

Oracle® Product Data Quality
Services for Excel Reference Guide
Version 5.5

August 2010

ORACLE®

Oracle Product Data Quality Services for Excel 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
<i>Intended Audience.....</i>	<i>v</i>
<i>Conventions Used in This Book.....</i>	<i>v</i>
<i>Related Information.....</i>	<i>v</i>
Chapter 1	1
Introduction	1
Toolbar and Menus Explained.....	3
<i>Jobs Menu</i>	<i>4</i>
<i>DataLens Tools Menu</i>	<i>6</i>
<i>DataLens Help Menu</i>	<i>8</i>
<i>Excel Cell Context-Sensitive Menu</i>	<i>8</i>
Chapter 2	11
Configuring the Server and Application.....	11
Adding a New Oracle DataLens Server.....	13
Updating or Removing an Oracle DataLens Server.....	15
Logging Into Oracle Product Data Quality.....	17
Application Options.....	18
Chapter 3	20
Transforming Your Data	20
Configuring a DSA.....	21
<i>Data Service Settings Tab.....</i>	<i>22</i>
<i>Advanced Settings Tab.....</i>	<i>24</i>
<i>Setting Up a DSA Job.....</i>	<i>26</i>
Unique Record Identifiers	27
<i>Creating the Unique Identifier Column.....</i>	<i>27</i>
<i>DSA Outputs.....</i>	<i>28</i>

The DSA Output Named <i>Output</i>	28
Running a DSA to Process Records.....	29
<i>Using Real Time Records Processing</i>	29
<i>Using Batch Mode Records Processing</i>	31
<i>Removing Result Data</i>	33
Viewing the Job Status.....	34
Chapter 4	36
Getting More from Your Data	36
Consolidating Excel Files	37
Changing and Grouping Data	40
<i>Filling Columns with Data</i>	40
<i>Character Analysis and Replacement</i>	41
Analyzing the Character Set	41
Replacing Control Characters.....	41
Searching for Character Hex Values.....	41
Removing Embedded HTML Tags.....	42
<i>Grouping Records</i>	43
Using Regression Testing.....	45
<i>Creating or Updating the Regression Set</i>	45
Adding Cells to the Regression Set	46
<i>Comparing Worksheets</i>	47
<i>Comparing Selected Records</i>	49
<i>Comparing Workbooks</i>	49
<i>Compare Job Results</i>	51
Data Sampling	52
Source Data Backup	53
Merging Data Lens Standardizations	54
Oracle DataLens AutoBuild Application.....	58
Chapter 5	59
Installing the Software.....	59
Installing the Oracle Product Data Quality Client Software.....	60
Installing Services for Excel	62
<i>Software Prerequisites</i>	62
<i>Hardware Prerequisites</i>	62

<i>Oracle DataLens Server Prerequisites</i>	62
Required Oracle DataLens Server Software.....	62
Optional DataLens Application Software	63
<i>Installing the Application</i>	63
Installing on Windows 32-bit:	63
Installing on Windows 64-Bit	65
Restoring Installation Settings from a Previous Installation	68

Preface

About this Book

This reference guide is intended to explain the basic capabilities of the Oracle Product Data Quality Services for Excel.

To understand all of the advanced features presented, you must use this reference 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 including all of the sample files.

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

Oracle Product Data Quality Application Studio Reference Guide

In addition, Oracle Product Data Quality Services for Excel training is encouraged.

Intended Audience

You should have a basic understanding of the Oracle Product Data Quality Technology, including the functionality of the Oracle Product Data Quality Knowledge Studio and the Oracle Product Data Quality Application Studio applications.

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

- Customers
- Sales Consultants
- Implementation Personnel
- Software Engineers
- Knowledge Engineers

Conventions Used in This Book

The following typographical conventions 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 ease of use information.

Note: Indicates additional or supplemental information.

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 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 Server 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 Guide* 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

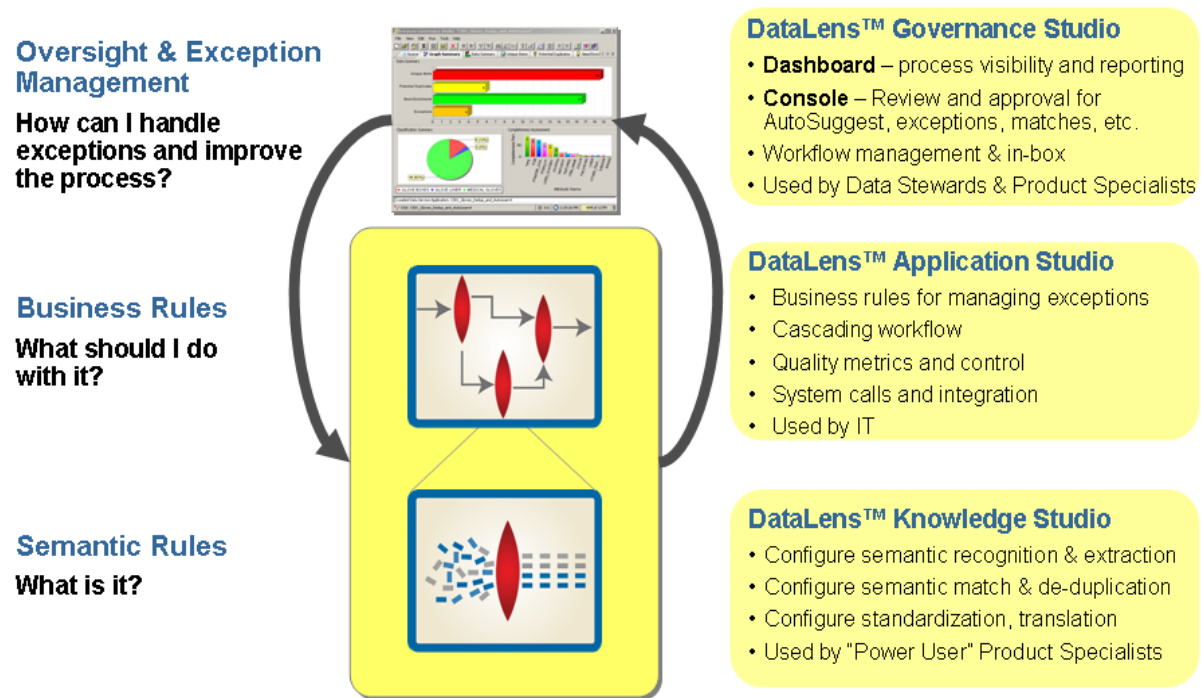
In this chapter

- Toolbar and Menus Explained

3

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.



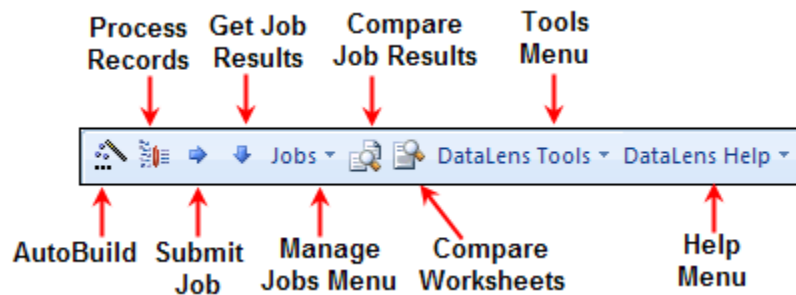
The Oracle Product Data Quality Services for Excel application is designed to provide enhanced, tailored, spreadsheet-based transformations of your data and is used in conjunction with your Oracle DataLens Server platform. Services for Excel enables you to make your corporate data business-ready using Microsoft Excel spreadsheet software.

Services for Excel interfaces directly to the Oracle DataLens Server to execute Data Service Applications (DSAs). This is just one way that DSAs can be used to standardize, match, de-duplicate, and convert enterprise data.

The Services for Excel application is designed to leverage the power of the entire suite of products delivered in Oracle Product Data Quality. It is set of powerful tools and commands installed into Excel and accessed using the custom toolbar and context-sensitive menu. For installation information, see Installing the Software on page 59.

Toolbar and Menus Explained

The following briefly describes the Services for Excel toolbar buttons and menus from left to right:



Note: The **AutoBuild** button is only visible when the appropriate Preference option is set. For more information, see [Configuring the Server and Application on page 11](#).

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

The buttons that are on the toolbar are the most often used functions so are available for quick access; each buttons is on one of the menus as well.

Jobs Menu

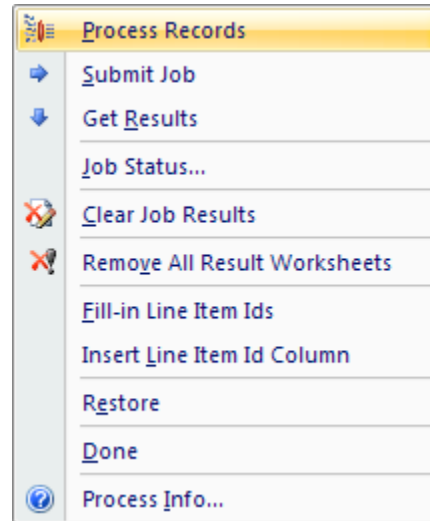
The **Jobs** button and associated menu items provide a set of convenient features for running, monitoring, and managing jobs.

Process Records

Executes a DSA job immediately and runs it synchronously. A job progress dialog is displayed during execution and you are notified when the job is complete. For more information, see Running a DSA to Process Records on page 29.

Submit Job

Starts a background job using the configured DSA. After a job has been submitted, you can save the Excel workbook, close Excel, and then reopen the same workbook later to retrieve the job results. You can retrieve job results only once.



Get Results

Retrieves the results from a submitted DSA jobs. If the job has not run to completion, a pop-up message informs you that the job is still running. If no job is running or is complete, then you are prompted for the job id to use when retrieving results.

Job Status...

Use to view the status and associated details of all your completed and currently executing jobs. For more information, see [Viewing the Job Status](#) on page 34.

Clear Job Results

Clears the data results across all output spreadsheets. All record data is deleted. The formatting for these spreadsheets is not affected.

Remove All Result Worksheets

Removes all of the result worksheets produced by the currently selected DSA; the input worksheet is not affected. You are prompted to confirm the deletion of the output worksheets.

Fill-in Line Item Ids

Analyzes the values in column A for each record to determine whether there is a valid unique id present. For more information, see [Unique Record Identifiers](#) on page 27.

Insert Line Item Id Column

Adds an identifier column to your input data. When selected, a new column A is inserted in front of all other existing columns, the column is titled *Id*, and a set of unique identifiers is added to this column. For more information, see [Unique Record Identifiers](#) on page 27.

Restore

Restores the source data spreadsheet to its prior state if a hidden backup worksheet is available. For more information, see [Source Data Backup](#) on page 53.

Done

Removes all hidden backup worksheets and Data Services for Excel configuration worksheets. You are prompted to confirm the deletion of these worksheets. For more information, see [Source Data Backup](#) on page 53.

Process Info...

Provides information on the currently configured DSA and any outstanding job results that have not been retrieved.

DataLens Tools Menu

The **Tools** button and associated menu items provides a selection of tools that the DataLens expert can use to work with results. The data profiling button and associated menu items provide a set of features for creating and managing profiling reports

Create Regression Set

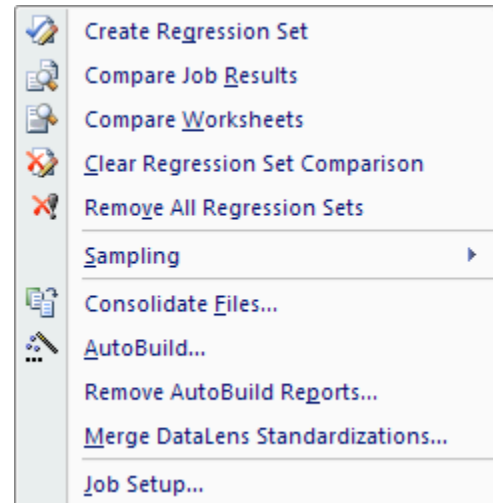
Used to create a regression set worksheet. The active worksheet is copied to a new worksheet with the same name as the active worksheet with a '_RSET' suffix.

Compare Job Results

Removes all regression test set worksheets in the active workbook that contain the '_RSET' suffix.

Compare Worksheets

Used to identify the differences between the results of two different DSA jobs.



Clear Regression Set Comparison

Clears all comparison highlighting on the active worksheet that was created by a regression test.

Remove All Regression Sets

Removes all regression test set worksheets in the active workbook.

Sampling

This menu contains the following two options:



Create Key Record Sample

Takes the records found in the active worksheet and creates a new worksheet, with a '_KSET' suffix, containing a random sample of records. For more information, see Data Sampling on page 52.

Create Baseline Record Sample

Takes the records found in the active worksheet and creates a new worksheet, with a '_BSET' suffix, containing a statistically significant random sample of records. For more information, see Data Sampling on page 52.

Consolidate Files...

Used to consolidate like or similar data records from several different files into a single data file.

AutoBuild...

Used to create a new data lens from customer data schema information that is implicit or found in customer database exports.

The **AutoBuild** option is only visible when the appropriate Preference option is set. For more information, see *Configuring the Server and Application* on page 11.

Typically, the export information is output in Excel spreadsheets and tab-separated files. This application is a step-by-step wizard that enables you to configure the key schema components: categories that can be converted to item definitions, columns, and labels that can be converted to item definition attributes, and record values that can be converted to valid attribute values captured as data lens phrases and terms. For more information, see *Oracle DataLens AutoBuild Application* on page 58.

Remove AutoBuild Reports...

Used to remove the three separate reports that provide detailed metric data and the semantic knowledge structure of a data lens created using the AutoBuild Application. For more information, see the *Oracle Product Data Quality AutoBuild Reference Guide*.

Merge DataLens Standardizations...

Allows you to merge data lens standardization types. For more information, see *Merging Data Lens Standardizations* on page 54.

Job Setup...

Configures how your source data will be processed and by which DSA. For more information, see *Configuring a DSA* on page 21.

DataLens Help Menu

Help for Services for Excel

Opens a list of Oracle Product Data Quality documents for your selection in a browser.



Preferences

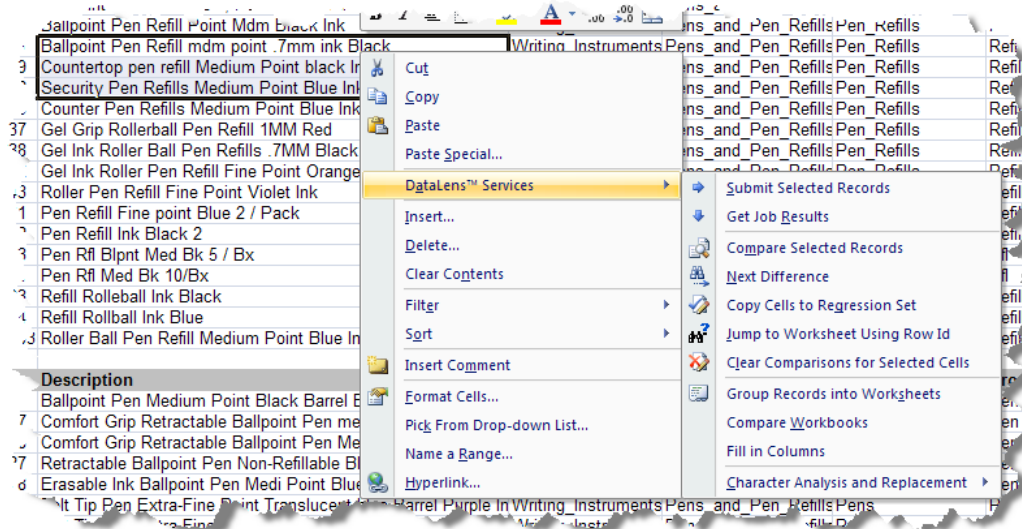
Allows you to view or change the Services for Excel product configuration. For more information, see [Configuring the Server and Application](#) on page 11.

