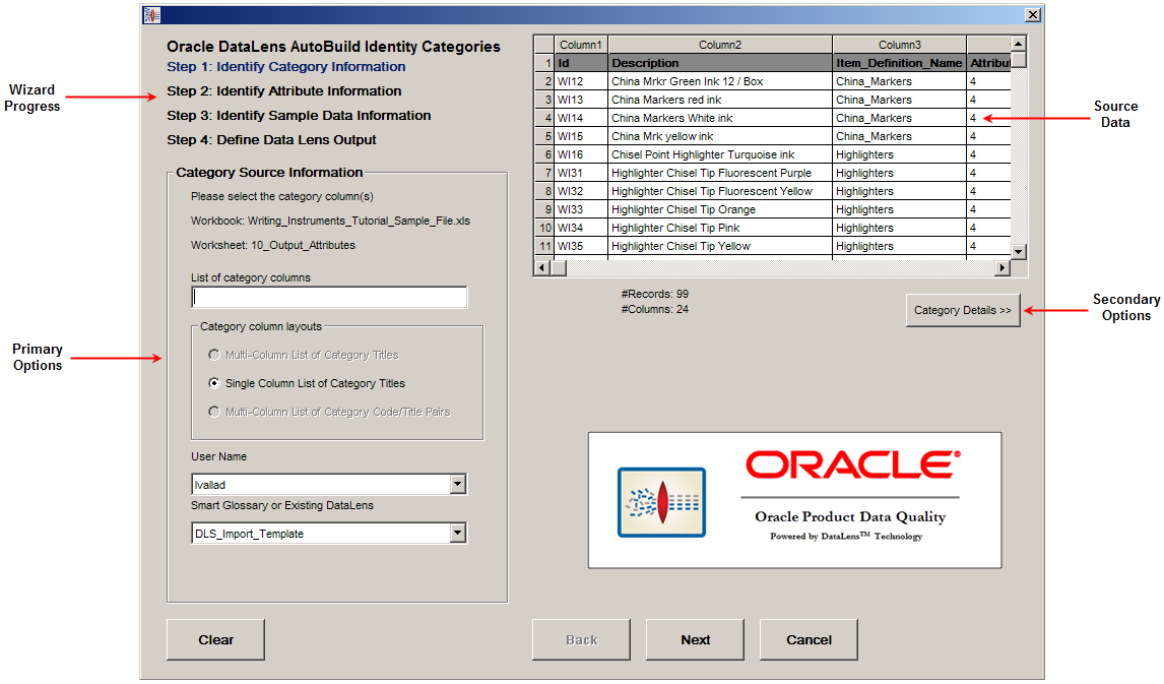
If this toolbar does not appear in Excel, the product may not be installed. For installation instructions and more information, see the *Oracle Product Data Quality Services for Excel Reference Guide* or contact Professional Services for assistance.

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 the Identify Category Information.



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

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

Each step of the wizard contains the four following four areas:

Wizard Progress

The step of the wizard that is active is colored blue

Source Data

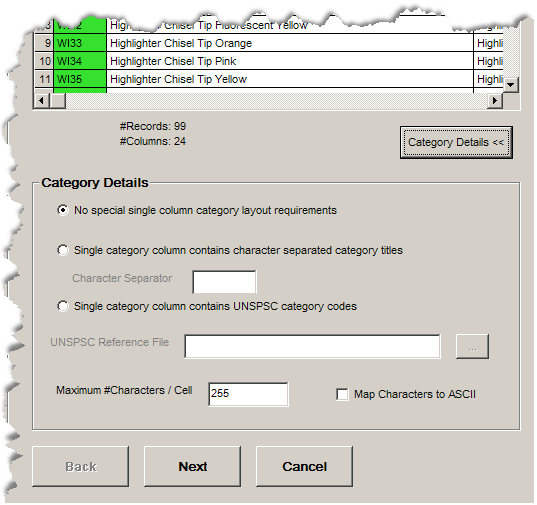
The data from the active Excel spreadsheet is displayed. The columns can be resized as in Excel and you can use the scrollbars to view more data. The columns are colorized as you make selections.

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 lens creation process with the active Excel workbook.

Back

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

Next

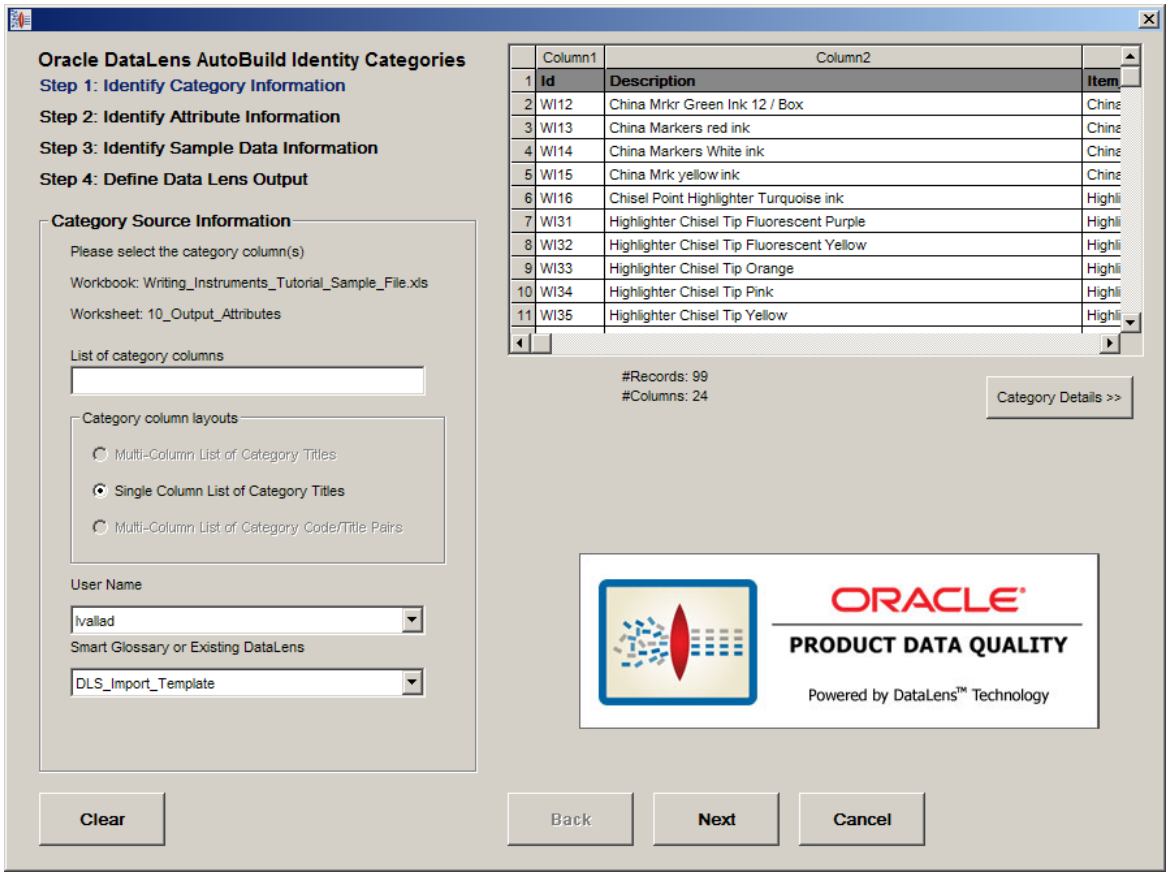
Advances the wizard to the next step in the process.

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 spreadsheet, 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:

List of category columns Textbox

Use this textbox to specify the columns that are to be used for the item definition hierarchy. 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 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).

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 data table is updated colorizing your category column selections in green.

Category column layout Radio Buttons

There are three possible category column layouts to choose from 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** textbox 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** textbox.

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.

User Name List Box

AutoBuild populates this drop-down box with the list of Oracle DataLens Server local users that have a corresponding login on the server. By default, the first user name is selected for you. If you have more than one local user name, then you can select the user name you want to use for the current AutoBuild session. The data lens that is created is located in the specified user's lens directory structure in …\appdata\datalens\login.

Smart Glossary or Existing Data Lens List Box

AutoBuild allows you to select the Smart Glossary 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.

Category/Attribute Data Table

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

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

#Records: 99
#Columns: 24

Category Details >>

Data Table

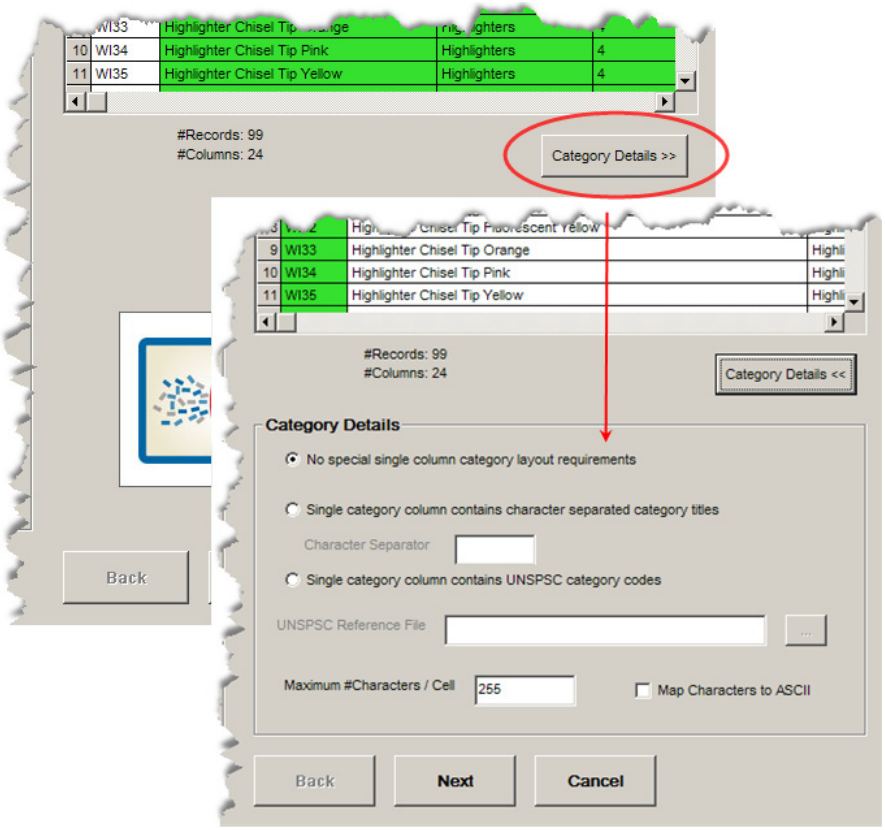
The data table displays the first 30 rows and 80 columns of the structured Excel spreadsheet data source. As the columns are identified in the **List of category columns** textbox, they are colorized green as in the preceding example.

Data Source Metrics

The total number of records and columns found in the structured data source appears immediately below the data table.

Category Details

The **Category Details** section is used 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** textbox. These options are used as follows:

No special single column category layout requirements radio button

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

Single category column contains character separated category titles Radio Button

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

Character Separator Textbox

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

Single category column contains UNSPSC category codes Radio Button

This option indicates that the single identified category column contains UNSPSC category codes. AutoBuild is UNSPSC aware and can build the complete UNSPSC

hierarchy for you automatically from a UNSPSC reference file that you supply with the following:

Note: The Oracle Product Data Quality ships with several version 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 the local directory containing the set of UNSPSC reference files delivered with the product. You can select one of these files or select your own.

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.

Maximum #Characters Textbox

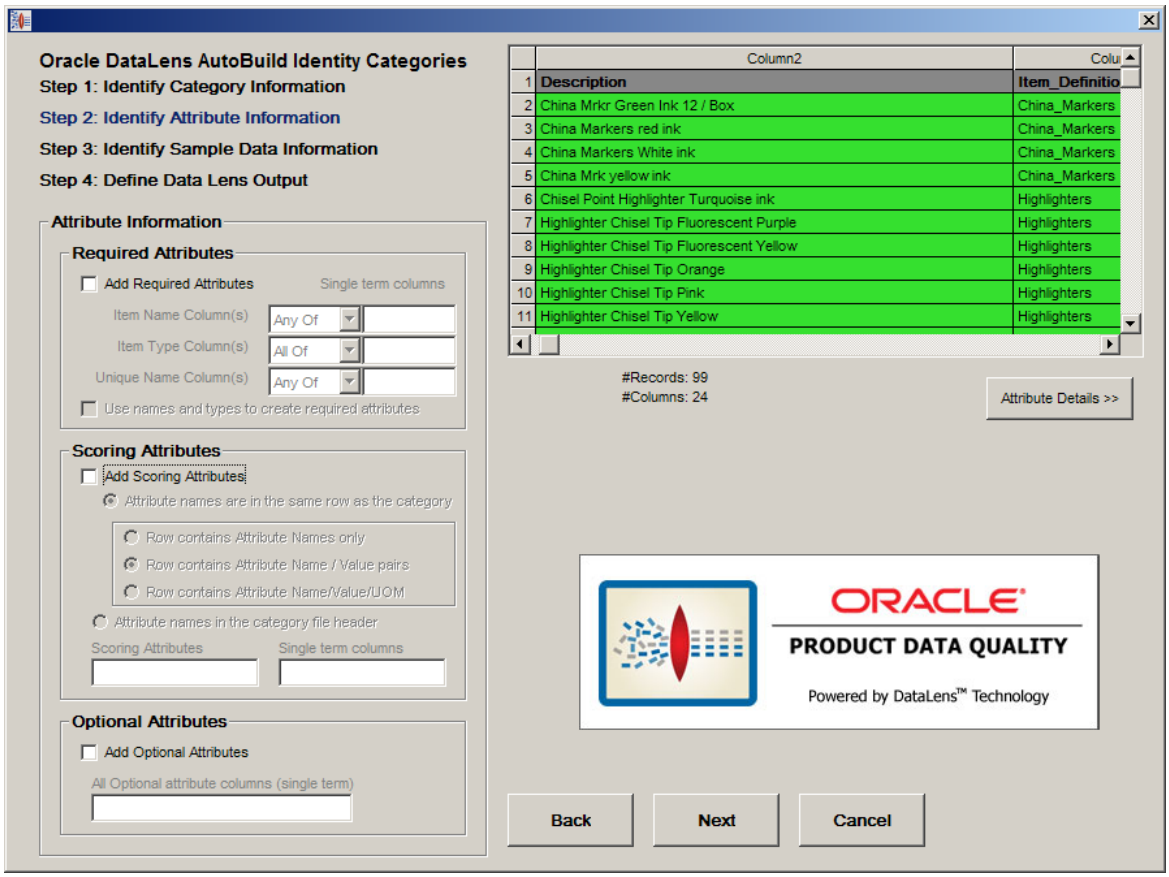
Use this textbox to control the maximum number of characters in an attribute value that AutoBuild will attempt to process. If the number of characters in an attribute value is greater than the number found in this textbox, AutoBuild ignores the attribute value. This option must be set to a number between 1 and 255 for MS Excel 2003 and between 1 and 2048 for MS Excel 2007. The default value is set to 255.

Map Characters to ASCII Check box

This option will map the data source diacritical characters (for example, ` and ~) to their ASCII equivalents. It is useful when the input data is a romance language containing diacritical characters and the standard English equivalent is needed.

