

Oracle® Product Data Quality

What's New in OPDQ

Releases 5.0, 5.1, 5.5.03, 5.5.03.02, 5.6, 5.6.0

E20610-04

March 2011

Welcome to Oracle Product Data Quality (OPDQ). The following sections describe new and changed functionality in OPDQ releases 5.0 through 5.6.0.

- ["Release 5.0 Features"](#) on page 1
- ["Release 5.1 Features"](#) on page 2
- ["Release 5.5 Features"](#) on page 3
- ["Release 5.5.03 Features"](#) on page 4
- ["Release 5.5.03.02 Features"](#) on page 4
- ["Release 5.6 Features"](#) on page 5
- ["Patch Release 5.6.0 Features"](#) on page 9

Release 5.0 Features

The following sections outline the changed and new features in release 5.0 of OPDQ.

Application Studio

This section describes changed features in Application Studio.

Automatic Match

The automatic matching functionality has been enhanced to improve product-specific matching capabilities as follows:

- Identity match provides clean-up for key-based matching.
- Semantic match extracts and standardizes for functional matching.

Services for Excel

The AutoBuild functionality has been enhanced to allow you to leverage existing information to reverse engineer the semantic model from available meta data and sample data to improve initial data lens development time. These enhancements include the ability to:

- Leverage existing information for operation immediately upon installation
- Term variant creation to increase recognition immediately upon installation
- Greater flexibility to import meta data rules
- Default Match Types

- Incremental processing to augment existing data lenses with additional metadata.

Governance Studio

This section describes new features to Governance Studio.

Governance Console

This release introduces the Governance Console feature to enable you to more effectively apply and maintain data governance standards.

Provides the ability to incorporate manual effort and ongoing learning into integrated process as follows:

- Item level validation, approval and edit
- Task-specific UIs for exception processing
 - Matching
 - Enrichment and correction
- Allows manual intervention while maintaining process integrity and efficiency

Governance Dashboard

This release introduces the Governance Dashboard feature to enable you to more effectively apply and maintain data governance standards.

Provides process visibility and control to drive continuous improvement of the following:

- Aggregate quality and productivity metrics
- Data quality by source
- Process effectiveness
- Identifies source quality and process improvement opportunities

Release 5.1 Features

The following sections outline the changed and new features in release 5.1 of OPDQ.

Application Studio

This section describes changed and new features to Application Studio.

AutoAbbreviate

The AutoAbbreviate functionality has been changed to provide the ability to repurpose and republish data into any form. It includes the ability to progressively abbreviate descriptions to fit the fixed-line lengths often required by legacy and many Enterprise Resource Planning systems. This enables you to use a systematic process to create shortened descriptions.

Semantic Index

This release introduces the semantic index functionality. It enables highly-scalable semantic matching and de-duplication of unstructured, non-standard data. This task is

often required in Product Information Management (PIM), master data management (MDM), and system consolidation situations.

Semantic indexing allows for faster and more scalable semantic matching.

Governance Studio

This section describes changed and new features to Governance Studio.

AutoSuggest

This release introduces the AutoSuggest functionality that provides system-based suggestions to expand recognition rules based on real operational data. Using inferences from the underlying semantic model, the AutoSuggest makes suggestions, that are used to enrich the model, so the system gets "smarter" through use thus allowing you to develop knowledge rapidly.

Knowledge Studio

This release introduces the Prediction functionality to provide a design time version of AutoSuggest so that it is available to the data analysts using the Knowledge Studio. This changed feature allows you to develop knowledge rapidly.

Release 5.5 Features

The following sections outline the changed and new features in release 5.5 of OPDQ.

Application Studio

This section describes changed and new features to Application Studio.

Dynamic Match

A new Semantic Index type was created to provide dynamic semantic matching:

- Eliminates index rebuild and the need for multiple semantic indexes if match rules change, as long as, the total set of attributes does not change.
- Allows you to experiment with matches.
- Dynamically operating algorithm that leverages multiple match types and weights in real time.

Output Classifications

This release introduces the ability to specify from one to multiple classifications from a single Item Definition Transformation, which simplifies processing.

