

Oracle® Enterprise Data Quality for Product Data

Services for Excel Reference Guide

Release 11g R1 (11.1.1.6)

E29143-03

February 2013

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

E29143-03

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

Primary Author: Lorna Vallad

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

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

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

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

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

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

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

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

Contents

Preface	v
Audience	v
Documentation Accessibility	v
Related Documents	vi
Conventions	vi
1 Introduction	
Toolbar and Menus Explained	1-1
Jobs Menu	1-1
Test Menu	1-2
DataLens Tools Menu	1-3
DataLens Help Menu	1-4
Excel Cell Context-Sensitive Menu	1-4
2 Installing and Configuring Services for Excel	
Preparing for Installation	2-1
Hardware and Software Requirements	2-1
Enterprise DQ for Product Prerequisites	2-1
Required Oracle DataLens Server Software	2-1
DataLens Client Software	2-1
Establishing a Windows Administrator User Account	2-2
Installing Services for Excel	2-2
Known Issues and Workarounds	2-2
Logging into the Oracle DataLens Server	2-3
Changing the Oracle DataLens Server	2-4
Configuring Application Options	2-4
3 Transforming Your Data	
Configuring a DSA	3-1
Setting Up a DSA Job	3-5
Unique Record Identifiers	3-6
Creating the Unique Identifier Column	3-6
DSA Outputs	3-7
The DSA Output Named <i>Output</i>	3-7
Running a DSA to Process Data Records	3-7

Using Real Time Records Processing	3-8
Using Batch Mode Records Processing.....	3-10
Removing Result Data.....	3-11
Viewing the Job Status	3-11

4 Getting More from Your Data

Loading Metadata Files	4-1
Changing and Grouping Data	4-2
Grouping Records	4-2
Filling Columns with Data.....	4-4
Character Analysis and Replacement	4-5
Analyzing the Character Set.....	4-5
Replacing Control Characters	4-5
Searching for Character Hex Values	4-5
Removing Embedded HTML Tags	4-6
Using Regression Testing	4-7
Creating or Updating the Regression Set	4-7
Removing Regression Set Worksheets.....	4-8
Adding Cells to the Regression Set	4-8
Comparing Selected Records.....	4-9
Comparing Job Results.....	4-10
Using Data Sampling.....	4-11
Backing Up Source Data	4-12
Using the AutoBuild Application	4-12

Preface

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

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

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

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

Audience

You should have a basic understanding of the Enterprise DQ for Product Technology, including the functionality of the Enterprise DQ for Product Knowledge Studio and Application Studio applications.

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

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

Documentation Accessibility

For information about Oracle's commitment to accessibility, visit the Oracle Accessibility Program website at <http://www.oracle.com/pls/topic/lookup?ctx=acc&id=docacc>.

Access to Oracle Support

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

Related Documents

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

- The *Oracle Enterprise Data Quality for Product Data Getting Started Guide* provides information about how to get started with EDQP and use the Application Launch Pad to start client applications.
- The *Oracle Enterprise Data Quality for Product Data Oracle DataLens Server Installation Guide* provides detailed Oracle DataLens Server installation instructions.
- The *Oracle Enterprise Data Quality for Product Data Application Studio Reference Guide* provides information about creating and maintaining Data Service Applications (DSAs).
- The *Oracle Enterprise Data Quality for Product Data AutoBuild Reference Guide* provides information about creating initial an data lens based on existing product information and data lens knowledge.
- The *Oracle Enterprise Data Quality for Product Data Knowledge Studio Reference Guide* provides information about creating and maintaining data lenses.
- The *Oracle Enterprise Data Quality for Product Data Governance Studio Reference Guide* provides information about building projects to analyze your transformed data, create reports to show the quality of your data, and identify missing attributes.
- The *Oracle Enterprise Data Quality for Product Data Glossary* provides definitions to commonly used Enterprise DQ for Product technology terms.

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

http://docs.oracle.com/cd/E35636_01/index.htm

Conventions

The following text conventions are used in this document:

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

Introduction

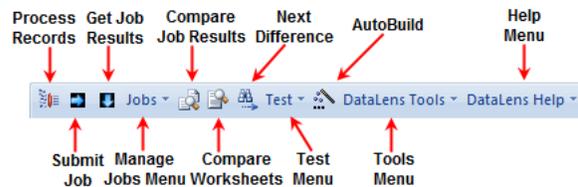
The Enterprise DQ for Product (EDQP) Services for Excel application is designed to provide enhanced, tailored, spreadsheet-based transformations of your data and is used in conjunction with your Oracle DataLens Server platform. Services for Excel enables you to make your corporate data business-ready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

Services for Excel interfaces directly to the Oracle DataLens Server to execute Data Service Applications (DSAs). This is just one way that data can be delivered to a DSA 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 Enterprise DQ for Product. 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 and Configuring Services for Excel](#)" on page 2-1.

Toolbar and Menu Explained

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



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

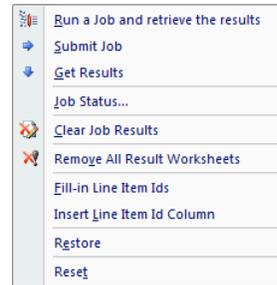
The buttons that are on the toolbar are the most often used functions so are available for quick access; each button is also on one of the menus.

Jobs Menu

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

Run a Job and retrieve the results

Executes a synchronous DSA job to immediately process data records, and returns the results. 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 Data Records"](#) on page 3-7.



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 job. 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. You can obtain the job id using the **Job Status...** option. A worksheet is automatically created for each DSA job run.

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 3-11.

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 created by retrieving DSA job results; 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 3-6.

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 3-6.

Restore

Restores the source data worksheet to its prior state if a hidden backup worksheet is available. For more information, see ["Backing Up Source Data"](#) on page 4-12.

Reset

Removes all hidden backup worksheets and Data Services for Excel configuration worksheets. You are prompted to confirm the deletion of each these worksheets. For more information, see ["Backing Up Source Data"](#) on page 4-12.

Test Menu

The **Test** button and associated menu items provides a selection of utilities that you can use for regression testing and data comparison.

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

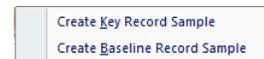
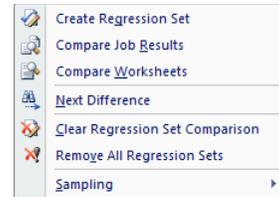
Clears all comparison highlighting on the active worksheet that was created by a regression test.

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 ["Using Data Sampling"](#) on page 4-11.

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 ["Using Data Sampling"](#) on page 4-11.

DataLens Tools Menu

The **Tools** button and associated menu items provides a selection of tools that you can use to open other Enterprise DQ for Product applications, access AutoBuild and create associated reports, and set various Services for Excel options.

Load Meta-Data Files...

Used to consolidate like or similar metadata records from several different files into a single metadata file. For more information, see ["Loading Metadata Files"](#) on page 4-1.

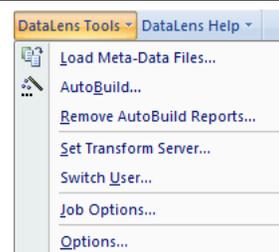
AutoBuild...