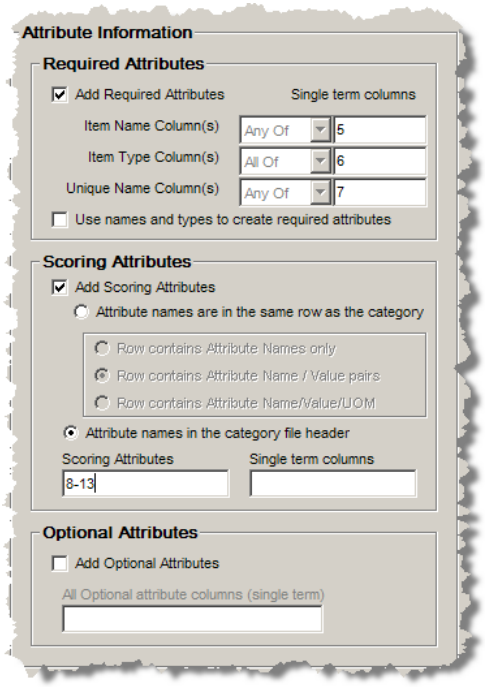
Step 2: Identify Attribute Information

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. Typically, the identifying attributes are item names, types, and unique brand names. The controls in this section are used as follows:

Add Required Attributes

This checkbox activates the controls in this box so that you can select which columns of the source data are the required attributes.

Item Name, Item Type, and Unique Name Columns Textboxes

Use these textboxes 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 any combination of 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 data table is updated by colorizing 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. 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).

Using 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 information 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 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 **Add Required Attributes** is selected	**Required** **All Of:** Item_Name: Item_Type:
Only item name columns are defined	**Required** **All Of:** **Any Of:** Name_Attr1: Name_Attr2: Attributes that are unique to my sibling categories if any …

Only item type columns are defined

Required

All Of:

All Of:

Type_Attr1:

Type_Attr2:

Attributes that are unique to this category alone if any

…

One or more columns of both item names and item types are define

Required

All Of:

Any Of:

Name_Attr1:

Name_Attr2:

Attributes that are unique to my sibling categories if any

…

All Of:

Type_Attr1:

Type_Attr2:

Attributes that are unique to this category alone if any

…

One or more brand name columns are defined	**Required** **Any Of:** **All Of:** Item_Name: Item_Type: **Any Of:** Brand: [a_brand]\[some_brand_name1] [a_brand]\[some_brand_name2] …
One or more brand name columns and item name and item type columns are defined	**Required** **Any Of:** **All Of:** **Any Of:** Name_Attr1: Name_Attr2: Attributes that are unique to my sibling categories if any … **All Of:** Type_Attr1: Type_Attr2: Attributes that are unique to this category alone if any … **Any Of:** Brand: [a_brand]\[some_brand_name1] [a_brand]\[some_brand_name2] …

Logic generated this option is *not* selected

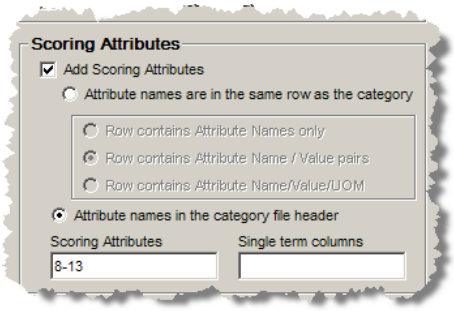
The following examples describe the required logic attributes that are generated by AutoBuild when the **Use names and types to create required attributes** check box is *not* selected:

Example	Required logic generated for each item definition
If no item name or item type columns are defined	**Required** **All Of:** Item_Name: Item_Type:
If only item name columns are defined	**Required** **All Of:** Item_Name: [i_column1] [i_column2] … Item_Type:
If only item type columns are defined	**Required** **All Of:** Item_Name: Item_Type: [t_column1] [t_column2] …

If one or more columns of both item names and item types are defined	**Required** **All Of:** Item_Name: [i_column1] [i_column2] … Item_Type: [t_column1] [t_column2] …
If one or more brand name columns are defined	**Required** **Any Of** **All Of** (*item name and item type logic as described above)* Brand: [a_brand]\[some_brand_name1] [a_brand]\[some_brand_name2] …

Scoring Attributes Box

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

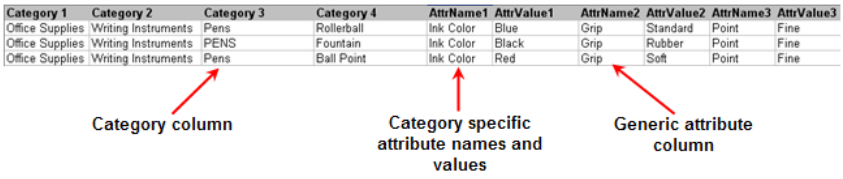


Add Scoring Attributes Check Box

Select this option to add scoring (form, fit, and function) attributes to your new data lens Item Definitions.

Attribute names are in the same row as the category Radio Button

Select this option when you attribute information is located in the same row as the associated categories. In this case, one or more columns that contain the different category levels and in the same spreadsheet, 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," etc. The attribute value information found in each row is specific to the category defined by that row as in the following example:

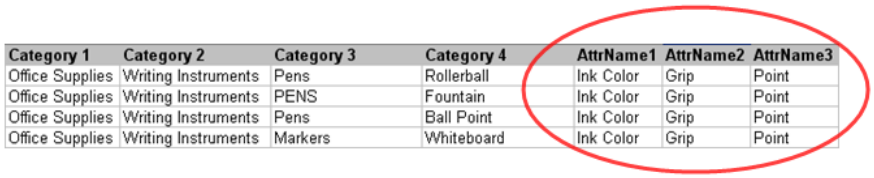


Category 1	Category 2	Category 3	Category 4	AttrName1	AttrValue1	AttrName2	AttrValue2	AttrName3	AttrValue3
Office Supplies	Writing Instruments	Pens	Rollerball	Ink Color	Blue	Grip	Standard	Point	Fine
Office Supplies	Writing Instruments	PENS	Fountain	Ink Color	Black	Grip	Rubber	Point	Fine
Office Supplies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

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

Row contains Attribute names Radio Button

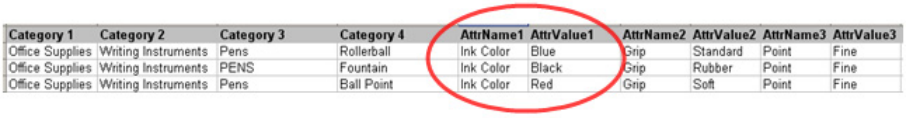
Select this option to instruct AutoBuild that only attribute names are available in the structured data source as in the following example:



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 Radio Button

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:



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

Row contains attribute name / value / UOM columns Radio Button

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

Attribute names in the category file header Radio Button

Select this option when the attribute names reside in the header row (typically row 1) of the structured data source. This spreadsheet structure is often used when the spreadsheet 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 spreadsheet 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 data 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
18	WI84	Pen Rfl Med Bk 10/Bx	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Rfl				
19	WI93	Refill Rolleball Ink Black	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rolleball	Ink Black		
20	WI94	Refill Rollball Ink Blue	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rollball	Ink Blue		
21	WI103	Roller Ball Pen Refill Medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller Ball	Blue Ink		
22										
23	Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color	Size	Ink_Type
24	WI6	Ballpoint Pen Medium Point Black Barrel Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Black Ink		
25	WI17	Comfort Grip Retractable Ballpoint Pen medium Point Blue ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Blue ink		
26	WI18	Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Red Ink		
27	WI27	Retractable Ballpoint Pen Non-Refillable Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink		
28	WI28	Erasable Ink Ballpoint Pen Medi Point Blue Barrel/Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Blue Ink		Erasable Ink
29	WI29	Felt Tip Pen Extra-Fine Point Translucent blue Barrel Purple In	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Felt Tip	Purple Ink		
30	WI30	Felt Tip Pen Extra-Fine Point Translucent black Barrel Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Felt Tip	Red Ink		
31	WI79	Pen Extra-Fine Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen		Blue Ink		
32	WI80	Pen Extra-Fine Point Green Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen		Green Ink		
33	WI88	Porous Pen Extra-Fine Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Porous	Red Ink		
34	WI91	Porous Pen Extra-Fine Tip White Barrel/Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Porous	Black Ink		
35	WI95	Ballpoint Pen Retractable Fine Point Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Black Ink		
36	WI96	Retractable Ballpoint Pen Fine Point Clear Barrel/Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink		
37	WI97	Retractable Ballpoint Pen Medium Point Black Barrel/ black in	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	black ink		
38	WI98	Retractable Ballpoint Pen Refillable Medium Point Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink		
39	WI99	Ballpoint Pen Rubber Grip Fine Point Retractable Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink		
40	WI100	Retractable Ballpoint Pen Rubber Grip Fine Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Red Ink		
41	WI101	Roller Ball Pen Gel ink Refillable Medium Blue	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Roller Ball			Gel ink
42	WI102	Roller Ball Pen Refillable Medium Red gel ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Roller Ball			gel ink
43	WI107	Pen , Ballpoint , Blue Ink, Medium Point	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Blue Ink		
44										
45	Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color	For_Sale_Packaging	
46	WI12	China Mrkr Green Ink 12 / Box	Writing_Instruments	Markers	China_Markers	China	Mrkr	Green Ink	12	
47	WI13	China Markers red ink	Writing_Instruments	Markers	China_Markers	China	Makers	red ink		
48	WI14	China Markers White ink	Writing_Instruments	Markers	China_Markers	China	Makers	White ink		
49	WI15	China Mrk yellow ink	Writing_Instruments	Markers	China_Markers	China	Mrk	yellow ink		
50										
51	Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color	For_Sale_Packaging	
52	WI16	Chisel Point Highlighter Turquoise ink	Writing_Instruments	Markers	Highlighters	Highlighter	Chisel Point	Turquoise ink		
53	WI31	Highlighter Chisel Tip Fluorescent Purple	Writing_Instruments	Markers	Highlighters	Highlighter	Chisel Tip	Fluorescent Purple		

Scoring Attributes Textbox

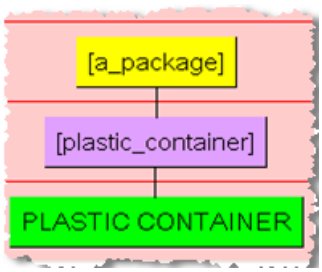
Use this textbox to specify the columns that are to be used to define the scoring attributes. Your attributes are added in lowest column number to highest column number order. To specify the list of attributes, you may use any combination of numbers, dashes, and commas to specify a single column, a range of columns, or a combination of range(s) and single column(s). 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

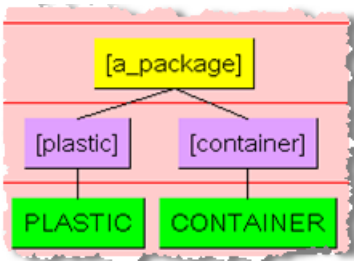
Use this option to force the creation of a single term under a phrase. When AutoBuild creates the data lens phrases and terms for identified attributes values, and the attribute values are identified as single term columns, then AutoBuild creates a single term to recognize the attribute value no matter how many words (tokens) are in the attribute value.

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 phrase and terms are built when the attribute value is two words 'PLASTIC' and 'CONTAINER' for either single term or multiple terms:



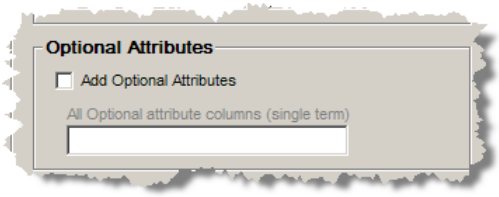
Single Term



Multiple Terms

Optional Attributes Box

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



Add Optional Attributes Check Box

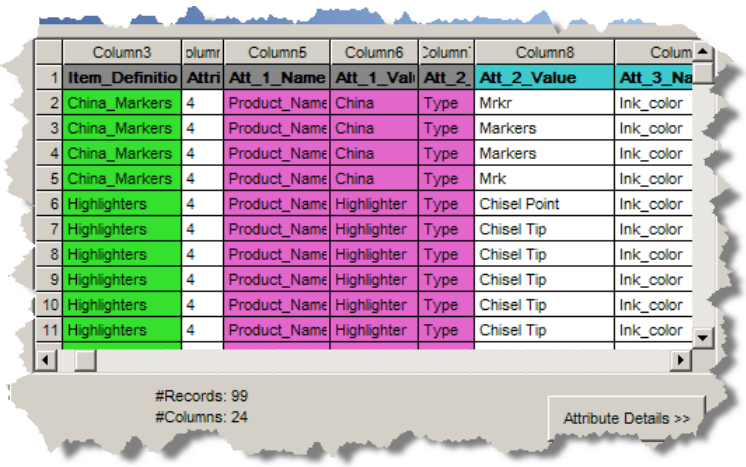
Select this option to add optional attributes and activate the **Add Optional Attributes columns (single term)** check box.

Add Optional Attributes columns (single term) Check Box

Select this option to add data lens standard optional attributes of '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. These attributes provide you with the optional attributes to add in your global attribute values once your new lens is generated. If you are using the 'DLS_Import_Template' Smart Glossary as the basis for your data lens (the default), then both the 'Units_per_Package' and 'URL' attributes have pre-defined phrases.

Category/Attribute Data Table

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



	Column3	olumr	Column5	Column6	Column7	Column8	Colum
1	Item_Definitio	Attri	Att_1_Name	Att_1_Vali	Att_2_	Att_2_Value	Att_3_Na
2	China_Markers	4	Product_Name	China	Type	Mrkr	Ink_color
3	China_Markers	4	Product_Name	China	Type	Markers	Ink_color
4	China_Markers	4	Product_Name	China	Type	Markers	Ink_color
5	China_Markers	4	Product_Name	China	Type	Mrk	Ink_color
6	Highlighters	4	Product_Name	Highlighter	Type	Chisel Point	Ink_color
7	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip	Ink_color
8	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip	Ink_color
9	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip	Ink_color
10	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip	Ink_color
11	Highlighters	4	Product_Name	Highlighter	Type	Chisel Tip	Ink_color

Data Table

The data table displays the first 30 rows and 80 columns of the structured Excel spreadsheet data 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:

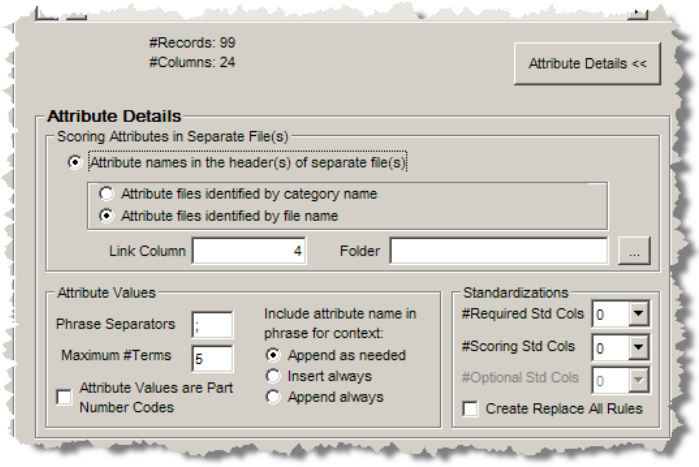
Green	category
Violet	item name, item type, and brand names
Blue	scoring attributes

Data Source Metrics

The total number of records and columns found in the structured data source appears immediately below the data table.