About Services for Excel

Provides an informational message about the Services for Excel product including versions, application path, installation path, user name, network domain, and the date.

Excel Cell Context-Sensitive Menu

Services for Excel provides a context-sensitive menu to help automate the task of working with specific subsets of records on a spreadsheet. The following menu appears when you right-click an Excel cell or selection of cells:



This menu is used as follows:

Submit Selected Records

Use this menu item to process the selected records with the currently selected DSA. All records with at least one cell in the active Excel region selected are processed.

Get Job Results

The same as the **Get Results** option on the **Jobs** menu. For more information, see Jobs Menu.

Compare Selected Records

Compares just the selected records to the corresponding records in the regression test set. All records with at least one cell in the active Excel region selection are compared.

Next Difference

Finds the next highlighted difference on the active worksheet. The difference highlighting is created as a result of running a regression test.

Copy Cells to Regression Set

Copies the acceptable values found on the active worksheet to the regression set worksheet. The typical use is to copy the highlighted cells from a regression test to the regression set worksheet as progressions or valid results. All actively selected Excel cells are copied to the corresponding regression set. All comparison highlighting is removed from the copied cells.

Jump to Worksheet Using Row Id

Finds the corresponding row in the target worksheet using the Id for the actively selected row. You are prompted to select the desired target worksheet. This feature can be very convenient if you need to move between worksheets containing different information about the same record. For example, use this feature when you want to move from a DSA output spreadsheet back to the original input record on another spreadsheet, or to find the corresponding record on the regression test worksheet.

Clear Comparisons for Selected Cells

Clears all of the comparison highlighting found in the selected Excel cells.

Group Records into Worksheets

Groups your records into a set of separate worksheets based on the values found in one of the input columns. For more information, see Grouping Records on page 43.

Compare Workbooks

Compares two selected workbooks for differences. A report is created on the differences that are found. The comparison function first notes whether the two workbooks have the same number of worksheets. Then, the individual worksheet names are compared from left to right as they appear in the workbooks. Finally, the contents of worksheets with the same name are compared and the first cell difference is reported.

Fill in Columns

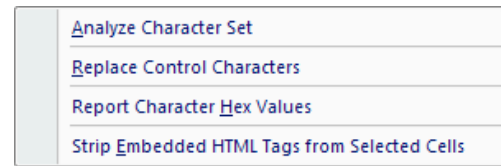
Completes a column of data based on the data in the column and adjacent cells. For more information, see [Filling Columns with Data](#) on page 40.

Character Analysis and Replacement

This menu contains the following two options:

Analyze Character Set

Analyzes the characters found in all the selected Excel cells and reports any control characters or Unicode spaces found. For more information, see [Character Analysis and Replacement](#) on page 41.



Replace Control Characters

Replaces all control characters found in the selected Excel cells with spaces. For more information, see [Character Analysis and Replacement](#) on page 41.

Report Character Hex Values

Searches for all hex values in the selected Excel cells and reports them. For more information, see [Character Analysis and Replacement](#) on page 41.

Strip Embedded HTML Tags from Selected Cells

Removes all HTML text that are embedded and any control characters in the selected Excel cells. For more information, see [Character Analysis and Replacement](#) on page 41.

Chapter 2

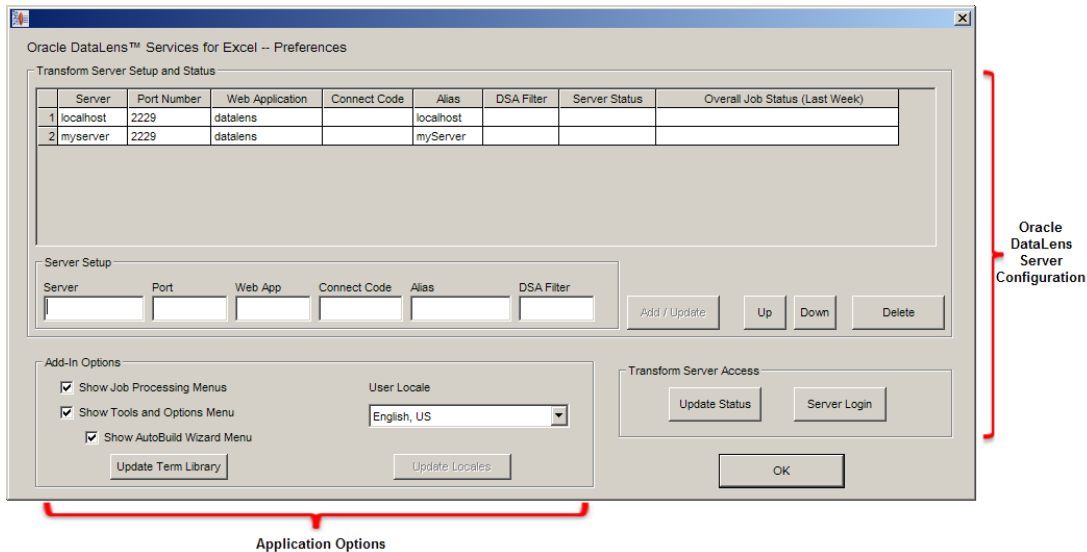
Configuring the Server and Application

In this chapter

- Adding a New Oracle DataLens Server 13
- Updating or Removing an Oracle DataLens Server 15
- Logging Into Oracle 17
- Application Options 18

The Oracle DataLens Server must be configured to run your DSA jobs before you can use the Services for Excel to transform your data. The basic Services for Excel global options and Oracle DataLens Server setup and maintenance tasks are configured using the same option.

To configure your Oracle DataLens Server or application options, click **DataLens Help** and select **Preferences**.



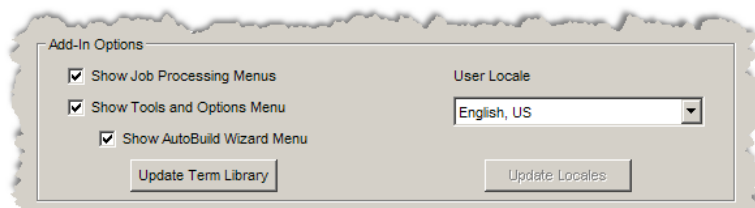
The **Oracle DataLens Services for Microsoft Excel - Preferences** dialog box appears. All changes made using this dialog box take effect immediately after clicking **OK**. The servers listed in the **Transform Server Setup and Status** table are listed in the order in which they will be used. The first server in the list is the default.

This dialog box interfaces directly with the Oracle DataLens Server and all changes effect only the configuration of Services for Excel application on your local machine.

The **Transform Server Setup and Status** and **Transform Server Access** sections of the **Oracle DataLens Services for Microsoft Excel - Preferences** dialog box are used to add, update, or delete Oracle DataLens Servers.

Adding a New Oracle DataLens Server

Use the **Server Setup** subsection of the **Transform Server Setup and Status** section to add a new Oracle DataLens Server.



Enter the following information in the respective textboxes:

Server

The name or IP address for the Oracle DataLens Server. All server names are forced to lowercase for consistency.

Port

The port used by the Oracle DataLens Server, which is 2229 by default.

Note: The port number defaults to 2229 if none is entered. The port number 2229 is the web services port number officially assigned to the Oracle DataLens Server by IANA (see <http://iana.org/assignments/port-numbers>).

Web App

The default is 'datalens' and cannot be changed. The Web App name, 'datalens', is used to access the server as in the following example:

<http://localhost:2229/datalens>

Connect Code

Typically, this is blank as license connect codes are not usually required; this is defined in your license if necessary.

Alias

The alias name to use for this server when displaying server information in the system dialogs. The server URL value is used by default.

DSA Filter

Only display DSAs whose names contain this text. Blank by default, which displays all DSAs.

Click **Add / Update** to add the new server.

	Server	Port Number	Web Application	Connect Code	Alias	DSA Filter	Server Status	Overall Job Status (Last Week)
1	localhost	2229	dataLens		localhost			
2	myserver	2229	dataLens		myServer			
3	prodserver	2229	dataLens		Production Server			

Server Setup

Server: Port: Web App: Connect Code: Alias: DSA Filter:

Add / Update Up Down Delete

Add-In Options

☒ Show Job Processing Menu
☒ Show Tools and Options Menu
☐ Show AutoBuild Wizard Menu

User Locale:

Update Term Library Update Library

Transform Server Access

Update Status Server Login

OK

The information in the **Transform Server Setup and Status** section columns is not editable. To change the information for an existing server, you must update it as described in Updating or Removing an Oracle DataLens Server.

Updating or Removing an Oracle DataLens Server

To update an existing Oracle DataLens Server definition, select the desired server from the list of servers in the **Transform Server Setup and Status** section by clicking on the row number.

Transform Server Setup and Status

	Server	Port Number	Web Application	Connect Code	Alias	DSA Filter	Server Status	Overall Job Status (Last Week)
1	localhost	2229	datalens		localhost			
2	newserver	2229	datalens		newserver			

Server Setup

Server	Port	Web App	Connect Code	Alias	DSA Filter
newserver	2229	datalens		newserver	

Add / Update Up Down Delete

Change the information in the **Server Setup** subsection as necessary, and then click **Add / Update** to save your changes.

Transform Server Setup and Status

	Server	Port Number	Web Application	Connect Code	Alias	DSA Filter	Server Status	Overall Job Status (Last Week)
1	localhost	2229	datalens		localhost			
2	newserver	2229	datalens		newserver			

Server Setup

Server	Port	Web App	Connect Code	Alias	DSA Filter
DSA SERVER	2229	datalens		DSA SERVER	Motor

Add / Update Up Down Delete

Transform Server Setup and Status

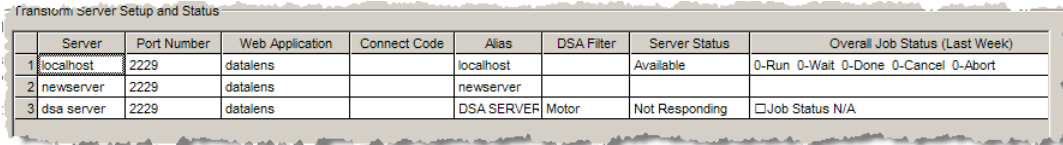
	Server	Port Number	Web Application	Connect Code	Alias	DSA Filter	Server Status	Overall Job Status (Last Week)
1	localhost	2229	datalens		localhost			
2	newserver	2229	datalens		newserver			
3	dsa server	2229	datalens		DSA SERVER	Motor		

You can use the **Up** and **Down** buttons to change the order of the servers in the list. The first server in the list is the default. You can change the default and server use order using these buttons.

Servers are deleted by selecting one or more from the list and clicking **Delete** to remove the selected server from the list.

Note: Deleting a server only removes it from use within the Services for Excel application; it does not remove it from the Oracle DataLens Server.

The weekly status statistics for each server is displayed in the **Overall Job Status (Last Week)** column. The **Update Status** button is used to refresh the status of one or more selected servers.

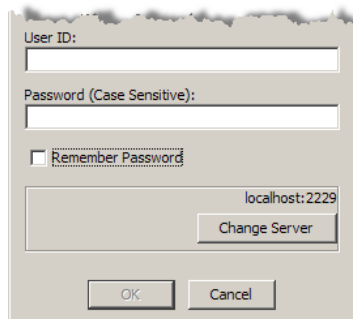


The screenshot shows a window titled "Transform Server Setup and Status" containing a table with the following data:

	Server	Port Number	Web Application	Connect Code	Alias	DSA Filter	Server Status	Overall Job Status (Last Week)
1	localhost	2229	datalens		localhost		Available	0-Run 0-Wait 0-Done 0-Cancel 0-Abort
2	newserver	2229	datalens		newserver			
3	dsa server	2229	datalens		DSA SERVER	Motor	Not Responding	<input type="checkbox"/> Job Status N/A

Logging Into Oracle Product Data Quality

You can use the **Server Login** button to log in to the administrative web page of the selected server.



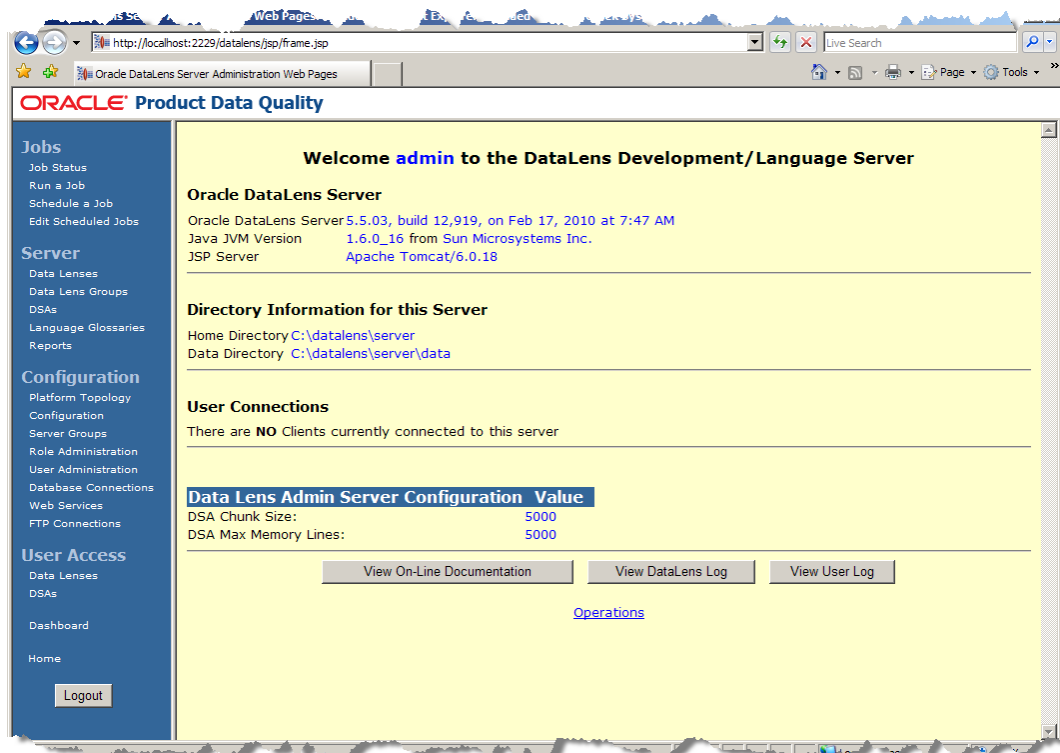
The login dialog box contains the following fields and controls:

- User ID:** A text input field.
- Password (Case Sensitive):** A text input field.
- Remember Password:** A checkbox.
- localhost:2229:** A label indicating the current server.
- Change Server:** A button to switch servers.
- OK:** A button to confirm login.
- Cancel:** A button to cancel login.

The **Oracle Product Data Quality Login** dialog box appears.

Enter your user name and password and click **OK**. You can avoid entering your password every time you logon by selecting the **Remember Password** checkbox.

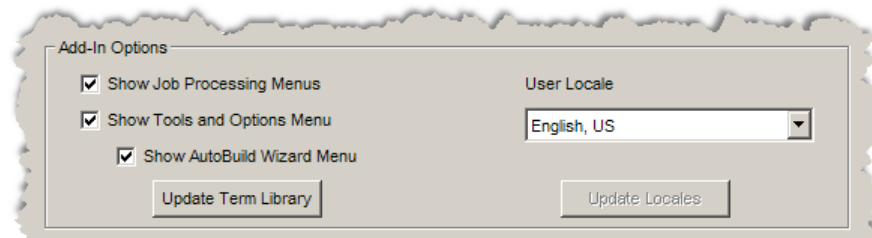
Your default browser is started and the Oracle DataLens Server Administration web page is displayed.