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

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 ["Using the AutoBuild Application"](#) on page 4-12.

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 *Oracle Enterprise Data Quality for Product Data AutoBuild Reference Guide*.



Set Transform Server...

Identifies the Oracle DataLens Server that will be used to transform data. For more information, see ["Changing the Oracle DataLens Server"](#) on page 2-4.

Switch User...

Allows you to change the user that is currently logged into the Oracle DataLens Transform Server and you can which server will transform your data. For more information, see ["Logging into the Oracle DataLens Server"](#) on page 2-3.

Job Options...

Configures how your source data will be processed and by which DSA. For more information, see ["Configuring a DSA"](#) on page 3-1.

Options...

Allows you to set options for global use in the Services for Excel. For more information, see ["Configuring Application Options"](#) on page 2-4.

DataLens Help Menu

Product Guide

Opens a list of Enterprise DQ for Product documents for your selection in a browser.

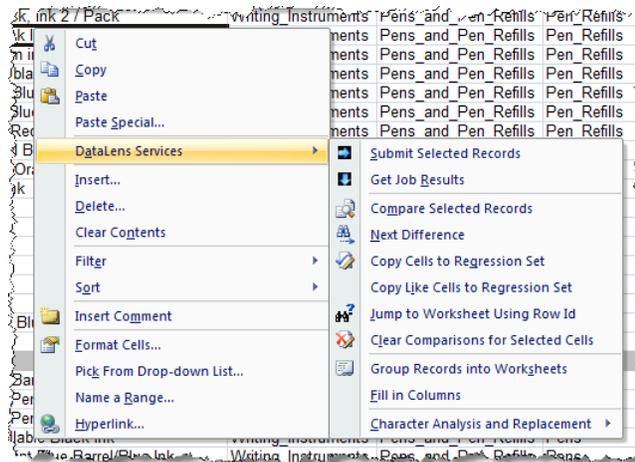


Help About

Provides an informational message about the Services for Excel product. Use the **Environment** button to view and save (to a text file) information about your environment including versions, locale, application data path, user names, transform server, 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:



Note: If you select the entire worksheet, the Services for Excel context sensitive menu is not active.

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](#)" on page 1-1.

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 highlighting indicates a difference in results between the regression base and the recently run data.

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. For more information, see "[Using Regression Testing](#)" on page 4-7.

Copy Like Cells to Regression Set

Copies values like the selected, highlighted cell found on the active worksheet to the regression set worksheet. If more than one cell is selected, only the first cell is used. All comparison highlighting is removed from the copied cells. For more information, see "[Using Regression Testing](#)" on page 4-7.

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 worksheet 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 4-2.

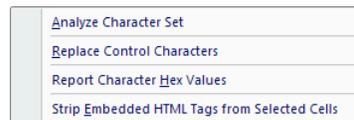
Fill in Columns

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 4-4.

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 4-5.

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 4-5.

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 4-5.

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 4-5.

Installing and Configuring Services for Excel

This chapter describes how to prepare for and install EDQP Services for Excel, log into the Oracle DataLens Server or change this server, and how to configure the application options.

Preparing for Installation

This section provides detailed software and hardware prerequisites, and how to obtain the Services for Excel software in preparation for installing.

Hardware and Software Requirements

You must ensure that the following Oracle DataLens Server hardware and software requirements are observed. These requirements represent the certified and supported server configurations.

Verify that you have met the minimum server requirements using the *Oracle Enterprise Data Quality for Product Data Hardware and Software Specification* found at the Oracle Enterprise Data Quality for Product Data Documentation Web site:

http://download.oracle.com/docs/cd/E20593_01/index.htm

Enterprise DQ for Product Prerequisites

The following Enterprise DQ for Product prerequisites 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. For more information, see *Oracle Enterprise Data Quality for Product Data Oracle DataLens Server Installation Guide*.

DataLens Client Software

Application Studio, which is used to create and maintain DSAs. Governance Studio, which is used by to manage and process data using the Oracle DataLens Server. Knowledge Studio, which is used to create and maintain data lenses. These client applications are installed as part of the Oracle DataLens Server installation and can be accessed from the EDQP Welcome Launch Pad. For more information, see *Oracle Enterprise Data Quality for Product Data Getting Started Guide*.

Establishing a Windows Administrator User Account

A Windows Administrator user account *must* be used to install or uninstall Services for Excel so that the necessary dynamic-link libraries (DLLs) are properly registered or unregistered.

Contact your system administrator for assistance in creating a new or identifying an existing Windows Administrator user account.

Installing Services for Excel

The Services for Excel installer is a standard Windows Installer executable file and installs on both 32-bit and 64-bit Windows operating system client machines running Microsoft Excel 2003, 2007, or 2010. Any previous versions of the software is automatically uninstalled as part of the installation. The Services for Excel installer file is obtained from the Oracle DataLens Server Welcome Web page.

Installing Services for Excel

Use the following steps to install the product on any supported client machine:

1. Ensure that Excel is closed.
2. Start your Oracle DataLens Server to use the Welcome Launch Pad. For details, see *Oracle Enterprise Data Quality for Product Data Getting Started*.
3. Click the **AutoBuild** button to download the appropriate Services for Excel installation file for your version of Excel.

Depending on the Download and Security settings in your browser, you may be asked to save the installer file, run it, or it automatically downloads.

4. If prompted by a Security Warning dialog, click **Run** to run the file and continue the installation.

Otherwise, locate the downloaded **DataLensServicesForExcelSetup.exe** file and double-click it to start the installation.

The installer begins, identifies the Excel version, installs accordingly, completes the installation, and starts Excel.

Note: If a Warning is displayed prompting you to overwrite an existing `DLS_Import_Template.project-date` file, click **Yes** to update the file.

5. 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 or 2010, 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.

Known Issues and Workarounds

The following sections describe known issues and any workarounds that may occur when installing or using Services for Excel.

Excel files do not open when you double-click an Excel workbook or short-cut to an Excel file

When opening Excel and an Excel file by double-clicking on it, Excel starts but the file is not opened. This is a known issue with the Excel application and can be corrected by following the instructions at the Microsoft Support Web site at

<http://support.microsoft.com/kb/211494>

Services for Excel Add-in does not appear in Excel 2003 on a 32-bit Windows 7 system after a successful Installation

This is a result of a problem with the 32-bit Microsoft Office 2003 Primary Interop Assemblies (PIA) including the necessary .Net Programmability Support not being installed and is a known issue. You can correct this issue by downloading and installing following the Microsoft Office 2003 PIA using the instructions at the Microsoft Download Web site at

<http://www.microsoft.com/en-us/download/details.aspx?id=20923>

Logging into the Oracle DataLens Server

You can log in or change the user that is currently logged into the Oracle DataLens Server and change the server to use to transform your data.

Note: You must log in to the server for all job processing activities. The login applies to all spreadsheets within the workbook. If you want to use a different login then use the **Switch User** option on the **DataLens Tools** menu. You can also open an entirely new workbook to log in as a different user.

