# Oracle® Product Data Quality

PIM Connector User's Guide

Version 5.5

September 2010

ORACLE

Oracle Product Data Quality PIM Connector User's 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 ..... iii
About this Book ..... iii
*Intended Audience* ..... *iii*
*Conventions Used in This Book* ..... *iv*
*Related Information* ..... *v*
Chapter 1 ..... 1
Introduction ..... 1
Chapter 2 ..... 4
Creating the Initial Semantic Model ..... 4
*1. Determine Domain* ..... *5*
*2. Item Class Import Process from Excel* ..... *8*
*3. AutoBuild Process in Excel* ..... *10*
*4. Review Semantic Model Report* ..... *13*
*5. Review Semantic Model* ..... *16*
*6. Refine Semantic Model* ..... *16*
*7. Setup Standards and Match Rules* ..... *17*
*8. Test a Data Lens on Sample Data and Create Base File* ..... *18*
Chapter 3 ..... 20
Importing and Mapping an Alternate Catalog ..... 20
*1. Define Alternate Catalog* ..... *20*
*2. Alternate Catalog Import Process from Excel* ..... *23*
*3. AutoBuild Process in Excel* ..... *24*
*4. Review Semantic Model Report* ..... *27*
*5. Review Alternate Catalog* ..... *28*
Creating Semantic Key Cache ..... 30
Chapter 4 ..... 31
Cleansing and Matching Batches ..... 31

*Introduction* ..... *31*
*1. Create Batch* ..... *31*
External Data Source ..... 32
*Create a Batch and import new items into the batch.* ..... *32*
Production Pull Data Source ..... 35
*Create a Batch from production PIM data* ..... *35*
*Process Interface Batch* ..... *35*
*2. Graph Summary* ..... *37*
*3. Ready for Load* ..... *37*
*4. Alternate Catalog Review* ..... *39*
*5. Dups Within Batch* ..... *41*
*6. Match Against PIM* ..... *44*
*7. Items for Enrichment* ..... *45*
*8. Exceptions* ..... *46*
*9. Style Matches* ..... *47*
*10. Release Batch* ..... *48*
*11. Error Codes* ..... *49*
Appendix A ..... 50
Installing the Client Software ..... 50

# Preface

## About this Book

This guide highlights the core process steps and features of Oracle Product Data Quality.

By extending the Oracle Product Information Management Data Hub with data quality capabilities, Oracle enables customers to improve the accuracy, completeness, and integrity of their master product data.

This guide uses the Vision Demo Database as a means to provide access to a common set of data, which is available to use as examples and validate the setup and configuration of the Oracle DataLens Server. It uses the Item Class Category Name of Capacitors under the Item Class of Passives.

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

You must have the Oracle Product Data Quality client software installed on your computer.

### Intended Audience

You should have a basic understanding of the DataLens Technology.

This manual is designed for the new users of the Oracle Product Data Quality:

- Business Analysts
- Subject matter experts (SMEs)
- IT administrators

## 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:

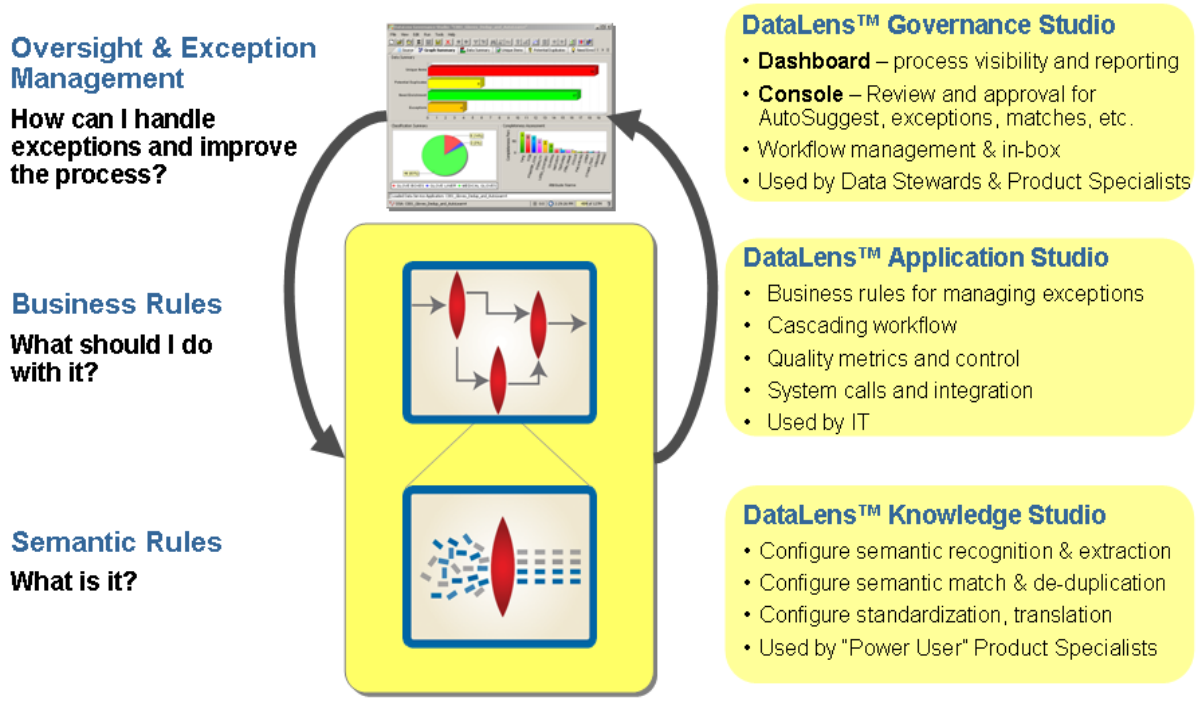
- The *Oracle Product Data Quality Application Studio Reference Guide* provides information about creating and maintaining Data Service Applications (DSAs).
- The *Oracle Product Data Quality AutoBuild Reference Guide* provides information about creating initial data lens based on existing product information and data lens knowledge.
- 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.

# Chapter 1

## Introduction

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.



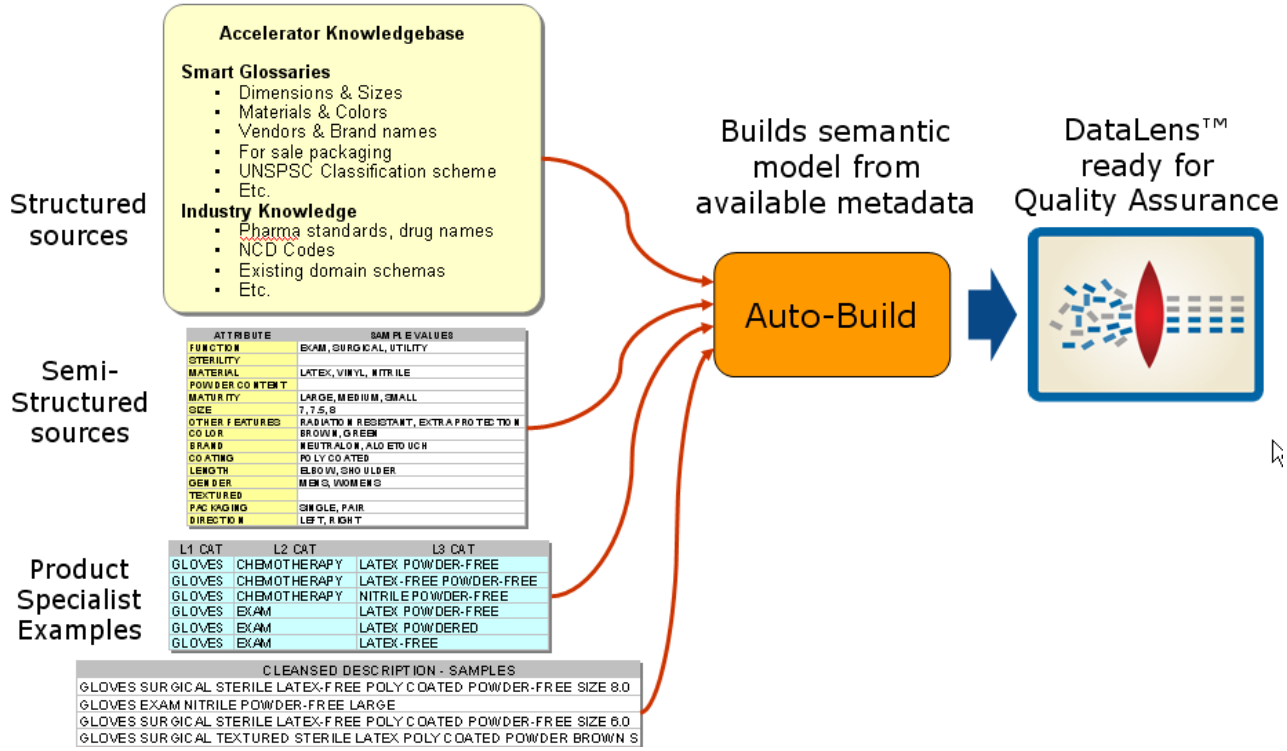

The process uses Oracle Product Data Quality AutoBuild, which is a component of the DataLens Technology to quickly leverage the existing Product Data Hub set of information and the Oracle Product Data Quality Smart Glossaries to create initial Semantic Model (data lens) specific to your enterprise content.

Oracle Product Data Quality ships with a set of application components to allow the user to test and ensure proper system configuration. The default data lens to process items is called 001_Capacitors_Demo_V2. This data lens contains semantic models that work with the Capacitors Item Class category in the Vision Database installed with most Oracle Product Information Management (PIM) Systems. The remainder of the document uses these samples shipped with the product.

AutoBuild constructs a data lens by examining the available product data examples. Given sufficient information, AutoBuild can accomplish the following.

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

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



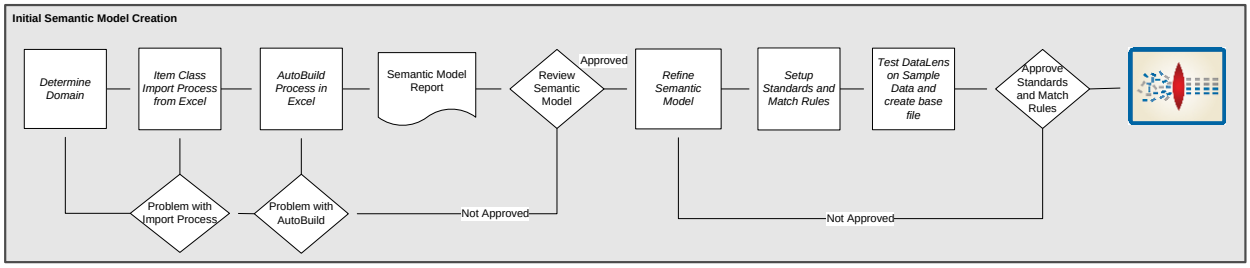
# Chapter 2

## Creating the Initial Semantic Model

You create an initial Semantic Model through the AutoBuild program, which is extracts the metadata from the Oracle Item Class tables and samples from the Product Data Hub Production tables. The process takes a single input parameter for Item Class Category Name and uses another parameter called Display Type Format. The Display Type Format is contained in Profile Options value called 'SemanticModel". For more information about configuring this option, see the *Oracle Product Data Quality Information Management Implementation Guide*.

The Item Class category will determine which sections of the Item Class Hierarchy will be extracted and the Display Format will determine which attribute metadata to extract to create the initial Semantic Model.

The Product Data Hub contains both information about the Item Classification and Attributes (Metadata) as well as sample data. The AutoLearn process leverages this information to create the initial semantic model. A Semantic Model is equivalent to a fully configured Data Lens with the Item Definition, attributes, phrases and terms as well as the standardizations and match rules. The following figure illustrates the process.



# 1. Determine Domain

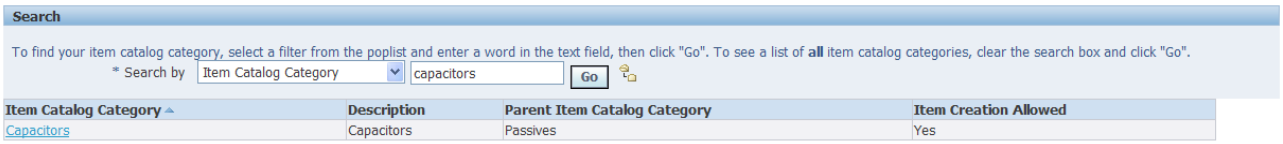
The user must determine what domain (Item Class) they would like to use to create the Semantic Model. This requires the setup of an Oracle Item Class Category Name.

Oracle Item Class Category Name Setup

The Semantic Model creation process relies on each Item Class Category Name being setup correctly and named correctly. The Semantic Model is associated with the Item Catalog hierarchy. The following shows the Capacitor hierarchy as laid out in the Vision Operations Item Catalog.

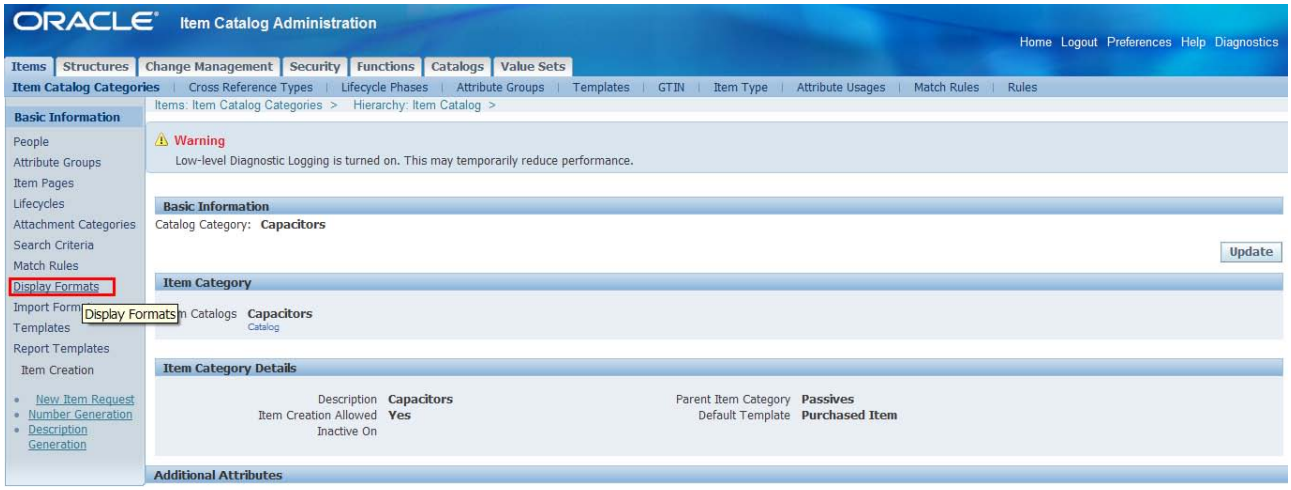
- PLM High Tech
    - Electronic Components
        - Passives
            - Capacitors

From the Setup Workbench, search for the Item Catalog Category for example, "Capacitors".