Workflow Enhancements

This release introduces the ability to use dedicated secondary Data Service Applications (DSAs) for additional processing, which includes the following options:

- Apply DSA to process a set of rows.
- Re-Run DSA to re-process a set of rows.
- Completion DSA to complete a process and save the results.

Governance Studio

This section describes changed and new features to Governance Studio.

Apply Augmentations

The application of the rule changes from AutoSuggest to enrich the semantic model has been automated thus allowing you to develop knowledge rapidly.

Interface Enhancements

The following features have been changed to enhance usability and data remediation processing:

- Field level edit restrictions
- List values
- Hidden columns
- Normalized view and output column renaming to simplify results
- Category filtering on all output to allow you to filter results by a category
- Messages
- Renaming output columns
- Project reuse through saving as a new project from an existing project

Survivorship and Merge

This release introduces the capability to create blended rows with lineage for loading into an application (for example, an item master or Product Information Management (PIM)).

Release 5.5.03 Features

The following sections outline the changed and new features in release 5.5.03 of OPDQ.

Application Studio

This section describes new features in Application Studio.

Ngram Processing and Quick Lookup DSA

This release introduces a set of DSA widgets to create solutions that handle non-category specific global data for systems migration and consolidation as well as clean-up, match, merge and de-duplication. These features enable statistical (global)-based processing for faster development and when the high confidence match techniques are not appropriate:

- Confidence index provides scoring process that includes exact and fuzzy matches.
- Quick Lookup DSA shows the matches in the context of their original descriptions.

Release 5.5.03.02 Features

The following sections outline the changed and new features in release 5.5.03.02 of OPDQ.

Application Studio

This section describes changes to Application Studio.

Step Description Truncation

The truncation of step descriptions occurs at run time to provide the first 255 characters of the description on the DSA Job Status page giving context for which DSA is running while simplifying the input to the internal database. The step descriptions are displayed in their entirety in the DSA so that you can create large descriptions for their step names and descriptions.

Certification

The Db2 9.x as a valid data source in Application Studio is certified in this release.

Governance Studio

This section describes changed and new features to Governance Studio.

Graphing

The graphing capabilities have been enhanced to improve usability and performance.

Results Persistence

This release introduces the ability to persist the results of an Apply secondary DSA, and then create a new Governance Studio project from those results, which improves usability and performance.

Oracle DataLens Server

This release introduces the Job Continuation feature. This feature allows large batch jobs to continue to run, new API jobs to start and run even when Transform Servers lose connectivity with the internal database repository (the Oracle DataLens Administration Server is down).

An option to allow you to disable DSA processing on the Oracle DataLens Administration Server is introduced.

Performance

Various memory and performance enhancements have been implemented.

Release 5.6 Features

The following sections outline the changed and new features in release 5.6 of OPDQ.

Application Studio

This section describes changed and new features to Application Studio.

Automatic Abbreviation of Item Definition Descriptions

The automatic abbreviation of Item Definition descriptions (AutoAbbreviate) functionality as defined in Item Definition Transformations is enhanced. It more intelligently abbreviates descriptions, without reducing the readability, by applying an

understanding of the important terms in the description. AutoAbbreviate prevents important attributes from being abbreviated though you must create abbreviations for required attributes. Additionally, it no longer abbreviates any units of measure, part numbers, or model numbers by not abbreviating any words associated with attributes containing numbers.

Ngram Creation

The creation of Matching Ngrams is enhanced to provide the following capabilities:

- Selection of the type of Ngrams to produce
- Avoid saving the duplicate Ngrams
- Truncation of Soundex (RSIndex) algorithms to a specific number of characters
- Specify the minimum number of characters created by per Ngram type

Restricting DSAs

The ability to restrict the use of DSAs by OPDQ users is introduced. Users can be restricted from using and viewing DSAs on your various types of servers by using these new options in conjunction with the use of the Application Studio options on **Role Administration** page in the **Oracle DataLens Servers Administration** Web pages.

Transformation Changes

The ODBC database option is deprecated in this release.

The ability to output Attribute and Group IDs in Item Definition Transformations is introduced.

Certification

Apache Tomcat version 6.0.29 and Java Development Kit (JDK) 6 Update 21 (1.6.0_21) are certified in this release. The improved native 64-bit support in Tomcat and enhanced memory management in JDK 21 result in better overall performance.