To change the user or Oracle DataLens Server, click **DataLens Tools** and select **Switch User...**



The screenshot shows a login dialog box titled "ORACLE ENTERPRISE DATA QUALITY FOR PRODUCT DATA Powered by DataLens Technology". It includes the following fields and controls:

- User ID: [Empty text box]
- Password (Case Sensitive): [Empty password box]
- Remember Password
- Oracle DataLens Administrator (Name or IP Address): localhost
- Port: 2229
- Use HTTPS
- OK and Cancel buttons at the bottom.

The Oracle Enterprise Data Quality for Product Data Login dialog box appears.

Switching Users

Enter a valid user name and password and click **OK**. You can avoid entering this password every time you logon as this user by selecting the **Remember Password** check box.

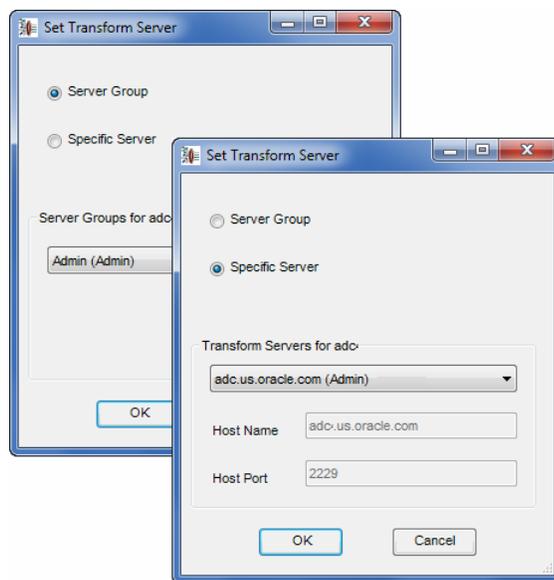
Changing the Oracle DataLens Administration Server

Click the **Change Server** button. Enter the name or IP Address of the Oracle DataLens Administration Server that you want to use to transform your data. If the server does not run on the default port number 2229, enter the correct port number. Click **OK** to change servers.

Changing the Oracle DataLens Server

You can specify a server group or a specific server to use to transform your data.

To change the user or Oracle DataLens Server, click **DataLens Tools** and select **Set Transform Server...**



Set the Oracle DataLens Transform Server by selecting one of the following:

Server Group

Select the server group from the **Oracle DataLens Server Group** list, and then click **OK** to change the server group.

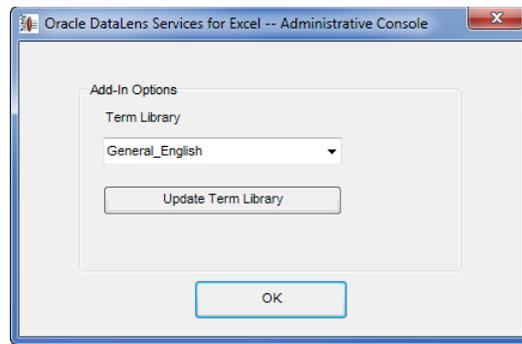
Specific Transform Server

Select a server from the **Oracle DataLens Server Group** list. If the server runs does not run on the default port number 2229, enter the correct port number. Click **OK** to change servers.

Configuring Application Options

You can configure a Term Library to augment and enhance your data lens grammar. A Term Library is composed of the term variants found in an existing data lenses. You can either create a new Term Library or add to an existing Term Library .

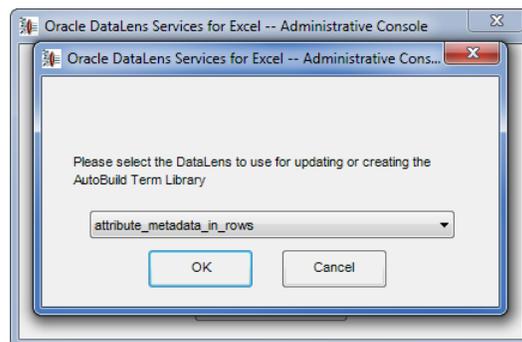
1. On the Services for Excel toolbar, click **DataLens Tools**, and then select **Options**.



2. Select **<Create New>** to create a new Term Library or select a Term Library from the list to update an existing library.

The Term Library you create or update is used by AutoBuild by extracting terms from the data lens that you select. This is particularly useful when you have extensive term variant refinements (in fact, term variant refinement are the only type of refinement used in Term Libraries) in a data lens that you want to leverage in new data lens.

3. Click **Update Term Library**, and then select the data lens that you want to use to create an initial one or update the existing Term Library.
4. Click **OK**.



The Term Library is created or updated to include these new terms and variations. The extracted terms are case insensitive, all underscores, plurals, and duplicates (containing identical variants) are removed.

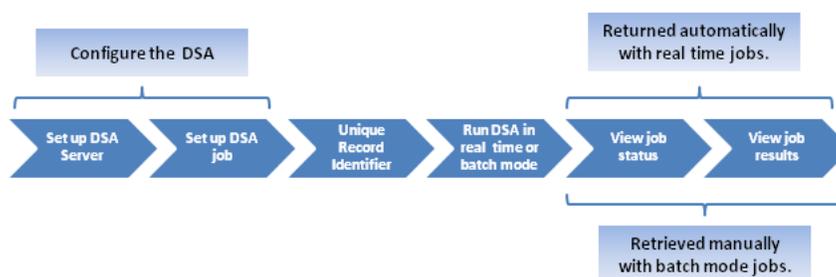
After creating a new Term Library it is automatically stored in the correct location required by AutoBuild, and then you can update the Term Library one data lens at a time. In Step 1 of the AutoBuild wizard, you can make your Term Library selection by clicking the **Category Details** button.

Upon completion, a report is displayed that details all of the terms that were updated for you to review and save.

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

Transforming Your Data

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



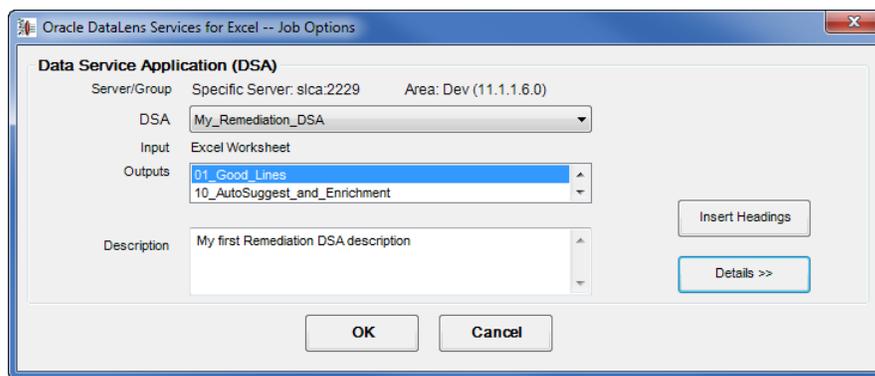
All of these steps are described in this chapter.

Configuring a DSA

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

Always start by opening the desired Excel workbook, and then selecting the specific worksheet 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 Options...** option.



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

The basic job options and server information is displayed.