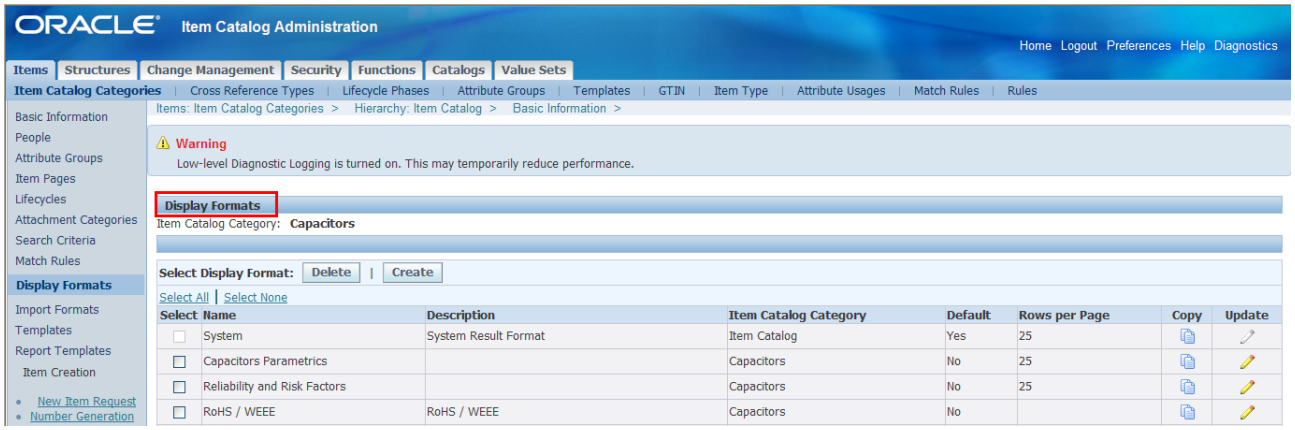



Select the "Capacitors" hyperlink.

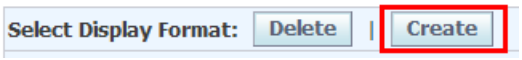
From the left side menu, select "Display Formats" to create the Semantic Model definition created from a set of attributes available in the Product Data Hub.



The Display Formats page allows you to create or edit (update) a Display Model.

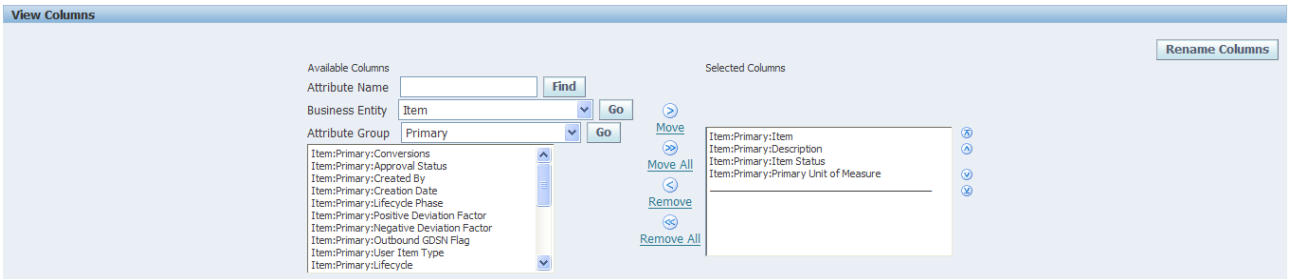



1. Create a Display Format



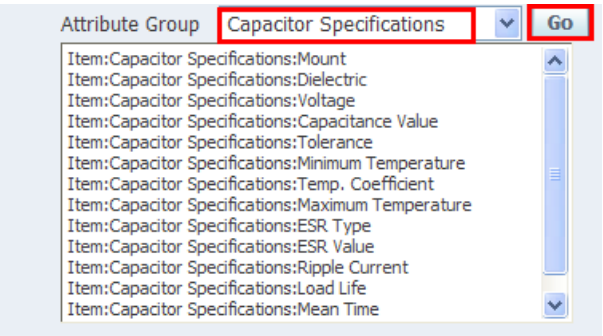

2. Name the model: "SemanticModel"
3. Provide a Description: "DataLens Semantic Model Attributes"
4. Select Number of Rows Displayed: "100"
5. Set as Default: check the box

6. Create the Display Format by selecting Attributes from Available Columns section on the left side (there are a set from the Primary Attribute Group by default in the right panel).



a. "Business Entity" should be "Item"

b. Each Attribute Group has a set of Attributes values, for example by choosing Capacitor Specifications and clicking "Go", you will see the set of attributes associated with that Attribute Group.



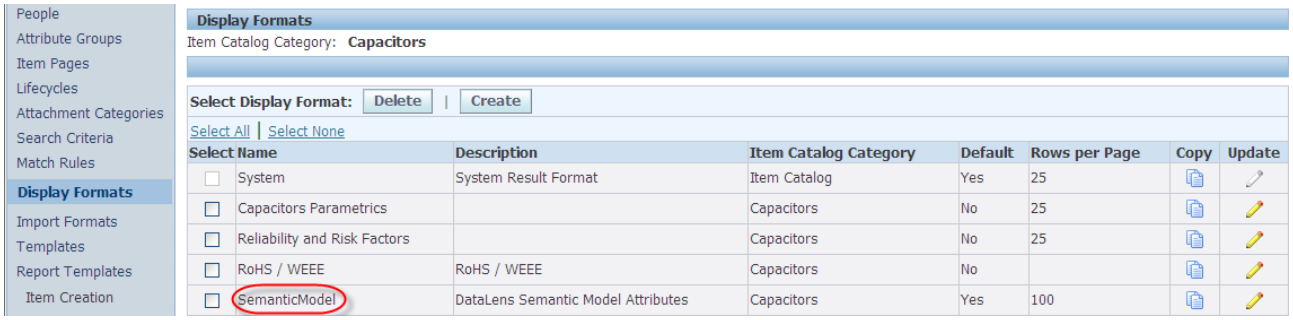

c. Highlight an individual attribute and select the > Move button to move them to the Selected Columns panel.

7. Select all of the Attributes from the Item:Capacitor Specifications using Move All button. The attributes should include the ones as shown in the table below.

| **Business Entity** | **Attribute Group** | **Attribute** |
|---|---|---|
| Item | Capacitor Specifications | Capacitance |
| Item | Capacitor Specifications | Dielectric |
| Item | Capacitor Specifications | ESR_Type |
| Item | Capacitor Specifications | ESR_Value |
| Item | Capacitor Specifications | Load_Life |
| Item | Capacitor Specifications | Maximum_Temperature |
| Item | Capacitor Specifications | Mean_Time |

| | | |
|---|---|---|
| Item | Capacitor Specifications | Min_Temperature |
| Item | Capacitor Specifications | Mount |
| Item | Capacitor Specifications | Ripple_Current |
| Item | Capacitor Specifications | Temp_Coefficient |
| Item | Capacitor Specifications | Tolerance1 |
| Item | Capacitor Specifications | Voltage |

8. Select the "Apply" button



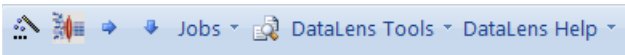

A new Display Format called SemanticModel with all of the necessary attributes is created for input into the AutoBuild process. The set of attributes should correlate to the spreadsheet output tab 10_AutoBuild in the CreateSemanticModels spreadsheet. Note: the AutoBuild spreadsheet tab has attribute name / value pairs where the attributes names are in the attr_name column and they run vertically.

## 2. Item Class Import Process from Excel

You employ the CreateSemanticModels Template Excel spreadsheet to import the Item Class (Category) and Attribute Metadata from the Oracle Product Information Management Data Hub system to create the import files for AutoBuild.

Note: The Template relies on the Oracle Product Data Quality Services for Excel.

1. Open Sheet 1 from the CreateSemanticModels Template Excel spreadsheet
2. Use the Services for Excel Add-Ins tab



3. Press the "Submit Job to transform server" icon to launch the ORA_AU_Autobuilder DSA to create the AutoBuild input files for specific categories for use by Oracle Apps Unlimited
4. Enter the Item Class Category parameter (for example, "Capacitors") associated with the Semantic Model that you are building and click "OK".
5. The job will be submitted to the server for processing.
6. Click on the down arrow "Get Job Results from transform server" icon to retrieve the results from the completed job.
7. When the job is retrieved, it creates two additional tabs, 10_Autobuild and 20_SampleData.

a. 10_Autobuild tab contains the following columns

| **Columns** | **Definition** |
|---|---|
| parent_catalog_group_id | System ID |
| parent_catalog_group | Item Class Category Parent Hierarchy Node Name |
| catalog_group_id | System ID |
| catalog_group | Item Class Category Name |
| attr_id | System ID |
| attr_name | Attribute name |
| attr_value | Attribute value |

b. 20_SampleData tab contains the following columns

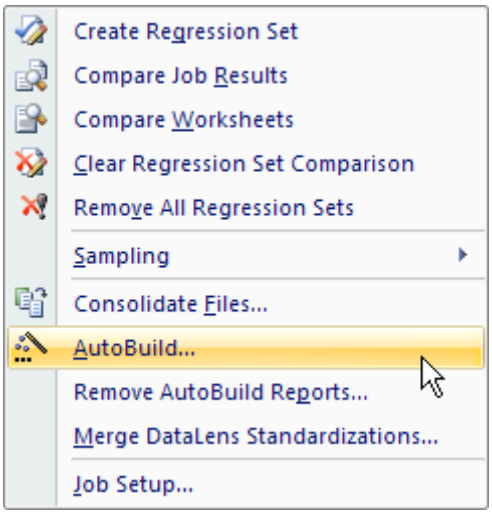
| **Columns** | **Definition** |
|---|---|
| catalog_group | Item Class Category Name |
| sku | Part number |
| description | Attribute description |

Two output tabs labeled tab 10_AutoBuild and 20_SampleData will be created automatically in the CreateSemanticModelss spreadsheet. The 10_AutoBuild spreadsheet should have at least one of the attributes (it may be with no data) listed from the Display Format in the Semantic Model created in the first step.

# 3. AutoBuild Process in Excel

AutoBuild constructs a data lens from the available product data metadata and example data (the server job pulls a distinct set of data) shown in the 10_AutoBuild tab. Given the correctly formatted information, AutoBuild will accomplish the following.

Start the AutoBuild process from the 10_AutoBuild tab.




1. Construct a full Item Definition hierarchy based on the Item Catalog Category, complete with required, scoring, and optional attributes
2. Construct rich term and phrase recognition rules
3. Provide an initial set of classification rules
4. Create Match Rules (Exact and Close Match)
5. Create Standardizations Types

AutoBuild provides an easy-to-use graphical wizard interface to a four step process to build the data lens.

The pre-configured AutoBuild settings consist of the following values:

| Steps | Values |
|---|---|
| **Step 1: Identify Category Information** | |
| List of category columns | 1-4 |
| Smart Glossary | DLS_Import_Template |

**Step 2: Identify Attribute Information**

**Required Attributes**

| | |
|---|---|
| Add Required Attributes | checked |
| Item Name Column(s) | 4 |

**Scoring Attributes**

| | |
|---|---|
| Add Scoring Attributes | checked |
| Attribute names in same row | selected |
| Row contains attribute name/value columns | selected |
| Scoring Attributes | 6,7 |
| Single term columns | 6,7 |
| **Standardizations** | (Attribute Details section) |
| Create Replace All Rules | checked |

**Step 3: Identify Sample Data Information**

**Select Sample Data Source**

| | |
|---|---|
| Sample data file | selected |
| Use current file | selected |
| Sample Data Worksheet | 20_SampleData |

**Sample Data Rows and Columns**

| | |
|---|---|
| Id Column | 2 |
| Group By Column | 1 |
| List of Sample Data Columns | 3 |

**Step 4: Identify Sample Data Information**

**Output**

**Data Lens Options**

| | |
|---|---|
| Generate / Update Data Lens | checked |
| Open New Data Lens | selected |

| | |
|---|---|
| Data Lens Name | <data lens name> |
| **Output Details** | |
| Create Match Rules | checked |
| **Define Exact Match** | |
| Exact_Match | 4 |
| Close_Match | 1 |
| Generate Report | Checked (should be checked) |

The initial data lens will be built with the Item Definition (Category) and attribute structure shown in the pre-configured AutoBuild settings. A set of reports added to the CreateSemanticModels spreadsheet as tabs. These reports include AutoBuild Meta Data Report, AutoBuild Summary Report, and AutoBuild Details Report.