Attribute Details

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



Scoring Attributes in Separate File(s) Box

The following section describes each option found in the **Scoring Attributes in Separate File(s)** box.

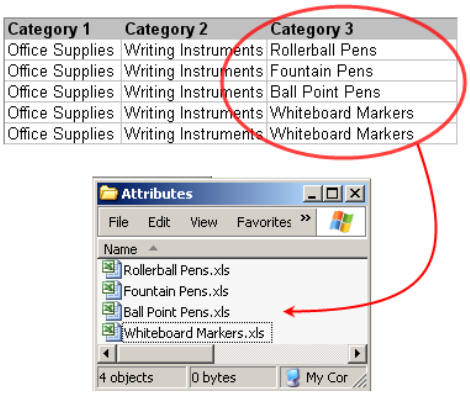
Attribute names in the header(s) of separate file(s) Radio Button

Select this option when your attribute information is located in multiple separate workbooks where each separate workbook provides the attribute information for one category.

For each workbook, it is assumed that the first worksheet contains the attribute information, and that the attribute names are present in the column headers. The separate workbooks can be linked to by either the lowest level category name, or a file name associated with each category using the remaining controls in this section.

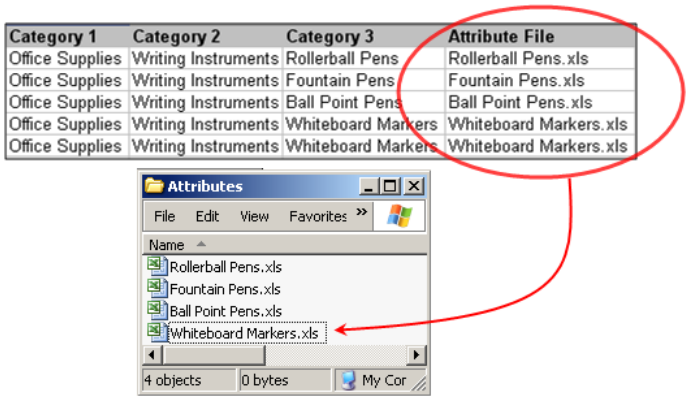
Attribute files identified by category name Radio Button

Select this option when separate attribute files are named corresponding to your lowest-level category names as in the following example:



Attribute files identified by the file name Radio Button

Select this option when your separate attribute files are identified by a file name column in the spreadsheet containing the category information as in the following example:



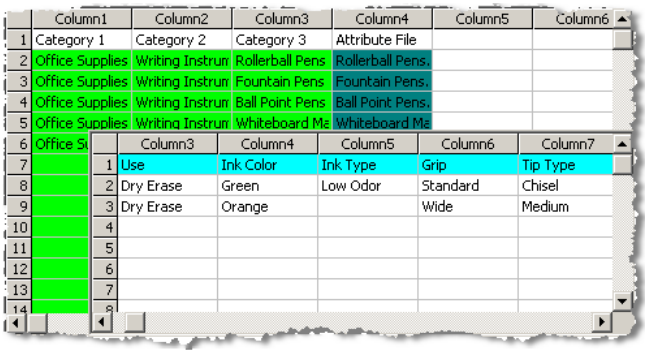
Selection of this option activates the **Link Column** textbox.

Link Column Textbox

Enter the number of the column that contains the attribute data that you want linked (or merged) together from the separate attribute files.

Folder Textbox

Use to locate and select, or enter the attribute directory (folder) path name that you want to use. When the **Folder** textbox is updated with the directory path, the **Category/Attribute Data** table is updated by opening one of the attribute files and colorizing your attribute columns in blue as in the following example:



Tip: The scroll bars for each set of data (spreadsheet) operate independently so that you can view all of the data in each file.

Attribute Values Box

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

Phrase Separators Textbox

Use this textbox to list all of the phrase characters that AutoBuild should apply to the attribute values. 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 the defined character separators. By default, the phrase separate character list is set to a semi-colon (;).

Maximum #Terms Textbox

Use this textbox 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 terms in an attribute value is greater than the number found in this textbox, AutoBuild ignores the attribute value. This option must be set to a number between 1 and 50. The default value is set to five.

Include attribute name in phrase for context Radio Buttons

Select one of the three radio buttons 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 associated phrase structure so that attribute values can be uniquely recognized without causing ambiguities between phrases.

Append as needed

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

Append always

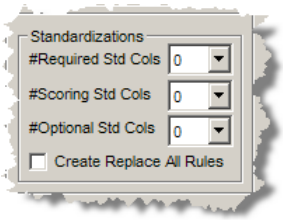
Select this option to append the attribute name as part of the phrase structure for all attribute values.

Insert always

Select this option to insert the attribute name as part of the phrase structure for all attribute values.

Standardizations Box

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



#Required/Scoring/Optional Std Cols Listboxes

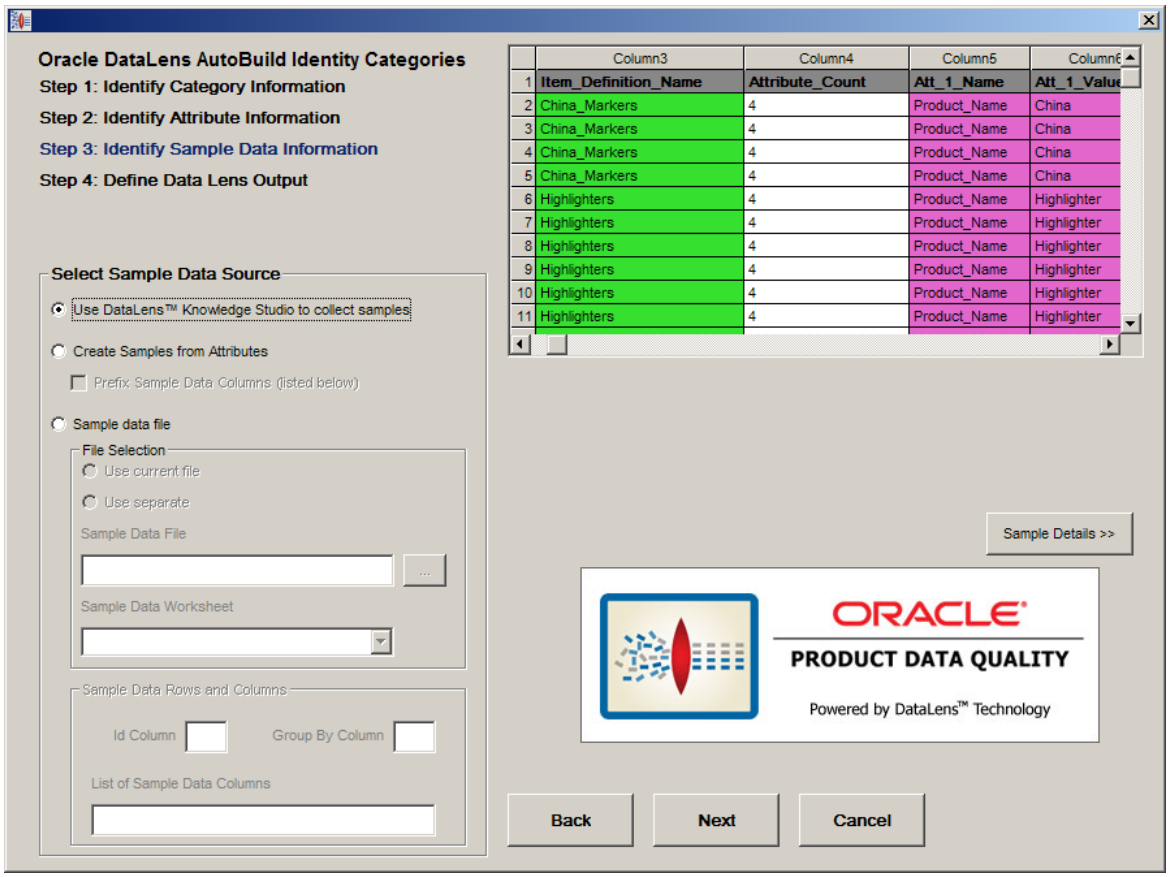
These options allow you to select of the number of standardization columns that appear to the right of each associated attribute column that contains the corresponding standardized attribute values. For each identified standardization column, the column header is used to create a unique standardization 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 Knowledge Studio to collect samples radio button

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 radio button

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 data that you need to process via the Oracle DataLens Server. When this radio button is selected, the **Sample Data Rows and Columns** box and the following check box are 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 data. When this checkbox is checked, use must use the **List of Sample Data Columns** textbox to define the columns to be prefixed with the attribute values.

Sample data file Radio Button

Select this option when you want to collect your sample data from a file. When this radio button is selected, the **File Selection** box becomes active so that you can identify the file containing the sample data. As you make your file selections, the example sample data file display is updated to provide feedback on the options you have selected (you must first select a file, please see the description below). The file you select can be the file that you are currently using as the source of example-structured data.

File Selection Box

This box contains options for you to use in selecting the sample data.

Use current file Radio Button

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

Use separate Radio Button

Use this option to activate the **Sample Data File** textbox to enter the full file path to the currently selected Excel workbook.

Sample data file Textbox

This textbox 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 Drop-down List

Use this drop-down 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 the details of the rows and columns used when creating the sample data.

Id Column Textbox

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. In this case, AutoBuild must know how to group the supplied attribute values into valid item descriptions. In all other cases, the id column is optional. This value must be a single column number.

Group By Column Textbox

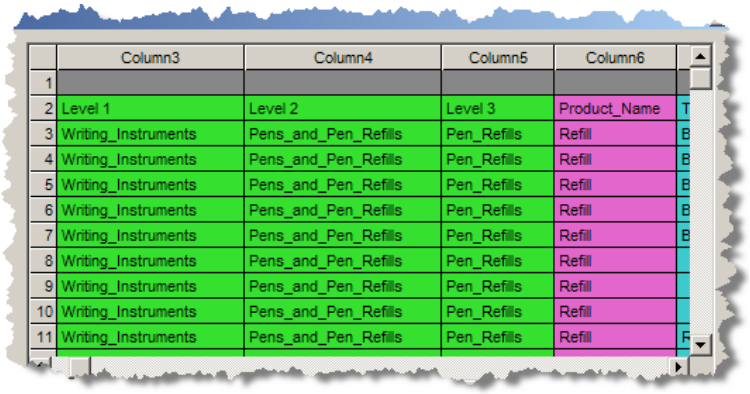
Use this option to identify a column to be used to create sample files that are grouped by the unique values found in the specified column. By default, AutoBuild creates the data lens standard 100 record, randomly selected sample files from your identified source file. When this option is specified, 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 Textbox

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). You confirm your column list entry by selecting the **Tab** key or selecting another user interface option.

Category/Attribute Data Table

The following describes the **Category/Attribute Data** 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

Data Table

The data table displays the first 30 rows and 80 columns of the structured Excel spreadsheet data 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:

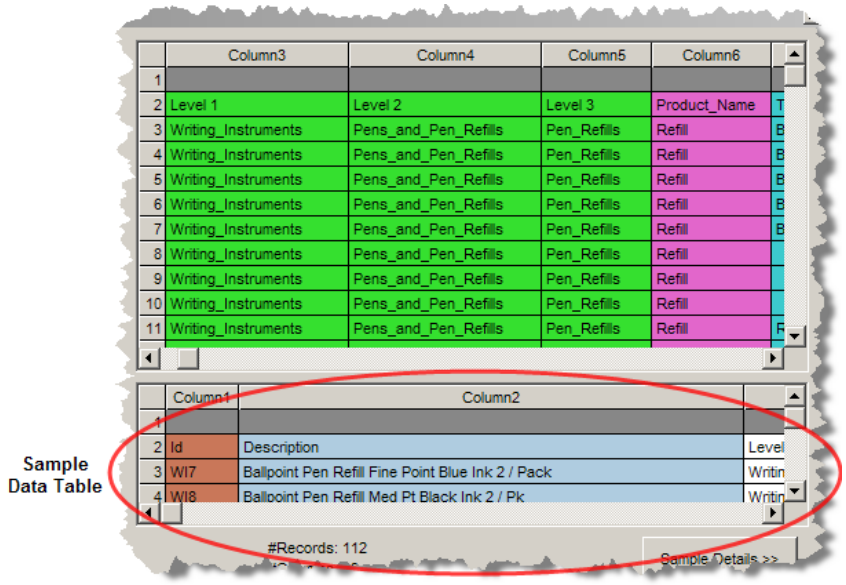
Green	category
Violet	item name, item type, and brand names
Blue	scoring attributes

Data Source Metrics

The total number of records and columns found in the structured data source appears immediately below the data table.

Sample Data Table

The following describes the Sample Data Table that is displayed when the **Sample data file** option is selected. This table operates like the **Category/Attribute Data** table.



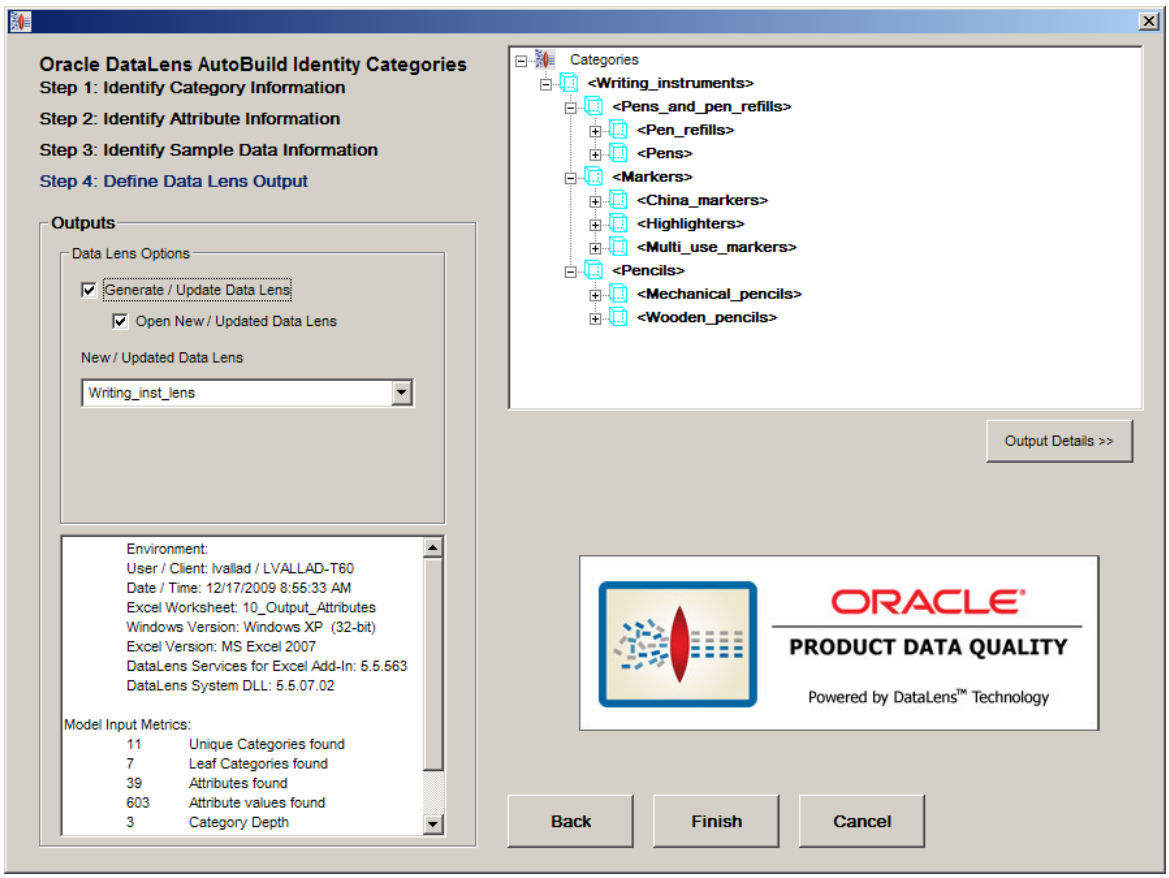
Sample Details

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

By default, these options provide detailed information only and are not active for modification. The two **Sample Details** check boxes describe the type of sample files to be created as defined in this step of the AutoBuild wizard.