DSA List

This list is automatically populated with the names of the DSA available on the server for processing data. You can select a different DSA from the list of available DSAs.

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. The use of DSAs can be restricted to specific users using the **Oracle DataLens Server Web** pages and Application Studio. All DSAs are displayed in the **DSA** list including restricted DSAs. If you select a restricted DSA and do not have permission to run it, an error is displayed and a job is not created when you attempt to process your data records. For more information about how to restrict a DSA, see *Oracle Enterprise Data Quality for Product Data Application Studio Reference Guide* and *Oracle Enterprise Data Quality for Product Data Oracle DataLens Server Administration Guide*.

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

Result Set List

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 and section title displayed in the **Result Set for Output** section.

Description Field

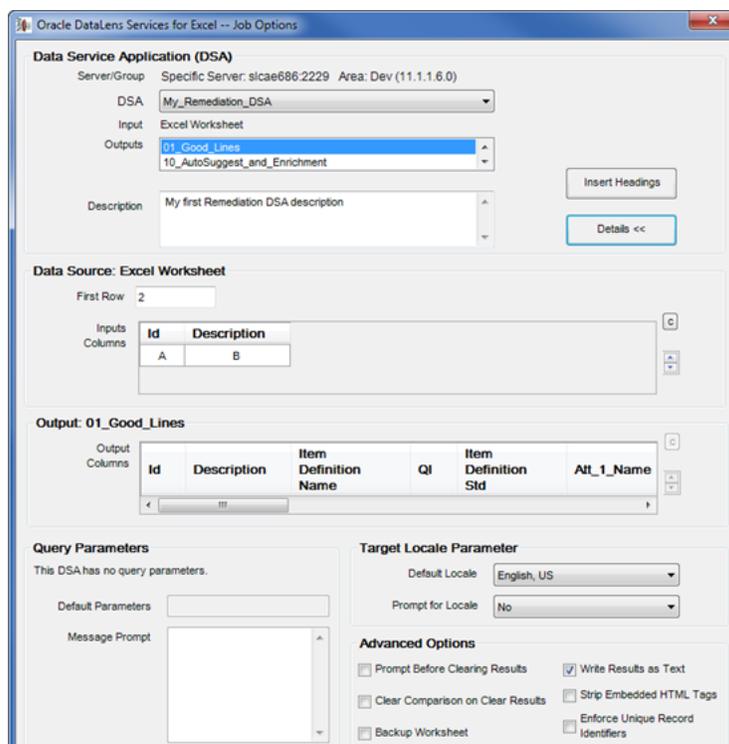
This field is informational and supplies the description of the DSA that you are using to process your data. It is blank if no description for the DSA exists.

Insert Headings Button

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

Details Button

All other job options are initialized with default values and typically do not require your attention though you can change them by clicking **Details >>**.



The options you can change are described as follows:

Data Source: Section

The type of data source is displayed as part of the title of this section and is based on the properties of the selected DSA. The data sources can be Excel, file input, or 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. In this example, the title is **Data Source: Excel Worksheet** thus indicating the data source is Excel.

First Row Field

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 worksheet match the input data columns expected by the DSA you have selected by shifting or editing these columns.

Result Set for Output Section

The selected Result Set is displayed as part of the title of this section. In this example, the title is **Result Set for Output LOAD** thus indicating the selected 'LOAD' Result Set.

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. The 'Output' result set is written directly back to the input worksheet.

Query Parameters Section

This section contains any DSA query parameter options.

Default Parameters Field

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.

Message Prompt Field

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.

Target Locale Parameter Section

This section contains language translation options.

Default Locale List

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 List

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**.

Advanced Options Section

This section contains advance job option settings.

Prompt Before Clearing Results Check Box

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

Clear Comparison on Clear Results Check Box

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

Backup Worksheet Check Box

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

Write Results as Text Check Box

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.

Strip Embedded HTML Tags Check Box

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.

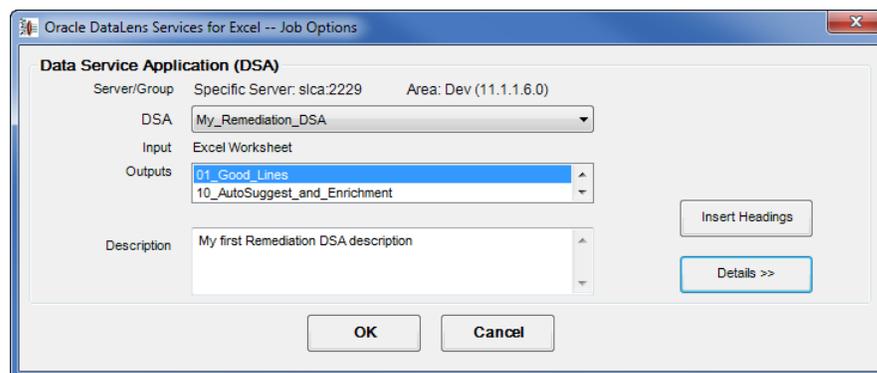
Enforce Unique Record Identifiers Check Box

Enables the verification that column A contains unique identifiers before the job is run to avoid job errors.

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 Options....**



3. 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.
4. Modify any other options on this tab or the **Advanced Settings** tab.
5. Select **OK** on either tab to accept your changes.
6. 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

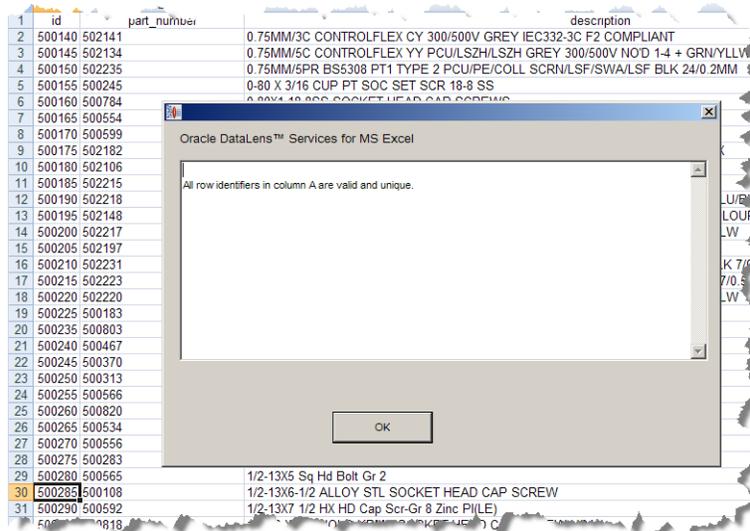
The Services for Excel Add-In makes it easy to add a unique identifier column if one does not exist. From the **Services for Excel** toolbar, click **Jobs**, and then select **Insert Line Item Id Column**. This automatically inserts a new column A to the left of your existing data columns and then inserts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

	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	CF085122	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 **Job Options** dialog. For more information, see "[Advanced Options Section](#)" on page 3-4.

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 worksheet 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 worksheet headers in row 1 are automatically created from the DSA output columns. While the data values are overwritten each time results are retrieved, any worksheet 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 Data 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.