# 4. Review Semantic Model Report

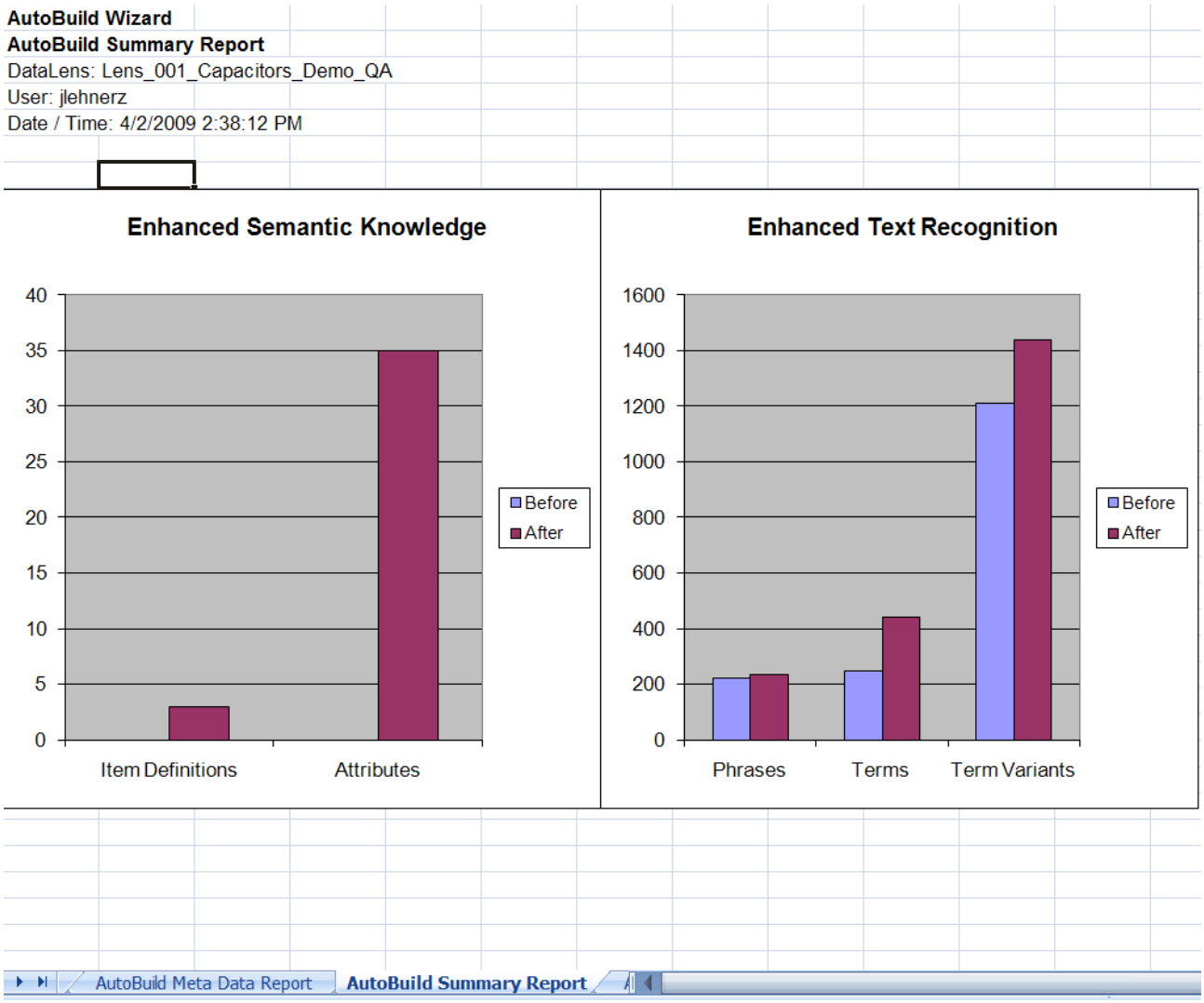
Review the AutoBuild reports to validate the data lens creation.

AutoBuild Meta Data Report defines the Item Definition and the associated Required / Scoring Attributes, terms and data. The report has three sections: Attribute Phrases and Terms. You should validate that each of them correlate with the 10_AutoBuild tab.

| | | | |
|---|---|---|---|
| **AutoBuild Wizard** | | | |
| **AutoBuild Meta Data Report** | | | |
| DataLens: Lens_001_Capacitors_Demo_QA | | | |
| User: jlehnerz | | | |
| Date / Time: 4/2/2009 2:38:12 PM | | | |
| Item Definitions | | | |
| Item Definitions with Attributes | | | |
| Phrases | | | |
| Terms | | | |
| **Item Definitions** | **New Meta-Data** | | |
| | **Renamed Meta-Data** | | |
| | Existing Meta-Data | | |
| **Passives** | | | |
| | **Capacitors** | | |
| | | **Capacitor_Networks** | |
| **Item Definitions with Attributes** | **New Meta-Data** | | |
| | **Renamed Meta-Data** | | |
| | Existing Meta-Data | | |
| Capacitors | Required Attributes | Item_Name | [i_capacitors] |
| | Scoring Attributes | Capacitance | /[a_capacitance] |
| | | Dielectric | [a_dielectric] |
| | | Voltage | /[a_voltage] |
| | | Tolerance1 | [a_tolerance1] |
| | | Capacitor_package_size | [a_capacitor_package_size] |
| | | Capacitor_case_size | [a_capacitor_case_size] |
| | | Temp_coefficient | [a_temp_coefficient] |

Sheet1 AutoBuild Meta Data Report

AutoBuild Summary Report provides two sets of graphs that provide a before and after view of the Semantic Knowledge and Text Recognition. The initial lens uses the DLS_Import_Template Smart Glossary as the starting point (before).



AutoBuild Details Report is a high level summary report showing the knowledge creation results.

| | |
|---|---|
| **AutoBuild Wizard** | |
| **AutoBuild Details Report** | |
| DataLens: Lens_001_Capacitors_Demo_QA | |
| User: jlehnerz | |
| Date / Time: 4/2/2009 2:38:12 PM | |
| | |
| | |
| | |
| Totals | |
| Performance | |
| Item Definition Additions | |
| Item Definitions Not Found | |
| Attribute Additions | |
| Attributes Not Found | |
| Phrase Refinements | |
| Term Refinements | |
| Classification Refinements | |
| AutoBuild Warnings | |
| | |
| | |
| | |
| | |
| **Totals** | |
| Input Meta-Data Metrics | |
| Unique Categories found | 3 |
| Leaf Categories found | 1 |
| Attributes found | 34 |
| Attribute values found | 504 |
| Category Depth | 3 |
| Unique Phrases found | 16 |
| Phrase Productions found | 122 |
| Unique Terms found | 192 |
| Category Sample Records generated | 156 |
| Category Sample Files generated | 1 |
| Random Sample Records generated | 156 |
| Random Sample Files generated | 1 |

AutoBuild Summary Report | **AutoBuild Details Report**

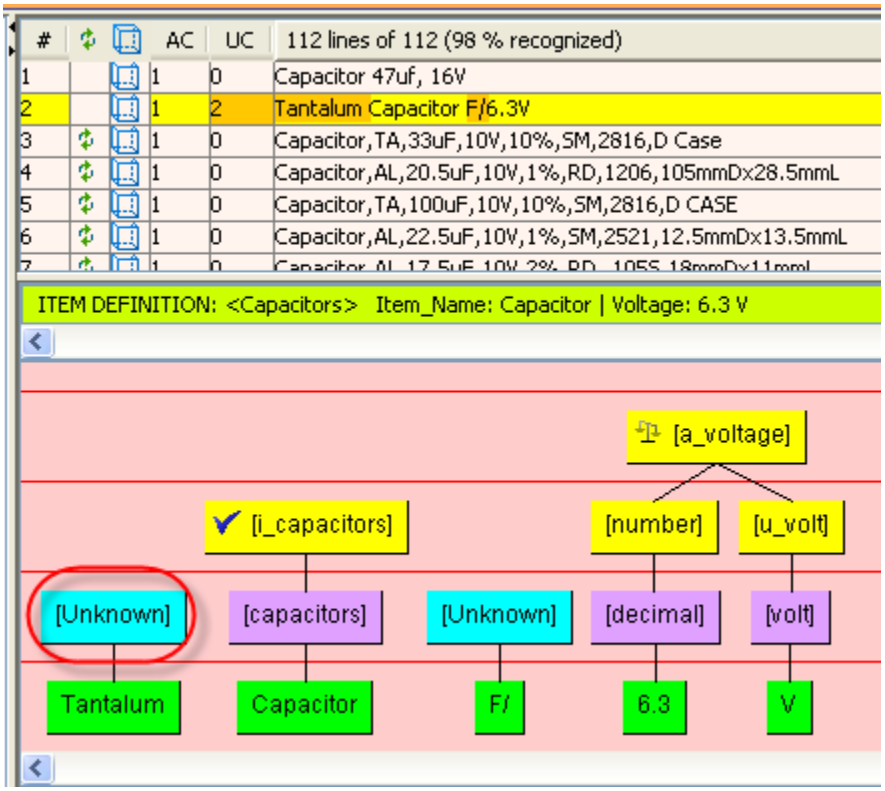
These reports should correctly represent the results of the AutoBuild knowledge building process and should correspond to the data lens knowledge. For more information on the AutoBuild process, see the *Oracle Product Data Quality AutoBuild Reference Guide*.

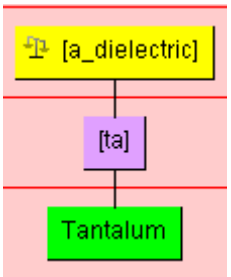
# 5. Review Semantic Model

Review the data lens to confirm proper creation. The data lens should have the correct Item Definition, Attributes Phrases, and Terminology as defined by the AutoBuild process.

# 6. Refine Semantic Model

You should review the data lens to define any undefined terminology to complete the process. The following diagram shows an example of the term "Tantalum" not being recognized and shown as "Unknown". A Subject Matter Expert would know that this is a variant of 'ta', which is a type of dielectric. The user would make the addition to refine the data lens.

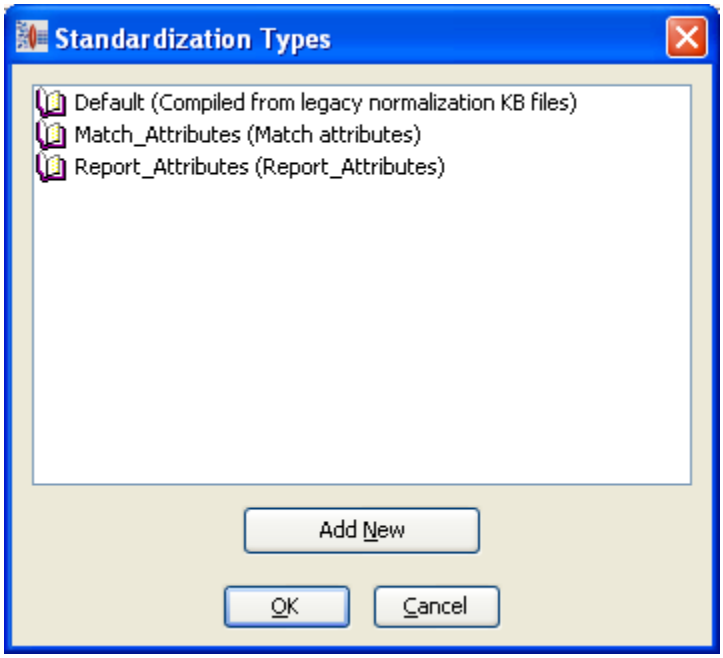


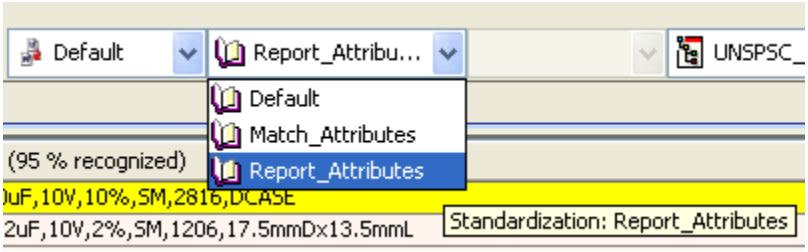

For more information, see the *Oracle Product Data Quality Knowledge Studio Reference Guide*

# 7. Setup Standards and Match Rules

Setup and review the data lens standardizations. There should be three Standardizations: Default, Match_Attributes, and Report_Attributes. You can see them by selecting the Standardization Types from the DataLens menu.

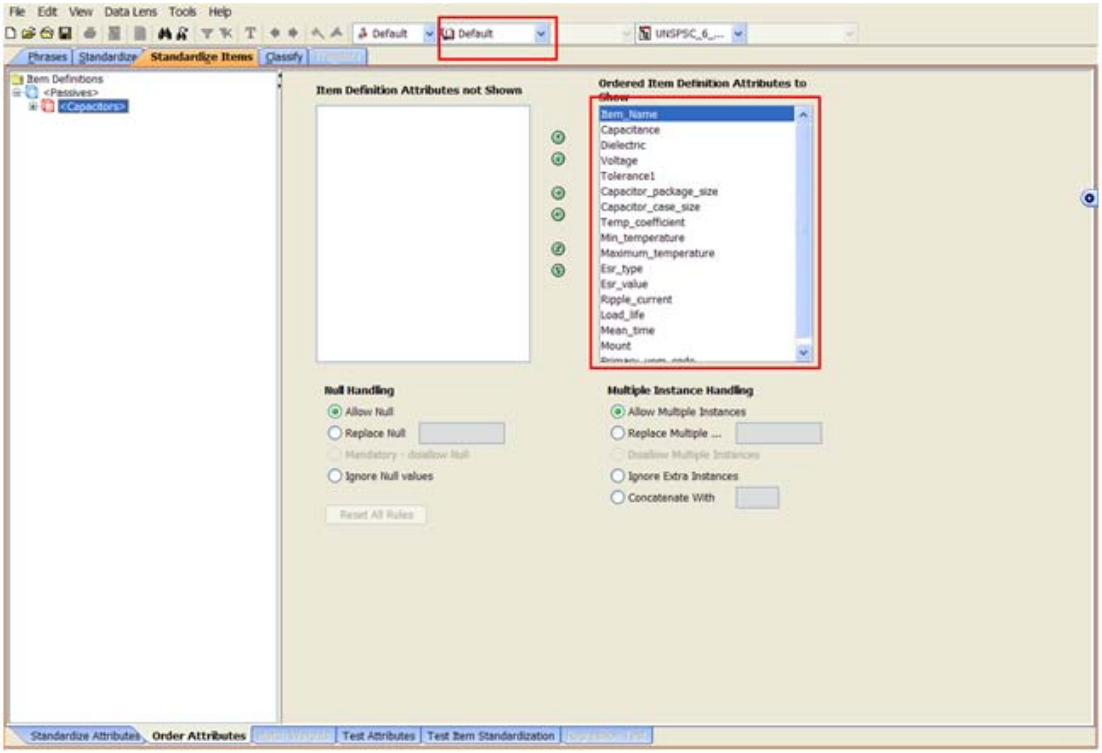



You can also view the Standardizations from the program toolbar.




Note: The attribute order is the same order as the name/value attribute order in the AutoBuild tab, which is based on the Display Format order. For the Match_Attributes standardization, there should be two Match weights created (Exact_Match with 4 attributes and Close_Match with 1 attribute).