For more information about the use of the Oracle DataLens Server Administration web page, see the *Oracle Product Data Quality Oracle DataLens Server Administration Guide*.

Application Options

The **Add-In Options** subsection of the **DataLens Services for Microsoft Excel - Preferences** dialog box allows you to control how the Services for Excel Add-In toolbar buttons and menu options, and the context-sensitive menu are displayed in the application.



The controls in the **Add-In Options** subsection are used as follows:

Show Job Processing Menus

Enables the **Jobs** menus and associated buttons on the toolbar, as well as background checking for job completion. When you select this option, the system queries the Oracle DataLens Server every second for job completion and displays a completion notification.

Show Tools and Options Menu

Enables the **DataLens Tools** menus and associated buttons on the toolbar. Selecting this option activates the following options for selection:

Show AutoBuild Wizard Menu

Enables the **AutoBuild** button on the toolbar and associated menu options on the **DataLens Tools** menu.

Update Term Library Button

Creates or updates the terminology library uses by AutoBuild. The terms are extracted from the data lens that you select. This is particularly useful when you have extensive refinements (for example, term variants) in a data lens that you want to leverage in new lens. You would update the term library to include these new terms and variations. The resulting AutoBuild data lens can more easily process new data.

The extract terms are case insensitive, all underscores, plurals, and duplicates (containing identical variants) are removed.

Update Locale Listbox

Allows you to change the user interface language locale for the locales that were shipped with the product. This list is populated with the locales that are configured in your Oracle DataLens Server.

Update Locales Button

This button is not active because it is used for administrative purposes by Professional Services only.

Tip: To use the same options for other Services for Excel users, you can copy the options file to each end user's computer. The options file, `system.xml`, is located in the installation folder, which is typically the folder, `C:\datalens\applications\config`.

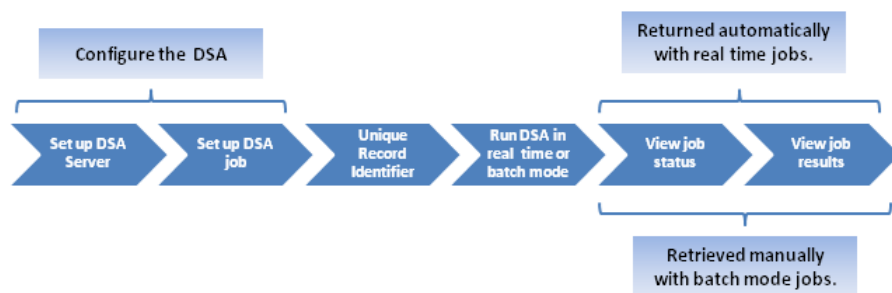
Chapter 3

Transforming Your Data

In this chapter

Configuring a DSA	21
Unique Record Identifiers	27
• Running a DSA to Process Records	29
• Viewing the Job Status	34

Transforming your data records using a DSA with Services for Excel is a simple step-by-step process as shown in the following diagram:



All of these steps are described in this chapter.

Configuring a DSA

Each Excel spreadsheet must be configured to use a DSA for data transformation. When you save the Excel workbook, your spreadsheet configurations are saved along with the workbook as a hidden spreadsheet for future reuse by the Services for Excel application.

Always start by opening the desired Excel workbook, and then selecting the specific spreadsheet containing the data that you want to process using a DSA.

From the **Services for Excel** toolbar, click **DataLens Tools**, and then select the **Job Setup...** option.

The **Job Setup** dialog box is displayed.

The options highlighted with bold blue text labels require your selection and all required selections are the default tab, **Data Service Settings**. All other options are initialized with default values and typically do not require your attention.

The following sections describe each of the tabs in the **Job Setup** dialog box.

Data Service Settings Tab

This is the main tab of the **Job Setup** dialog box and is used to set up the basic parameters of your transformation service. Typically, you only need to set up the required parameters (indicated by the bold blue text labels) on this tab.

Data Service Settings Advanced Settings

Data Service Application (DSA)

Server: localhost

DSA: 01_Gloves_Demo_Create_IM

Result Set: LOAD

10_Table_Created

☐ Write Results as Text

☐ Strip Embedded HTML Tags

Insert Headings

Data Source

Data Source: Excel

First Row: 2

ID	DESCRIPTION	MANUFACTURE	PART_NUMBER
A	B	C	D

Result Set for Output LOAD

Data Target: LOAD

Output Columns:

ID	DESCRIPTION	MANUFACTURE	PART_NUMBER
A	B	C	D

The controls on this tab are described as follows:

Server Listbox

This list is automatically populated with the name or IP addresses of your available Oracle DataLens Servers. You should contact your Oracle DataLens Server Administrator for the correct server to use. Typically, your administrator will configure just one server for you to use.

Write Results as Text Checkbox

Enables the writing of all DSA results to Excel using Excel 'Text' cell format. For example, this feature could be used to prevent the Excel auto-formatting feature from turning ratios and part numbers into dates or to prevent the loss of leading zeros.

DSA Listbox

This list is automatically populated with the name of the DSA that you are using to process your data. You can select a different DSA from the list of available DSAs.

Note: DSAs are created using the Application Studio. After you deploy a DSA to an Oracle DataLens Server, it is available to Services for Excel and other external applications.

Strip Embedded HTML Tags Checkbox

Enables the stripping of embedded HTML tags often found in product data. This feature does not strip tags from full HTML web pages. For example, it will remove HTML break and list tags often found inside product data stored in databases.

Result Set Listbox

This list is automatically populated to display each of the DSA outputs, which are known as *result sets*. Selecting a result set changes the information displayed in the **Result Set for Output LOAD** section of this tab.

Insert Headings

This button will insert headings on the spreadsheet(s) corresponding to the input and output fields of the selected DSA. This will overwrite any headings that might have appeared on the spreadsheet.

Description Textbox

This textbox is informational and supplies the description of the DSA that you are using to process your data. It is blank if no description for the DSA exists.

Data Source Listbox

This list is automatically populated based on the properties of the selected DSA. The data sources can be Excel, File input, or they may be generated directly by the DSA. If the data source is the DSA, then the data may be coming from an external data source, such as a database or XML file.

First Row Textbox

This supplies a single valid Excel row to the system and represents the starting point for the input data. Your DSA reads the Excel input columns starting with this row.

Input Columns Table

This table is automatically populated with an editable list of the Excel columns that are to be sent to the Oracle DataLens Server for processing. This is a list of valid Excel column names. The inputs are the expected input fields from the DSA while the column names are Excel alphabetic columns. From the previous illustration, that would be 'ID', 'DESCRIPTION', 'MANUFACTURER', and 'PART_NUMBER' with the corresponding columns of 'A, B, C, and D'.

Note: The default Excel Column list has the same number as the input fields and is in alphabetical order starting with A.

You can edit the automatically assigned column letters by clicking the specific column letter cell in the table, clicking **Backspace** to delete the active column letter, and then entering a new valid Excel column letter or letters. You can navigate between cells from left to right and right to left by using the arrow keys on your keyboard.

The arrow buttons to the right of the table allow you to move all of the columns as a group to the right or left. All of the column letter assignments can be cleared at once by selecting the clear button above the arrow buttons to the right of the table.

You can ensure that the columns in your Excel spreadsheet match the input data columns expected by the DSA you have selected by shifting or editing these columns.

Data Target Listbox

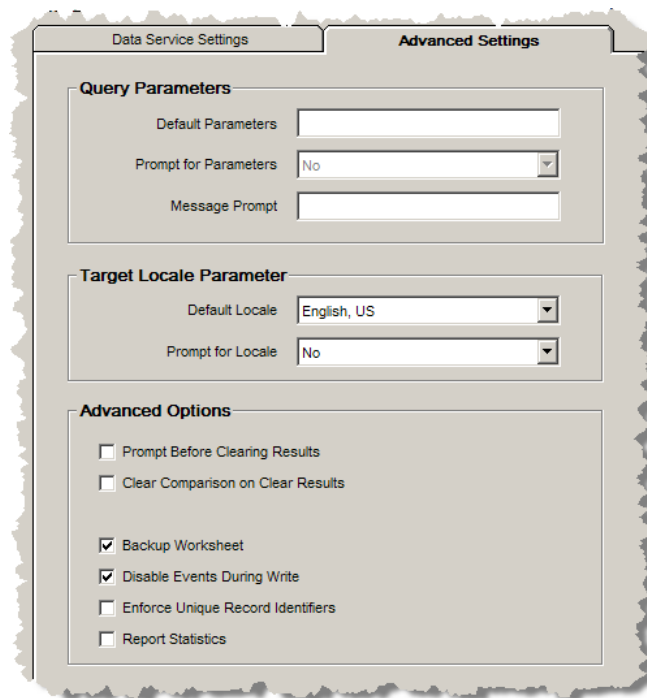
This list is automatically with the name of the worksheet within your workbook where the DSA results will be stored for the actively selected output. The result records will be written either to the active worksheet (when the DSA result set is named 'Output') or to a worksheet by the same name as the DSA output. Each target worksheet can be viewed by selecting the associated result set name.

Output Columns Table

This table is automatically populated with an editable table of Excel columns that are to be populated with the outputs from a DSA. This is a list of valid Excel Column names derived from the selected DSA. When the special 'Output' result set is selected, you can edit the output column letters for the Data Source as previously described. (Note: the "Output" result set is written directly back to the input worksheet.)

Advanced Settings Tab

This tab contains settings for the DSA runtime parameters and the overall DSA performance. Some DSAs require user-supplied runtime parameters for the **Target Locale**, in the case of language translation, or database queries. You can provide default parameters for each of these settings, as well as, specifying whether to prompt for specific values. Additionally, you can supply your own prompt message.



The screenshot shows a dialog box with two tabs: "Data Service Settings" and "Advanced Settings". The "Advanced Settings" tab is active. It contains three sections:

- Query Parameters:**
 - Default Parameters: A text input field.
 - Prompt for Parameters: A dropdown menu with "No" selected.
 - Message Prompt: A text input field.
- Target Locale Parameter:**
 - Default Locale: A dropdown menu with "English, US" selected.
 - Prompt for Locale: A dropdown menu with "No" selected.
- Advanced Options:**
 - ☐ Prompt Before Clearing Results
 - ☐ Clear Comparison on Clear Results
 - ☒ Backup Worksheet
 - ☒ Disable Events During Write
 - ☐ Enforce Unique Record Identifiers
 - ☐ Report Statistics

The controls on this tab are described as follows:

Default Parameters Textbox

Use when a DSA receives its input from a database query and the user must supply one or more parameters to the query.

To enter multiple default parameters, use the pipe character (|) to separate the list of parameters. Ensure that you avoid putting extra spaces around the pipe characters.

Prompt for Parameters Listbox

When set to **Yes**, the user is prompted for the database query parameters for each job. If you supplied default parameter(s), then those defaults are displayed to the user. The user can choose to accept the defaults or enter their own query values.

Message Prompt Textbox

Use to enter a help message to be displayed during runtime when the user is prompted for the database query parameter(s). The message prompt can be used to inform the user of the expected parameters and their valid values.

Default Locale Listbox

The default locale is used when a DSA needs a run-time language translation setting. A different locale can be selected from the list of available locales.

Prompt for Locale Listbox

Use if your DSA has a run-time translation setting and you want the user to select the translation locale with each separate run by setting this parameter to **Yes**. When set to **Yes**, the user is prompted for the target locale for each job. The user receives no prompts when this parameter is set to **No**.

Prompt Before Clearing Results Checkbox

Provides a dialog confirmation before clearing results prior to running a job. By default, this is off.

Clear Comparison on Clear Results Checkbox

When creating regression tests, this will clear comparisons. By default, this setting is off.

Use Progress Display Checkbox

Provides a progress display dialog as a job is running. By default, this setting is on.

Backup Worksheet Checkbox

Creates a backup copy of the worksheet before overwriting it with the job results. By default, this setting is turned off.

Disable Events During Write Checkbox

Disables Excel events when DSA results are written to spreadsheets. This option is used to prevent other pre-existing workbook macros from interfering with the update of results from a job. By default, this setting is on.

Report Statistics Checkbox

Provides a dialog with statistical information after having completed a DSA job. By default, this setting is off.

Setting Up a DSA Job

To set up a DSA, use the following steps:

1. Open your Excel workbook and select the tab associated with the Excel worksheet that contains your data.
2. From the **Services for Excel** toolbar, click **DataLens Tools**, and then select **Job Setup....**

Oracle DataLens™ Services for Excel – Job Setup

Bold Blue labels indicate required selections.

Data Service Settings | Advanced Settings

Data Service Application (DSA)

Server: localhost

DSA: 01_Gloves_Demo_Create_IM

Result Set: LOAD

10_Table_Created

☒ Write Results as Text

☐ Strip Embedded HTML Tags

Insert Headings

Data Source

Data Source: Excel

First Row: 2

Inputs Columns	ID	DESCRIPTION	MANUFACTURE	PART_NUMBER
A	B	C	D	

Result Set for Output LOAD

Data Target: LOAD

Output Columns

OK Cancel

3. Select the Oracle DataLens Server you want to use from the list.
4. Select the DSA you want to use from the list.

The Data Source will automatically be entered by the system. Input and output columns are also automatically populated when the DSA is selected.

5. Modify any other options on this tab or the **Advanced Settings** tab.
6. Select **OK** on either tab to accept your changes.
7. Save your Excel workbook.

Your DSA configuration is now saved with the active workbook; this configuration will be available the next time you open this workbook.

Unique Record Identifiers

As record data is routed through a DSA, records may be routed by the DSA to a variety of output steps. Some outputs represent exceptions that must be reviewed by product experts. Other output record sets are ready for production use in your enterprise and may be targeted to several different enterprise applications. In order to effectively track records as they flow through a DSA each record *must* carry with it a unique identifier. The Services for Excel application assumes that the first column of the input data, also known as column 'A' in Excel, contains the unique identifier.