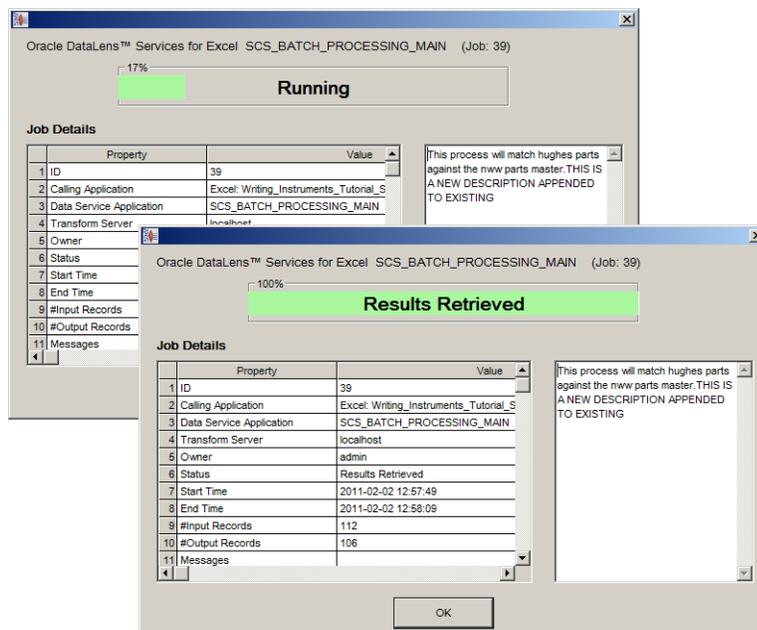
The worksheet columns that are used for input to the job and where the results are placed back into your Excel workbook after the job has completed are all determined by the design of the DSA. For more information, see *Oracle Enterprise Data Quality for Product Data Application Studio Reference Guide*.

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 **Run Job** button on the 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 Mrkr Green Ink 12 / Box	
11	W113	China Markers red ink	
12	W114	China Markers White ink	
13	W115	China Mrk yellow ink	
14	W116	Chisel Point Highlighter Turquoise ink	
15	W117	Comfort Grp Retractable Ballpoint Pen medium Point Blue ink	
16	W118	Comfort Grp 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 processing this data, the results are placed in new worksheets as in the following:

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

Sheet1 | 60 Exceptions | 87 Confirmed Cross Matches | 90 Ready for Load

Input Data Worksheet Result Set Worksheets

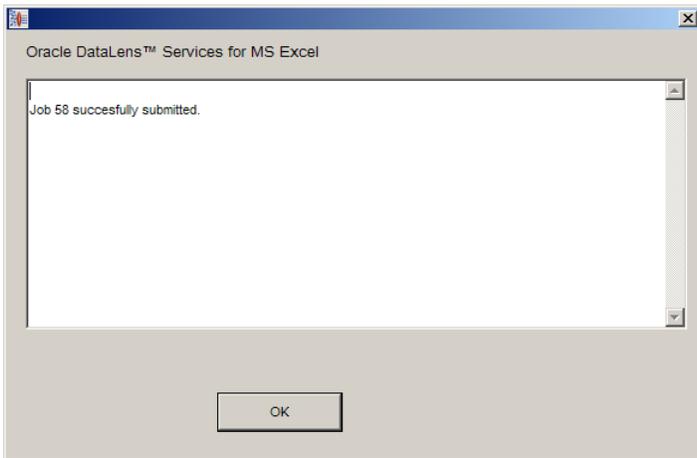
The original input data has been transformed, exceptions are identified, and subsequent columns contain any extracted attribute information.

Using Batch Mode Records Processing

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

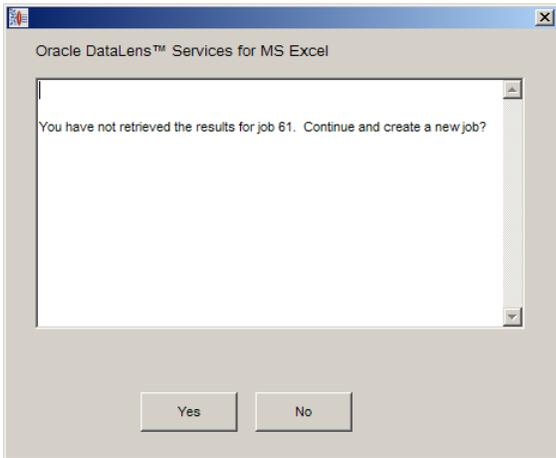
While a batch mode job is running, you can use Excel for your other tasks. The results of job can be retrieved later. After a job is submitted, you can even save the Excel workbook and close Excel, and when you return to the same workbook, you can retrieve the job results.

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:



To check on the status of your job, view the job status as described in "[Viewing the Job Status](#)" on page 3-11. 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 pages to check on job status. For more information, see *Oracle Enterprise Data Quality for Product Data Oracle DataLens Server Administration Guide*.

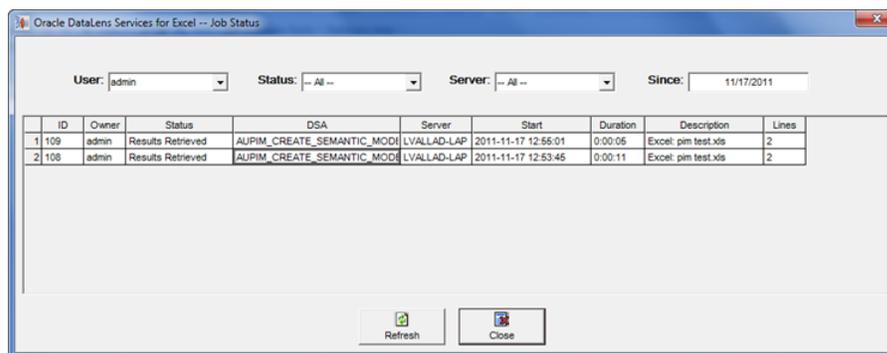
Removing Result Data

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....**



Any job that was created within the Services for Excel or the Oracle DataLens Server is displayed including those that are pending or have failed. By default, only the jobs of user that is logged into the server are listed.

The view can be changed using the **User**, **Status**, **Server**, and **Since** lists. These controls filter the displayed data based on your selections. For example, you could use the **Since** list to change the date that you want to set as the starting point for data retrieval. Alternatively, selecting **Failed** from **Status** filters all failed jobs from the data retrieved.

If a job is running, you can select it and click **Cancel** to stop it.

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

Caution: When a DSA returns the results of a processing job (synchronous or asynchronous), you *must* retrieve the data immediately, particularly for those jobs that return large amounts of data. Failure to do so can eventually result in an Oracle DataLens Server crash once the server memory is exhausted. The memory can be exhausted from many jobs where the results are not retrieved. In order to improve response time, the server keeps all job results in memory when the job results are less than the DSA maximum memory setting, until the user or calling application retrieves the results or the server is restarted. For information about setting these values, see *Oracle Enterprise Data Quality for Product Data Oracle DataLens Server Administration Guide*.