For example, default Standardization would have the following attribute order:

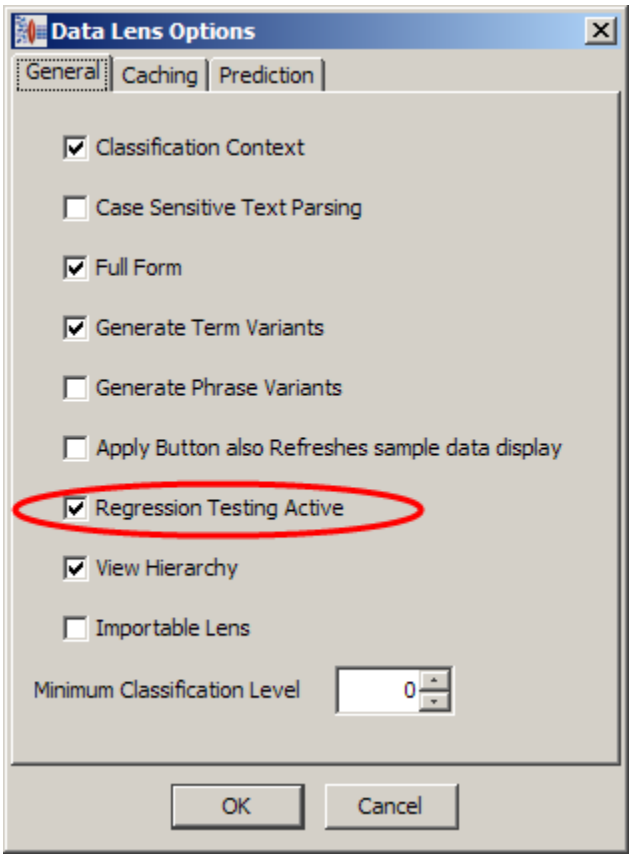



The user should review and change the attribute ordering, create phrase standardizations and replacements appropriate to the data lens.

## 8. Test a Data Lens on Sample Data and Create Base File

The purpose of regression testing is to validate that any maintenance to the data lens has not created any ambiguities or issues compared to the phrases and terms that were previously defined in a prior version of the data lens. This important step should be performed after changes have been made to a mature data lens.

Regression testing can be accessed from the phrases, standardization and classification tabs. The user should set the "Regression Testing Active" option from the DataLens Options to create the base file to enable Regression testing.



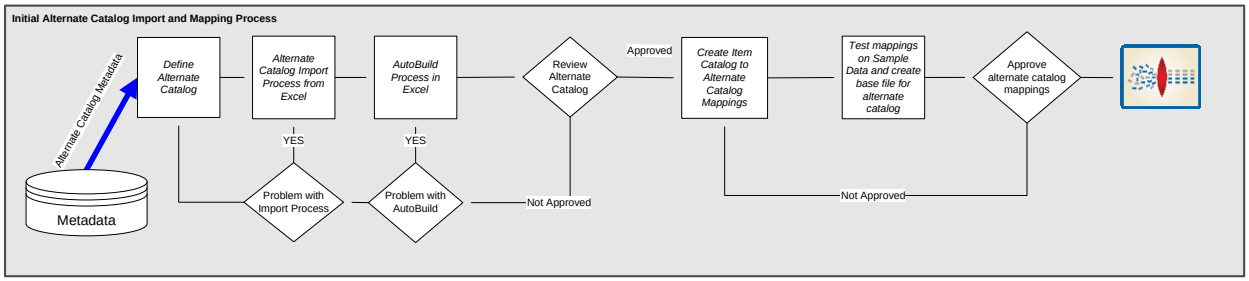

For more information about setting up and using Regression tests, see the *Oracle Product Data Quality Knowledge Studio Reference Guide*.

Once you are happy with the data lens works, you should contact your Administrator to associate this new data lens with the cleansing and matching process.

# Chapter 3

## Importing and Mapping an Alternate Catalog

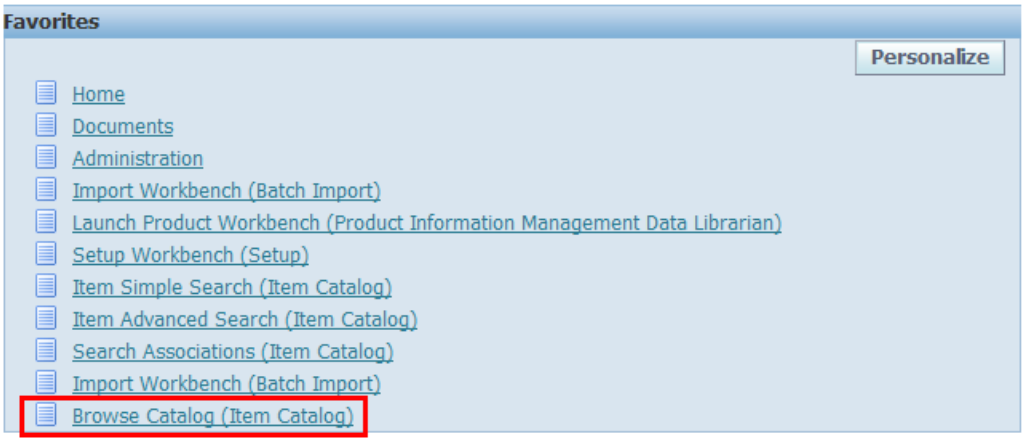
You create an alternate catalog mapping through the AutoBuild program, which extracts the Oracle Alternate Catalog Hierarchy structures metadata from the Product Data Hub Production tables. The process takes a single input parameter for Item Class Category Name. The Item Class category will determine which sections of the Item Class Hierarchy will be extracted to create an export that can be used by AutoBuild to create a schema file.



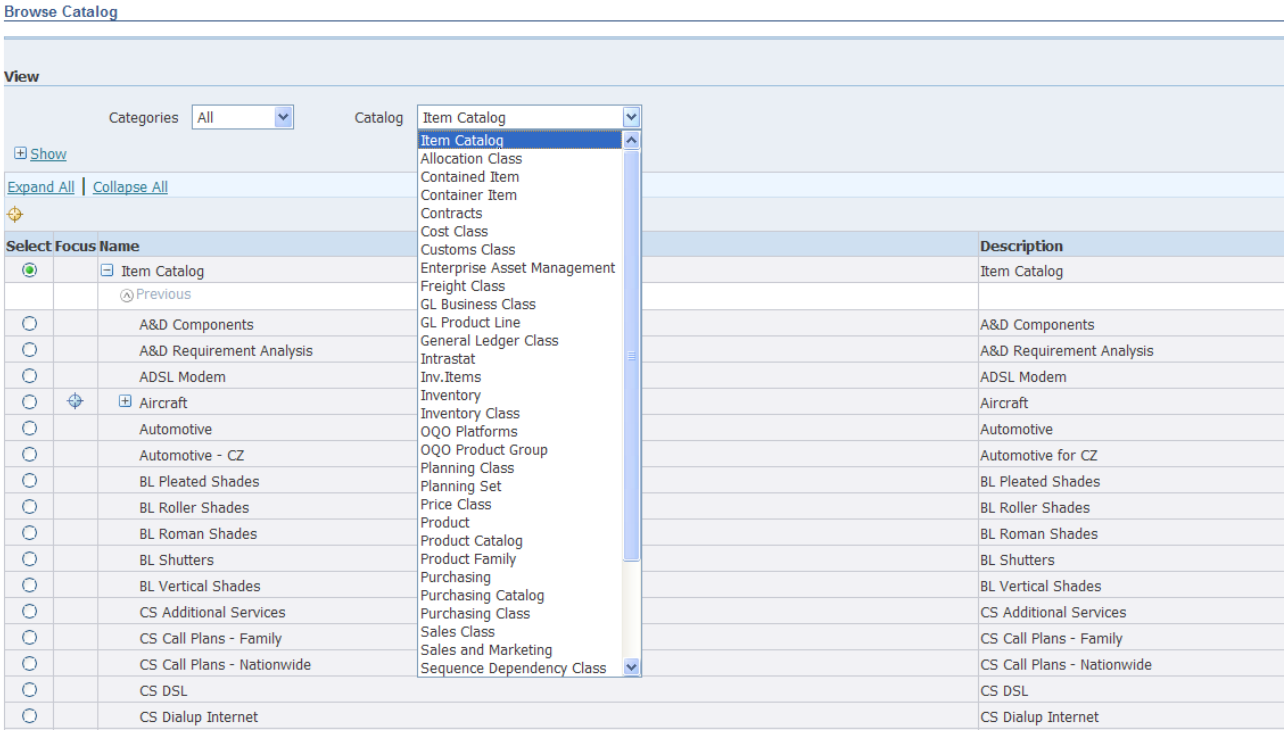

### 1. Define Alternate Catalog

The user must determine what domain (other catalog) they would like to associate with the previously created Semantic Model. This requires the user to know what catalogs exist in the Oracle Product Data Hub in order to select the one to associate with the existing Oracle Item Class Category Name.

Select Browse Catalog.



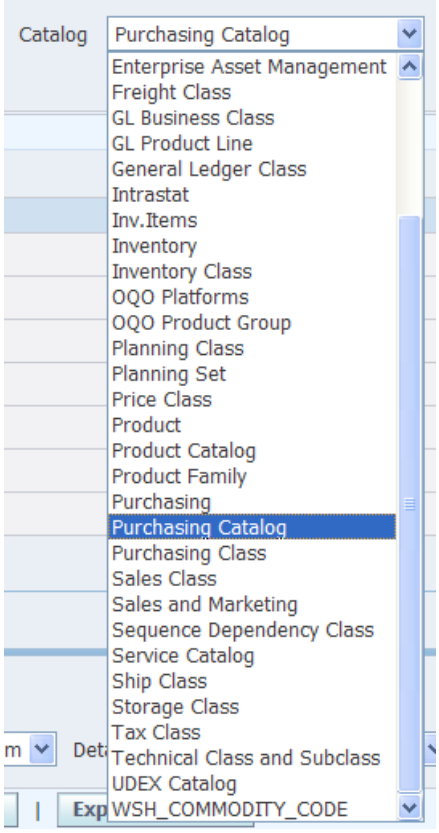

The Oracle Product Data Hub has a set of preconfigured catalogs. From the home page, you can browse the catalogs and view the active catalogs from the "Catalog" drop down menu.



The following shows the Capacitor hierarchy as laid out in the Vision Operations Item Catalog.

- Item Catalog
    - PLM High Tech
        - Electronic Components
            - Passives
                - Capacitors

The default configuration includes the Purchasing Catalog as an alternate catalog for the Capacitors Semantic Model. You should select "Purchasing Catalog" from the Catalog list.



The Purchasing Catalog has a slightly different hierarchy with one less level.

- Purchasing Catalog
    - Electronic Components
        - Passives
            - Capacitors

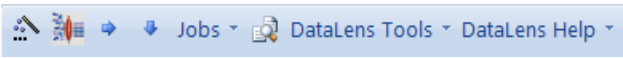
| | |
|---|---|
| Purchasing Catalog | Purchasing Catalog |
| Electromechanical Components | Electromechanical Components |
| Electronic Components | Electronic Components |
| Discretes | Discretes |
| IC's | IC's |
| Optoelectronics | Optoelectronics |
| Passives | Passives |
| Capacitors | Capacitors |
| Circuit Protection | Circuit Protection |
| Encoders | Encoders |
| Frequency Controlers | Frequency Controlers |
| Inductors | Inductors |
| RC Networks | RC Networks |
| Resistors | Resistors |
| Signal Transformers | Signal Transformers |

An alternative catalog is a new user defined classification to add to the Semantic model (data lens) to allow the association of individual item with more than one catalog.

## 2. Alternate Catalog Import Process from Excel

The user employs the "ImportAlternateCatalogs.xls" Excel spreadsheet to import the Alternate Category from the Oracle Product Data Hub system to create the import files for AutoBuild.

1. Open Sheet 1 from the ImportAlternateCatalogs Template Excel spreadsheet
2. Use the Services for Excel Add-Ins tab:




3. Press the "Submit Job to transform server" icon to create the AutoBuild input files for specific alternate category for use by Oracle Apps Unlimited
4. Enter the Alternate Catalog name parameter. As an example enter "Purchasing Catalog" and click "OK" and the job will be submitted to the server for processing.

a. You can use any Catalog that is listed in Oracle's E-Business, but there should a logical relationship for the selection.
   i. Purchasing Catalog
   ii. Product Catalog
   iii. Etc…

5. Click on the down arrow "Get Job Results from transform server" icon to retrieve the results from the completed job.
6. When the job is retrieved, it creates one additional tab, 001_Alternate_Catalog, which contains the following columns.

| Columns | Definition |
|---|---|
| catalog | Alternate Catalog name |
| parent_catalog_id | System ID |
| parent_category_desc | Alternate Catalog hierarchy descriptions |
| category_id | System ID |
| catetory_name | Alternate Catalog hierarchy category names |

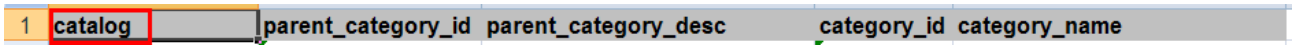
The output tab labeled tab 001_Alternate_Catalog will be input to the AutoBuild process.

## 3. AutoBuild Process in Excel

AutoBuild constructs a new user defined classification into a data lens from the available category information shown in the 001_Alternate_Catalog tab. Given the correctly formatted information, AutoBuild will construct a new user defined Classification with the same name as the catalog name (for example, Purchasing_Catalog) using the name in the catalog column. You should start the AutoBuild process from the 001_Alternate_Catalog tab.

Notes:

1. The "catalog" column name is a special name for the AutoBuild process and must be used to construct the classification (Alternate Catalog).



2. AutoBuild uses the numeric structure and relationship of the parent_category_id and category_id to define the classification hierarchy.

| 1 | catalog | parent_category_id | parent_category_desc | category_id | category_name |
|---|---|---|---|---|---|

3. For the purposes of this flow the Administrator must have included the data lens in the process.

Note: The process can automatically handle the creation of three alternate catalogs. If you require more than three alternate catalogs, then contact your Oracle DataLens Server Administrator.

AutoBuild provides an easy-to-use graphical wizard interface to a four-step process to build the data lens.