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 Section

The controls in this section are used as follows:

Generate / Update Data Lens Check Box

Select this check box to create or update a data lens when the **Finish** button is selected. It activates the following check box:

Open New / Update Data Lens Check Box

Select this check box to open your new or updated data lens in the Knowledge Studio once it is created.

New / Updated Data Lens Textbox

Use this textbox to name your new data lens or select an existing data lens to use as the basis for a name. Your name must be unique. If you use an existing data lens to begin to create a name, you must change it to be unique. AutoBuild does not allow you to over-write an existing lens.

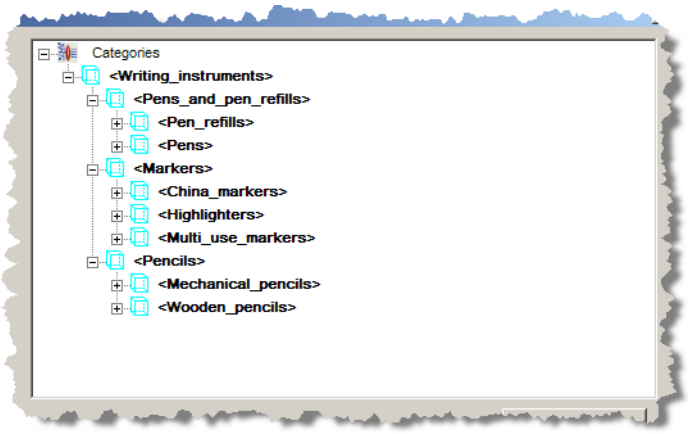
Report Textbox

This read-only textbox 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 textbox. For example, duplicate category names.

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 data 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 data correctly.

AutoBuild Item Definition Explorer

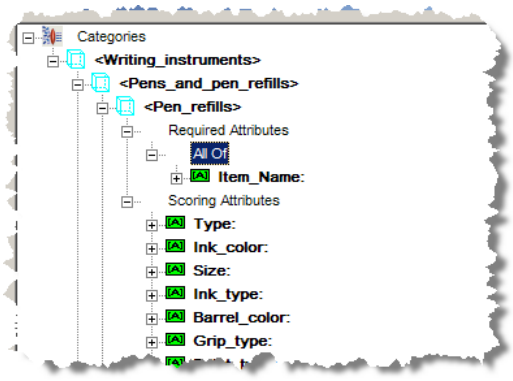
The Item Definition Explorer provides a way to review the lens model created by AutoBuild. The generated model reflects the structured data provided and your option selections.



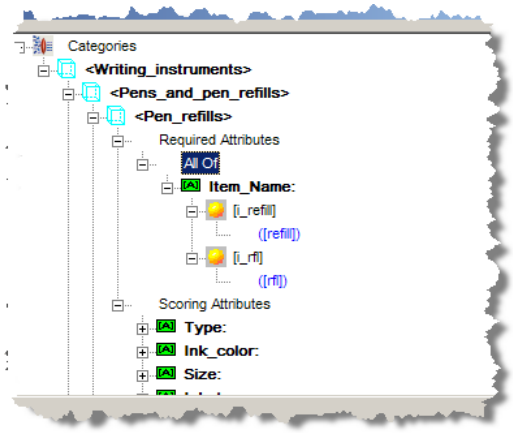
The **AutoBuild Item Definition Explorer** provides a way to review the result of your selections prior to generating the data lens. If you find inconsistencies in the model, you can return to previous AutoBuild wizard steps and make the necessary adjustments or corrections to ensure that your data lens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

Under each item definition node, you can view the Required, Scoring, and Optional attribute structures by opening each correspondingly labeled node. Each attribute has an associated green [A] icon.

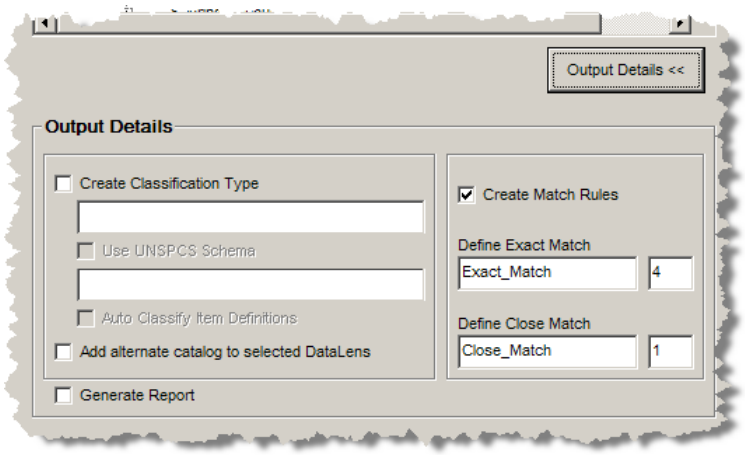


Finally, you can drill down deeper into each phrase and view the phrase productions as a set of one or more grouped term names. For each attribute, you can view the assigned phrase(s) if any. Each phrase has an associated lens phrase yellow ball icon as in the following example.



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

Name Textbox

Use this textbox to name your new classification type.

Use UNSPSC Schema checkbox

This checkbox is active when the **Single category column contains UNSPSC category codes** option is set in Step 1.

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 checkbox is selected, the read-only textbox below it displays the name of the selected UNSPSC reference schema file. For more information about configuring a UNSPSC schema, see Category Details.

Auto Classify Item Definitions Check Box

Use this textbox to ensure that AutoBuild generates a set of default classification rules by associating the item definitions to the categories in the new classification type.

Generate Report Check Box

Use this checkbox to request an overall AutoBuild metrics report that can be saved as a text file. This report contains all of the information found in the model metrics report described in Step 3: Identify Sample Data Information, as well as details on the generated lens, and a complete list of the option selections used during the AutoBuild session. For more information, see AutoBuild Report.

Create Match Rules Check Box

Select this check box to create default matching rules. A standardization type called 'Match_Attributes' is automatically created. A default attribute order for each item definition, containing only scoring attributes in the order specified in the spreadsheet columns, is automatically created. The match types 'Exact_Match' and 'Close_Match' are created based on the standardization type 'Match_Attributes'. When this checkbox is selected, the **Define Exact Match** and **Define Close Match** controls are activated.

Define Exact Match Controls

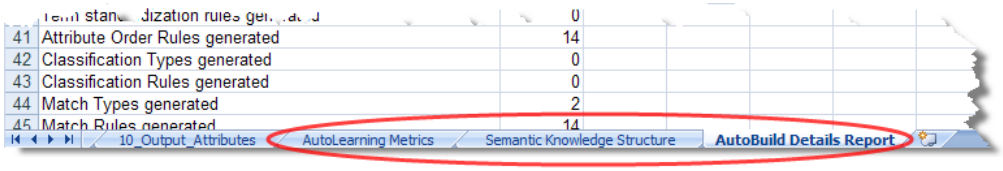
Defines the number of required attributes for the Exact Match rule per item definition.

Define Close Match Controls

Defines the number of required attributes for the Close Match rule per item definition.

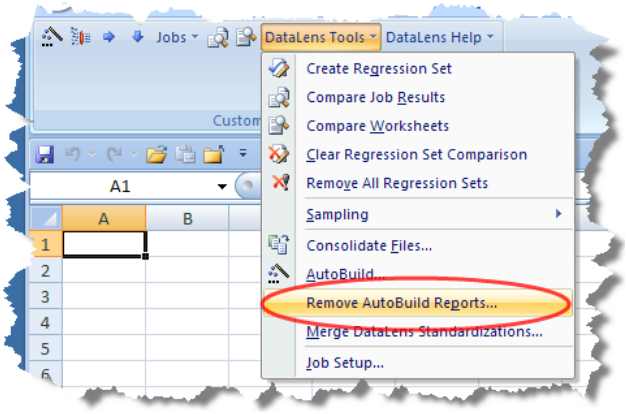
AutoBuild Reports

The AutoBuild application can automatically generate three separate reports that provide detailed metric data and the semantic knowledge structure of your data lens. These reports are 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. The reports are added to your Excel workbook as new sheets so that you can easily review, print, and maintain the information.



AutoBuild Reports

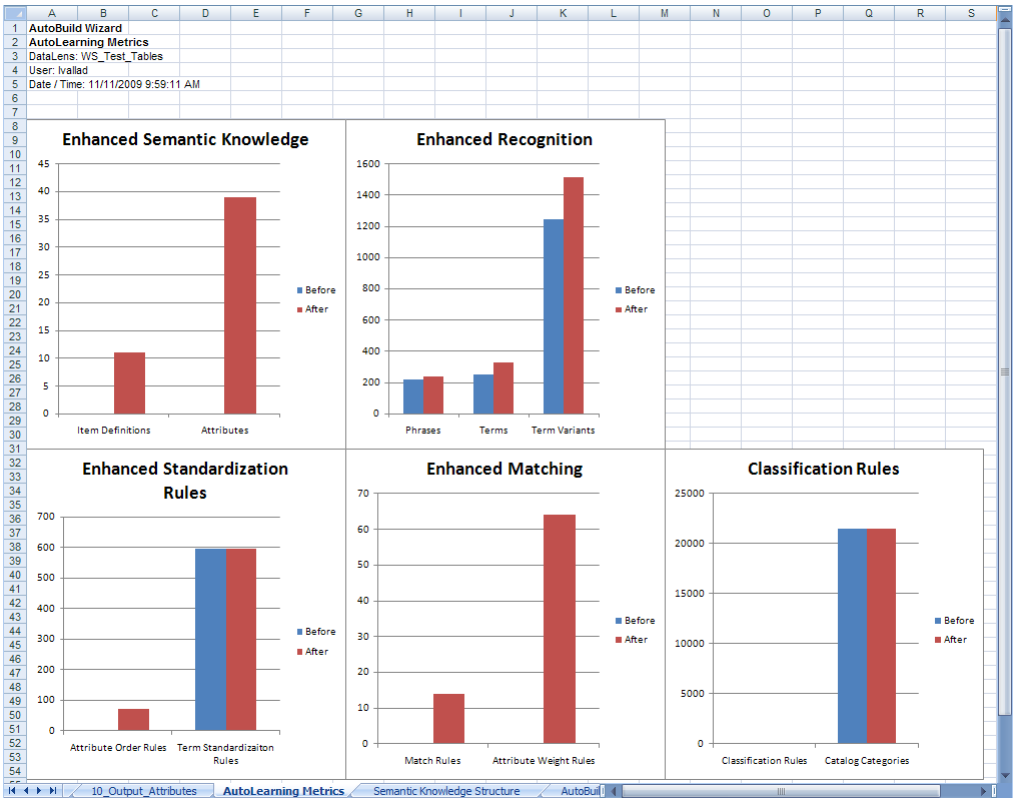
Reports are updated in your Excel workbook each time you run AutoBuild to generate a data lens. You can delete these reports using AutoBuild. Click the **DataLens Tools** button and select **Remove AutoBuild Reports...** as follows:



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

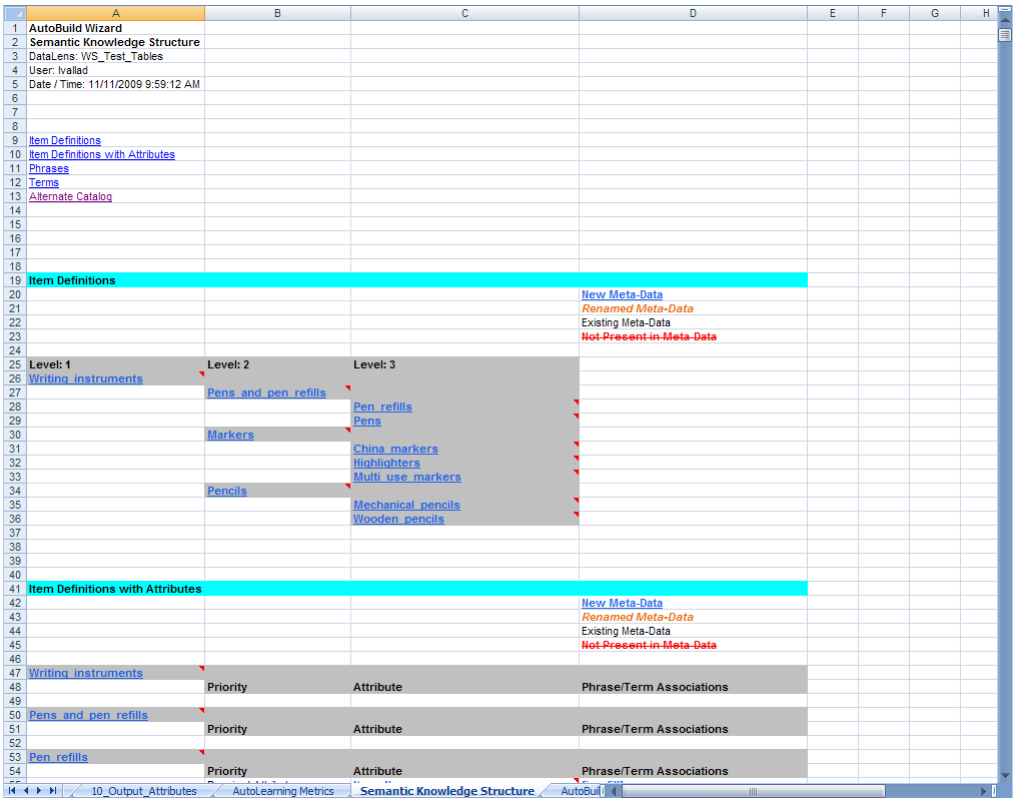
AutoLearning Metrics Report

This report summarizes all of the objects created in your data lens by AutoBuild. The report includes a bar graph showing the before and after states for each of the most important objects as in the following sample report:



Semantic Knowledge Structure

This report details the Item Definitions, phrases, terminology, and alternate catalog that comprise your data lens, which AutoBuild created as in the following sample report:



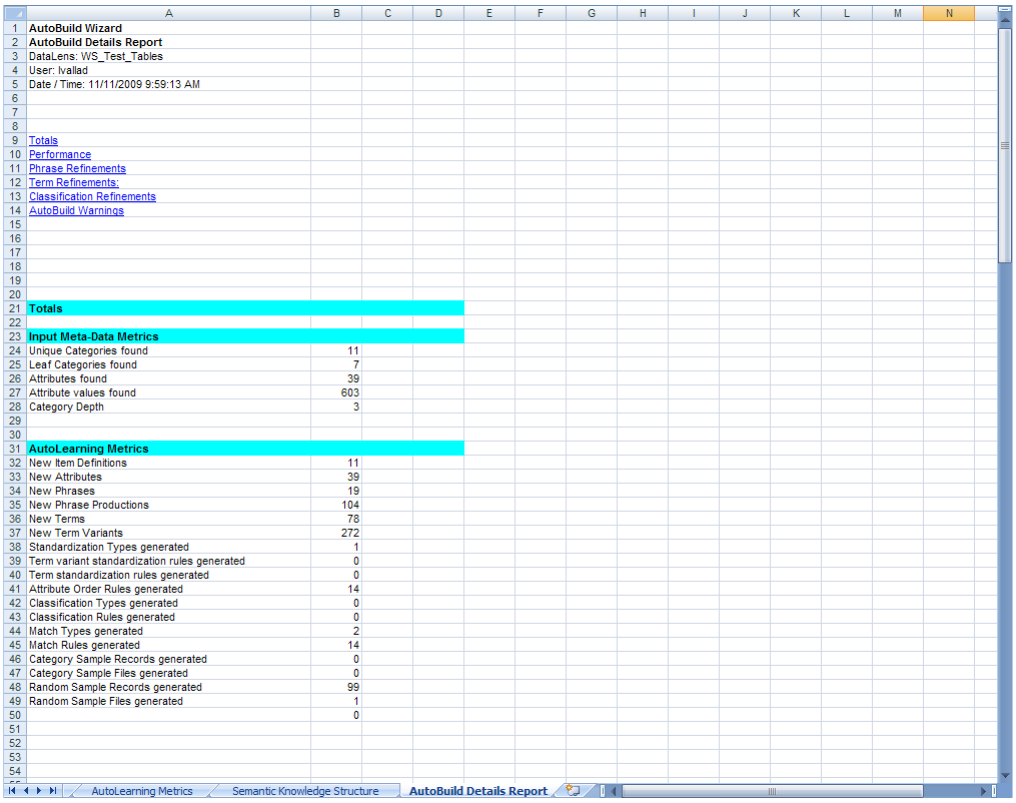
	A	B	C	D
1	AutoBuild Wizard			
2	Semantic Knowledge Structure			
3	DataLens: WS_Test_Tables			
4	User: lvallad			
5	Date / Time: 11/11/2009 9:59:12 AM			
6				
7				
8				
9	Item Definitions			
10	Item Definitions with Attributes			
11	Phrases			
12	Terms			
13	Alternate Catalog			
14				
15				
16				
17				
18				
19	**Item Definitions**			
20				New Meta-Data
21				*Renamed Meta-Data*
22				Existing Meta-Data
23				~~Not Present in Meta-Data~~
24				
25	**Level: 1**	**Level: 2**	**Level: 3**	
26	Writing_instruments			
27		Pens_and_pen_refills		
28			Pen_refills	
29			Pens	
30		Markers		
31			China_markers	
32			Highlighters	
33			Multi_use_markers	
34		Pencils		
35			Mechanical_pencils	
36			Wooden_pencils	
37				
38				
39				
40				
41	**Item Definitions with Attributes**			
42				New Meta-Data
43				*Renamed Meta-Data*
44				Existing Meta-Data
45				~~Not Present in Meta-Data~~
46				
47	Writing_instruments			
48		**Priority**	**Attribute**	**Phrase/Term Associations**
49				
50	Pens_and_pen_refills			
51		**Priority**	**Attribute**	**Phrase/Term Associations**
52				
53	Pen_refills			
54		**Priority**	**Attribute**	**Phrase/Term Associations**

10_Output_Attributes | AutoLearning Metrics | Semantic Knowledge Structure | AutoBuil

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

AutoBuild Details Report

This report details all actions taken by AutoBuild during the creation of your data lens. These actions include the time to create the data lens, phrase, term, and classification refinements, and any warnings that exist that were collected by AutoBuild as in the following sample report:



	A	B
1	**AutoBuild Wizard**	
2	**AutoBuild Details Report**	
3	DataLens: WS_Test_Tables	
4	User: lvallad	
5	Date / Time: 11/11/2009 9:59:13 AM	
9	Totals	
10	Performance	
11	Phrase Refinements	
12	Term Refinements:	
13	Classification Refinements	
14	AutoBuild Warnings	
21	**Totals**	
23	**Input Meta-Data Metrics**	
24	Unique Categories found	11
25	Leaf Categories found	7
26	Attributes found	39
27	Attribute values found	603
28	Category Depth	3
31	**AutoLearning Metrics**	
32	New Item Definitions	11
33	New Attributes	39
34	New Phrases	19
35	New Phrase Productions	104
36	New Terms	78
37	New Term Variants	272
38	Standardization Types generated	1
39	Term variant standardization rules generated	0
40	Term standardization rules generated	0
41	Attribute Order Rules generated	14
42	Classification Types generated	0
43	Classification Rules generated	0
44	Match Types generated	2
45	Match Rules generated	14
46	Category Sample Records generated	0
47	Category Sample Files generated	0
48	Random Sample Records generated	99
49	Random Sample Files generated	1
50		0

AutoLearning Metrics | Semantic Knowledge Structure | AutoBuild Details Report

You can use the links at the beginning of the report to navigate to the specific data indicated. The AutoBuild Warnings section should be reviewed so that you can adjust your data lens to improve it.

Chapter 3

The AutoBuild Process

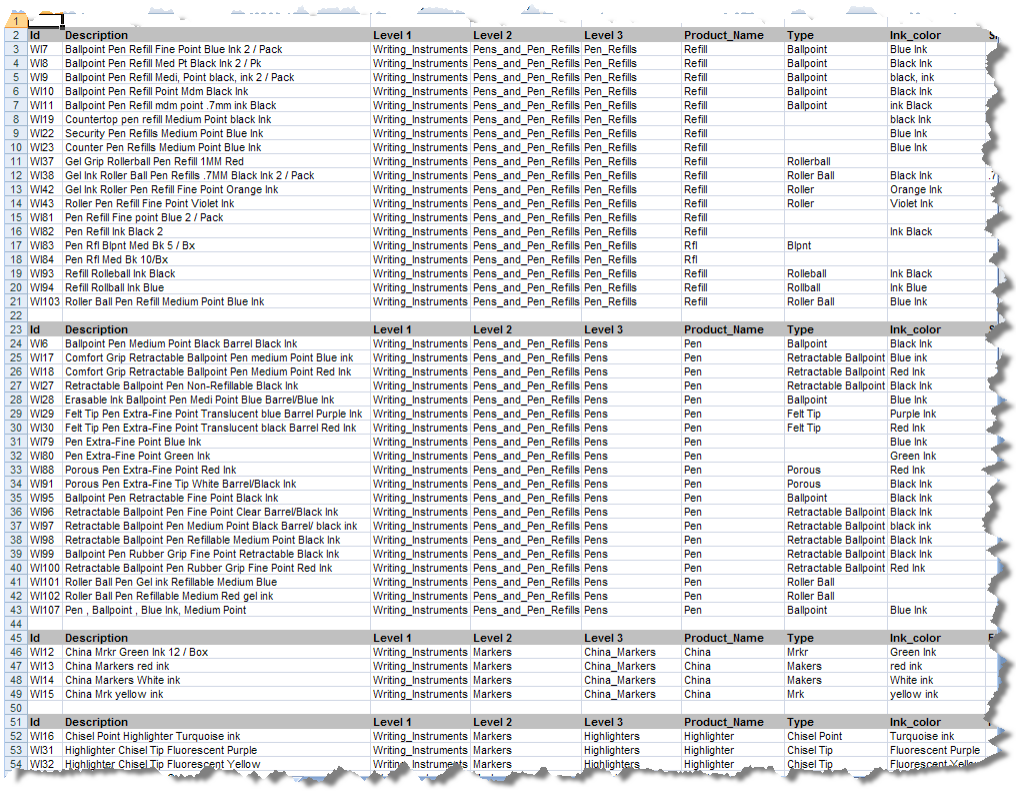
In this chapter

- Using the AutoBuild Process 48
- Finishing Your New Data Lens 60
- Data Pairs Data Lens Example 62
- Further Considerations 66

This chapter provides an end-to-end example of how to use AutoBuild to read structured data 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 the AutoBuild Process

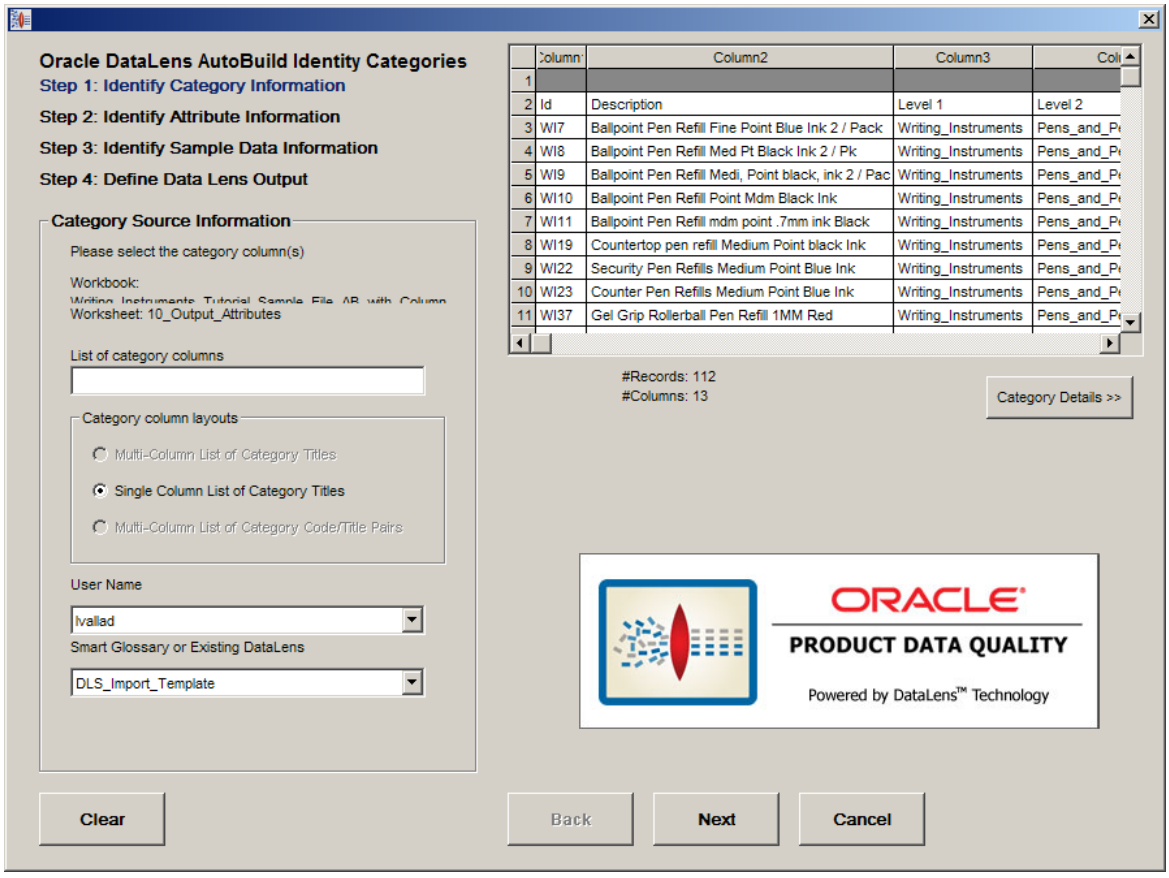
You start the AutoBuild process opening the Excel workbook file that contains your structured data. Once the workbook is open, ensure that AutoBuild is installed. For more information about installation or the AutoBuild Wizard, see The AutoBuild User Interface on page 12.



Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color
WI7	Ballpoint Pen Refill Fine Point Blue Ink 2 / Pack	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint	Blue Ink
WI8	Ballpoint Pen Refill Med Pt Black Ink 2 / Pk	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint	Black Ink
WI9	Ballpoint Pen Refill Medi, Point black, ink 2 / Pack	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint	black, ink
WI10	Ballpoint Pen Refill Point Mdm Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint	Black Ink
WI11	Ballpoint Pen Refill mdm point .7mm ink Black	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint	ink Black
WI19	Countertop pen refill Medium Point black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill		black Ink
WI22	Security Pen Refills Medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill		Blue Ink
WI23	Counter Pen Refills Medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill		Blue Ink
WI37	Gel Grip Rollerball Pen Refill 1MM Red	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rollerball	
WI38	Gel Ink Roller Ball Pen Refills .7MM Black Ink 2 / Pack	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller Ball	Black Ink
WI42	Gel Ink Roller Pen Refill Fine Point Orange Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller	Orange Ink
WI43	Roller Pen Refill Fine Point Violet Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller	Violet Ink
WI81	Pen Refill Fine point Blue 2 / Pack	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill		
WI82	Pen Refill Ink Black 2	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill		Ink Black
WI83	Pen Rfl Blpnt Med Bk 5 / Bx	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Rfl	Blpnt	
WI84	Pen Rfl Med Bk 10/Bx	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Rfl		
WI93	Refill Rolleball Ink Black	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rolleball	Ink Black
WI94	Refill Rollball Ink Blue	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Rollball	Ink Blue
WI103	Roller Ball Pen Refill Medium Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pen_Refills	Refill	Roller Ball	Blue Ink
Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color
WI6	Ballpoint Pen Medium Point Black Barrel Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Black Ink
WI17	Comfort Grip Retractable Ballpoint Pen medium Point Blue ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Blue ink
WI18	Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Red Ink
WI27	Retractable Ballpoint Pen Non-Refillable Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
WI28	Erasable Ink Ballpoint Pen Medi Point Blue Barrel/Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Blue Ink
WI29	Felt Tip Pen Extra-Fine Point Translucent blue Barrel Purple Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Felt Tip	Purple Ink
WI30	Felt Tip Pen Extra-Fine Point Translucent black Barrel Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Felt Tip	Red Ink
WI79	Pen Extra-Fine Point Blue Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen		Blue Ink
WI80	Pen Extra-Fine Point Green Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen		Green Ink
WI88	Porous Pen Extra-Fine Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Porous	Red Ink
WI91	Porous Pen Extra-Fine Tip White Barrel/Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Porous	Black Ink
WI95	Ballpoint Pen Retractable Fine Point Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Black Ink
WI96	Retractable Ballpoint Pen Fine Point Clear Barrel/Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
WI97	Retractable Ballpoint Pen Medium Point Black Barrel/ black ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	black ink
WI98	Retractable Ballpoint Pen Refillable Medium Point Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
WI99	Ballpoint Pen Rubber Grip Fine Point Retractable Black Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Black Ink
WI100	Retractable Ballpoint Pen Rubber Grip Fine Point Red Ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Retractable Ballpoint	Red Ink
WI101	Roller Ball Pen Gel ink Refillable Medium Blue	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Roller Ball	
WI102	Roller Ball Pen Refillable Medium Red gel ink	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Roller Ball	
WI107	Pen , Ballpoint , Blue Ink, Medium Point	Writing_Instruments	Pens_and_Pen_Refills	Pens	Pen	Ballpoint	Blue Ink
Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color
WI12	China Mrkr Green Ink 12 / Box	Writing_Instruments	Markers	China_Markers	China	Mrkr	Green Ink
WI13	China Markers red ink	Writing_Instruments	Markers	China_Markers	China	Makers	red ink
WI14	China Markers White ink	Writing_Instruments	Markers	China_Markers	China	Makers	White ink
WI15	China Mrk yellow ink	Writing_Instruments	Markers	China_Markers	China	Mrk	yellow ink
Id	Description	Level 1	Level 2	Level 3	Product_Name	Type	Ink_color
WI16	Chisel Point Highlighter Turquoise ink	Writing_Instruments	Markers	Highlighters	Highlighter	Chisel Point	Turquoise ink
WI31	Highlighter Chisel Tip Fluorescent Purple	Writing_Instruments	Markers	Highlighters	Highlighter	Chisel Tip	Fluorescent Purple
WI32	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.