Creating the Unique Identifier Column

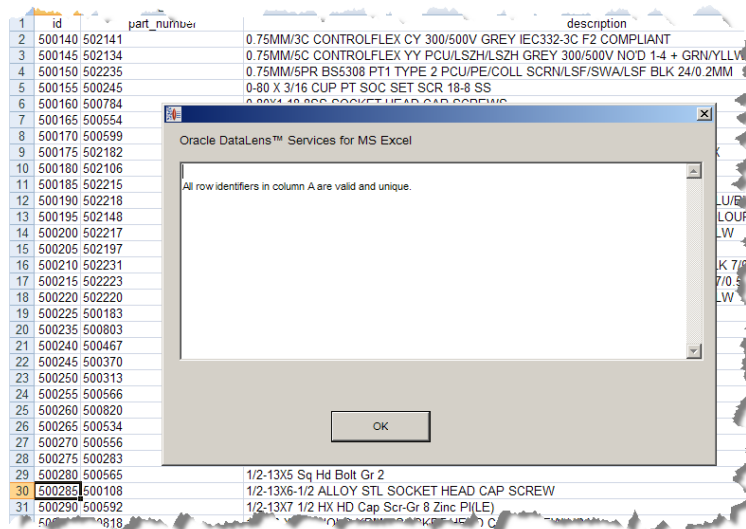
The Services for Excel Add-In makes it easy to add a unique identifier column if one does not exist. From the **Services for Excel** toolbar, click **Jobs**, and then select **Insert Line Item Id Column**. This automatically inserts a new column A to the left of your existing data columns and then inserts a unique set of record identifiers into column A, as in the following:

	A	B	C
1	Manufacturer	Part Number	Description
2	Avery	AVE49560	Mechanical Pencil .5MM Emerald Barrel
3	Bic	BICFRC21BE	Pencil .5MM Green Barrel Automatic
4	Cross	BICFRC21BK	Ballpoint Pen Medium Point Black Barrel Black Ink
5	Eversharp	CRO80041	Ballpoint Pen Refill Fine Point Blue Ink 2 / Pack
6	NIB-N		
7	Mont		
8	Pape		
9	Park		
10	Pente		
11	Pilot		
12	Sanf		
13	Wate		
14	Zebra		
15	Unite		
16	Autopoint		

	A	B	C	D
1	Ids	Manufacturer	Part Number	Description
2	1	Avery	AVE49560	Mechanical Pencil .5MM Emerald Barrel
3	2	Bic	BICFRC21BE	Pencil .5MM Green Barrel Automatic
4	3	Cross	BICFRC21BK	Ballpoint Pen Medium Point Black Barrel Black Ink
5	4	Eversharp	CRO80041	Ballpoint Pen Refill Fine Point Blue Ink 2 / Pack
6	5	NIB-National Industries for the Blind	CRO800421	Ballpoint Pen Refill Med Pt Black Ink 2 / Pk
7	6	Mont Blanc	CRO800422	Ballpoint Pen Refill Medi, Point black, ink 2 / Pack
8	7	Papermate	CRO800423	Ballpoint Pen Refill Point Mdm Black Ink
9	8	Parker Pen	CRO800425	Ballpoint Pen Refill mdm point .7mm ink Black
10	9	Pentel	CRO80045	China Mrkr Green Ink 12 / Box
11	10	Pilot	CRO81002	China Markers red ink
12	11	Sanford	CRO81012	China Markers White ink
13	12	Waterman	CRO8511	China Mrk yellow ink
14	13	Zebra Pen	CRO85112	Chisel Point Highlighter Turquoise ink
15	14	United Stationers	CRO8512	Comfort Grip Retractable Ballpoint Pen medium Point Bl
16	15	Autopoint	CRO85122	Comfort Grip Retractable Ballpoint Pen Medium Point B

If your data already has a unique identifier column, you should move that column so that it is the first data column in your spreadsheet.

Additionally, you can check the uniqueness of your column A identifiers to ensure proper record processing. From the **Services for Excel** toolbar, click **Jobs**, and then select **Fill-in Line Item Ids**.



A report about the uniqueness of your record identifiers is displayed. This report can be saved to a text file by clicking **Save Report** and entering a file name.

If there are id duplicates, you can the option of allowing the Services for Excel application to correct them by clicking **OK**; click **Cancel** to exit the dialog box without changing any records. In addition to fixing the duplicate identifiers, the application fills in any null or missing identifiers.

Note: The maximum valued unique identifier found in column A is used as the starting point for creating additional unique identifiers.

If desired, you can also enforce unique item identifiers from the Data Services Options dialog. For more information, see the Advanced Settings Tab section of Configuring a DSA.

DSA Outputs

Each record set associated with a DSA output is written to a separate Excel spreadsheet, which are known as a result set. Each Excel spreadsheet is named the same as the DSA output step. Whether you get results from Process Job or Get Job Results, the new results overwrite the previous results on the corresponding spreadsheet. The spreadsheet headers in row 1 are automatically created from the DSA output columns. While the data values are overwritten each time results are retrieved, any spreadsheet formatting is preserved so that any formatting you add is not lost.

The DSA Output Named *Output*

The Services for Excel application treats a DSA output named *Output* a little differently than any other named output step. DSA records routed to the Output step are written to columns on the input worksheet. By default, the data in these result records are written to columns to the right of the input columns. In other words, the results data are appended to each corresponding input record or row.

Running a DSA to Process Records

A DSA job is immediately executed on the Oracle DataLens Server. The job execution does the following:

1. Sends the input data found on the active worksheet in (columns A and B are used in the example in this section) to the configured Oracle DataLens Server for processing.
2. Retrieves the transformation results from the server after the job has completed.
3. Writes the retrieved results to the Excel worksheet(s) specified by the DSA.

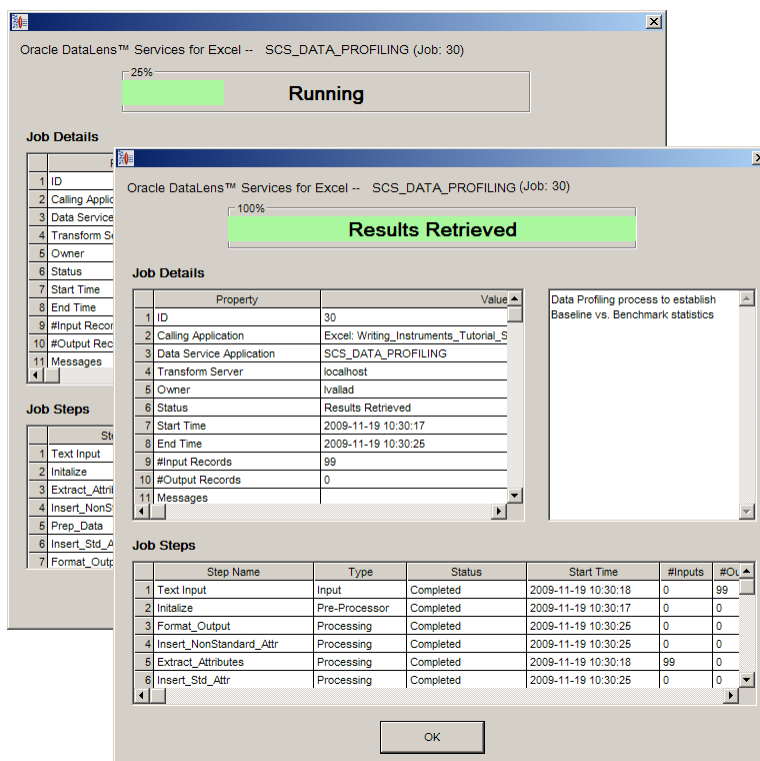
The worksheet columns that are used for input to the job and where the results are placed back into your Excel workbook after the job has completed are all determined by the design of the DSA. For more information, see the *Oracle Product Data Quality Application Studio Reference Guide*.

There are two ways that you can process data records using your DSA, real time (synchronous) and batch (asynchronous) mode. Both of these DSA run methods are described in this section.

Using Real Time Records Processing

To run a configured DSA, click the **Process Records** button on the **Services for Excel** toolbar.

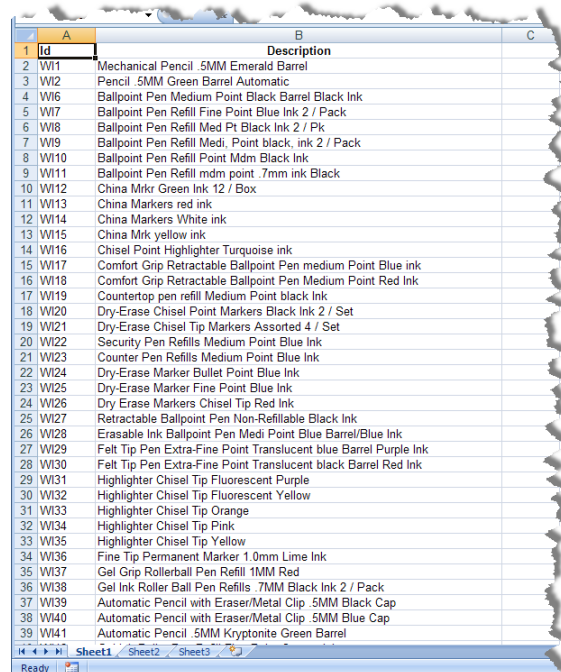
When processing the data records, job status messages are displayed to inform you of the progress and completion as in the following:



When the job is finished and the results have been retrieved, you should review the job statistics. Once you have done so, click **OK**. Your transformed data is included in your Excel workbook as defined by your DSA.

Note: A job is finished when 'Results Retrieved' is displayed rather than 'Complete'. For jobs that are processing a great deal of data, retrieving all results and finishing the job can take some time.

In the following example, the input data records contain a unique id and a description:



1	A	B	C
	Id	Description	
2	W11	Mechanical Pencil .5MM Emerald Barrel	
3	W12	Pencil .5MM Green Barrel Automatic	
4	W16	Ballpoint Pen Medium Point Black Barrel Black Ink	
5	W17	Ballpoint Pen Refill Fine Point Blue Ink 2 / Pack	
6	W18	Ballpoint Pen Refill Med Pt Black Ink 2 / Pk	
7	W19	Ballpoint Pen Refill Medi. Point black, ink 2 / Pack	
8	W110	Ballpoint Pen Refill Point Mdm Black Ink	
9	W111	Ballpoint Pen Refill mdm point .7mm ink Black	
10	W112	China Mkr Green Ink 12 / Box	
11	W113	China Markers red ink	
12	W114	China Markers White ink	
13	W115	China Mkr yellow ink	
14	W116	Chisel Point Highlighter Turquoise ink	
15	W117	Comfort Grip Retractable Ballpoint Pen medium Point Blue ink	
16	W118	Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink	
17	W119	Countertop pen refill Medium Point black Ink	
18	W120	Dry-Erase Chisel Point Markers Black Ink 2 / Set	
19	W121	Dry-Erase Chisel Tip Markers Assorted 4 / Set	
20	W122	Security Pen Refills Medium Point Blue Ink	
21	W123	Counter Pen Refills Medium Point Blue Ink	
22	W124	Dry-Erase Marker Bullet Point Blue Ink	
23	W125	Dry-Erase Marker Fine Point Blue Ink	
24	W126	Dry Erase Markers Chisel Tip Red Ink	
25	W127	Retractable Ballpoint Pen Non-Refillable Black Ink	
26	W128	Erasable Ink Ballpoint Pen Medi Point Blue Barrel/Blue Ink	
27	W129	Felt Tip Pen Extra-Fine Point Translucent blue Barrel Purple Ink	
28	W130	Felt Tip Pen Extra-Fine Point Translucent black Barrel Red Ink	
29	W131	Highlighter Chisel Tip Fluorescent Purple	
30	W132	Highlighter Chisel Tip Fluorescent Yellow	
31	W133	Highlighter Chisel Tip Orange	
32	W134	Highlighter Chisel Tip Pink	
33	W135	Highlighter Chisel Tip Yellow	
34	W136	Fine Tip Permanent Marker 1.0mm Lime Ink	
35	W137	Gel Grip Rollerball Pen Refill 1MM Red	
36	W138	Gel Ink Roller Ball Pen Refills .7MM Black Ink 2 / Pack	
37	W139	Automatic Pencil with Eraser/Metal Clip .5MM Black Cap	
38	W140	Automatic Pencil with Eraser/Metal Clip .5MM Blue Cap	
39	W141	Automatic Pencil .5MM Kryptonite Green Barrel	

After process this data, the results are placed in new worksheets as in the following:

Item ID	Batch Type	Type	Match Status	Org	Source System	Batch ID
W1	Mechanical Pencil .5MM Emerald Barrel	EXC	REPROCESS			
W2	Pencil .5MM Green Barrel Automatic	EXC	REPROCESS			
W6	Ballpoint Pen Medium Point Black Barrel Black Ink	EXC	REPROCESS			
W7	Ballpoint Pen Refill Fine Point Blue Ink 2 / Pack	EXC	REPROCESS			
W8	Ballpoint Pen Refill Med Pt Black Ink 2 / Pk	EXC	REPROCESS			
W9	Ballpoint Pen Refill Medi. Point black, ink 2 / Pack	EXC	REPROCESS			
W10	Ballpoint Pen Refill Point Mdm Black Ink	EXC	REPROCESS			
W11	Ballpoint Pen Refill mdm point .7mm ink Black	EXC	REPROCESS			
W12	China Mkr Green Ink 12 / Box	EXC	REPROCESS			
W13	China Markers red ink	EXC	REPROCESS			
W14	China Markers White ink	EXC	REPROCESS			
W15	China Mkr yellow ink	EXC	REPROCESS			
W16	Chisel Point Highlighter Turquoise ink	EXC	REPROCESS			
W17	Comfort Grip Retractable Ballpoint Pen medium Point Blue ink	EXC	REPROCESS			
W18	Comfort Grip Retractable Ballpoint Pen Medium Point Red Ink	EXC	REPROCESS			
W19	Countertop pen refill Medium Point black Ink	EXC	REPROCESS			
W20	Dry-Erase Chisel Point Markers Black Ink 2 / Set	EXC	REPROCESS			
W21	Dry-Erase Chisel Tip Markers Assorted 4 / Set	EXC	REPROCESS			
W22	Security Pen Refills Medium Point Blue Ink	EXC	REPROCESS			
W23	Counter Pen Refills Medium Point Blue Ink	EXC	REPROCESS			
W24	Dry-Erase Marker Bullet Point Blue Ink	EXC	REPROCESS			
W25	Dry-Erase Marker Fine Point Blue Ink	EXC	REPROCESS			
W26	Dry Erase Markers Chisel Tip Red Ink	EXC	REPROCESS			
W27	Retractable Ballpoint Pen Non-Refillable Black Ink	EXC	REPROCESS			
W28	Erasable Ink Ballpoint Pen Medi Point Blue Barrel/Blue Ink	EXC	REPROCESS			
W29	Felt Tip Pen Extra-Fine Point Translucent blue Barrel Purple Ink	EXC	REPROCESS			
W30	Felt Tip Pen Extra-Fine Point Translucent black Barrel Red Ink	EXC	REPROCESS			
W31	Highlighter Chisel Tip Fluorescent Purple	EXC	REPROCESS			
W32	Highlighter Chisel Tip Fluorescent Yellow	EXC	REPROCESS			
W33	Highlighter Chisel Tip Orange	EXC	REPROCESS			
W34	Highlighter Chisel Tip Pink	EXC	REPROCESS			
W35	Highlighter Chisel Tip Yellow	EXC	REPROCESS			
W36	Fine Tip Permanent Marker 1.0mm Lime Ink	EXC	REPROCESS			
W37	Gel Grip Rollerball Pen Refill 1MM Red	EXC	REPROCESS			
W38	Gel Ink Roller Ball Pen Refills .7MM Black Ink 2 / Pack	EXC	REPROCESS			
W39	Automatic Pencil with Eraser/Metal Clip .5MM Black Cap	EXC	REPROCESS			
W40	Automatic Pencil with Eraser/Metal Clip .5MM Blue Cap	EXC	REPROCESS			
W41	Automatic Pencil .5MM Kryptonite Green Barrel	EXC	REPROCESS			

Input Data Worksheet Result Set Worksheets

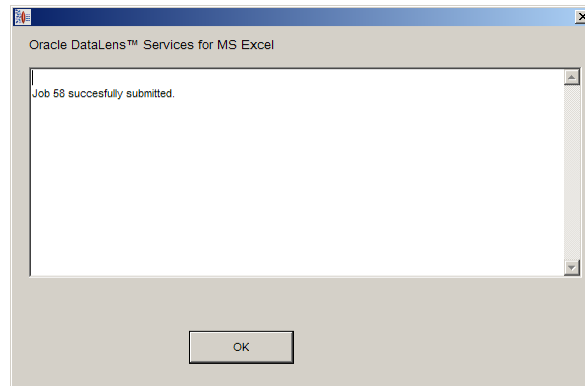
The original input data has been transformed, exceptions are identified, and subsequent columns contain any extracted attribute information.