The pre-configured AutoBuild settings consist of the following values:

| **Steps** | **Values** |
|---|---|
| **Step 1: Identify Category Information** | |
| List of category columns | 3,5 |
| Smart Glossary | use an existing data lens aside from Smart Glossaries (For example, DLS_ prefix) |
| Multi-Column List of Category Code/Title paris | Selected |
| **Step 2: Identify Attribute Information** | ***Nothing Entered*** |
| **Required Attributes** | |
| Add Required Attributes | |
| Item Name Column(s) | |
| **Scoring Attributes** | |
| Add Scoring Attributes | |
| Attribute names in same row | |
| Row contains attribute name/value | |
| Scoring Attributes | |
| Single term columns | |

| **Step 3: Identify Sample Data Information** | ***Nothing Entered*** |
|---|---|
| **Select Sample Data Source** | |
| Use Knowledge Studio | selected |
| Use current file | |
| Sample Data Worksheet | |
| **Sample Data Rows and Columns** | |
| Id Column | |
| Group By Column | |
| List of Sample Data Columns | |
| **Step 4: Identify Sample Data Information** | |
| **Output** | |
| **Data Lens Options** | |
| Generate / Update Data Lens | checked |
| Open New Data Lens | selected |
| Data Lens Name | <data lens name> |
| **Output Details** | |
| Create Classification Type | Checked (Grayed out) |
| Type Name | Catalog name (shown in catalog column) |
| Add alternate catalog | Checked (this will Gray out other options) |
| Create Match Rules | Grayed out |
| **Define Exact Match** | |
| Exact_Match | 4 (Grayed out) |
| Close_Match | 1 (Grayed out) |
| Generate Report | Checked (should be checked) |

*Note: the data lens can be anything to test the creation of the user defined classification, but must be correct data lens in order to affect the entire process as described in note number 3 above.

When the process completes, the selected data lens will have a new user defined classification. AutoBuild will also create a set of reports added to the spreadsheet as tabs. These reports include AutoBuild Meta Data Report, AutoBuild Summary Report, and AutoBuild Details Report.

## 4. Review Semantic Model Report

Review the AutoBuild reports to validate the data lens creation.

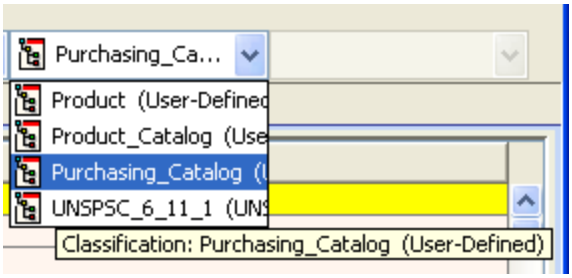
AutoBuild Meta Data Report defines the Item Definition and the associated Required / Scoring Attributes, terms and data. There are three sections (Attribute, Phrases and Terms) to the report. You should validate that each of them correlate with the data in the AutoBuild tab.

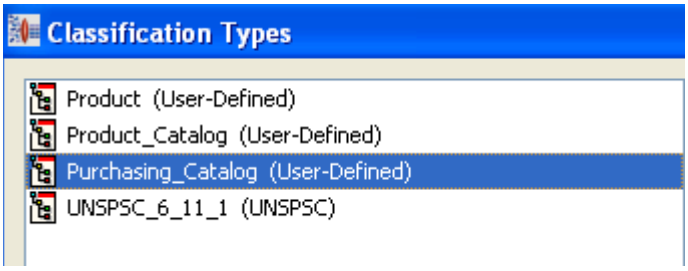
AutoBuild Summary Report provides two sets of graphs that provide a before and after view of the Semantic Knowledge and Text Recognition. The initial lens may be the DLS_Import_Template Smart Glossary or the target data lens as the starting point (before).

AutoBuild Details Report is a high level summary report showing the knowledge creation results. These reports should correctly represent the results of the AutoBuild knowledge building process and should correspond to the data lens knowledge.

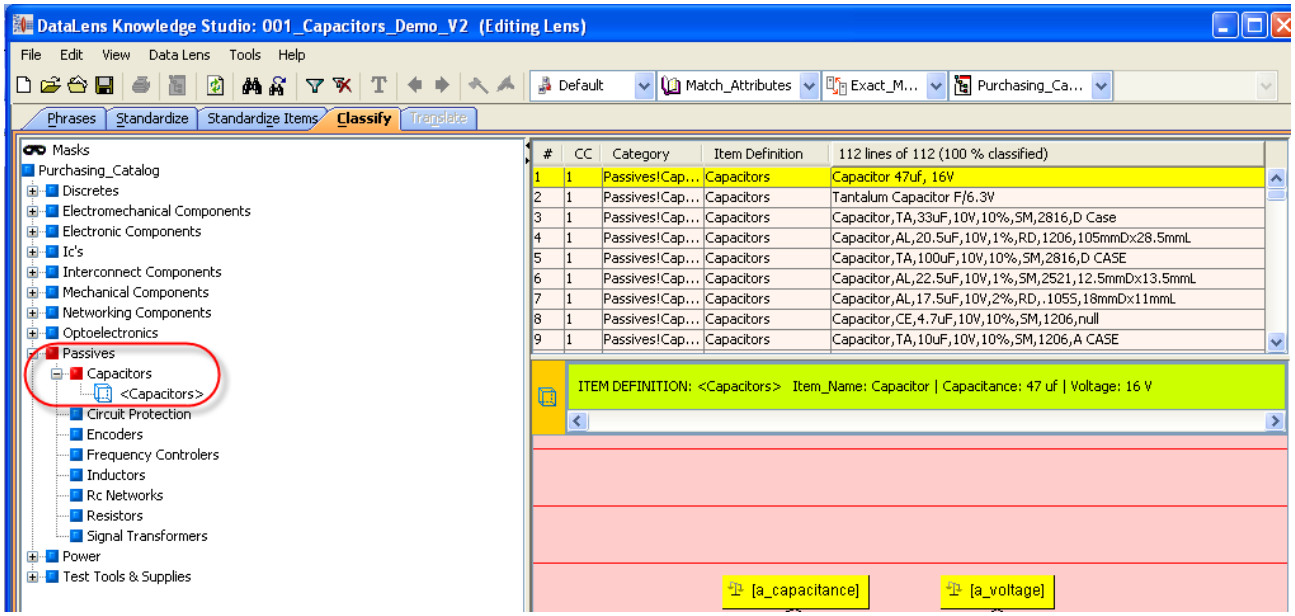
# 5. Review Alternate Catalog

You should review the data lens to confirm that the user-defined classification is correct. The first step is to determine if the alternate catalog is present in the data lens. The new alternate catalog will appear as a (User-Defined) type (for example, "Purchasing_Catalog). You can see this by clicking on the drop down from the toolbar or by selecting Classification Types from the DataLens menu.



The final step is to open up the Classify tab and relate the Semantic Model (Item Definition) to the Category. In this case, Capacitors is related to Capacitors to create an alternate catalog assignment. For more information about setting up and using Classification Types, see the *Oracle Product Data Quality Knowledge Studio Reference Guide*.



# Creating Semantic Key Cache

Oracle Product Data Quality uses patented Semantic technology to create a set of standardized attributes that are output to generate a set of master data or Semantic Key Cache. The cache represents a "semantic" index created from the extracted and standardized attribute information from the PIM.

The Semantic cache is used to rapidly match (compare) requested products to the items found in the cache. Any matches are returned to enable further processing such as form, fit or functional equivalents or duplicate detection to protect the PIM from the addition of identical records.

Open the Governance Studio and run the Update_Semantic_Cache job. It will ask for a Category as input. Enter the Category (for example, Capacitors). Once the Governance Studio job Update_Semantic_Cache runs, it will produce the Semantic Cache for that Category with typically no results shown in a Governance Studio tab for successful items. Two review tabs Unassigned_Semantic_Key and Unassigned_Item_Class are created to provide the user with information about exceptions.

At the end of the process, you will have a complete Semantic Key Cache. This is really an IT function and can only be confirmed by looking at the database. For more details, see the *Oracle Product Data Quality Information Management Implementation Guide*.

The process is typically run on a periodic basis (for example, nightly) to ensure that new updates to the PIM are included in the cache.

# Chapter 4

## Cleansing and Matching Batches

### Introduction

A newly created batch will be a status of "Active". When the batch is in process, the system will automatically change the batch status to "Pending" to ensure that the batch cannot be changed while it is in process. This is a critical step to ensure that the job can be correctly processed. The Governance Studio project will not produce any results and will generate an error on the Error Codes tab if the status is not set to "Active".

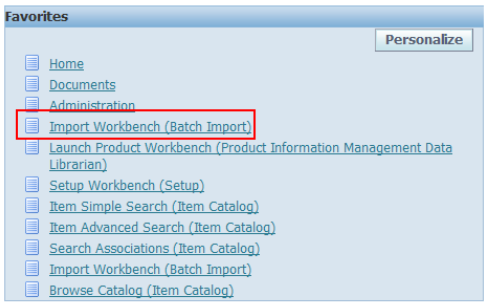
### 1. Create Batch

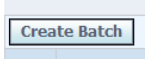
There are two types of batches that can be created for processing. The first type of batch is from an external source (outside the PIM). The second type of batch is a pull from production (within the PIM).

# External Data Source

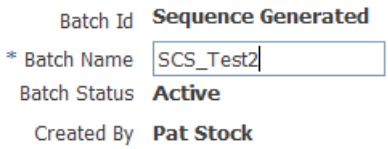
## *Create a Batch and import new items into the batch.*

1. From the Import Workbench…



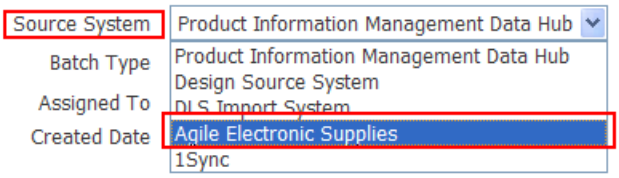

select "Create Batch" 

. Create a batch by entering a Batch Name




.

Choose "Agile Electronic Supplies" in the "Source System" dropdown…

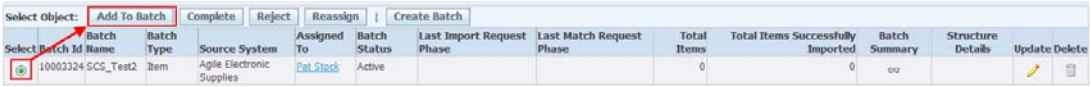



.

Accept the defaults and then select "Apply".

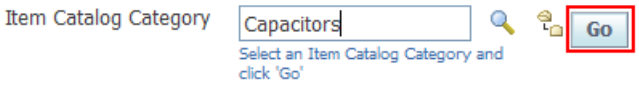
2. Search for the Batch so that you can add records. You can search by Batch Name, Id or even Created On date.

3. Select batch radio button and select "Add to Batch" to add records to the batch.

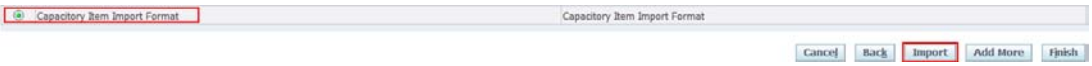



4. Select "Items" as the Object and press continue.

5. Enter "Capacitors" in the "Item Catalog Category" text field or look it up and press Go.




6. Select "Capacitory Item Import Format" and press "Import"




7. This brings up the EGO Bulkload Integrator using the Excel format to create the mechanism to import the records. Select the Excel format (for example, Excel 2003) and press "Next". Note: that you will probably warned from IE to allow the download of the file just below the tab in IE.

8. At this point, you will save the WebADI.xls spreadsheet (rename to the created batch id) to your computer.
9. Open the spreadsheet (Macros must be enabled), which will access Oracle system through Macros. You will get an Oracle dialog that provides status information as it creates the spreadsheet. The Oracle dialog will close automatically once it finishes.

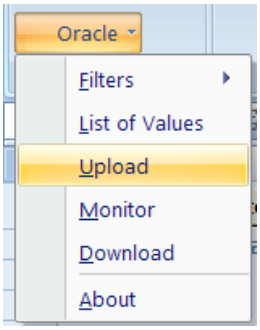
---

Note: You will need to *unprotect the sheet* to insert additional rows beyond the 10 available. The first lines of the spreadsheet have the information from the new batch creation. These include Batch Id, Batch Name and Source System. It generates the following spreadsheet format using the Agile Electronic Supplies format.

---

| **Column Header** | **Value** |
|---|---|
| Transaction Type | "SYNC" |
| Source System Reference | Unique Source Reference Id |
| Organization Code | "V1" |
| Item : Item | *Nothing should be entered (only used for existing PIM items* |
| Item : Description | Item description |

10. Use the Oracle menu Upload option to upload the rows to the newly created batch. It presents an Oracle upload dialog, keep the defaults, and select "Upload".




You will receive confirmation of a good upload and the number of rows uploaded.

11. From the Import Workbench, search for the batch by entering the Batch Id or date created and selecting "Go". This will bring up the active batches. Review the batch to confirm that the Total Items is now equal to the total uploaded.

| Select | Batch Id | Batch Name | Batch Type | Source System | Assigned To | Batch Status | Last Import Request Phase | Last Match Request Phase | Total Items | Total Items Successfully Imported | Batch Summary | Structure Details | Update | Delete |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ○ | 10003316 | SCS_NewBatch1 | Item | Agile Electronic Supplies | Pat Stock | Active | | | 40 | 0 | | | | |

12. Select the Update pencil (second to last column) to review the batch contents. They will appear in the "Confirmed" tab.

The new batch will have with a set of imported rows in the "Unconfirmed" tab because these are external to the PIM and must be validated (confirmed) by the user before an import is possible.

## Production Pull Data Source

### *Create a Batch from production PIM data.*

Create_PIMDH_Production_Batch Governance Studio project creates a new Batch, pulling data from the Production tables in the PIM.

1. Enter the name of the Item Category (for example, Capacitors).
2. When the job finishes, it will create a new batch with data from the Item Category entered in the first step.
3. Review the "Batch Creation Results" tab to see the results of the job run, and note or copy the new Batch Id or later use.
4. Review the newly created Batch in Oracle Batch Interface table. The new batch will use the following defaults:

| Field Name | Value |
|---|---|
| Batch Id | <New Batch Number> (System Generated) |
| Batch Name | DLS Production pull batch |
| Batch Type | Item |
| Source System | Product Information Management Data Hub |
| Assigned To | Albert Bakker |
| Batch Status | Completed |

Note: the batch is set to the "Active" status in order to allow the system to correctly process the batch; otherwise, you will receive an error.

The results will be an Oracle batch with a set of imported rows in the "Confirmed" tab created from the production PIM data for the selected item category. They are considered Confirmed because they are the PIM.

### *Process Interface Batch*

This section describes the capabilities of each of the Process Interface Batch tabs.

You will use the Batch Id to create a unique Governance Studio job to process the data pulled for the selected Category.