Getting More from Your Data

This chapter describes the following:

- ["Loading Metadata Files"](#) on page 4-1
- ["Changing and Grouping Data"](#) on page 4-2
- ["Using Regression Testing"](#) on page 4-7
- ["Using Data Sampling"](#) on page 4-11
- ["Backing Up Source Data"](#) on page 4-12
- ["Using the AutoBuild Application"](#) on page 4-12

Loading Metadata Files

With the Services for Excel metadata file load 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.

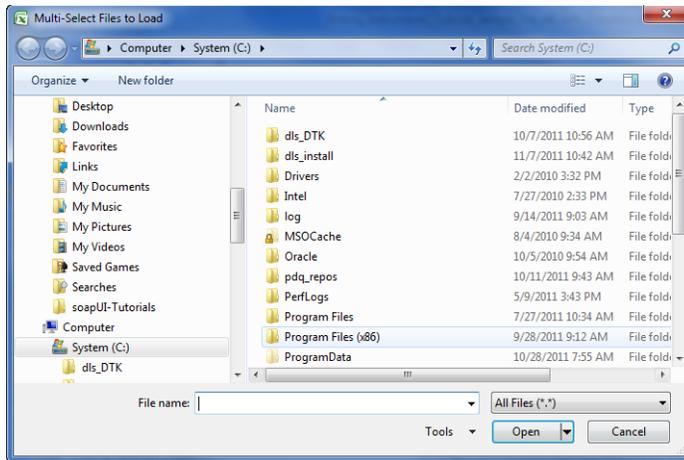
You can merge multiple source metadata Excel files into a single target Excel file of the same type. The file consolidation functionality makes the merging of distinct data files straightforward and easy.

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 a single target file that you want to merge the source files into.
3. Ensure that you create worksheets in the target Excel workbook that exactly match the source worksheet names. 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.
4. Open the target workbook.

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

By default, the system appends the records it finds in all of the source files to the Excel target file.



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 **Open**.

A progress dialog displays the overall progress.

Changing and Grouping Data

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

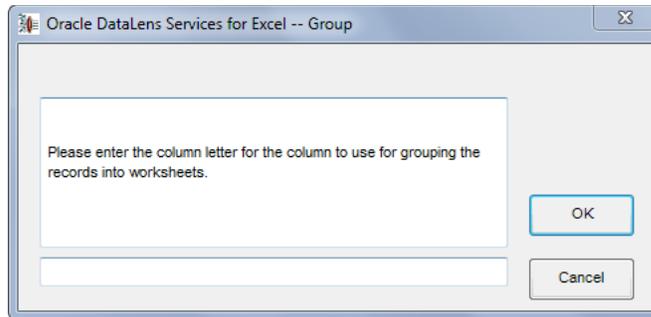
Grouping Records

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.

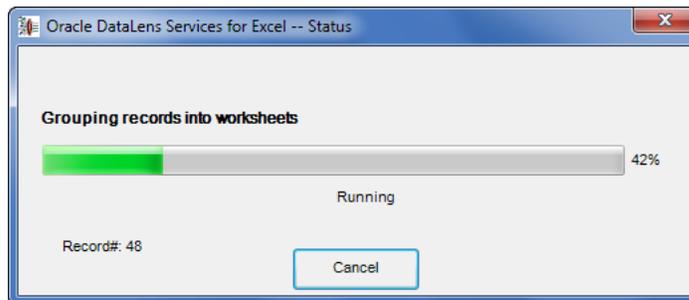
For example, the following data could be grouped by manufacturer.

1	A	B	C	D
	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	Everrsharp 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

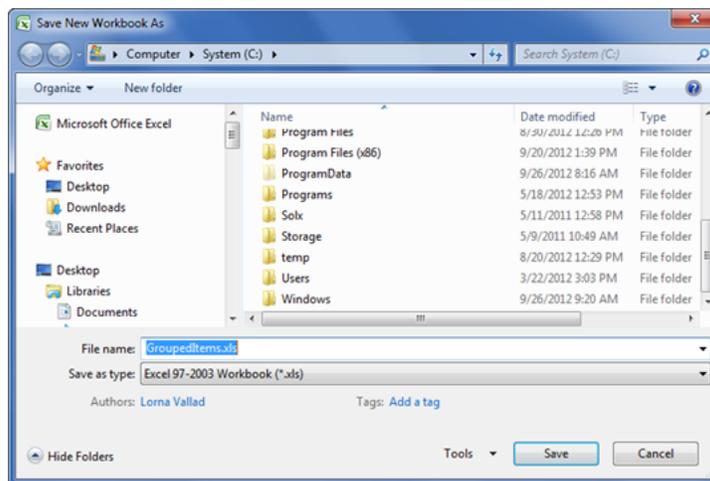
To group records, right-click on any cell then select **DataLens Services**, and then select **Group Records into Worksheets**.



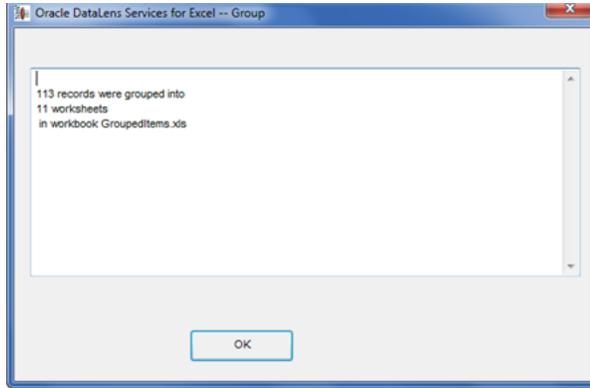
Enter the column you want to use as filter for grouping the records; it is case insensitive. Click **OK** to continue.



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. A confirmation message describing the number of records grouped, the number of tabs created, and the a name of the saved file.



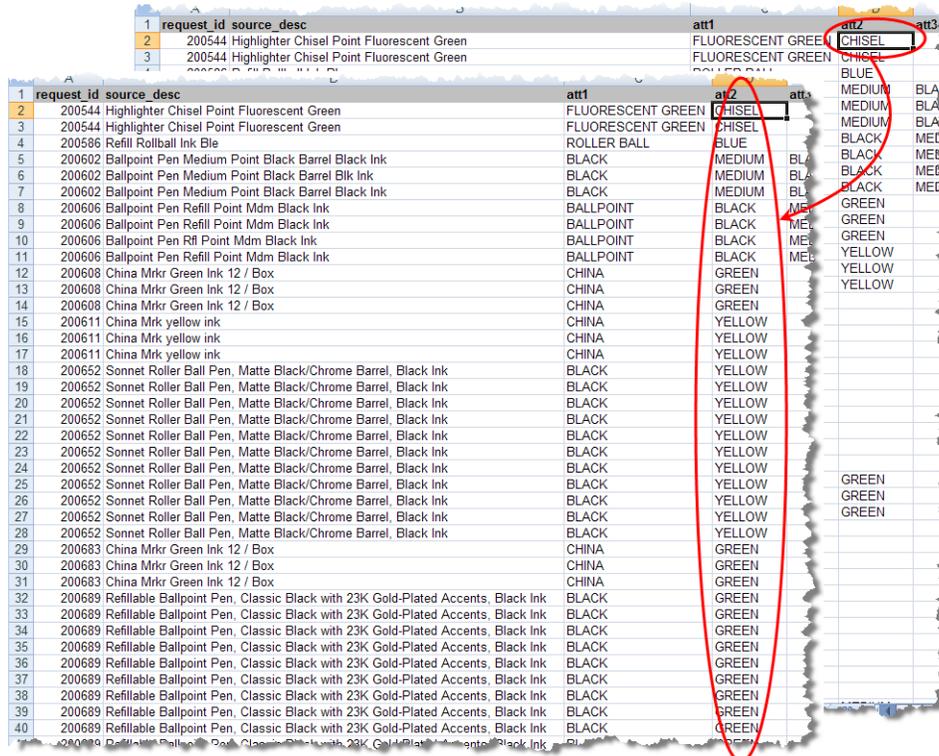
The new workbook containing a worksheet for each group of data will be similar to the following:



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**.