Using Batch Mode Records Processing

To run a configured DSA, click the **Submit Job** button on the **Services for Excel** toolbar.

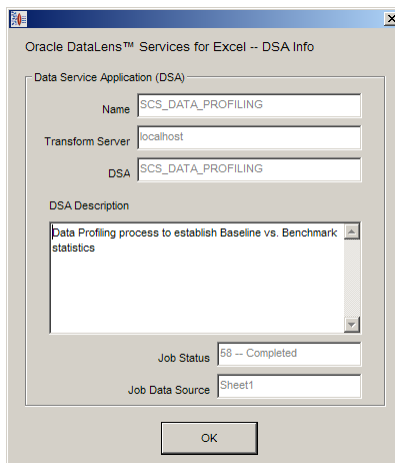
While a batch mode job is running, you can use Excel for your other tasks. The results of job can be retrieved later. After a job is submitted, you can even save the Excel workbook and close Excel, and when you return to the same workbook, you can retrieve the job results.

When you use the **Submit Job** button to start a job, the input records are first read from Excel and then the job is submitted to the Oracle DataLens Server for processing. If you have a large amount of data, a status message is displayed to show you the data read progress. When the job has successfully started to run at the Oracle DataLens Server, a status message is displayed:



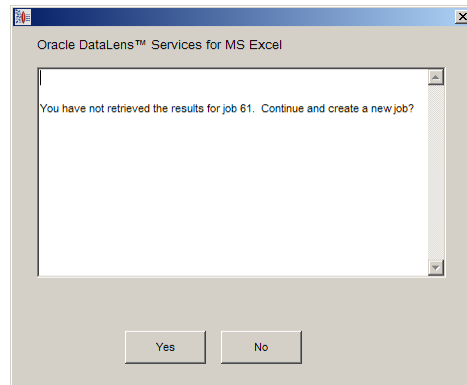
Your DSA is now executing the job in the background on the Oracle DataLens Server.

There are two ways to check on the status of your job. The first and easiest is the job process information by clicking **Jobs** on the **Services for Excel** toolbar, and then select **Process Info....** This dialog displays the status of your current job as well as information about the data source for the job and the DSA used to run the job.



The second way to view the job status as described in the Viewing the Job Status section. These results are displayed the same way as real time records processing.

You must review the job status before attempting to re-run your data, in real time or batch mode, or you are prompted to choose to overwrite the current results by creating a new job or you can click **No** so that you can review your results.



Tip: You can also use the [Oracle DataLens Server Administration web page](#) to check on job status. For more information, see the [Oracle Product Data Quality Oracle DataLens Server Administration Guide](#).

Removing Result Data

You can clear your job result data all at once from each of your result worksheets. From the **Services for Excel** toolbar, click **DataLens Tools**, and then select **Clear Job Results**. The data on all worksheets is deleted though the worksheets remain.

To remove all of the result worksheets, use the **Remove All Result Worksheets** option from the same menu.

Viewing the Job Status

To view the status of all of the jobs that you submitted (synchronous or asynchronous) to the Oracle DataLens server, click the Jobs button on the on the **Services for Excel** toolbar, and then select **Job Status....**

Oracle DataLens™ Services for Excel -- Job Status

User: Server: History:

Active Jobs

Job	Status	Owner	Calling Application	Data Service Application	Start Time	Duration	#Inputs	#Outputs

Completed Jobs

Job	Status	Owner	Calling Application	Data Service Application	End Time	Duration	#Inputs
1 61	Completed	invalid	Excel: MRO_Capabilities_Showcase_Source.xls	SCS_DATA_PROFILING	2010-02-23 13:47:36	0:00:04	412
2 60	Results Retrieved	invalid	Excel: MRO_Capabilities_Showcase_Source.xls	SCS_DATA_PROFILING	2010-02-23 13:47:13	0:00:06	412
3 59	Results Retrieved	invalid	Excel: MRO_Capabilities_Showcase_Source.xls	SCS_DATA_PROFILING	2010-02-23 13:46:59	0:00:06	412
4 58	Results Retrieved	invalid	Excel: MRO_Capabilities_Showcase_Source.xls	SCS_DATA_PROFILING	2010-02-23 13:41:53	0:00:06	412
5 57	Results Retrieved	invalid	Excel: MRO_Capabilities_Showcase_Source.xls	SCS_DATA_PROFILING	2010-02-23 13:36:02	0:00:13	412

Job Details

Property	Value
1 ID	
2 Calling Application	
3 Data Service Application	
4 Transform Server	
5 Owner	
6 Status	
7 Start Time	
8 End Time	
9 #Input Records	
10 #Output Records	
11 Messages	
12	
13	
14	

Job Steps

Step Name	Type	Status	Start Time	#Inputs	#Outputs

OK Refresh ☐ Auto Refresh

The Job Status dialog is displayed. At the top of the dialog box are three listboxes that you can use to change the displayed job status results as follows:

User:

Display one specific user or all users listed.

Server:

Display one specific Oracle DataLens Server or all servers listed.

History:

Display specific job history duration from the list of choices.

The job status results are filtered immediately after a selection is made from any of the listboxes.

The remaining sections of the **Job Status** dialog box are described as follows:

Active and Completed Jobs

Displays summary information about jobs that are active or that have completed.

Selecting a job in one of either of the active or completed lists redisplay the information in the **Job Details** section.

Job Details

Displays the job duration, the number of records processed by each step in the DSA, and any job execution messages.

Job Steps

Displays each step of the DSA including the type of step, status, start and end times, and input and output records processed.

Use the **Refresh** button to retrieve any additional job status information from the server and redisplay the job information.

If you select the **Auto-Refresh** checkbox on the dialog, the **Job Status** dialog box refreshes the job status every two seconds while it is open. The automatic refresh stops when you select a specific job to view its details.

Warning: When a DSA returns the results of a processing job (synchronous or asynchronous), you *must* retrieve or delete the data immediately, particularly for those jobs that return large amounts of data. Failure to do so can result in an Oracle DataLens Server crash once the server memory is exhausted. The memory can be exhausted from many small jobs that are not retrieved because each DSA job that is under the DSA maximum memory or chunk size set for the server is held in memory until it is retrieved or the server is restarted. For information about setting these values, see the *Oracle Product Data Quality Oracle DataLens Server Administration Guide*.

Chapter 4

Getting More from Your Data

In this chapter

- Consolidating Excel Files 37
- Changing and Grouping Data 40
- Using Regression Testing 45
- Data Sampling 52
- Source Data Backup 53
- Merging Data Lens Standardizations 54
- Oracle DataLens AutoBuild Application 58

Consolidating Excel Files

With the Services for Excel file consolidation functionality, you can merge data records from similar Excel workbooks into a single workbook. This can be very helpful in creating single file sources for processing by a DSA rather than processing the data in each file individually.

You can merge multiple Excel files into a single Excel file of the same type. The file consolidation functionality makes the merging of distinct data files straightforward and easy.

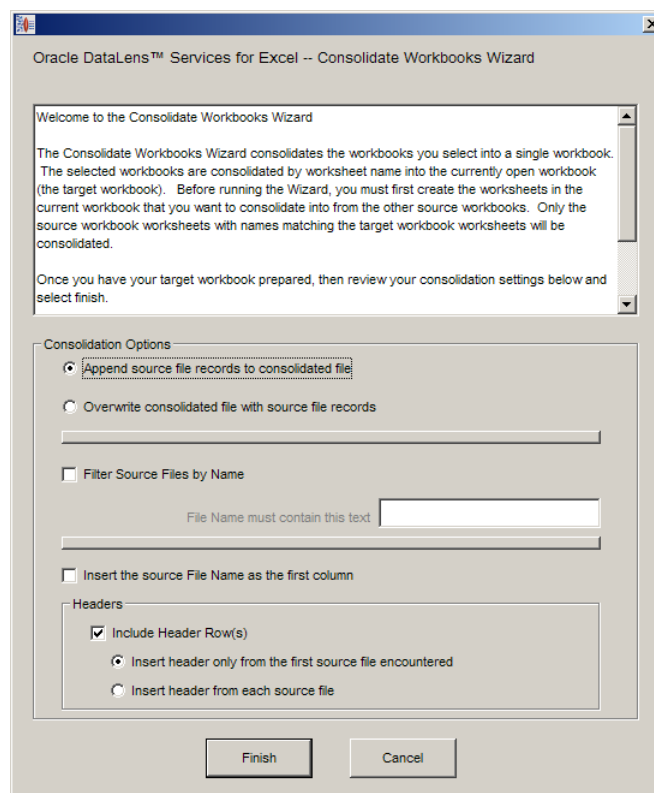
Before beginning, you should do the following:

1. Ensure that all the source files that you want to consolidate are located in a single folder.
2. Create the single target file that you want to merge the source files into.

Ensure that you create properly named worksheets in the target Excel workbook. The Excel workbooks are merged by using the target worksheet names. All source worksheets that match a target worksheet name are merged. All source worksheets that do not match a target worksheet name are ignored.

3. Open the target workbook.

To consolidate files, click the **DataLens Tools** button on the **Services for Excel** toolbar, and then select **Consolidate Files....**



The **Consolidate Workbooks Wizard** dialog box is initialized to the most common default consolidation settings. By default, the system appends the records it finds in all of the source files to the Excel target file.

Use the controls in this dialog box as follows:

Consolidation Options Radio Buttons

Select a radio button to indicate how you want your files consolidated. You can append all data records in the consolidation to the existing data or you can overwrite the contents of the current file with the contents of the consolidation file.

Filtering Source Files Checkboxes and Textbox

When there are many source files mixed in with many other files in the same folder, you can use the file filter function to select the desired files. For this function to be helpful, each of the desired files *must* contain common text in their names.

Select the checkbox to filter the source file names and enter the text that you want to use to filter. For example, you would enter 'item' if all of your source file names had this word somewhere in the file name.

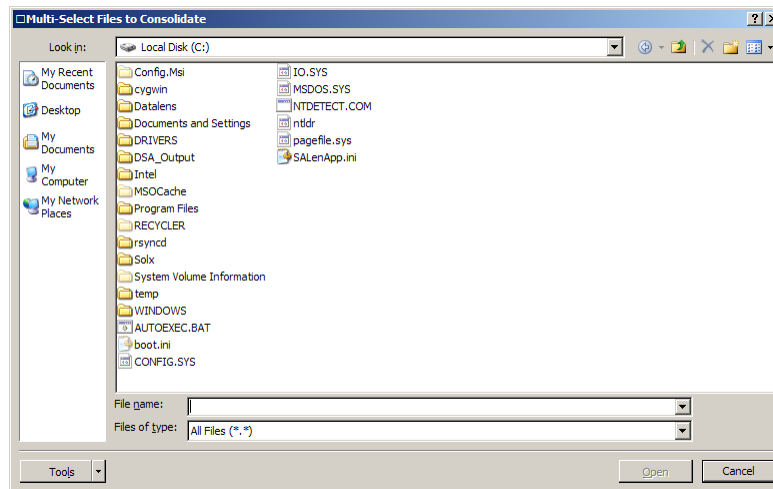
Insert the source File Name as the first column Checkbox

Select the checkbox to insert the source file name as the first column in the consolidated file if you need to track the original source for each record.

Header Checkbox and Radio Buttons

You have full control over how the header is created in the target from the source headers. You can include the source headers or not. If you choose to use the source file header, you can elect to use only the header from the first source file, or you can keep each header row from each source file. If you keep all the header rows, they will be appended to the target file in the order that they are encountered in the source files.

Click **Finish** to begin the file consolidation process.



You are prompted to locate and select the source files that you want consolidated. Use the **Shift** or **Ctrl+A** to select multiple files continuously or discontinuously respectively, and then click **OK**.

A final prompt is displayed so that you can decide whether you want to continue and update the active workbook file. Click **Yes** to begin the file consolidation or **No** to cancel. A progress dialog displays the overall progress if the amount of data records necessitates it.

Changing and Grouping Data

This section describes how you can change or group your data.

Filling Columns with Data

In addition to adding unique data (as described in [Creating the Unique Identifier Column](#) on page 27), you can add data to your source columns. This is particularly useful in adding data to your attribute category columns, which must contain data. The data is completed from the selected cell downward so it is most useful when you select the first cell in the column.

To fill in (complete) the data in a column, select a cell to serve as the starting point. Right-click on the selected cell, select **DataLens Services**, and then select **Fill in Columns**.

request_id	source_desc	att1	att2	att3
2	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN	CHISEL	
3	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN	CHISEL	
1	request_id source_desc	att1	att2	att3
2	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN	CHISEL	
3	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN	CHISEL	
4	200586 Refill Rollball Ink Ble	ROLLER BALL	BLUE	
5	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK	MEDIUM	
6	200602 Ballpoint Pen Medium Point Black Barrel Blk Ink	BLACK	MEDIUM	
7	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK	MEDIUM	
8	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT	BLACK	
9	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT	BLACK	
10	200606 Ballpoint Pen Rfl Point Mdm Black Ink	BALLPOINT	BLACK	
11	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT	BLACK	
12	200608 China Mrkr Green Ink 12 / Box	CHINA	GREEN	
13	200608 China Mrkr Green Ink 12 / Box	CHINA	GREEN	
14	200608 China Mrkr Green Ink 12 / Box	CHINA	GREEN	
15	200611 China Mrk yellow ink	CHINA	YELLOW	
16	200611 China Mrk yellow ink	CHINA	YELLOW	
17	200611 China Mrk yellow ink	CHINA	YELLOW	
18	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
19	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
20	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
21	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
22	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
23	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
24	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
25	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
26	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
27	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
28	200652 Sonnet Roller Ball Pen, Matte Black/Chrome Barrel, Black Ink	BLACK	YELLOW	
29	200683 China Mrkr Green Ink 12 / Box	CHINA	GREEN	
30	200683 China Mrkr Green Ink 12 / Box	CHINA	GREEN	
31	200683 China Mrkr Green Ink 12 / Box	CHINA	GREEN	
32	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
33	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
34	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
35	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
36	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
37	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
38	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
39	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	
40	200689 Refillable Ballpoint Pen, Classic Black with 23K Gold-Plated Accents, Black Ink	BLACK	GREEN	

As in the preceding example, the attribute category column was completed by selecting the first cell in the column then using the fill in function. It operates in much the same way as this similar functionality in Excel as it completes the cells with data based on the data in other cells.

Character Analysis and Replacement

When your data records contain control characters, hexadecimal values, or HTML tags, it is advantageous to remove these characters to ensure proper data processing. These characters can be misinterpreted by Excel so removing them is important.

Analyzing the Character Set

You can use Services for Excel to examine your data records to ascertain whether there are any control characters or Unicode spaces.

To do this, select the data records that you want to analyze and right-click then select **DataLens Services**, select **Character Analysis and Replacement**, and then select **Analyze Character Set**.

The selected cells are immediately examined and a report is provided upon completion.

Replacing Control Characters

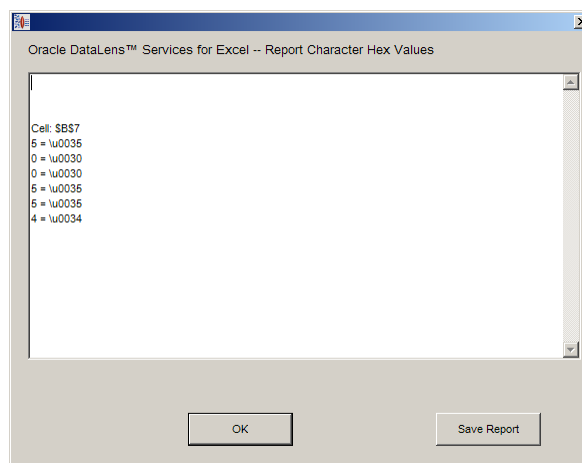
To replace control characters, select the data records that you want to analyze and right-click then select **DataLens Services**, select **Character Analysis and Replacement**, and then select **Replace Control Characters**.