1. Open the DataLens_Cleansing_and_Matching or a previously created project (for example, DataLens_Cleansing_and_Matching-10003377).
2. Select the "New Project From" option from the File Menu. You will be prompted to enter a Unique Id (for example, Batch Number plus a version number), this is where you enter the previously created Batch number (for example, 10003377_V1).

3. Press the run button
4. Enter the batch id (for example, 10003377) from the previously created Batch number as input to the Governance Studio job.

Note: The difference between a creating a job with data pulled from the PIM and data from an external source is that there is not a concept of checking for duplicates so those tabs are always be blank.

Once the Governance Studio job DataLens_Cleansing_and_Matching runs, it will produce a series of output tabs that provide the user with a way to process the batch using a series of processing tabs. An individual item uses the following process logic.

1. Error Codes – If the batch is not correctly prepared or not in the correct state to be processed.
2. Release Batch – Provides a mechanism to change the Oracle batch from "Pending" to "Active", allowing a user to review the completed changes.
3. Exceptions – If the record falls below a Quality Index of 20 (nothing recognized) then it becomes an exception.
4. Dups Within Batch – The process checks for duplicates within the original source batch and presents them to the user to determine what should be done.
5. Match on Mfg Part – The process checks for duplicates by comparing the source manufacture name and part number against an existing record in the PIM.
6. Match Against PIM – The process checks for the source with the records in the PIM using the Semantic index cache to determine if the source is a duplicate to an existing record in the PIM.
7. Items for Enrichment – If a source record is not a duplicate and does not meet the Quality Index threshold then present it to the user to enrich the records so that they may be of the necessary completeness to load into the PIM.
8. Ready for Load – Fully complaint, standardized and enriched records ready to be imported into the PIM
9. Alternate Catalog Review – All fully standardized records (meets the Quality Index threshold) may have an alternate catalog (classification) assigned based on the Semantic Model definition – see previous section.
10. Style Matches – A final check against the PIM to determine if this item could be associated with an existing Style in PIM.

Note: Once the processing is complete, you will have to run Release Batch process from within Governance Studio job to reset the batch status from "Pending" to "Active" in order to review and confirm the changes made during this process.

## 2. Graph Summary

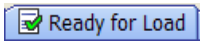
The Graph Summary provides a graphical summary of the state of the batch processing.

The Data Summary shows the totals by processing category. As you process items, the numbers will change and you can monitor this pane to verify the result. For example, if you enrich a record to the point where it can pass the quality threshold for import into the PIM, it will move from the Items for Enrichment category to the Confirmed New Review category.

The Classification Summary displays the total in a pie chart by Classification. If there is only one type then only one type will show.

The Completeness Assessment displays information about the attributes. It shows the coverage of the attributes across the record set. If a particular attribute has 100% coverage then that individual attribute count (completeness) will equal the record set total.

## 3. Ready for Load




The "Ready for Load" tab provides a view of the fully complaint, standardized and enriched records that met the quality requirements and are "import ready" and will be moved into the PIM through the interface table (batch) process. Once the user has reviewed and handled all of the items on the other tabs, bringing the totals to zero, the remainder will appear on the Confirmed New Review tab. The user checks from one to all of the rows and selects the green arrow to complete the process, which updates the Interface Batch with this set of standardized high quality data, using the Batch Id and the ITUI (Interface Table Unique Id) as the key.

1. Select the "Ready for Load" tab in the Governance Studio.
2. Click on from one to all of the rows in the tab.

| ✔ | interface_table_unique_id | organization_id | source_syste |
|---|---|---|---|
| ✔ | 67643 | 204 | Agile Electronic |
| ✔ | 67644 | 204 | Agile Electronic |
| ✔ | 67647 | 204 | Agile Electronic |
| ✔ | 67656 | 204 | Agile Electronic |
| ✔ | 67663 | 204 | Agile Electronic |
| ✔ | 67665 | 204 | Agile Electronic |
| ✔ | 67670 | 204 | Agile Electronic |

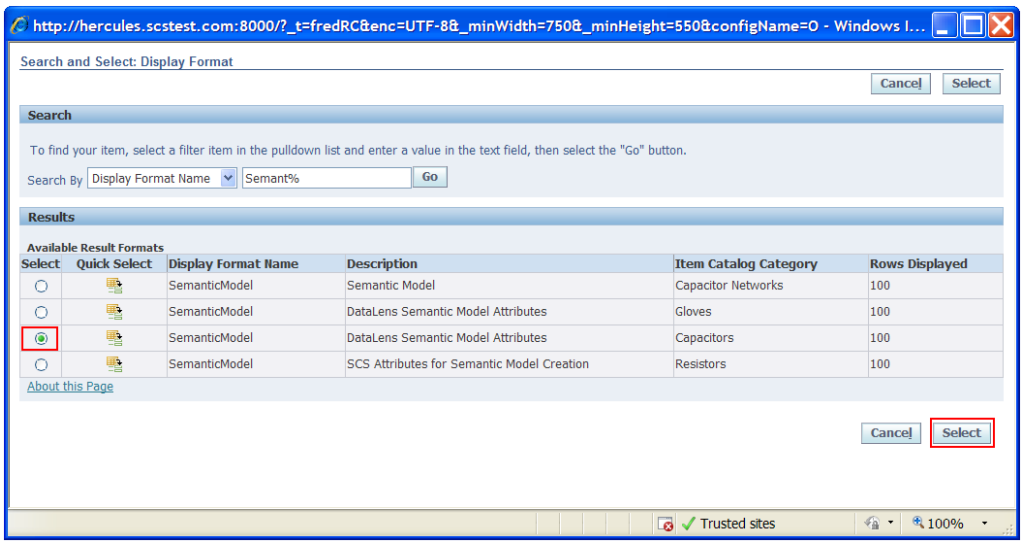
3. Click on the Green arrow to process the selected rows.
    a. The selected rows will turn blue, letting you know that the rows have been processed and that the rows are now unavailable for further processing.

4. Once the processing is complete, you will have to run the Release Batch process to reset the batch status from "Pending" to "Active" to review and confirm the changes made during this process. The job uses the batch id as input.
5. Confirm that the data is set correctly for the checked row(s) by going back into batch in the Oracle Item Import Workbench (Assumes batch set back to Active to review – see note in Introduction section).
    a. Search for the batch.
    b. Select "Update" to review the batch.

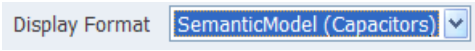
| Select | Batch Id | Batch Name | Batch Type | Source System | Assigned To | Batch Status | Last Import Request Phase | Last Match Request Phase | Total Items | Total Items Successfully Imported | Batch Summary | Structure Details | Update | Delete |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ◉ | 10003316 | SCS_NewBatch1 | Item | Agile Electronic Supplies | Pat Stock | Active | | | 40 | 0 | | | | |

    c. Select (for example, SemanticModel (Capacitors)) from Display Format drop down.

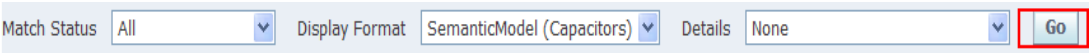
Note: I f this does not exist, select "More" and search for "SemanticModel". Select the correct "SemanticModel" (in the case where there are multiple and press "Select".



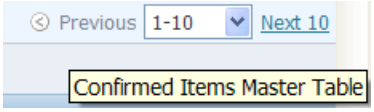

It will then appear in the list on the as shown above.




Select Go to apply it.



Use the previous and next buttons to review the items for correctness




The Standardized items are in place within the Oracle Interface Table for the checked rows from the Governance Studio job.

# 4. Alternate Catalog Review

This tab provides an opportunity to review and process the Alternate Catalog assignment at the item (row) level.

1. Select the "Alternate Catalog Review" tab in the Governance Studio.
2. Select from one to all of the rows in the tab.

| | interface_table_unique_id | organization_id | batch_id | item_number | Alt_Catalog_Name | Alt_Catalog_Category |
|---|---|---|---|---|---|---|
| ✓ | 67643 | 204 | 10003316 | KDL8298 | Purchasing Catalog | Capacitors |
| ✓ | 67644 | 204 | 10003316 | KDL8299 | Purchasing Catalog | Capacitors |
| ✓ | 67647 | 204 | 10003316 | BI10556 | Purchasing Catalog | Capacitors |
| ✓ | 67656 | 204 | 10003316 | BI10556 | Purchasing Catalog | Capacitors |
| ✓ | 67663 | 204 | 10003316 | KDL8301 | Purchasing Catalog | Capacitors |
| ✓ | 67665 | 204 | 10003316 | KDL8303 | Purchasing Catalog | Capacitors |
| ✓ | 67670 | 204 | 10003316 | KDL8308 | Purchasing Catalog | Capacitors |

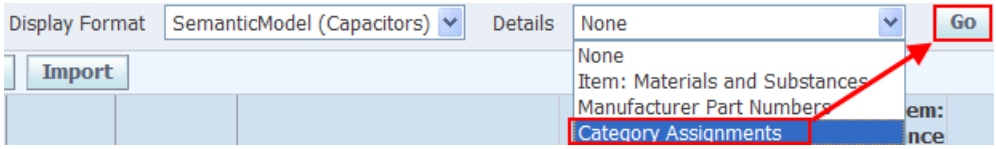
3. Click on the Green arrow to process the selected rows.

The selected rows will turn blue, letting you know that the rows have been processed.

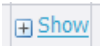
The rows become unavailable for further processing.

4. Once the processing is complete, you will have to run Release Batch process to reset the batch status from "Pending" to "Active" to review and confirm the changes made during this process. The job uses the batch id as input.
5. Confirm that the Catalog and Category are set correctly for the checked row(s) by going back into batch in the Oracle Item Import Workbench.
    a. Search for the batch.
    b. Select "Update" to review the batch.
    c. Select (for example, SemanticModel (Capacitors)) from Display Format drop down and select "Go" to apply it.
    d. Select "Category Assignments" from the Details drop down

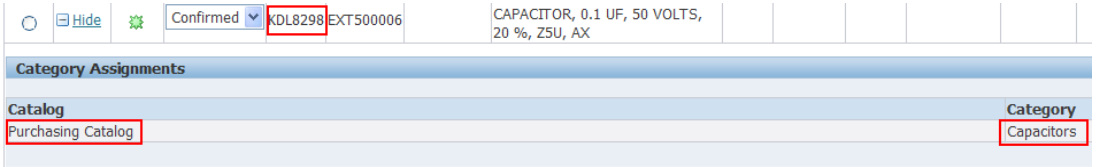
e. Select "Go" to apply it.




f. Choose "+Show" to open the row.

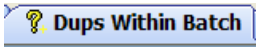



g. Confirm that the Catalog name (for example, "Purchasing Catalog") and Category (for example, "Capacitors") is in place for <u>all</u> of the rows you selected earlier in this section.

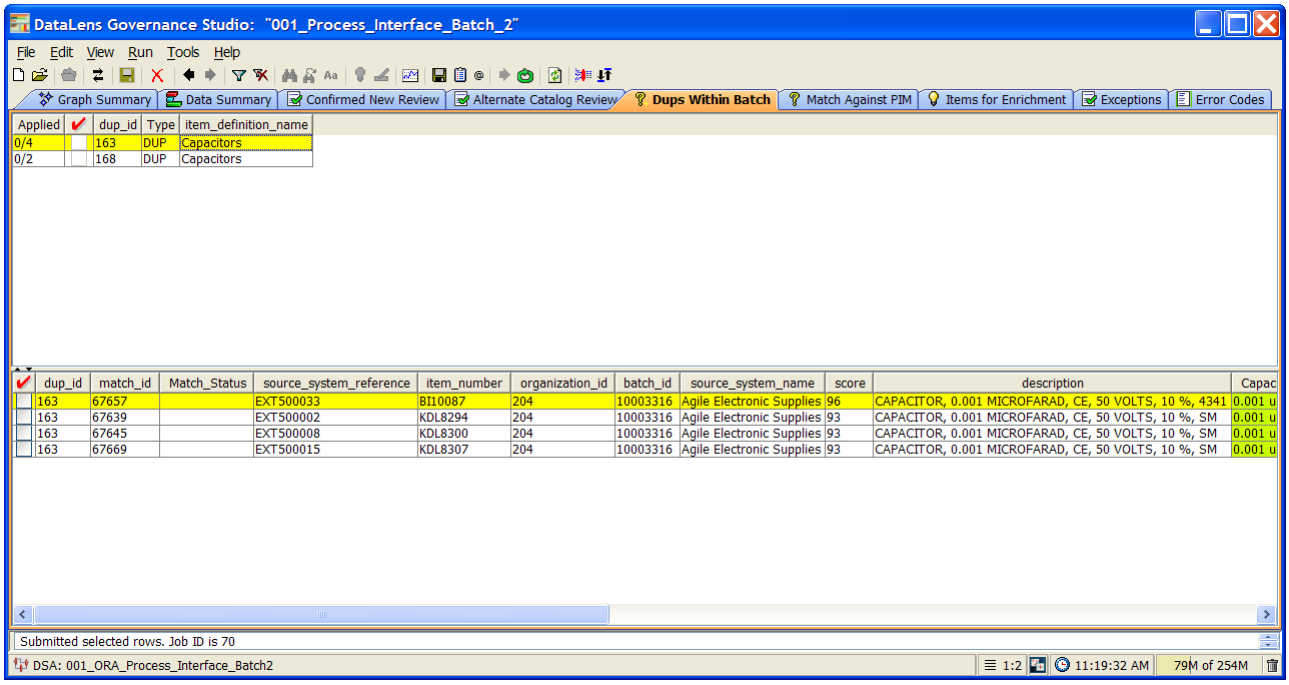



The Alternate Catalog is in place within the Oracle Interface Table for the checked rows from the Interface Batch from the Governance Studio job.

# 5. Dups Within Batch




The "Dups Within Batch" tab uses a split screen function that shows a set of record groups (rows) in the top pane, which are considered duplicates within the import batch. By selecting a row from the top pane, once a row is selected from the top pane, the system highlights the corresponding matched candidates for that row on the bottom pane.




The duplicate rows are showed in the bottom pane. The system also gives Data Stewards color coding to highlight key attributes participating in the match. The system will color the attributes based on the type of match rule that has been triggered on the attribute. Below is the breakdown of the color coding:

- **Green Attributes** – Denotes attributes that match to the parent row and that these attributes were set to required in the item specific match rules.
- **Blue Attributes** – Denotes attributes that match to the parent row and that were set to participating but not required in the item specific match rules.
- **Blank and un-highlighted** – Indicates that the matched row did not contain a value for that particular attribute and therefore that attribute did not participate in the match score.
- **Populated and un-highlighted** – Indicates that the matched row contains a value for that attribute but that the attribute value for the matched row did not match the attribute value for the parent row.