Open the **DataLens Tools** menu, and then click the **AutoBuild** button.



The first time the AutoBuild wizard is opened from a spreadsheet, 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**.

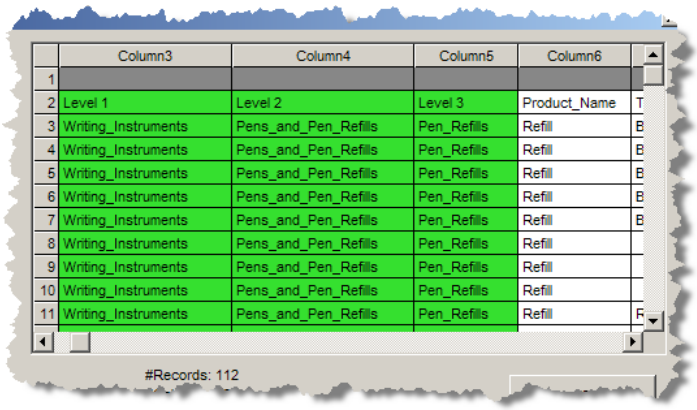
Identifying the Category Information

You must identify which of the columns of your structured data 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 spreadsheet 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** textbox. 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 colorized 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	F

#Records: 112

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 express 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.

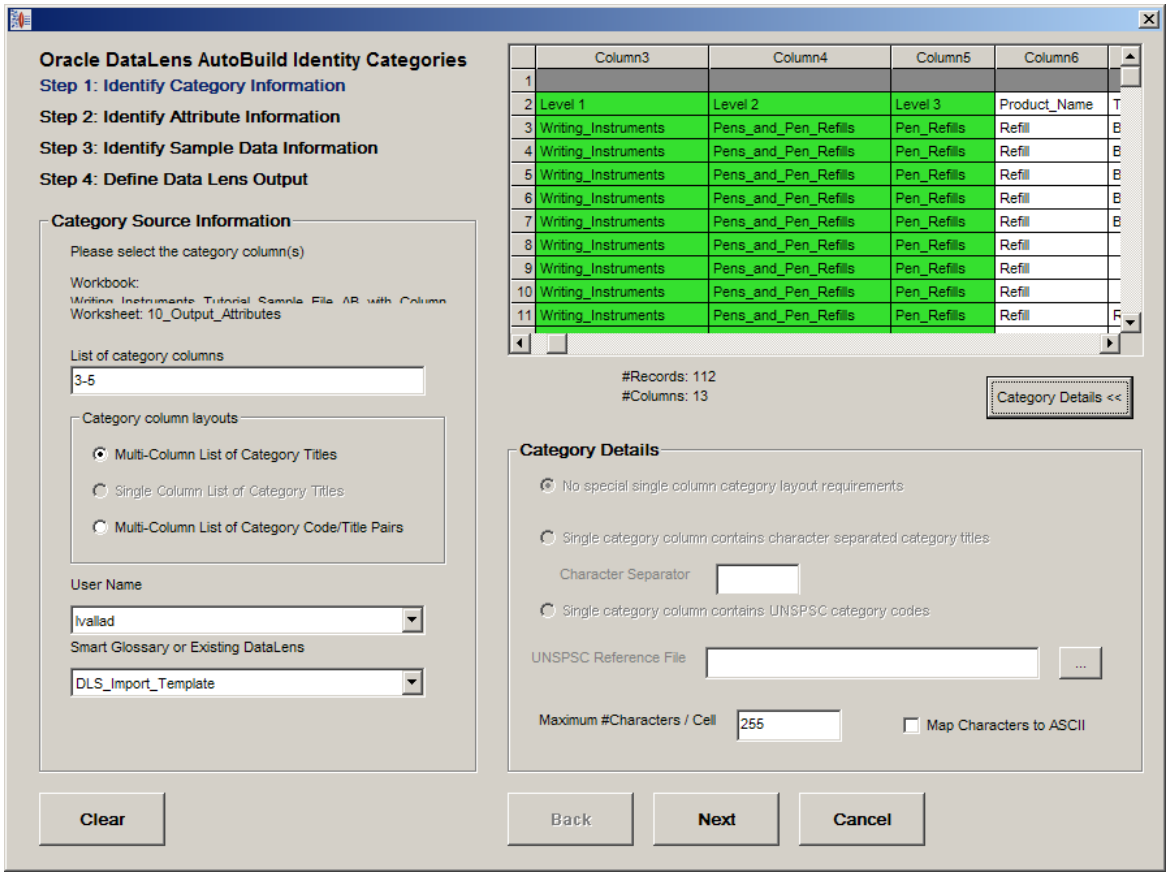
The third category layout is **Multi-Column List of Category Code/Title Pairs**. This layout is similar to the multi-column list previously described with the addition that each category is expressed as pair of columns. The first column contains a category code and the second column contains the category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

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

Selecting the User

Use the **User Name** drop-down list to select the user that you want to associate with the new data lens and that will be logged if the Knowledge Studio is opened upon completion of the wizard.

The example selections are as follows:



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

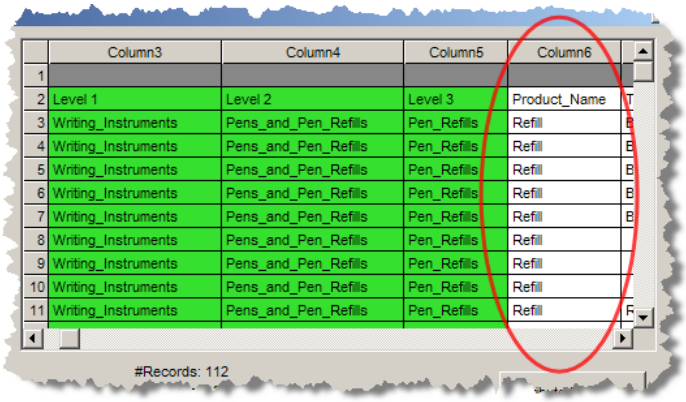
Identifying the Attribute Information

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.

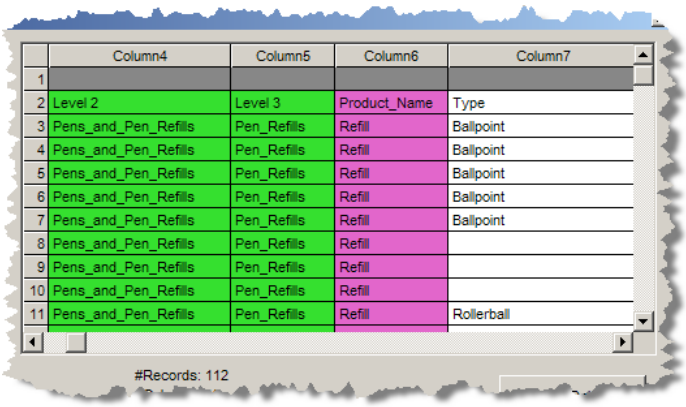
In this example, column 6 contains the product name as in the following:



	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

#Records: 112

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:



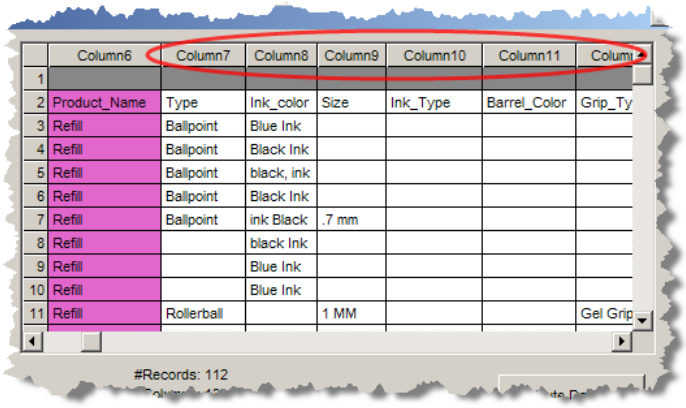
	Column4	Column5	Column6	Column7
1				
2	Level 2	Level 3	Product_Name	Type
3	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
4	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
5	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
6	Pens_and_Pen_Refills	Pen_Refills	Refill	Ballpoint
7	Pens_and_Pen_Refills	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

#Records: 112

Sometimes structured data 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	Colum
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

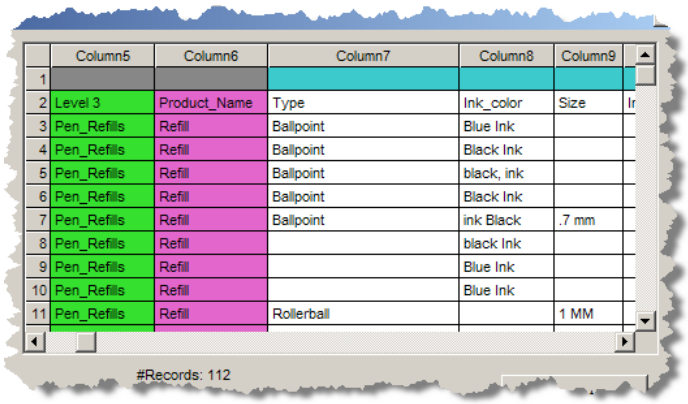
#Records: 112

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** textbox. You can supply a comma-separated list of column numbers, as well as column ranges.

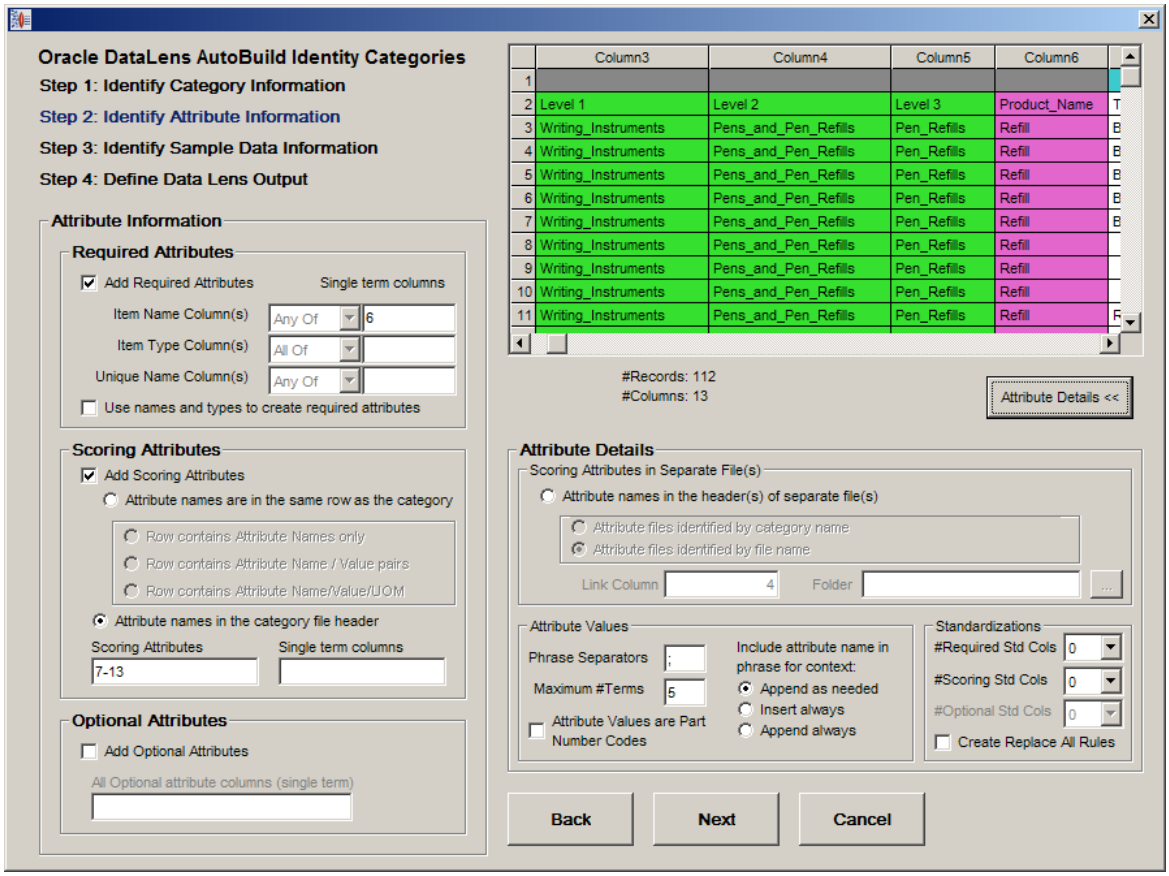
Once the list of attribute columns is accepted, the selected scoring attributes are colorized blue as previously described.



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

#Records: 112

The example selections are as follows:



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.

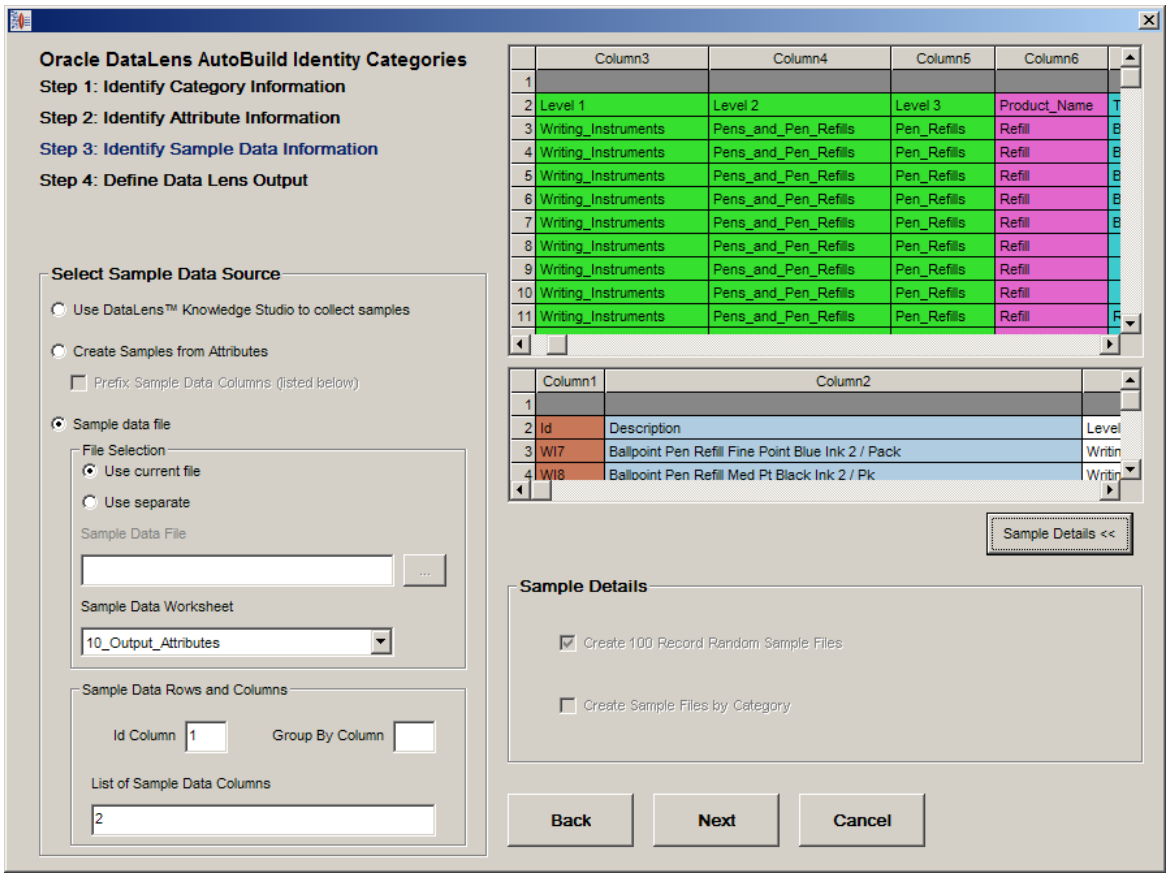
This example creates the samples from attributes so the appropriate option is selected, as well as the **Prefix Sample Data Columns (listed below)** checkbox.