Governance Studio

This section describes changed and new features to Governance Studio.

Apply Augmentations

The Apply Augmentations feature is changed so that the missing attribute suggestions are written to an Excel file. This file can then be reviewed and changed, and then imported into a given data lens in Knowledge Studio.

Filtering with Multiple Columns

The ability to filter data using more than one column is introduced. Filtering data with two or more columns is enabled using an 'and' criteria so that the data displayed can be reduced further.

Large Data Set Handling

Memory can be conserved in the Governance Studio by populating the tab when you select it, as well as, the ability to remove tab results from memory when a different tab is selected if that tab has more rows than a preset limit. The tab data reloads when you

select the tab again. The source data can be read directly from disk and streamed directly to the server without having to store the source data in the project.

Opening Projects with No Data Processed

The ability to open an existing project file without processing and displaying data results in any Output tabs is introduced. The new **Open Project - No Results** option on the **File** menu allows you to open a project that previously could not be opened because it exceeded the maximum memory allocation.

Restricted DSAs

DSA restriction is introduced. Users can be restricted from using and viewing DSAs so that only the DSAs that user has been allowed to use are displayed.

Support for Excel 2007 Files

Support for Excel 2007 files with the .xlsx extension is introduced.

Knowledge Studio

This section describes changes to Knowledge Studio.

Active and Inactive Item Definitions

The ability to use AutoBuild to create an entire Item Definition structure with all Item Definitions designated as inactive (for Production) so that a user in Knowledge Studio can activate individual Item Definitions as they are ready for production use is introduced. The data steward retains the ability to fully test all the interactions across all Item Definitions within the Knowledge Studio.

Copy Standardizations

Copying standardization rules from a source type to a destination within the Item Definition tree either individually or with children is introduced. It includes the ability to merge or replace and the ability to copy global standardizations from one type to another and from one Item Definition to another.

Data Parsing Enhancements

The data parsing functionality is enhanced to eliminate extra parsing by reducing duplicated tokenization, parsing, and item definition thus improving performance

Item Definitions

Item Definitions no longer automatically display the underlying attributes. You must double-click on an Item Definition to expand the full attribute structure for viewing and standardizing. This change reduces the amount of memory needed thereby improving performance.

Importing Automatically Suggested Knowledge From the Governance Studio

The Governance Studio Apply Augmentations feature creates an Excel file that contains missing attribute suggestion knowledge, which is now imported into the open data lens.

Translation

The Test Translation sub-tab is deprecated and replaced by the following:

- The Test Item Translation sub-tab is introduced. It allows you to review the translated phrases for the Item Definition for your source language sample data to validate your results.
- The Test Global Translation sub-tab is introduced. It allows you to review the translated phrases for the Item Definition for your source language sample data to validate your results.

Support for Excel 2007 Files

Support for Excel 2007 files with the .xlsx extension is introduced.

Oracle DataLens Server

Significantly better performance by allowing the server to automatically create multiple threads to maximize the CPU processing for lens transformation steps if Processor Percentage is set to value greater than 0. For example, if the server has 8 CPUs and the Processor Percentage is set to 50% then the server will automatically create four processing threads.

Improved Memory management by leveraging JDK 21 and removing explicit garbage collection requests to increase performance and memory processing.

The thread pool size can now be set from the server to allow better thread management in a single server environment with multiple applications.

In the event that the underlying database goes down, the server will automatically reconnect the entire db connection pool once the database is available again thus eliminating the need to restart the server.

Improved data lens load times by changing and simplify the data lens load processing steps.

Improved Internationalization support, through externalizing strings, install and run using source OS and adding 29 new supported languages in the Knowledge Studio.

Setting the `ServerID` for Transformation Servers is now automatic and set by the server to ensure each `ServerID` is unique and sequential.

Configuration Options

The following Configuration options are deprecated in this release:

- Apply Augmentation options 2 and 3 that enabled the automatic update of data lenses
- Unused Transform Map and DGD email addresses

The Oracle DataLens Server options are changed as follows:

- Max Batch Jobs is deprecated
- Processor Percentage is introduced
- Thread Pool Size is introduced

Role and User Administration

The ability to restrict the use of DSAs by OPDQ users is introduced. Users can be restricted from using and viewing DSAs on your various types of servers by using these new options in conjunction with the use of the Application Studio options on

Role Administration page in the **Oracle DataLens Servers Administration** Web pages.