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.

Note: If you select the entire worksheet, the Services for Excel context sensitive menu is not active.

Replacing Control Characters

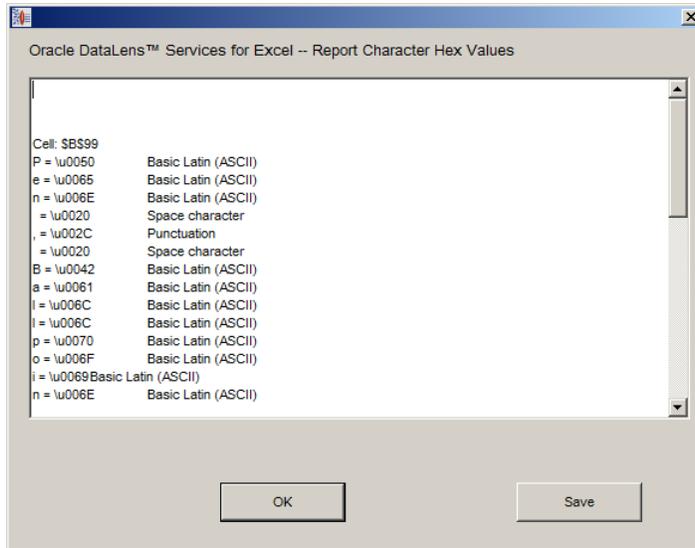
To replace control characters, select the data records that you want to analyze and right-click then select **DataLens Services**, select **Character Analysis and Replacement**, and then select **Replace Control Characters**.

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:

```

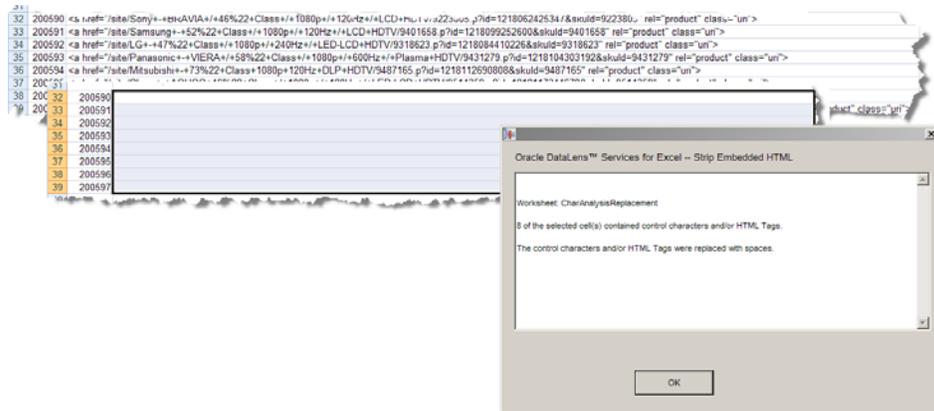
Cell: $B$3
R = \u0052Basic Latin (ASCII)
E = \u0045Basic Latin (ASCII)
S = \u0053Basic Latin (ASCII)
P = \u0050Basic Latin (ASCII)
C = \u0043Basic Latin (ASCII)
F = \u0046Basic Latin (ASCII)
 = \u0020Space character
A = \u0041Basic Latin (ASCII)
X = \u0058Basic Latin (ASCII)
3 = \u0033Basic Latin (ASCII)
0 = \u0030Basic Latin (ASCII)
0 = \u0030Basic Latin (ASCII)
O = \u004FBasic Latin (ASCII)
H = \u0048Basic Latin (ASCII)
M = \u004DBasic Latin (ASCII)
1 = \u0031Basic Latin (ASCII)
/ = \u002F Punctuation
    
```

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.

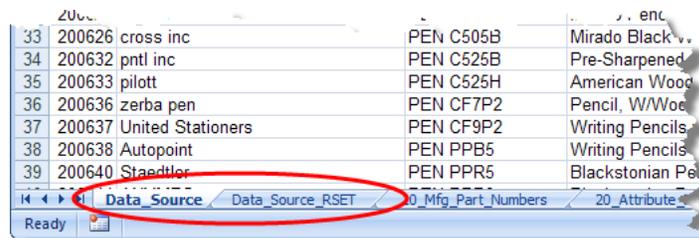
Using Regression Testing

The purpose of regression testing is to ensure that formerly acceptable results have not unexpectedly changed based on recent changes to a datalens or DSA. 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.

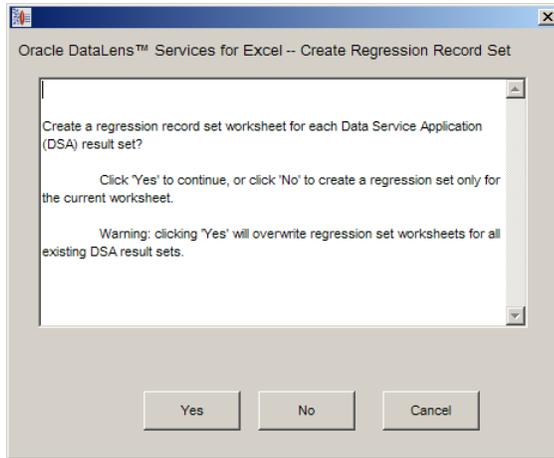
Creating or Updating the Regression Set

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 **Test**, 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 the `_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:

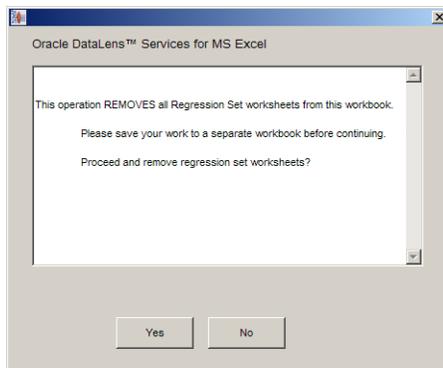


You can overwrite (or create) the existing regression record sets for the active worksheet and all result worksheets by selecting **Yes**; selecting **No** only overwrites the regression set for the active worksheet. Use the **Cancel** button to exit without changing any worksheets.