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 data 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** textbox.

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

The example selections are as follows:

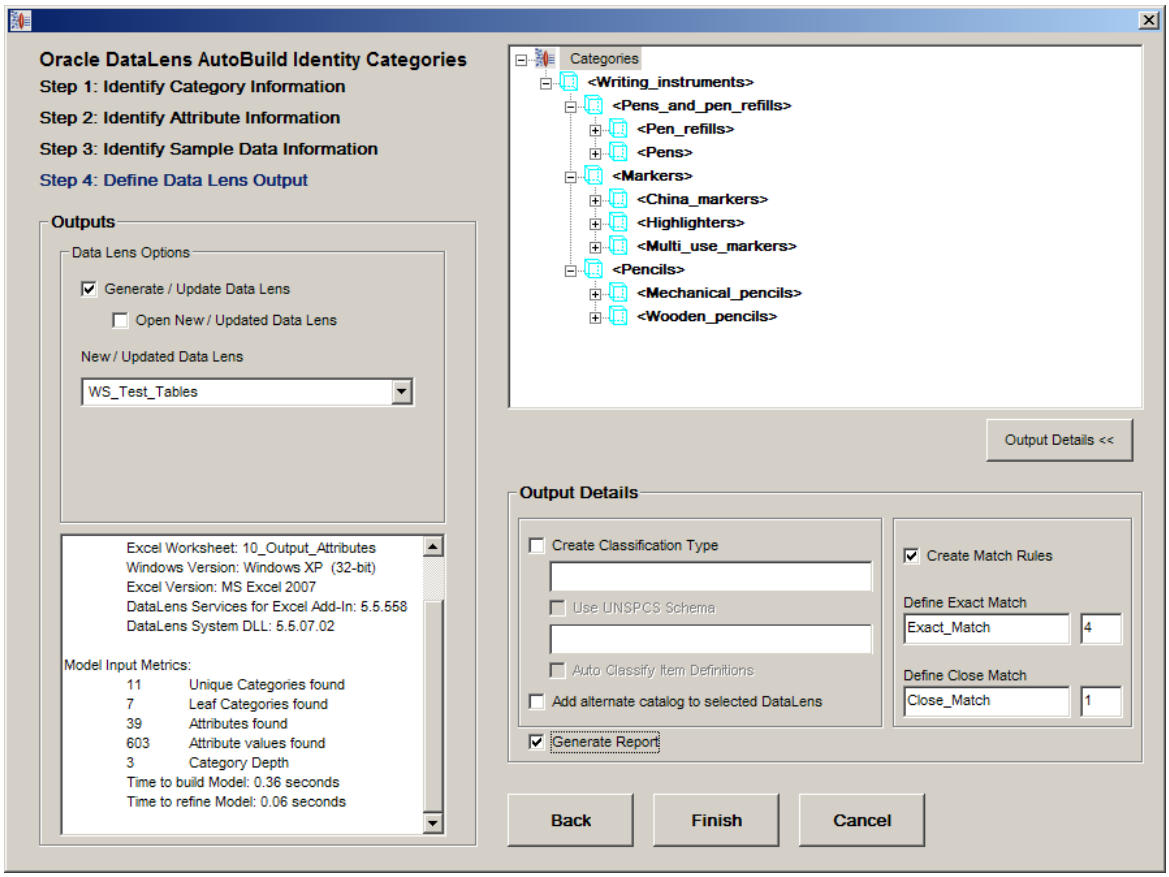


Click **Next** to advance to the final step.

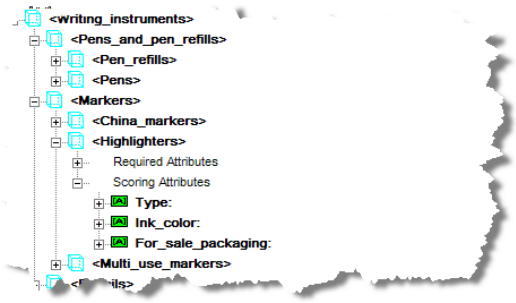
Before advancing, AutoBuild reads the category and attribute information that you have identified using the wizard and a model is generated containing the resulting Item Definitions, Item Definition attributes, attribute to phrase associations, and related phrases and terms.

Defining the Data Lens Output

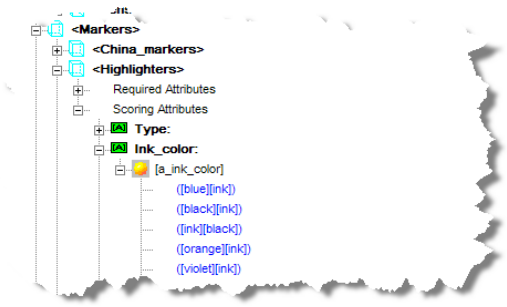
The final step of the wizard enables you to review the structure of the data lens that will be created, name it, and select any specific options before generating the data lens.



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 *child* Item Definitions. Expanding '**Highlighters**', and then the Scoring attributes allows you to view these attributes as in the following:

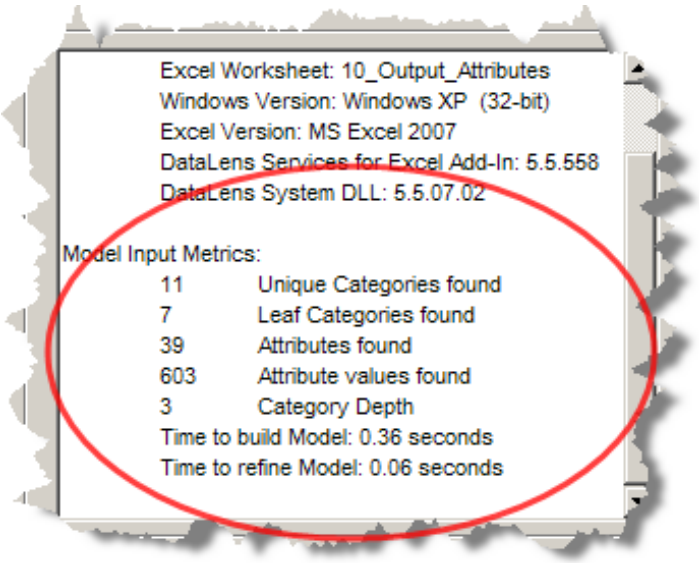


You can drill even further. Expanding the '**Ink_color**' attribute, and then the '**[a_ink_color]**' phrase allows you to view the entire phrase production as in the following:



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.

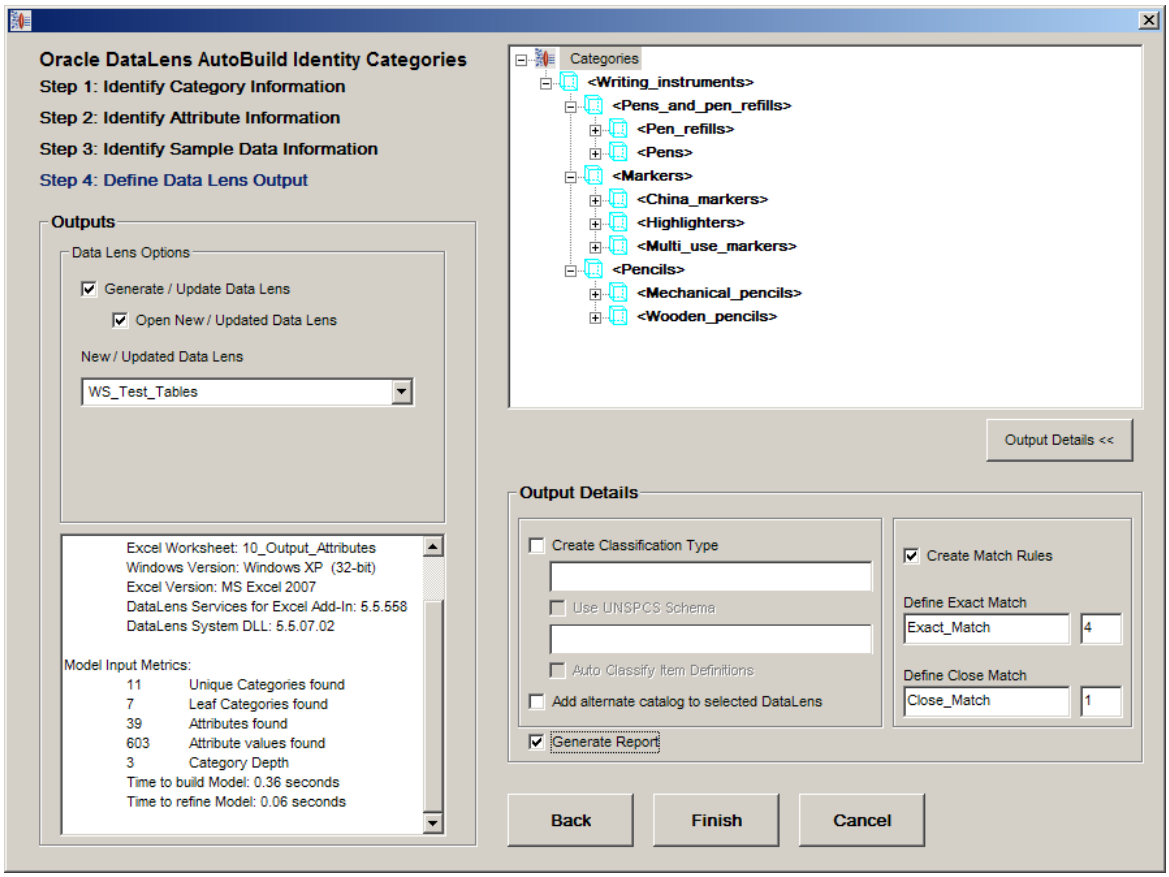
You should review the messages in the status pane to ensure that the metrics reported make sense and that there is no warning or error messages.



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 data examples. You should review the warnings to determine whether the structured input should be corrected.

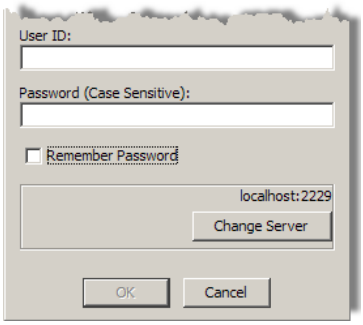
Because AutoBuild will not overwrite an existing lens, you must provide a unique data lens name each time you run AutoBuild.

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



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

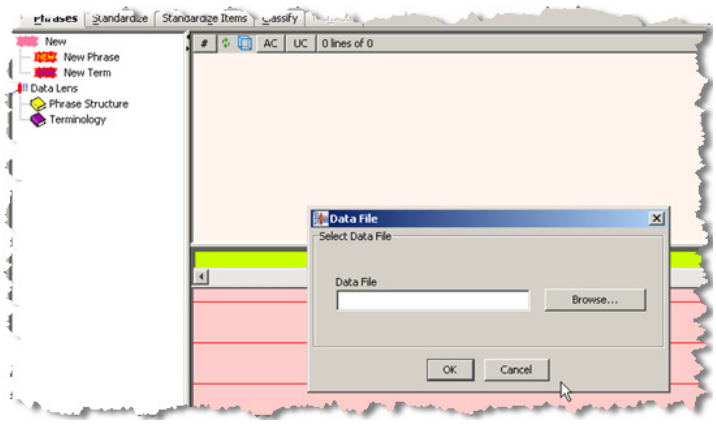
The new data lens is created from the category and attribute information provided in your structured data. When the AutoBuild process completes, the Oracle Product Data Quality login dialog box appears because the **Open New / Updated Data Lens** option was set in this example:



Enter your user name and password and click **OK**.

> Tip: You can avoid entering your password every time you logon by selecting the **Remember Password** checkbox.

The Knowledge Studio application automatically opens and requests that you select a sample file.



Select the sample data file that you want to use and click **OK** to open the data lens so that you can view the AutoBuild results.

You are now ready to modify and extend the 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 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.

   You should complete this step to limit the number of ambiguities to phrases that are not associated with attributes where multiple phrases reference the same term(s). This ensures that the remaining steps are successful.

2. Remove any remaining phrase level ambiguities.

   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 the required 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 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 lens regression test sets.

   Use the Knowledge Studio to create regression test set(s) in the lens so that future maintenance efforts can be proper 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 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 lens is added or updated.

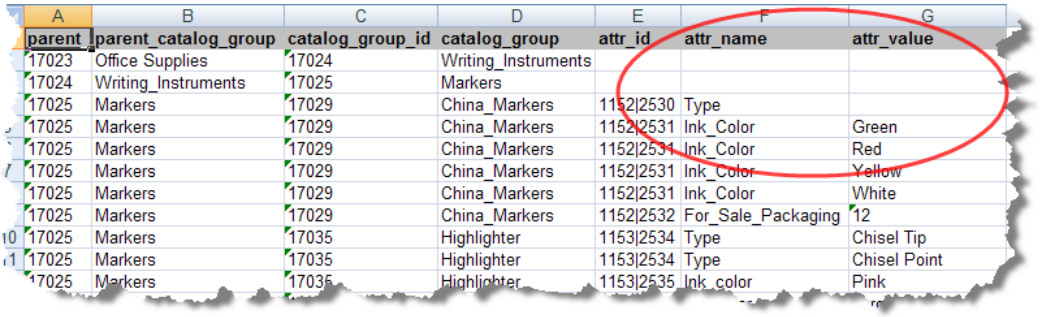
7. Improve the AutoBuild configuration.