The selected cells are immediately examined, any control characters that are found are replaced with spaces, and a report is provided upon completion.

Searching for Character Hex Values

To search for hex values so that you can modify your source data appropriately, select the data records that you want to analyze and right-click then select **DataLens Services**, select **Character Analysis and Replacement**, and then select **Report Character Hex Values**.

You should review Hex Values in situations where the characters are not visible because there is not an equivalent ASCII character, such as double byte languages or control characters.



The selected cells are immediately examined; any hex values found are reported. You can save the report to a folder so that you can use it to correct these hex values. The saved report is similar to the following:

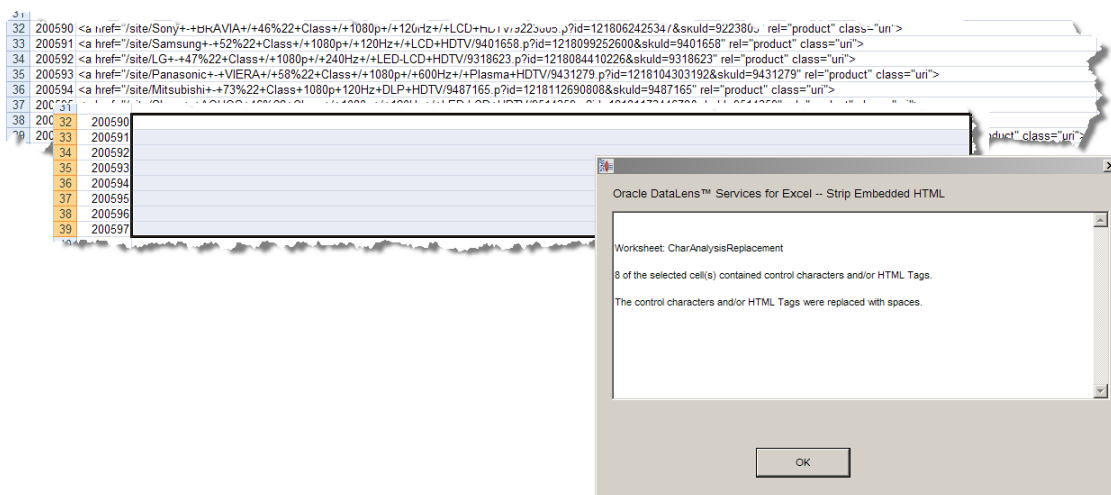
```
Cell: $B$7
5 = \u0035
0 = \u0030
0 = \u0030
5 = \u0035
5 = \u0035
4 = \u0034
```

The cell containing the hex values is identified as are each of the hex values.

Removing Embedded HTML Tags

When your data contains HTML tags, you must remove all tags to enable successful data processing. The tags and their contents including Java scripts, meta data, escape characters, etc. must be stripped from your data.

To remove all embedded HTML tags and control characters, select the data records cells that may contain HTML tags. Right-click the selected cells then select **DataLens Services**, select **Character Analysis and Replacement**, and then select **Strip Embedded HTML Tags from selected cells**.



The selected cells are parsed for HTML tags, which are removed if found. An informational message is displayed that details how many of the cells contained data that was replaced with spaces.

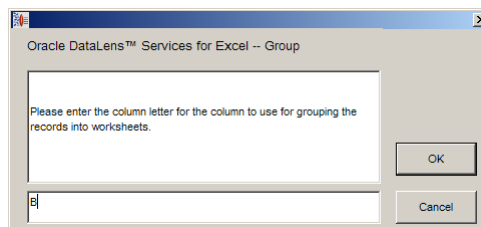
Grouping Records

The record grouping functionality allows you to group all of the records in the active worksheet into a set of separate worksheets based on the values found in one of the input columns. This enables closer analysis of your data in a more granular fashion.

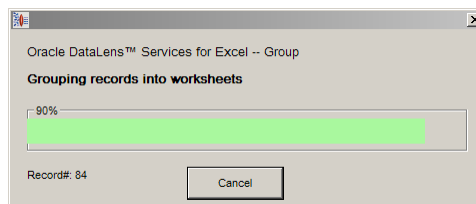
For example, the following data could be grouped by manufacturer.

	A	B	C	D
1	Id	Manufacturer	Part Number	Description
2	200501	Averie	AVE49560	Mechanical Pencil .5MM Emerald Barrel
3	200507	paper mate	CRO800423	Ballpoint Pen Refill Point Mdm Black Ink
4	200512	watermann	CRO8511	China Mrk yellow ink
5	200517	steadler	CRO85132	Dry-Erase Chisel Point Markers Black Ink 2 / Set
6	200523	MMM	NIB-014541172	Dry Erase Markers Chisel Tip Red Ink
7	200528	Eversharp Corporation	MNB-40SGCB	Highlighter Chisel Tip Fluorescent Purple
8	200534	Pilot Inc	NIB007542691	Gel Grip Rollerball Pen Refill 1MM Red
9	200544		?????????	Highlighter Chisel Point Fluorescent Green
10	200551	Dix Corp	PAP/56408	Low Odor Dry-Erase Marker Chisel Tip Green Ink
11	200559	eversharp corp	PAP52401N?	Marker Extra-Fine Tip Black Ink
12	200566	sandford	PENBFL5BPAK6	Chisel Point Highlighter Fluorescent Green
13	200572	steadler	PENBKS10EC	Pen Extra-Fine Point Green Ink
14	200576	avery corp	PIL-77215	Pen Rfl Med Bk 10/Bx
15	200580	integgra	PIL-77222	Porous Pen Extra-Fine Point Red Ink
16	200581	boon inter	PIL/77223	Porous Marker Extra-Fine Tip Blue Ink
17	200585	avry	PIL(77232)	Refill Rolleball Ink Black
18	200586	bic corp	PIL#77233	Refill Rollball Ink Blue
19	200587	cross inc	PIL_77234	Ballpoint Pen Retractable Fine Point Black Ink
20	200588		PIL_77234	Ballpoint Pen Retractable Fine Point Black Ink

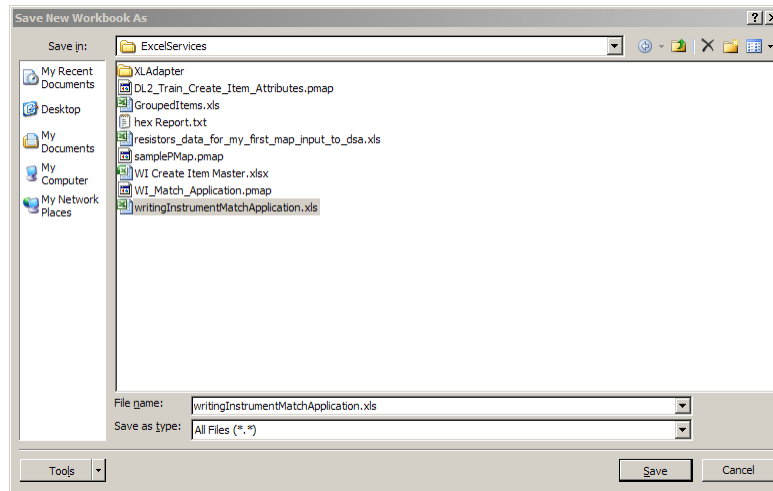
To group records, right-click on any cell then select **DataLens Services**, and then select **Group Records into Worksheets**.



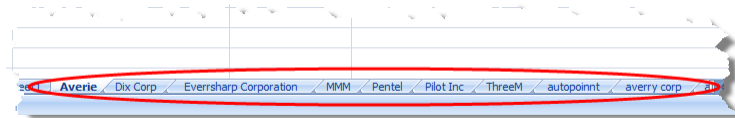
Enter the column you want to use as filter for grouping the records; it is case insensitive. Click **OK** to continue.



Your records are sorted into new worksheets that are created and named by the groups found in the selected column and stored in a new workbook.



You can use the default file name or enter your own, change the directory, and then click **Save** to complete the process. The new workbook containing a worksheet for each group of data will be similar to the following:



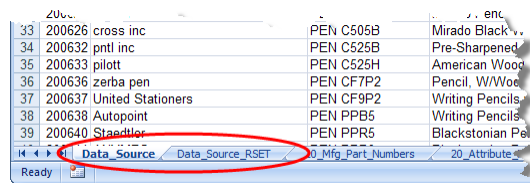
Using Regression Testing

The purpose of regression testing is to validate the quality of the data and changes in the output data. This important step should be performed after changes have been made to the source data.

A regression set is tied to the source data worksheet from which it is created. A good regression base is typically large, usually composed of more than a thousand records. You should recreate your regression set each time you change your data lens or DSA to ensure that recent data lens or DSA changes are apparent.

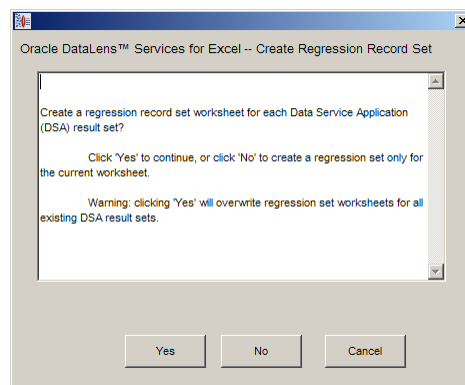
Creating or Updating the Regression Set

The first step to regression testing is to create the regression set of data. You can create a regression set of the source data or any result worksheet. To do this, from the **Services for Excel** toolbar, click **DataLens Tools**, and then select **Create Regression Set**.



The regression set is automatically created and stored in a new worksheet named the same as the source data worksheet with an '_RSET' suffix.

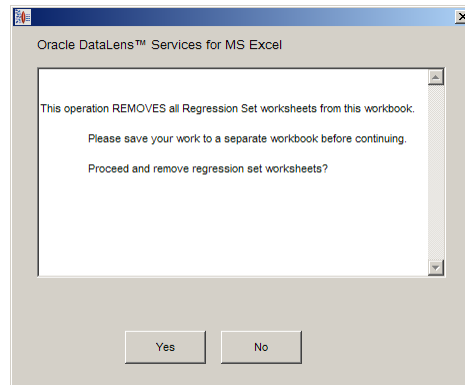
You use this same process to update an existing regression set and when an existing regression record set is detected, the following query appears:



You can overwrite (or create) the existing regression record sets for the active worksheet and all result worksheets by selecting **Yes**; selecting **No** only overwrites the regression set for the active worksheet. Use the **Cancel** button to exit without changing any worksheets.

The data comparison functions in the following sections can be used to compare your source data against your regression data.

All regression set worksheets can be removed at one time so that you can create a new set. To do this, from the **Services for Excel** toolbar, click **DataLens Tools**, and then select **Remove All Regression Sets**.



A confirmation query is displayed and when you click **Yes**, all regression worksheets are removed from your workbook.

Adding Cells to the Regression Set

Once you have reviewed and approved a set of cells in a result worksheet, you can update the corresponding regression worksheet so that you can track your work. This allows you to see your data testing progression, particularly in large regression sets.

To do this, select the cells of data that you want to copy. Right-click in the selection then select **DataLens Services**, and select **Copy Cells to Regression Set**.

22	200613	2	Exact Match Std Mfg	trend ent	Trend Enterprises	PAP90062
23	200614	3	Exact Match Std Mfg Std PN	vis org	Visual Organizers	PEN 07637
24	200618	3	Exact Match Std Mfg Std PN	pickett	Pickett	PEN 507HB
25	200620	3	Exact Match Std Mfg Std PN	pmcomp	PM Company	PEN 50B
26	200621	3	Exact Match Std Mfg Std PN	logitek	Logitech	PEN C25BPHB3K6
27	200633	3	Exact Match Std Mfg Std PN	pilot	Pilot	PEN C525H
28	200644	1	Match on Raw Mfg Raw PN	Avery	Avery	SAN02181
29	200645	1	Match on Raw Mfg Raw PN	Duck	Duck	SAN02199
30	200646	1	Match on Raw Mfg Raw PN	3M	3M	SAN03073
31	200647	1	Match on Raw Mfg Raw PN	Tombow	Tombow	UNV52905
32	200653	1	Match on Raw Mfg Raw PN	Mont Blanc	Mont Blanc	CRO8753
33	200654	1	Match on Raw Mfg Raw PN	Papermate	Papermate	KOH285
34	200655	2	Exact Match Std Mfg	Perker Pen	Parker Pen	KOH287
35	200656	3	Exact Match Std Mfg Std PN	Pentel	Pentel	NIB-003077885
36	200657	3	Exact Match Std Mfg Std PN	Pilot	Pilot	NIB-013174222
37	200658	3	Exact Match Std Mfg Std PN	Sanford	Sanford	NIB-013188641
38	200663	1	Match on Raw Mfg Raw PN	Koh-I-Noor	Koh-I-Noor	PAP64896
39	200666	2	Exact Match Std Mfg	Sparko	Sparco	PAR12182
40	200670	1	Match on Raw Mfg Raw PN	3M	3M	PIL70001
41	200677	3	Exact Match Std Mfg Std PN	Visual Organizers	Visual Organizers	ZEB83211--
42	200678	2	Exact Match Std Mfg	Faberr Kastel	Faber Castell	ZEB83711

The selected cells, or records, are copied to the same location in the regression set worksheet as in the preceding example.

Comparing Worksheets

This feature compares the records found in the active worksheet to the corresponding records in the regression set worksheet (same worksheet name with an '_RSET' suffix). You can compare source data or result worksheet against a regression worksheet.

A matching record is found by use of the record identifiers found in column A of both worksheets. Once a matching record is found, a cell-by-cell comparison of data values is done. The following example compares a result worksheet against the corresponding regression worksheet:

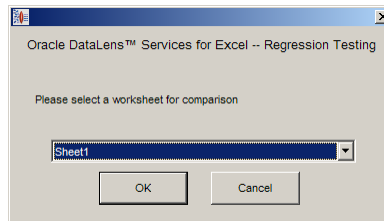
The screenshot displays an Excel spreadsheet with the following columns: **q_part_number**, **source_desc**, and **matched_part**. Two rows are highlighted in yellow, indicating differences. Red arrows point from these rows to the 'Regression Test Results' dialog box.

The dialog box, titled 'Oracle DataLens™ Services for Excel -- Regression Testing', shows the following information:

- Regression Test Results**
- Date/Time: 11/25/2009 2:02:37 PM
- Worksheet Comparison Summary**
- Workbook: writingInstrumentMatchApplication.xls
- Total Time: 0:00:00
- Worksheet -> 10_Mfg_Part_Numbers compared to -> 10_Mfg_Part_Numbers_RSET
- Unique id based comparison
- 2 Different Row(s) found
- Row: 3
- Row: 12
- Time to compare worksheets: 0:00:00 seconds
- Environment:**
 - User / Client: lvalled / LVALLAD-T60
 - Date / Time: 11/25/2009 2:02:37 PM
 - Excel Workbook: writingInstrumentMatchApplication.xls
 - Windows Version: Windows XP (32-bit)
 - Excel Version: MS Excel 2007
 - DataLens Services for Excel Add-In: 5.5.563
 - DataLens System DLL: 5.5.07.02
- Buttons: OK, Save Report

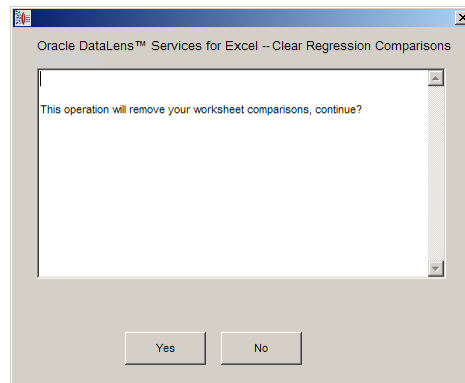
The comparison highlights any cells that contain differences with a yellow background and the report lists all differences, by row, so that you can easily find them. Use the **Save Report** button to save the information to a text file for later review or printing.