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

Removing Regression Set Worksheets

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 **Test**, 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**.

request_id	source_desc	attributes
23	200613 2_Exact Match Std Mfg Std Mfg	Visual Organizers PAPER062
24	200614 3_Exact Match Std Mfg Std Mfg	Visual Organizers PEN 07637
25	200620 3_Exact Match Std Mfg Std Mfg	Pickett Pickett PEN 507HB
26	200621 3_Exact Match Std Mfg Std Mfg	pmcomp PM Company PEN 50B
27	200633 3_Exact Match Std Mfg Std Mfg	logytek Logitech PEN C25BPHB3K6
28	200633 3_Exact Match Std Mfg Std Mfg	pilot Pilot PEN C525H
29	200644 1_Match on Raw Mfg Raw PN	Avery Avery SAN02181
29	200645 1_Match on Raw Mfg Raw PN	Duck Duck SAN02193
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 CR08753
33	200654 1_Match on Raw Mfg Raw PN	Papermate Papermate KOH285
34	200655 2_Exact Match Std Mfg	Perker Pen Perker Pen KOH287
35	200656 3_Exact Match Std Mfg Std PN	Pentel Pentel NB-003077885
36	200657 3_Exact Match Std Mfg Std PN	Pilot Pilot NB-013174222
37	200658 3_Exact Match Std Mfg Std PN	Sanford Sanford NB-013189641
38	200663 1_Match on Raw Mfg Raw PN	Koh-I-Noor Koh-I-Noor PAF64896
39	200666 2_Exact Match Std Mfg	Sparko Sparko 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	Faber-Kastell Faber-Kastell ZEB83171

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

Additionally, you can select one of the highlighted result cells and Services for Excel can locate similar or 'like' cells of data to your regression set. To do this, select one cell of data and right-click then select **DataLens Services**, and select **Copy Like Cells to Regression Set**.

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
1	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
2	200544 Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
3	200586 Refill Rollball Ink Blue	ROLLER BALL
4	200602 Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
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	CHINA
13	200608 China Mrk	CHINA
14	200611 China Mrk	CHINA
15	200611 China Mrk	CHINA
16	200611 China Mrk	CHINA
17	200611 China Mrk	CHINA
18	200611 China Mrk	CHINA
19	200611 China Mrk	CHINA
20	200611 China Mrk	CHINA

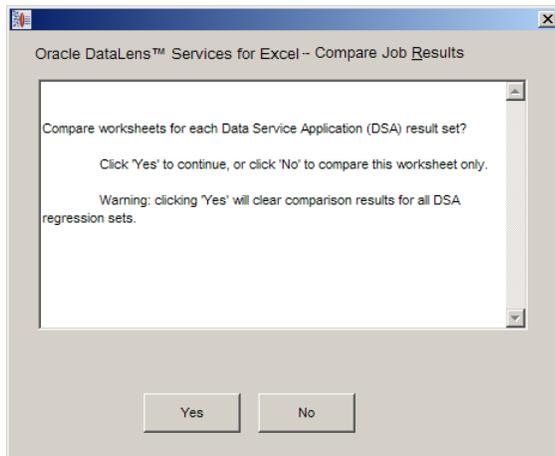
To clear the comparison highlighting, select the rows you want to return to normal then select **DataLens Services**, and then select **Clear Comparisons for Selected Cells**. The colorization is removed from the selected cells as in the following example:

request_id	source_desc	attr
200544	Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
200544	Highlighter Chisel Point Fluorescent Green	FLUORESCENT GREEN
200586	Refill Rollball Ink Ble	ROLLER BALL
200602	Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
200602	Ballpoint Pen Medium Point Black Barrel Blk Ink	BLACK
200602	Ballpoint Pen Medium Point Black Barrel Black Ink	BLACK
200606	Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
200606	Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
200606	Ballpoint Pen Rfl Point Mdm Black Ink	BALLPOINT
200606	Ballpoint Pen Refill Point Mdm Black Ink	BALLPOINT
200608	China Mrkr Green Ink 12 / Box	CHINA
200608	China Mrkr Green Ink 12 / Box	CHINA
200608	China Mrkr Green Ink 12 / Box	CHINA
200611	China Mrk yellow ink	CHINA
200611	China Mrk yellow ink	CHINA
200611	China Mrk yellow ink	CHINA

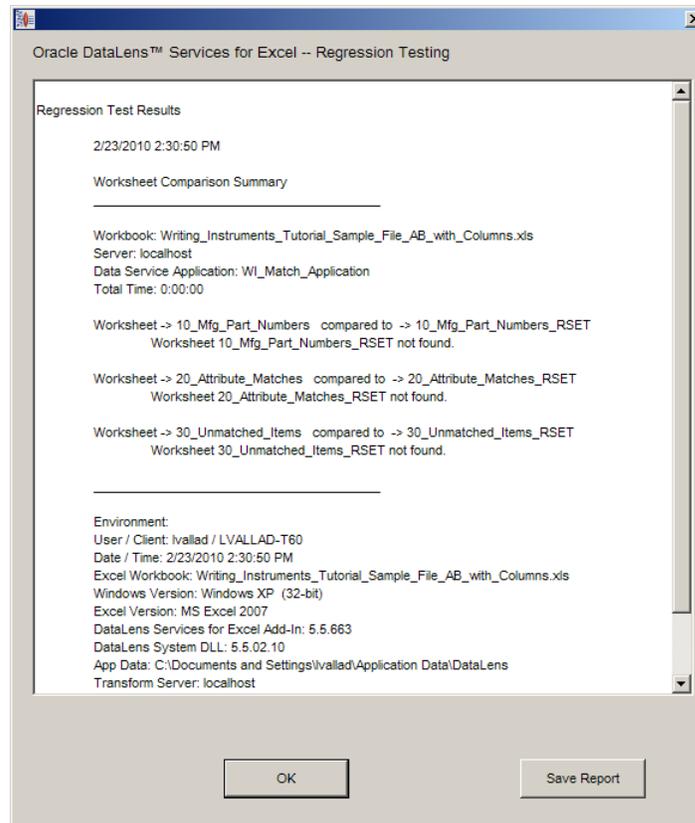
The cell in row six was selected and so the highlighting was removed.

Comparing 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 **Test**, 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.

Using 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 **Test** 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 **Test** button, select **Sampling**, and then select **Create**

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.

Backing Up Source Data

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 Options Section](#)" on page 3-4.

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.

Reset

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.

Using the AutoBuild Application

The purpose of the 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.

With AutoBuild, you can leverage existing Excel worksheet product information and Enterprise DQ for Product 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 Enterprise DQ for Product applications to recognize, standardize, classify, and translate your data as well as for further transformation in accordance with comprehensive business processes that you define.

To start AutoBuild, click the **DataLens Tools** button, and then select **AutoBuild...** or click the **AutoBuild** button on the toolbar.

Note: The following message may be displayed:

No data lenses were found.
Please make sure the Knowledge Studio is installed and check out the Smart Glossaries.

This means that Services for Excel is installed though the EDQP Client Software is not installed. For installation instructions, see "[Installing and Configuring Services for Excel](#)" on page 2-1.

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