   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

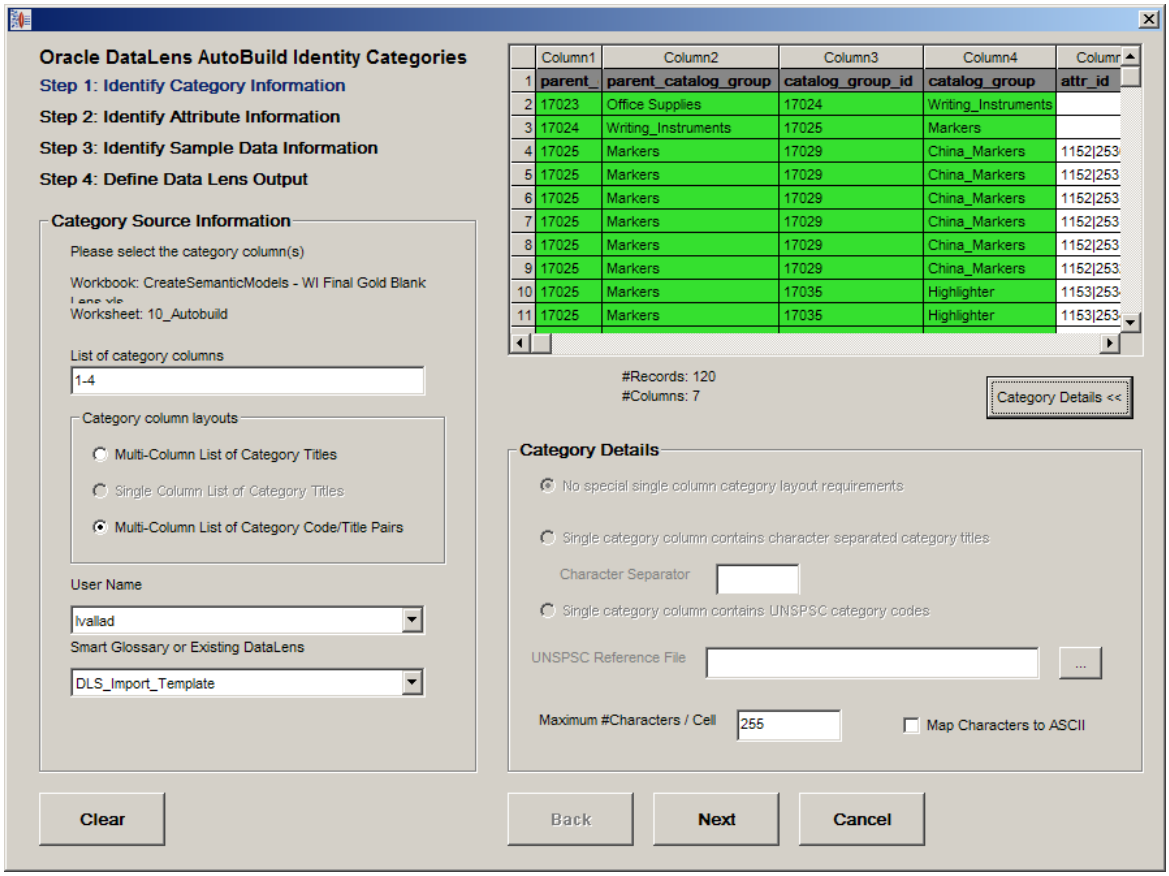
This section provides an example of using data that contains both the attribute name and value in each record. This paired data is used in the AutoBuild process in the same manner as described in the Using the AutoBuild Process section though the Attribute Information options are different.

An example of a paired data Excel spreadsheet is as follows:

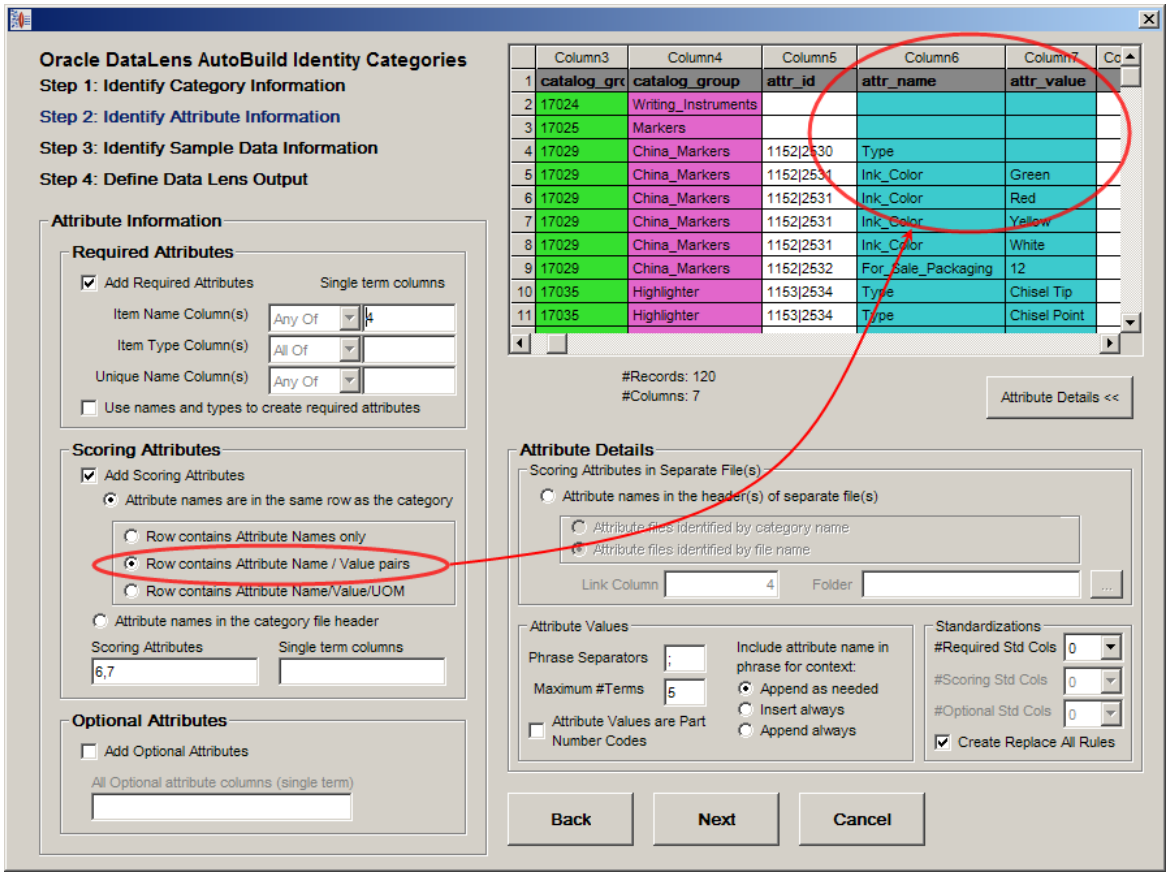


	A	B	C	D	E	F	G
	parent_	parent_catalog_group	catalog_group_id	catalog_group	attr_id	attr_name	attr_value
	17023	Office Supplies	17024	Writing_Instruments			
	17024	Writing_Instruments	17025	Markers			
	17025	Markers	17029	China_Markers	1152|2530	Type	
	17025	Markers	17029	China_Markers	1152|2531	Ink_Color	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
10	17025	Markers	17035	Highlighter	1153|2534	Type	Chisel Tip
11	17025	Markers	17035	Highlighter	1153|2534	Type	Chisel Point
	17025	Markers	17035	Highlighter	1153|2535	Ink_color	Pink

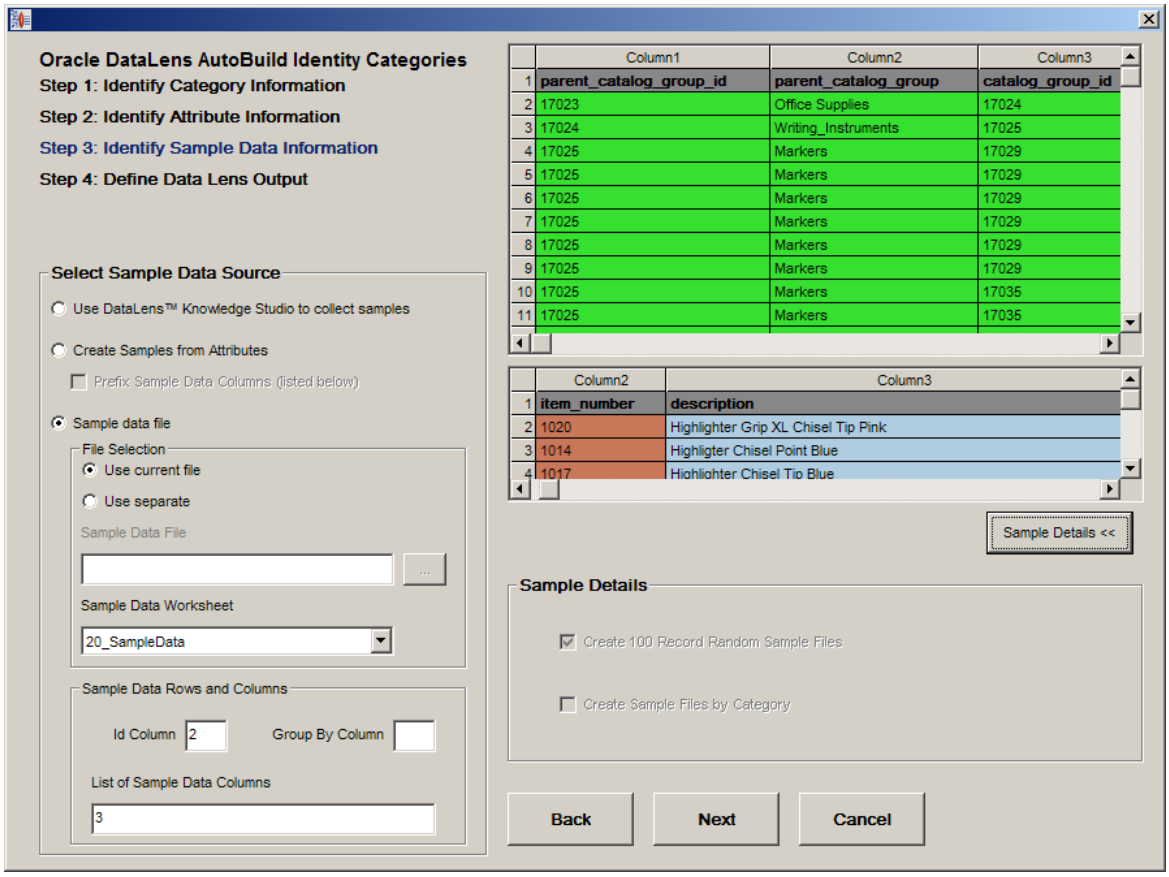
Selecting the category information could look like the following:



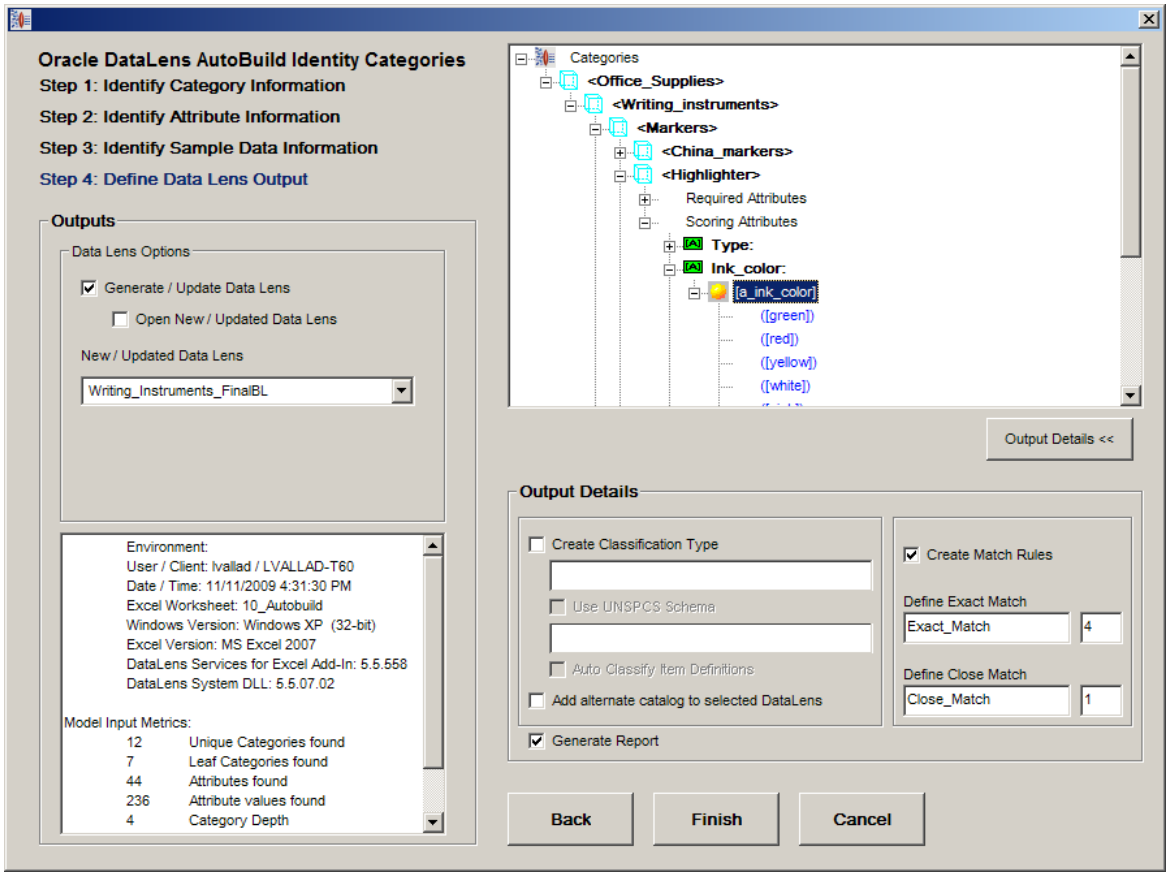
To indicate that the source data is paired occurs in Step 2 of the AutoBuild wizard as in the following:



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



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



Your data can be organized into columns in a myriad of ways in an Excel spreadsheet. 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.

Iteration

Auto-Build always creates a new data lens and does not overwrite an existing lens so you should use an iterative process to find the optimal AutoBuild settings for your data.

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 export the Item Definitions. Open an existing data lens and import the exported AutoBuild generated Item Definitions then use a common Smart Glossary.

Create Examples

When there is no well-structured, *gold* data, creating just a few example records per category can have a greatly decrease data lens development time. Additionally, you should use the data lens terminology importing function to close the gaps in data recognition.

***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 DataLens attribute centric matching.

Appendix A

Installing the Software

Installing the Software

Oracle Product Data Quality uses a concept called Java Webstart to initially install and maintain the current version of the software on your client desktop. The process requires you to access the Oracle DataLens Server to initiate the connection and download the software. If you have the correct license to access the software, you install the Oracle Product Data Quality client software through Java Webstart by browsing to the installation page for the Oracle DataLens Server as follows:

Browse to the following URL:

http://<*server*>:2229/datalens/datalens.jnlp

where <*server*> is the hostname of the Oracle DataLens Server to download the software.

If you do not have a supported java environment on the target installation machine, you must download and install a Java runtime environment from:

http://<*server*>:2229/datalens/datalens.html

Accepting the Security Warning

A security warning is displayed.

Oracle Product Data Quality files are digitally signed by a trusted source. To avoid the security dialogue in the future you can select the **Always trust content from this publisher** checkbox.

Click **Run** to continue.