The bottom child pane view allows you to choose either a single override row or create a combined value (blended) row by choosing a primary row with values from other rows. This

survivorship process ends up with the one row or combined row that represents the set of values to be input into the system as the master record. The following steps should be used to create the master record:

1. Select the override row by right clicking on the row and selecting "Create Override Row" form menu option. After the selection, the following will occur:
    a. The override row will turn orange and the new master row will appear at the bottom on the set in red with a Match Status of "MO" (Manual Override). The override row checkbox will be checked and disabled.
    b. All other rows will be automatically checked, but available to uncheck to exclude from processing.

| ✓ | dup_id | match_id | Match_Status | source_system_reference | item_number | organization_id | batch_id | source_system_name | score | description | Capac |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ☐ | 163 | 67657 | | EXT500033 | BI10087 | 204 | 10003316 | Agile Electronic Supplies | 96 | CAPACITOR, 0.001 MICROFARAD, CE, 50 VOLTS, 10 %, 4341 | 0.001 u |
| ☐ | 163 | 67639 | | EXT500002 | KDL8294 | 204 | 10003316 | Agile Electronic Supplies | 93 | CAPACITOR, 0.001 MICROFARAD, CE, 50 VOLTS, 10 %, SM | 0.001 u |
| ☐ | 163 | 67645 | | EXT500008 | KDL8300 | 204 | 10003316 | Agile Electronic Supplies | 93 | CAPACITOR, 0.001 MICROFARAD, CE, 50 VOLTS, 10 %, SM | 0.001 u |
| ☐ | 163 | 67669 | | EXT500015 | KDL8307 | 204 | 10003316 | Agile Electronic Supplies | 93 | CAPACITOR, 0.001 MICROFARAD, CE, 50 VOLTS, 10 %, SM | 0.001 u |
| ✓ | 163 | 67657 | MO | EXT500033 | BI10087 | 204 | 10003316 | Agile Electronic Supplies | 96 | CAPACITOR, 0.001 MICROFARAD, CE, 50 VOLTS, 10 %, 4341 | 0.001 u |

2. Optionally, right click an individual cell from any of the other rows to create an override cell by right clicking from the desired cell and selecting "Set Override Cell". The override cell will turn orange and that row will be automatically checked (if unchecked) and disabled while the cell in the original override row will be deselected and return to the original color. Finally, the value will appear in the master row. Note: if you want to enter a value, you must first enter the value prior to selecting that cell as an override cell; otherwise the change will not be reflected in the master row.

| Capacitance | Dielectric | Voltage | Tolerance1 | Capacitor_package_size |
|---|---|---|---|---|
| 0.001 uF | CE | 50 VLT | 10 % | 4341 |
| 0.001 uF | CE | 50 VLT | 20 % | |
| 0.001 uF | CE | 50 VLT | 10 % | |
| 0.001 uF | CE | 50 VLT | 10 % | |
| 0.001 uF | CE | 50 VLT | 20 % | 4341 |

The user can also change the override cell selection by right clicking an orange highlighted cell and selecting another cell with "Set Override Cell". Users can select an unlimited number of override cells to create the blended record. To remove an override cell, the user needs to right click the cell from the red row and select "Set Override Cell". The user can also remove the override row selections by right clicking any row and selecting "Remove Override Row". Selecting "Remove Override Row", resets the entire duplicate set and returns the set to the original state.

Once the user completes the selection process, the user selects the Re-run button to process the row(s).

**Expected Result:** The Override row (orange / red row) IS reprocessed as a new row and could end up back in any of the tabs, but it is very unlikely that it will end up in the Enrich or Exception tabs because in order to participate in the Dedup process, the QI needed to be above the threshold. The other checked rows in the set, which Match Status is not changed by the user, should be updated in the interface tables with a match status of Ignore.

The user may also override the system by entering a match status and sending the rows to be reprocessed. They may be one of the following:

1. Confirmed New: After reviewing the information, the user may conclude that although the system thinks these are duplicates, they are not. The row(s) will be included as new items in the PIM.
2. Unconfirmed: (default value) - The user must confirm the match status in order to import the item.
3. Ignore / Blank: Temporarily hold this item from import. The next time the same item data is uploaded within a batch, it appears in the Excluded tab for that batch.
4. Exclude: The row(s) will *not* be included as new items in the PIM and will be moved to an excluded region so that if the same item data row is loaded, it will be automatically moved to the Excluded tab.
5. Reprocess: Allow the user to update the data so that the row can be reprocessed and removed as a duplicate.

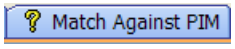
Once the user completes the selection process, the user selects the Re-run button to process the row(s). The only instance that the data in the PIM will actually be enriched is when the user selects the "Confirmed New" Match Status. The system will ignore all other rows that the user enriched (for example, unselected, and selected and any match type including blank).

Processing steps:

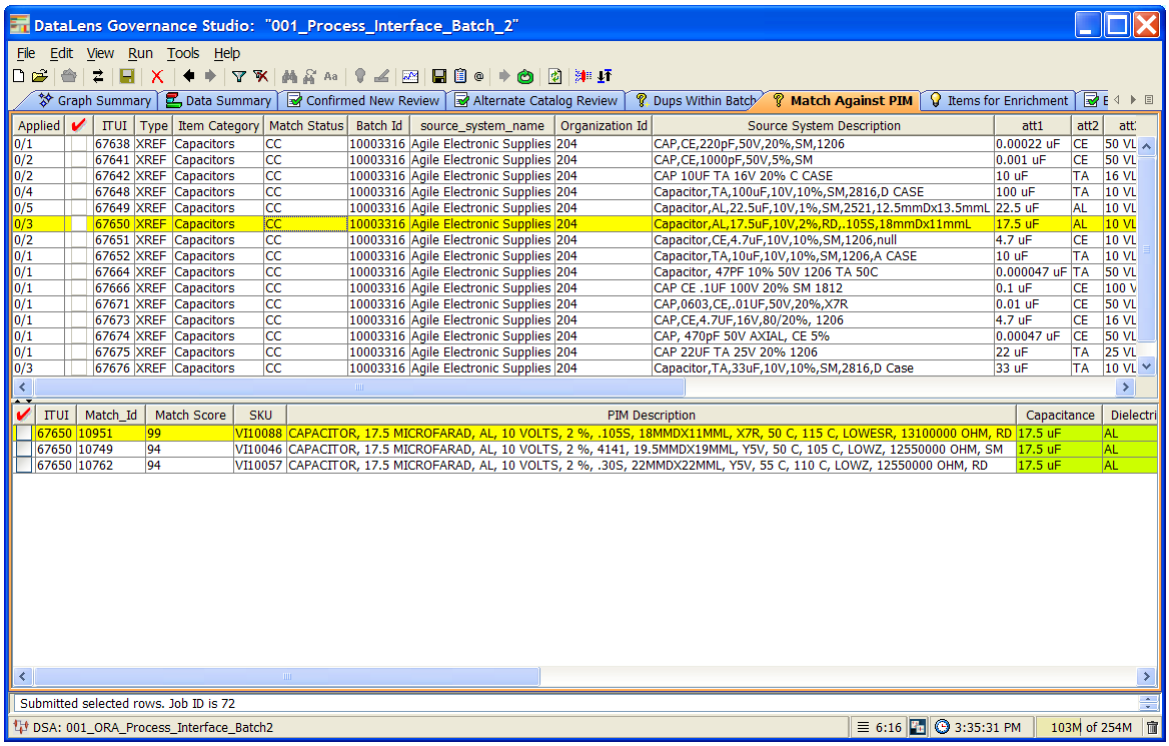
1. Select the "Dups Within Batch" tab in the Governance Studio.
2. Make the process selections as described above.
3. Click on the Green Re-run button to process the selected rows.
4. For duplicate processing, confirm that the override row appears in one of the Process Interface Batch Governance Studio job tabs. If the meets the quality standards and completeness, it will appear in the Confirmed New Review tab; otherwise, it may appear Items for Enrichment tab or Match Against PIM tab.
5. For override processing, once the processing is complete, you will have to run Release Batch process to reset the batch status from "Pending" to "Active" to review and confirm the changes made during this process. The job uses the batch id as input.
6. Confirm that the data is set correctly for the checked row(s) by going back into batch in the Oracle Item Import Workbench (Assumes batch set back to Active to review – see note in Introduction section).
    a. Search for the batch.
    b. Select "Update" to review the batch.
    c. Select (for example, SemanticModel (Capacitors)) from Display Format drop down and select Go and select Apply.
    d. Use the previous and next buttons to review the items for correctness

The override row is present on one of Process Interface Batch Governance Studio job tabs or for overridden rows in the Oracle Batch with the correct status.

# 6. Match Against PIM




The "Match Against PIM" tab uses a split screen function that shows a set of record groups (rows) in the top pane from the incoming records which are functionally equivalent (duplicate) with record(s) in the Oracle Product Data Hub. It allows the system to identify these incoming records as a cross-reference to an existing Oracle Product Data Hub record. Once a row is selected from the top pane, the system highlights the corresponding matched candidates for that row on the bottom pane. The user selects the row in the bottom pane that is the preferred match for the cross-reference.




| Applied | ✔ | ITUI | Type | Item Category | Match Status | Batch Id | source_system_name | Organization Id | Source System Description | att1 | att2 | att |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0/1 | | 67638 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP,CE,220pF,50V,20%,SM,1206 | 0.00022 uF | CE | 50 VL |
| 0/2 | | 67641 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP,CE,1000pF,50V,5%,SM | 0.001 uF | CE | 50 VL |
| 0/2 | | 67642 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP 10UF TA 16V 20% C CASE | 10 uF | TA | 16 VL |
| 0/4 | | 67648 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | Capacitor,TA,100uF,10V,10%,SM,2816,D CASE | 100 uF | TA | 10 VL |
| 0/5 | | 67649 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | Capacitor,AL,22.5uF,10V,1%,SM,2521,12.5mmDx13.5mmL | 22.5 uF | AL | 10 VL |
| 0/3 | | 67650 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | Capacitor,AL,17.5uF,10V,2%,RD,.105S,18mmDx11mmL | 17.5 uF | AL | 10 VL |
| 0/2 | | 67651 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | Capacitor,CE,4.7uF,10V,10%,SM,1206,null | 4.7 uF | CE | 10 VL |
| 0/1 | | 67652 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | Capacitor,TA,10uF,10V,10%,SM,1206,A CASE | 10 uF | TA | 10 VL |
| 0/1 | | 67664 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | Capacitor, 47PF 10% 50V 1206 TA 50C | 0.000047 uF | TA | 50 VL |
| 0/1 | | 67666 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP CE .1UF 100V 20% SM 1812 | 0.1 uF | CE | 100 V |
| 0/1 | | 67671 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP,0603,CE,.01UF,50V,20%,X7R | 0.01 uF | CE | 50 VL |
| 0/1 | | 67673 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP,CE,4.7UF,16V,80/20%, 1206 | 4.7 uF | CE | 16 VL |
| 0/1 | | 67674 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP, 470pF 50V AXIAL, CE 5% | 0.00047 uF | CE | 50 VL |
| 0/1 | | 67675 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | CAP 22UF TA 25V 20% 1206 | 22 uF | TA | 25 VL |
| 0/3 | | 67676 | XREF | Capacitors | CC | 10003316 | Agile Electronic Supplies | 204 | Capacitor,TA,33uF,10V,10%,SM,2816,D Case | 33 uF | TA | 10 VL |

| ✔ | ITUI | Match_Id | Match Score | SKU | PIM Description | Capacitance | Dielectri |
|---|---|---|---|---|---|---|---|
| | 67650 | 10951 | 99 | VI10088 | CAPACITOR, 17.5 MICROFARAD, AL, 10 VOLTS, 2 %, .105S, 18MMDX11MML, X7R, 50 C, 115 C, LOWESR, 13100000 OHM, RD | 17.5 uF | AL |
| | 67650 | 10749 | 94 | VI10046 | CAPACITOR, 17.5 MICROFARAD, AL, 10 VOLTS, 2 %, 4141, 19.5MMDX19MML, Y5V, 50 C, 105 C, LOWZ, 12550000 OHM, SM | 17.5 uF | AL |
| | 67650 | 10762 | 94 | VI10057 | CAPACITOR, 17.5 MICROFARAD, AL, 10 VOLTS, 2 %, .30S, 22MMDX22MML, Y5V, 55 C, 110 C, LOWZ, 12550000 OHM, RD | 17.5 uF | AL |

The matched row(s) are showed in the bottom pane. The system also gives Data Stewards color-coding to highlight key attributes participating in the match. The system will color the attributes based on the type of match rule that has been triggered on the attribute. Below is the breakdown of the color-coding:

- **Green Attributes** – Denotes attributes that match to the parent row and that these attributes were set to required in the item specific match rules.
- **Blue Attributes** – Denotes attributes that match to the parent row and that were set to participating but not required in the item specific match rules.
- **Blank and un-highlighted** – Indicates that the matched row did not contain a value for that particular attribute and therefore that attribute did not participate in the match score.

- **Populated and un-highlighted** – Indicates that the matched row contains a value for that attribute but that the attribute value for the matched row did not match the attribute value for the parent row.

The user selects a row in the top pane to see a set of candidate row(s) in the bottom pane to create a cross reference row. Once the cross reference row is checked, the user can send the cross reference row to the Oracle Interface Table.

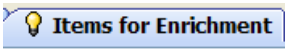
Note: the ITUI (Interface Table Unique) ID / Match_Id relationship is used for the cross reference. The Match Status is set to "Confirmed Match" which denotes a cross reference to an existing item in the PIM.

The system will automatically enter the PIMDH Item number (for example, VI10585) in the PIMDH Item field and will set the confirm status to "CM" (Confirmed Match).

Once the user completes the selection process, the user selects the Re-run button to process the row(s).

The cross reference appears in the batch. Note: the cross reference row(s) are removed from the list and the only way to confirm is by reviewing the Oracle batch.

# 7. Items for Enrichment




The "Items for Enrichment" tab with AutoSuggest output is identified by the Lightbulb icon to the left of the tab name. Data in these tabs may access the Governance Studio AutoSuggest feature that automatically predicts attributes that may have been missed due to misspellings and abbreviations or missing grammar. When this tab is not empty and there are suggestions available the "Suggest values for missing attributes" Lightbulb in the button bar above the output tabs will be visible. If no suggestions were found the icon in the button bar will be grayed out.