If the active worksheet is a regression or data source sheet, you are prompted to select the worksheet you want to compare the active worksheet with as in the following:



The comparison operates in the same manner as previously described.

You can remove the comparison highlighting or the comparisons on the regression worksheet only and leave the highlighting in the result worksheet. To do this, from the **Services for Excel** toolbar, click **DataLens Tools**, and then select **Clear Regression Set Comparisons**.



Select **Yes** to remove only the comparison highlighting on the result worksheet; click **No** to remove only the comparison data on the regression worksheet.

Comparing Selected Records

You can compare a selection of records to quickly isolate data of interest and avoid comparing all the records in the regression set. To do this, select the rows of interest on a source data or result worksheet then right-click, select **DataLens Services**, and then select **Compare Selected Records**.

The records are compared and any differences are highlighted in yellow as in the following example, which selected the records 2 through 10:

request_id	source_desc	att1
2	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
3	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
4	200586 Refill Rollball Ink Blue	ROLLER BALL
5	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
6	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
7	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
8	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
9	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
10	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
11	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
12	200608 China Mrk Green Ink 12 / Box	CHINA
13	200608 China Mrk Green Ink 12 / Box	CHINA
14	200608 China Mrk Green Ink 12 / Box	CHINA
15	200611 China Mrk yellow ink	CHINA
16	200611 China Mrk yellow ink	CHINA
17	200611 China Mrk yellow ink	CHINA

To clear the comparison highlighting, select the rows you want to return to normal then select **DataLens Services**, and then select **Clear Comparisons for Selected Cells**. The colorization is removed from the selected cells as in the following example:

request_id	source_desc	att1
2	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
3	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
4	200586 Refill Rollball Ink Blue	ROLLER BALL
5	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
6	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
7	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
8	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
9	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
10	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
11	200606 Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
12	200608 China Mrk Green Ink 12 / Box	CHINA
13	200608 China Mrk Green Ink 12 / Box	CHINA
14	200608 China Mrk Green Ink 12 / Box	CHINA
15	200611 China Mrk yellow ink	CHINA
16	200611 China Mrk yellow ink	CHINA
17	200611 China Mrk yellow ink	CHINA

The cell in row six was selected and so the highlighting was removed.

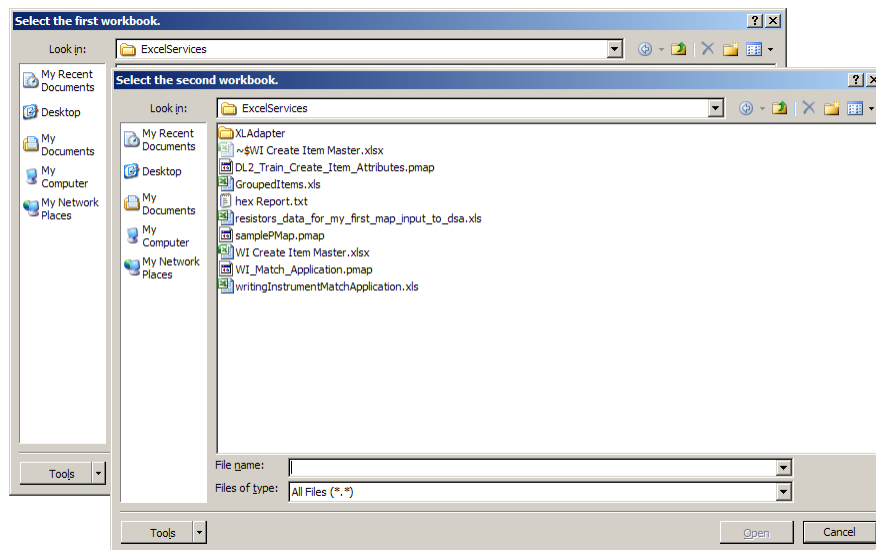
Comparing Workbooks

When comparing entire workbooks, the comparison is high-level, so may not contain very much detail; it operates as follows:

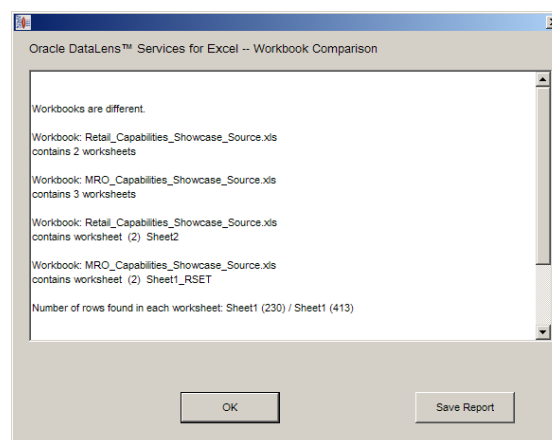
1. Identifies whether the two workbooks have the same number of worksheets.

2. The individual worksheet names are compared from left to right as they appear in the workbooks.
3. The contents of worksheets with the same name are compared and the first cell difference is reported.

To compare two workbooks, right-click anywhere in the active worksheet, select **DataLens Services**, and then select **Compare Workbooks**.



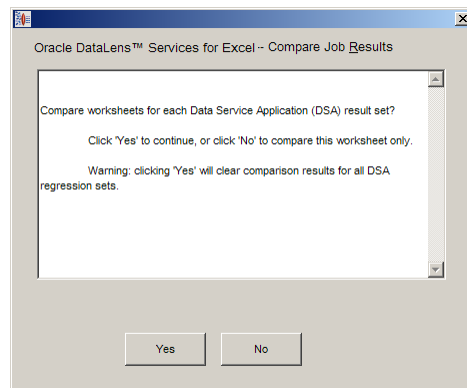
Locate and select the first workbook, and then the second workbook. A report is created detailing the differences that are found as in the following sample report:



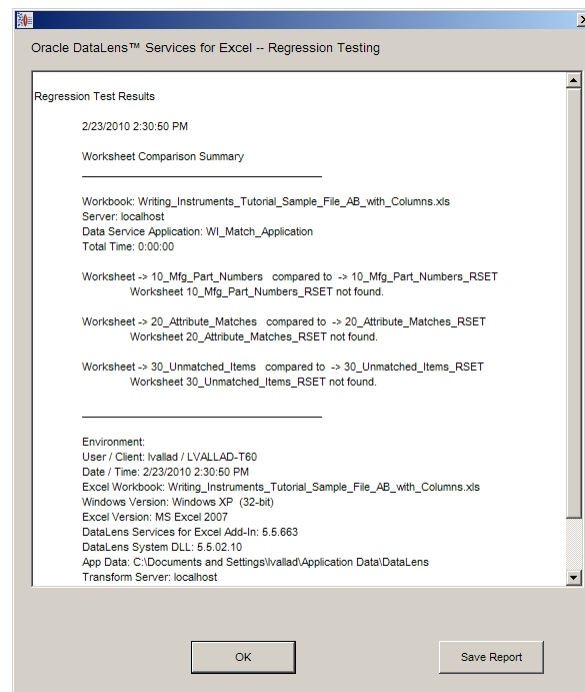
The report lists all differences found between the workbooks. Use the **Save Report** button to save the information to a text file for later review or printing.

Compare Job Results

You can compare the processed data on all result worksheets or just one. In an active results worksheet, from the **Services for Excel** toolbar, click **DataLens Tools**, and then select **Compare Job Results**.



Select **Yes** to compare all result worksheets and clear any previous comparison results or select **No** to compare the active worksheet only.



The report lists all differences found between the result worksheet and its regression worksheet divided by each sheet. Use the **Save Report** button to save the information to a text file for later review or printing.

Data Sampling

It can be advantageous to create a sampling of your data records to test or debug your DSA, or visually inspect your processed data. By identifying a subset of data records, you can save processing time. The data samples are created with a random set of records. You can use these sampling methods:

Key Record Sample

The *key* record set is created with at least 20 randomly selected records, though not more than 1% of the total number of records, taken from the active worksheet. You can use a key record set worksheet to test the basic functionality of the selected DSA and associated data lenses. When the data source contains less than 20 records, all records are used.

To create a key sample, select the worksheet that for which you want a key data sample. Click the **DataLens Tools** button, select **Sampling**, and then select **Create Key Record Sample**.

The data sample is automatically created and is stored in a sheet that has the same name as the active worksheet with `_KSET` appended to the end. For example, if the active worksheet name were `Source_Data` then the key sample worksheet created would be named `Source_Data_KSET`.

Baseline Record Sample

The *baseline* record set is created with at least 1000 randomly selected records, though not more than 8% of the total number of records, taken from the active worksheet. It is used to test your data rather than the data transformation process. When the data source contains less than 1000 records, all records are used.

To create a key sample, select the worksheet that you want a key data sample of. Click the **DataLens Tools** button, select **Sampling**, and then select **Create**

Baseline Record Sample.

The data sample is automatically created and is stored in a sheet that has the same name as the active worksheet with `_KSET` appended to the end. For example, if the active worksheet name were `Source_Data` then the key sample worksheet created would be named `Source_Data_KSET`.

When you have created a data sampling worksheet, you can then set up a DSA job to process the data so that you can review it more easily.

If you attempt to create another sampling of the same data, you are prompted to choose whether you want to overwrite your existing sample.

You can use a standard worksheet deletion in Excel to remove any data sampling worksheets that you create.

Source Data Backup

You can create a backup of your source data worksheet to ensure that you do not lose any of your work. This backup worksheet created and updated automatically when you select the **Backup Worksheet** option when setting up your DSA processing job. For more information about setting this option, see Advanced Settings Tab on page 24.

Note: This backup worksheet is a hidden worksheet to avoid accidental deletion.

The following functions, found on the **Jobs** menu, are used in conjunction with the backup worksheet:

Restore

Restores the source data worksheet using the latest backup worksheet. Since a restore potentially overwrites the active worksheet, if there is an associated hidden backup worksheet, you are prompted to proceed with overwriting the data.

Done

Removes the backup worksheet, regression worksheet, and all other hidden worksheets created by the Services for Excel application. This is useful if you no longer want to use the workbook and are archiving the contents or if you want to share the workbook with others without providing all of the processing information.

Merging Data Lens Standardizations

The ability to merge standardization types within a data lens is a powerful tool supplied by Services for Excel application. You can add the phrase and terminology standardizations from one type to another within a data lens easily with this function. For example, you could build a global standardization type for use throughout your data lens to ensure that the phrase and terminology structures are standard in all standardization types.

Note: You should ensure that you have opened, saved, and checked in your data lens before attempting to merge your standardizations to be able to recover the previous version if necessary.

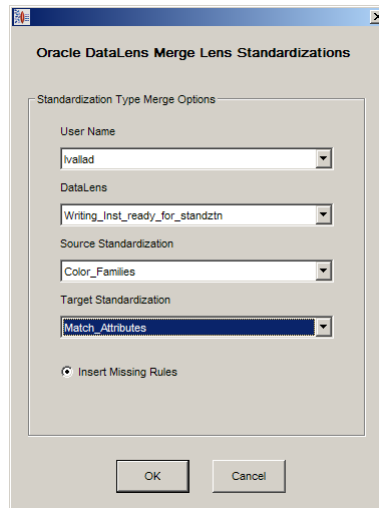
The merging of standardization types includes a number of sections and operates in a top-down processing order to merging all top-level sections of the source data lens standardizations into the target data lens standardization. If conflicts occur, the standardizations in the source data lens are used and those in the target are ignored.

Note: The exception is that the attribute order is not currently supported.

The following data lens sections are affected during the merge:



From the **Services for Excel** toolbar, click **DataLens Tools**, and then select **Merge DataLens Standardizations....**



Use the **AutoBuild: Merge DataLens Standardizations** dialog box as follows:

User Name

Select the user name that this DSA job should be run as; the default is the user that is logged into the Oracle DataLens Server.

DataLens

Select the data lens that contains the standardization types that you want to merge.

Source Standardization

Select the standardization type that contains the phrase and terminology structures that you want to merge into another type.

Target Standardization

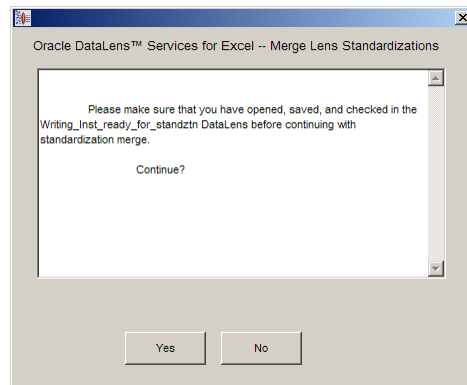
Select the standardization type into which the standardization information will be merged.

Insert Missing Rules

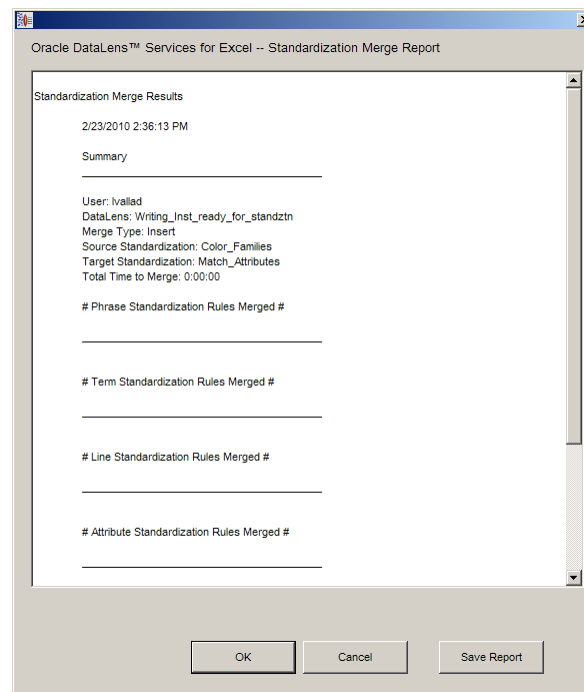
Selected by default, this option ensures that any missing phrase or terminology rules from the source standardization type are merged into the target.

Clearing this radio button can result in incomplete rules. In addition, it allows a global standardization to be created and applied to each data lens so that only the data lens's grammar is standardized, but the grammar is not expanded.

Clicking **OK** begins the merging process and displays a confirmation message.



Clicking **Yes** continues the process; clicking **No** discontinues the process and no changes are made. Once the merging process is complete, a merge report similar to the following is displayed.



You can review the details about each rule that was merged, as well as general merge statistics. The merge report can be saved to a text file by clicking **Save Report** and entering a file name.

Oracle DataLens AutoBuild Application

The purpose of the Oracle DataLens AutoBuild application is to extract implicit and explicit taxonomy information from existing product information and then use that extracted information to create a data lens specific to your enterprise content. The application relies upon product information contained in Excel spreadsheets and is available with the Services for Excel add-in to Oracle Product Data Quality.