Patch Release 5.6.0 Features

The following sections outline the changed and new features in release 5.6.0 of OPDQ.

AutoBuild

This section describes changed and new features to AutoBuild.

- The AutoBuild Wizard has been modified to streamline the process of building or updating a data lens.
- The ability to add to an existing data lens is enhanced.
- The ability to activate Item Definitions in generated (new or updated) data lenses is introduced; inactive is the default.
- Populating data lens descriptions or DSA comment fields with category and attribute identifiers, which are stored in the `system_id` and `group_id` internal fields though are not visible in the Knowledge Studio.
- Selecting a Term Library to add new terms and full-form term variants.
- Unused options have been removed. The following options are introduced:
 - Step 1:
 - **Generate a new DataLens**
 - **Add to an existing DataLens**
 - **Add alternate catalog to Existing DataLens**
 - **List of Term Libraries**
 - Step 2:
 - Include attribute name in phrase for context - **Never**
 - **Map values directly to each attribute**
 - Step 4:
 - **Add alternate catalog to selected DataLens**
 - **Use alias for ids**
 - **Activate Item Definitions**

For detailed information about these options and the wizard, see *AutoBuild Reference Guide*.

Services for Excel

This section describes changed and new features to Services for Excel.

- The installation and uninstallation processes have been simplified with use of scripts for both 32-bit and 64-bit client systems.
- The following files and folder are deprecated:
 - `ExcelServicesResource.xml`

- system.xml
- common folder
- Configuration information is now contained in the `dlsforexcel_cfg.xml` file. It is located in the `application_data/DataLens/config` directory, where `application_data` is the users application directory for the version of Windows that is running. This directory also contains other configuration files necessary to the application including `dlsforexcel_termlib_General_English.xml`, which is used to add terminology (term) variants to term libraries.
- Log files are located in the `application_data/DataLens/log` directory, where `application_data` is the users application directory for the version of Windows that is running. The information logged includes installation specifics, job status, server interaction, and any errors. These log files can be reviewed to ascertain errors and informational purposes.
- Translation property files are located in the `application_data/DataLens/locale` directory, where `application_data` is the users application directory for the version of Windows that is running. This directory is examined for existing property files that can be used for translation and populates the **Target Locale** option, which is defaulted to English if no translation property files exist.
- The ability to restrict the use of DSAs is introduced. The use of DSAs can be restricted to specific users. The use of DSAs can be restricted to specific users using the **Oracle DataLens Server Administration** Web pages and Application Studio. All DSAs are displayed when selecting a DSA including restricted DSAs. If a restricted DSA is selected and the user does not have permission to run it, an error is displayed.
- The toolbar has been enhanced to add a **Test** menu to allow easier access to regression testing and data comparison utilities.
- The ability to select a DataLens Server Group is introduced.
- Job processing, and obtaining status and results has been modified to closely mimic that of Governance Studio to provide more consistent use in OPDQ.
- Logging into the Oracle DataLens Server has been modified to use the OPDQ Launch Pad to provide more consistent use in OPDQ. You must log in to the server for all job processing activities. The log in applies to all spreadsheets within the open workbook. You can change the user logged in or open an entirely new workbook to login as a different user.
- The ability to create or update Term Libraries is introduced to allow you to easily add new variant for use in data lens processing.

Upgrading Oracle DataLens Servers

The ability to upgrade your Oracle DataLens Server from OPDQ releases, 5.0.01, 5.0.02, 5.1, and 5.5.x to 5.6 is introduced. The OPDQ upgrade methodology is to install the 5.6 Oracle DataLens Administration Server, while maintaining the ability to roll back to your existing server.

Note: Assistance for upgrades can be obtained by contacting Oracle Consulting Services. Upgrades should be a planned migration to ensure the retention of your Server Group Topology, data repository, DSAs, data lenses, and data.

Oracle Product Data Quality What's New in OPDQ, Releases 5.0, 5.1, 5.5.03, 5.5.03.02, 5.6, 5.6.0
E20610-04

Copyright © 2010, 2011 Oracle and/or its affiliates. All rights reserved.

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.