After clicking on an AutoSuggest output tab you will also see a list of rows of data that were designated for output by the respective output step in the DSA. To view suggestions click on the AutoSuggest Lightbulb in the button bar. Suggestions are now highlighted in green.

Several operations are allowed in the AutoSuggest table to process your data further and make it easier to edit, move, copy, or accept suggestions. In addition to the normal table functions available to output tables, an AutoSuggest table allows you to accept or reject suggestions by double-clicking over them. If a user double clicks a suggestion then the cell is highlighted in pink showing that the attribute has been rejected.

When you have completed reviewing the suggestions by rejecting the bad ones, keeping the good ones and editing the cells to add (enrich) the data, you may check the rows that you want to reprocess by clicking on the row and clicking on the re run icon. After enriching, the item will follow the standard process and end up in one of the tabs depending on the outcome (for example, If the enrichment was sufficient then it will be of high enough quality to appear in the Confirmed New Review tab).

Select the "Items for Enrichment" tab in the Governance Studio.

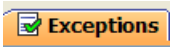
Make the process selections as described above.

Click on the Green Re-run button to process the selected rows.

Confirm that the enriched row appears in one of the Process Interface Batch Governance Studio job tabs. If the meets the quality standards and completeness, it will appear in the Confirmed New Review tab; otherwise, it may appear Items for Enrichment tab or Match Against PIM tab.

The enriched record is present on one of Process Interface Batch Governance Studio job tabs.

# 8. Exceptions

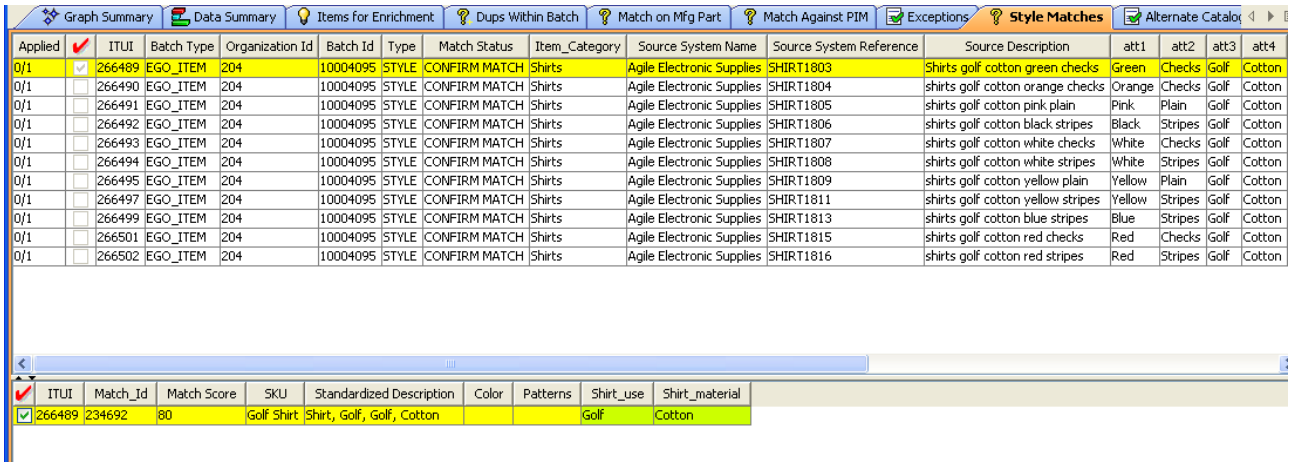



The Exceptions tab presents the users with items that were unrecognized (did not meet the quality threshold). The user has the opportunity to review and / or edit the description then make a decision about what to do with the item.

1. Select the "Exceptions" tab in the Governance Studio.
2. Make a Match Status selection from the following set of selections:
    a. IGNORE – Ignore the row
    b. EXCLUDE – Exclude the row
    c. CONFIRMED NEW – Confirm new row (allow into PIM)
    d. UNCONFIRMED – Unconfirmed row that must be manually reviewed by the user in the Oracle Batch.
    e. REPROCESS – Reprocess the row
3. Click on the Green Re-run button to process the selected rows.
4. Confirm that the row appears in the Process Interface Batch with the correct status or in one of the Process Interface Batch Governance Studio job tabs if reprocessed. If the meets the quality standards and completeness, it will appear in the Confirmed New Review tab; otherwise, it may appear "Items for Enrichment" tab or "Match Against PIM" tab.

# 9. Style Matches

The "Style Matches" tab provides a view of the records that match an existing record Style Item in the PIM based on a particular set of Style attributes. The tab uses a split screen function that shows a set of record groups (rows) in the top pane from the incoming records, which match an existing Style in the Oracle Product Data Hub. Once a row is selected from the top pane, the system highlights the corresponding matched candidate for that row on the bottom pane. There should be only one Style match if the system is setup correctly. The user selects the row in the bottom pane that is the match.




| Applied | ✔ | ITUI | Batch Type | Organization Id | Batch Id | Type | Match Status | Item_Category | Source System Name | Source System Reference | Source Description | att1 | att2 | att3 | att4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0/1 | ☑ | 266489 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1803 | Shirts golf cotton green checks | Green | Checks | Golf | Cotton |
| 0/1 | ☐ | 266490 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1804 | shirts golf cotton orange checks | Orange | Checks | Golf | Cotton |
| 0/1 | ☐ | 266491 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1805 | shirts golf cotton pink plain | Pink | Plain | Golf | Cotton |
| 0/1 | ☐ | 266492 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1806 | shirts golf cotton black stripes | Black | Stripes | Golf | Cotton |
| 0/1 | ☐ | 266493 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1807 | shirts golf cotton white checks | White | Checks | Golf | Cotton |
| 0/1 | ☐ | 266494 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1808 | shirts golf cotton white stripes | White | Stripes | Golf | Cotton |
| 0/1 | ☐ | 266495 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1809 | shirts golf cotton yellow plain | Yellow | Plain | Golf | Cotton |
| 0/1 | ☐ | 266497 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1811 | shirts golf cotton yellow stripes | Yellow | Stripes | Golf | Cotton |
| 0/1 | ☐ | 266499 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1813 | shirts golf cotton blue stripes | Blue | Stripes | Golf | Cotton |
| 0/1 | ☐ | 266501 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1815 | shirts golf cotton red checks | Red | Checks | Golf | Cotton |
| 0/1 | ☐ | 266502 | EGO_ITEM | 204 | 10004095 | STYLE | CONFIRM MATCH | Shirts | Agile Electronic Supplies | SHIRT1816 | shirts golf cotton red stripes | Red | Stripes | Golf | Cotton |

| ✔ | ITUI | Match_Id | Match Score | SKU | Standardized Description | Color | Patterns | Shirt_use | Shirt_material |
|---|---|---|---|---|---|---|---|---|---|
| ☑ | 266489 | 234692 | 80 | Golf Shirt | Shirt, Golf, Golf, Cotton | | | Golf | Cotton |

The matched row is shown in the bottom pane. The system also gives Data Stewards color coding to highlight key attributes participating in the match. The system will color the attributes based on the type of match rule that has been triggered on the attribute. Below is the breakdown of the color coding used on this tab:

- **Green Attributes** – Denotes attributes that match to the parent row and that these attributes were set to required in the item specific match rules.

The user selects a row in the top pane to see a set of candidate row in the bottom pane to apply the Style. Once the Style row is checked, the user can send the Style row to the Oracle Interface Table.

> Note: the ITUI (Interface Table Unique) ID / Match_Id relationship is used for the match cross reference. The Match Status is set to "Confirmed Match" which sets the Style information for that record.

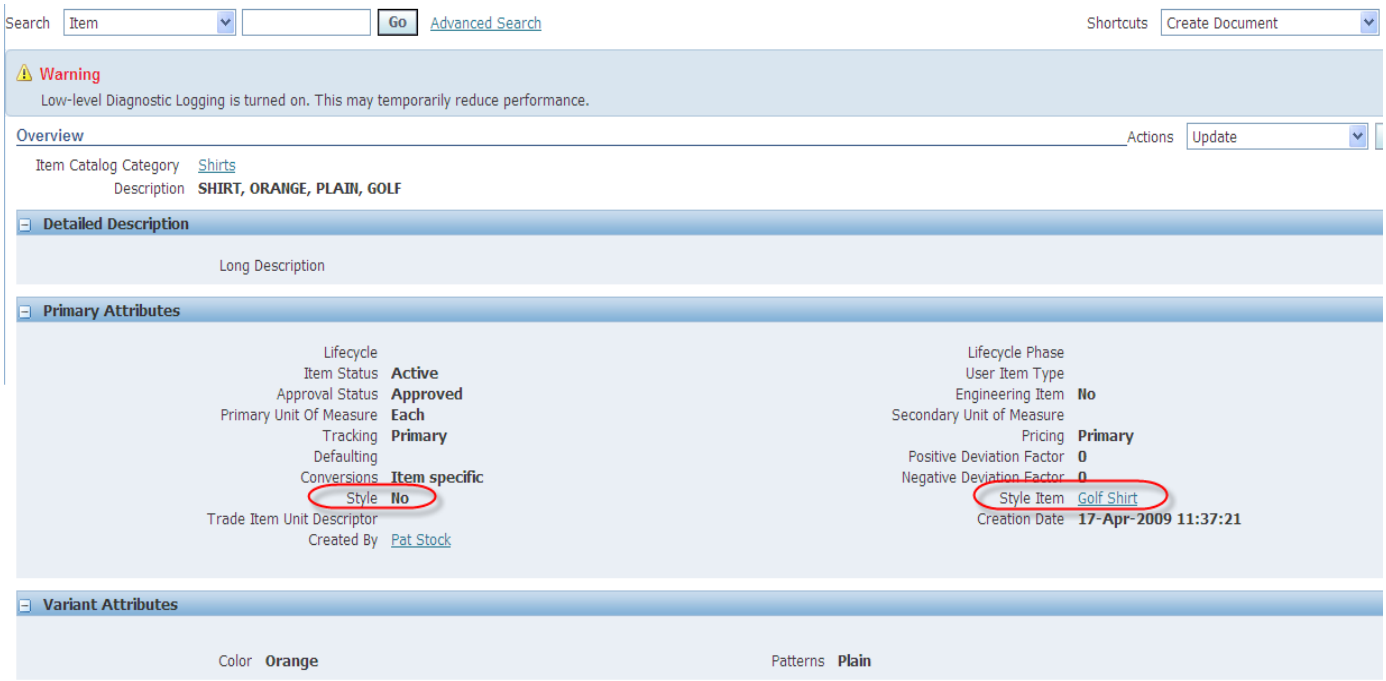
Once the user completes the selection process, the user selects the green arrow button to process the selected row(s).

The item will get an internal field called style_item_id set to the match_id in the batch.

Note: the cross reference row(s) are removed from the list and the only way to confirm is by reviewing the item in PIM.

After the matches are applied, the processed rows will be disabled and turn blue. Once the row(s) are imported, the changes can be verified by reviewing the items have been linked to the matched style as shown below.

Note: There will be no visible indication in the Import Workbench UI about the Style association.




# 10. Release Batch

The "Release Batch" tab allows the user to reset the batch from "Pending" to "Active" so the Oracle batch can be reviewed and imported. It also inactivates the current Governance Studio job to save the contents and prevent changes.

1. Select the "Release Batch" tab in the Governance Studio.
2. Click on the row.
3. Click on the Green check button to process the selected row.

It will complete the current batch in the Governance Studio and set the project to read only (disabling most of the interface functions) and set from "Pending" to "Active". Note: to reactive the batch, use the File – "New Project From" to reopen and process using a version number and the batch number as input.

# 11. Error Codes

The Error Codes tab surfaces any errors from running the job. A typical error would be trying to run the job against a batch with an Active status when it should have been a Completed status.

# Appendix A

## Installing the Client Software

Oracle Product Data Quality uses a concept called Java Web Start 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.

You download and install the Oracle Product Data Quality client applications using Java Web Start by browsing to the installation page for your Oracle DataLens Server as follows:

1. Using Microsoft Internet Explorer, browse to one of the following URLs as appropriate for your server:

   Note: If you setup a different port number for your application server other than 2229, you must use that port number in the following URL when browsing to the Oracle DataLens Server to download the client applications.

   **32-bit**

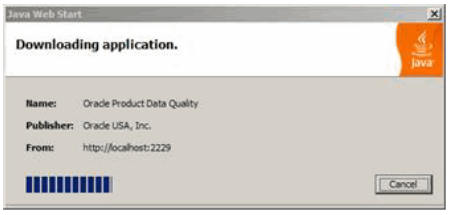
   http://<server>:2229/datalens/datalens.html

   **64-bit**

   http://<server>:2229/datalens/datalens64.html

   Where <*server*> is the hostname of the Oracle DataLens Server

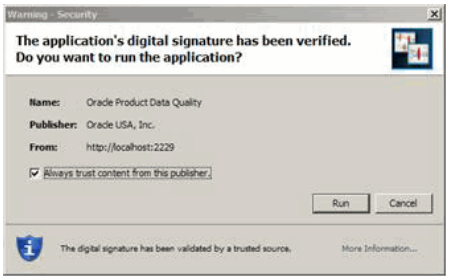
The application download and installation begins. If you do not have a supported Java environment on the target installation machine the Java Web Start program automatically redirects you to a Java download site and begins a Java Runtime installation.



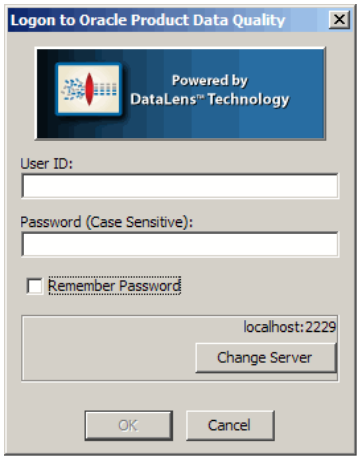

2. If the preceding Java Web Start message is not displayed, you must initiate a connection and download the software by browsing to:

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

   Oracle Product Data Quality files are digitally signed by a trusted source so the following security warning is displayed.




3. To avoid the security dialogue in the future you can select the **Always trust content from this publisher** check box.

4. Click **Run** to continue and complete the installation.

   The Oracle Product Data Quality log on dialog is displayed.