With AutoBuild, you can leverage existing Excel spreadsheet product information and Oracle Product Data Quality Smart Glossaries to create initial data lenses specific to your business requirements. You can then enhance the knowledge captured in your data lenses by using the full suite of Oracle Product Data Quality applications to recognize, standardize, classify, and translate your data as well as for further transformation in accordance with comprehensive business processes that you define.

For details on using the AutoBuild application, see the *Oracle Product Data Quality AutoBuild Reference Guide*.

Chapter 5

Installing the Software

Installing the Oracle Product Data Quality 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

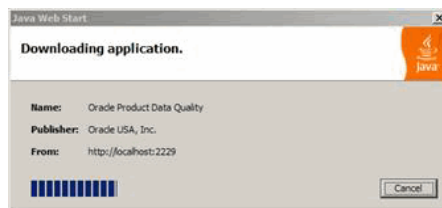
<http://<server>:2229/datalens/datalens.html>

64-bit

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

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

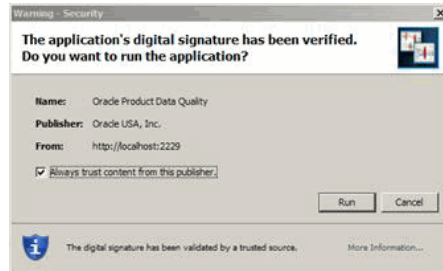
The application download and installation begins. If you do not have a supported Java environment on the target installation machine the Java Web Start program automatically redirects you to a Java download site and begins a Java Runtime installation.



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.



Installing Services for Excel

This section provides detailed software and hardware prerequisites, and installation instructions for the Services for Excel.

Software Prerequisites

Services for Excel has been tested and certified to run on the latest service releases of the following software:

- Windows XP Professional
- Microsoft Internet Explorer 6.0 or greater
- Microsoft Office 2007, 2003 or Office XP
- Adobe Acrobat Reader

Hardware Prerequisites

Services for Excel has been tested and certified to run on the following hardware configurations:

Intel Pentium IV + 2.8 - 3.0 GHz processors (recommended) or better with the following:

- 1 GB RAM
- Standard Hard Drive - 1 GB of free disk space
- CD reader
- Network connectivity

Intel Centrino 1.6 GHz +processors or better with the following:

- 1 GB RAM
- Standard Hard Drive - 1 GB of free disk space
- CD reader
- Network connectivity

Oracle DataLens Server Prerequisites

The following Oracle DataLens Server products must be installed before using Services for Excel.

Required Oracle DataLens Server Software

- Oracle DataLens Server—provides access to DSAs to execute a job immediately and run it synchronously.

Optional DataLens Application Software

DataLens Application Studio—used to create and maintain DSAs.

DataLens Governance Studio—used by Data Stewards, Product Specialist, and Commodity Specialist to manage and process data using the Oracle DataLens Server.

Installing the Application

Install Services for Excel using section appropriate for your client's OS.

Installing on Windows 32-bit:

Use the following steps to install the product:

1. Ensure that Excel is closed.
2. Obtain and unzip Oracle Product Data Quality as described in *Oracle Product Data Quality Oracle DataLens Server Installation Guide* and make it available in a shared directory on a server so that client workstations can access to the installation files.
3. Ensure that the `c:\dls_install\install` directory created in the previous step contains `dlsForExcel_xp.zip` as this file contains all installation files.
4. Locate and retrieve (if necessary) the `dlsForExcel_xp.zip`, and then unzip it in the same directory.
5. Uninstall any existing version of Services for Excel using the **Windows Control Panel**.

If you cannot uninstall Services for Excel using Control Panel, execute the following steps:

- a. Locate **MS-DOS Command Prompt** (`cmd.exe`), right-click on it, and then select **Run as administrator**.
 - b. Execute the following commands:

```
cd c:\DataLens\applications\lib
Regsvr32 /u DataLensServicesForExcel.dll
Regsvr32 /u SCSServices.dll
```

A confirmation for each `Regsvr32` command is displayed indicating that the DLL was successfully unregistered.
6. Start Excel and ensure that the **Services for Excel** Add-In has been removed.
 7. Exit Excel.
 8. Locate **MS-DOS Command Prompt** (`cmd.exe`), right-click on it, and then select **Run as administrator**.
 9. Enter the following commands to run the installer:

```
c:\dls_install\install
DataLensServicesForExcel_Install_XP32bit.bat
```

The installer begins and the following is displayed:

```
Oracle DataLens Services for Excel Installation program
Copyright (c) 2001, 2010, Oracle and/or its affiliates. All rights reserved.
...
...
This installation is for Windows XP ONLY
If you are not running Windows XP, please close this window now.
Otherwise, please select Enter to continue.
...
...
Press any key to continue . . .
```

10. Press **Enter** to continue the installation on your 32-Bit XP system.

```
Microsoft (R) File Expansion Utility Version 5.1.2600.0
Copyright (C) Microsoft Corp 1990-1999. All rights reserved.

Expanding datalensservicesforexcel.cab to c:\windows\system32\SCSServices.dll.

Microsoft (R) File Expansion Utility Version 5.1.2600.0
Copyright (C) Microsoft Corp 1990-1999. All rights reserved.

Expanding datalensservicesforexcel.cab to
c:\windows\system32\DataLensServicesFo
rExcel.dll.

Finished installing the Oracle DataLens Services for Excel.
...
You must now launch the WebStart:
"http://<server>:<port>/datalens/datalens.jnlp"
before you are able to access the Excel Adapter.
...
Please close this window after you have read this information.
Press any key to continue . . .
```

11. Press **Enter** to complete the installation.

12. Start Excel.

13. Ensure that the **Services for Excel** Add-In toolbar is present.

The Services for Excel toolbar is identical in all supported versions of though they are presented differently. When using Excel 2007, you must click the **Add-Ins** tab to access the toolbar. The appearance of the toolbar indicates a successful installation of the Services for Excel product.

If the appropriate toolbar is not displayed, contact Professional Services for assistance.

The Services for Excel 32-bit application installation is complete.

Installing on Windows 64-Bit

1. Ensure that Excel is closed.
2. Uninstall any existing version of Services for Excel using one of the following:

For Version 5.5.03 and earlier:

Uninstall using the **Windows Control Panel**.

For Version 5.5.03 and later:

- a. Locate the **DataLensServicesForExcel.dll** file and note the directory. For example, this file is located in C:\Windows\Downloaded Program Files directory on Windows 7.
- b. Locate the **SCSServices.dll** file and note the directory. For example, this file is located in C:\Windows\SysWOW64 directory on Windows 7.
- c. Locate **MS-DOS Command Prompt** (cmd.exe), right-click on it, and then select **Run as administrator**.
- d. Change directories to the directory located in step a.
- e. Execute the following command:

```
Regsvr32 /u DataLensServicesForExcel.dll
```

- f. Change directories to the directory located in step b.
- g. Execute the following commands:

```
Regsvr32 /u SCSServices.dll
```

A confirmation for each `Regsvr32` command is displayed indicating that the DLL was successfully unregistered.

3. Start Excel and ensure that the **Services for Excel** Add-In has been removed.
4. Exit Excel.

5. Locate **Internet Explorer**, right-click on it, and then select **Run as administrator**.

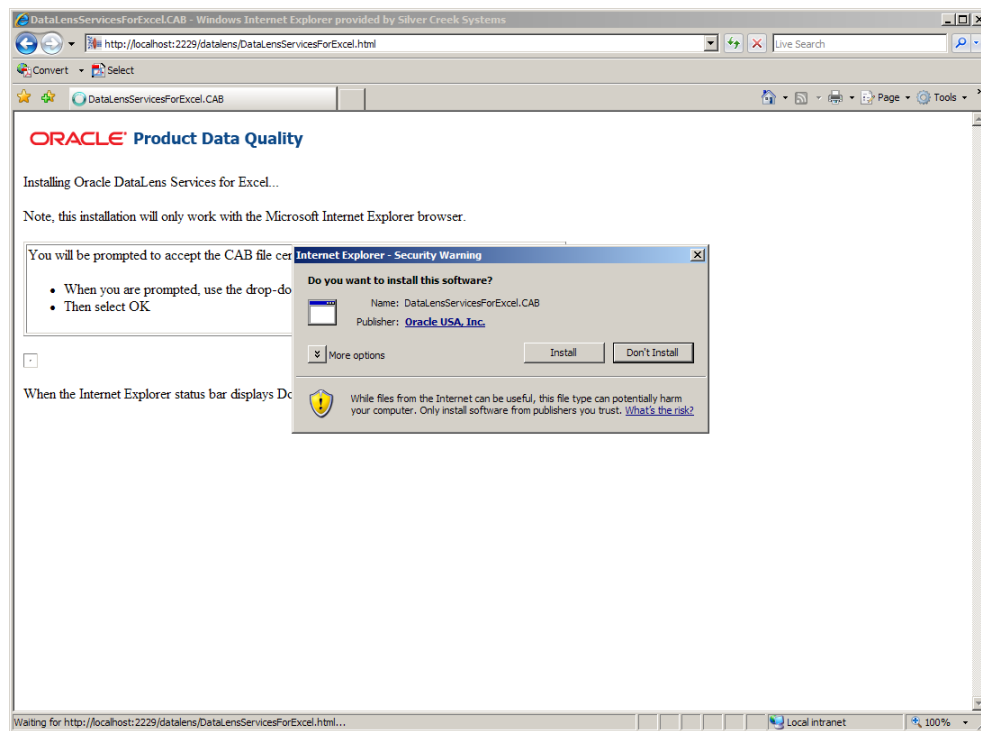
Important: Users with full Administrator privileges are not able to register the Services for Excel MS objects; only an Administrator user can register the necessary MS objects.

6. Browse to the following Web site:

Note: The following URL is case sensitive.

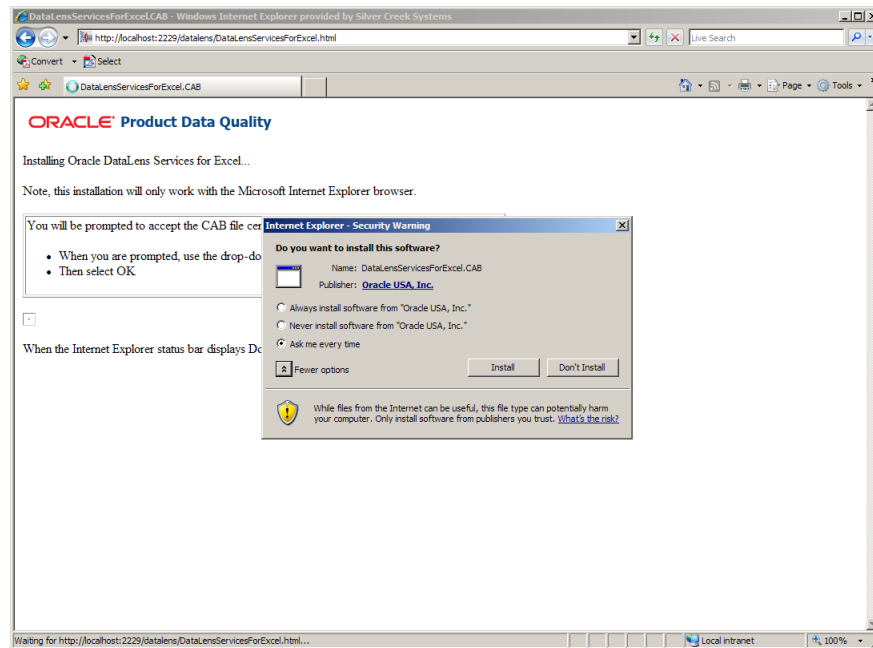
<http://server:2229/datalens/DataLensServicesForExcel.html>

Where *server* is the name of your Oracle DataLens Server. If you setup a different port number for your application server other than 2229, you must use that port number in this URL.

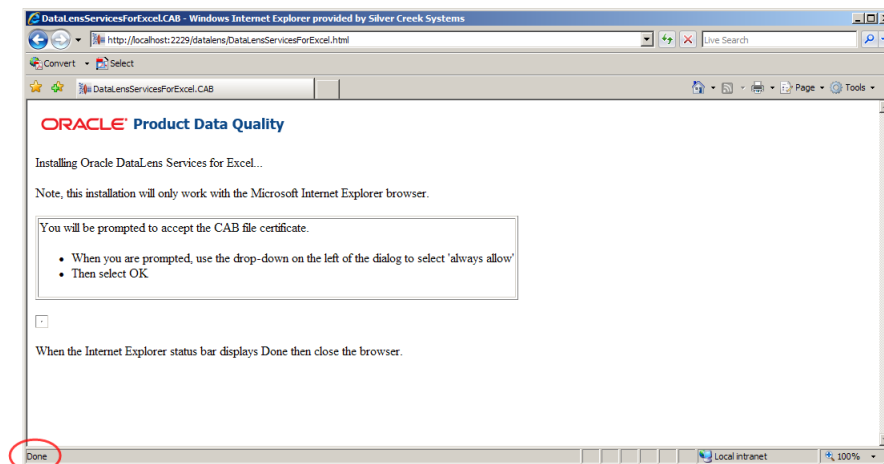


Note: This page is not missing a graphic; it is a placeholder for an object tag that registers MS objects.

- Click the **More Options** arrow.



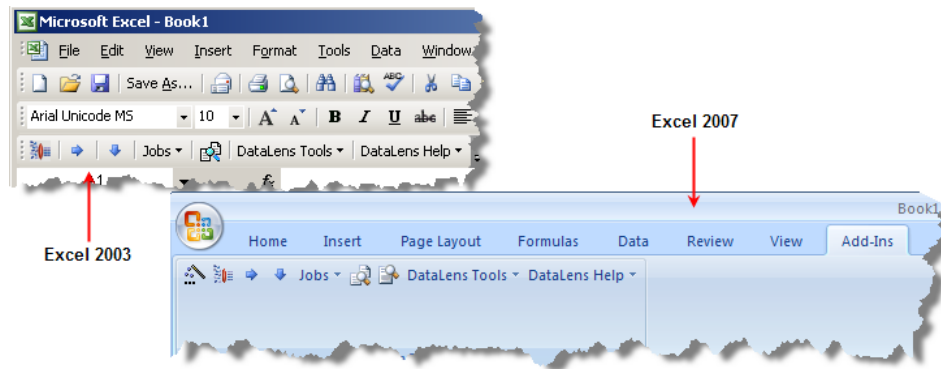
- Select **Always install software from "Oracle USA, Inc."** so that you are not prompted by security warnings throughout the software installation.
- Click **Install** to install the software.



The **DataLensServicesForExcel.dll** and **SCSServices.dll** files are retrieved from the Oracle DataLens Server and are registered.

- Ensure that **Done** is displayed in your browser because this indicates that the installation completed.
- Start Excel.

12. Ensure that the **Services for Excel** Add-In toolbar is displayed as in the following:



The Services for Excel toolbar is identical in both Excel 2003 and 2007 though they are presented differently. When using Excel 2007, you must click the **Add-Ins** tab to access the toolbar. The appearance of the toolbar indicates a successful installation of the Services for Excel product.

If the appropriate toolbar is not displayed, contact Professional Services for assistance.

The Services for Excel 64-bit application installation is complete.

Restoring Installation Settings from a Previous Installation

Optionally, you can restore the settings from a previous installation, by copying the following files:

ExcelServicesResources.xml
system.xml
Term_Library_en_US.xml

From the c:\datalens\applications\config directory to one of the following:

Windows XP

c:\Documents and Settings\<Username>\Applications\DataLens\config\

Windows Vista

c:\Users\<Username>\AppData\DataLens\config\

Windows 7

c:\Users\<Username>\AppData\Roaming\DataLens\config\