Oracle Product Data Quality Overview
(adapted from Oracle PDQ Corporate Datasheet)

Alec Towlson
Oracle Corporation
Enterprise PLM EMEA

Keywords:

Products; Services; Attributes; Synchronization; Master Data; End-To-End Process; De-duplication; Standardization; Advanced Matching; Merging; Cleansing; Cross Referencing; Semantic Recognition; Data Mastering; Data Centralization; Supply Chain Optimization; Spend Management; Inventory Management; New Product Introduction; Retail Promotion Management; Trading Community Support, Ecommerce; Product Catalogue Management; GDSN; Distributors; Manufacturers; Buy-Side; Sell Side; Retailers; Inside; Supplier Management; PLM; PIM; Product MDM; Online Search

Introduction

Even the best Product Information Management (PIM), Master Data Management (MDM) or data integration initiative can be derailed by poor product data quality. ***Oracle Product Data Quality*** is built on the latest semantic technology to handle the extreme variability of product data and overcome the limitations of traditional customer data quality solutions. ***Oracle Product Data Quality*** can be integrated with any application or process, and is pre-integrated at a semantic level with the Oracle Product Hub to reduce the time and cost to deploy and operate your MDM system for product data, or any other system of record – while also extending its capability and benefits.

Product Data Quality – Specific Problems, Specialized Solution

Data quality has been recognized for some time as a critical component of any Master Data Management or data integration initiative, but it is all too often under-planned and under-funded and ends up as the greatest single risk factor in the project. When dealing with product data this problem is compounded as traditional data quality tools that were designed for customer (name & address) data struggle with the extreme variability of typical product data. Consequently enterprises are forced to use manual effort or custom code to 'clean-up' their data – neither of which is likely to be a good solution.

Oracle Product Data Quality was built from the ground up to tackle these problems and to integrate easily with any existing application or process – from PLM to procurement and from online retail to asset management and many more. It can be called in real-time or batch mode and can provide PDQ services from simple standardization to match, merge, transformation, translation and more.

Oracle Product Data Quality is pre-integrated with Oracle Product Hub making it the first and only MDM for product data solution to offer a dedicated product data quality capability that is both highly capable of dealing with the specific nuances of product data and is fully integrated to the core PIM system.

Oracle Product Data Quality – understands how to apply rules based on the meaning of data content values rather than just data content structure

Product data covers tens of thousands of product categories, each of which has a different semantic definition or context (vocabulary, inferences, validations etc.). To make matters worse, within a single category the format of the data can be infinitely variable (word order, punctuation, synonyms, national language, abbreviation etc.). Such variation is beyond the scope of traditional data quality solutions and a different technology approach is required. *Oracle Product Data Quality* incorporates patented DataLens™ technology specifically to deal with these issues:

• **Semantic-based** – Semantic recognition technology handles pattern and category variations without additional programming.

• **Auto-learning** – Semantic inference is applied to understand unrecognized data. If confirmed, inferences create new rules and the system continues to learn as a natural consequence of ongoing operation.

• **Scalability** – Category-based semantic models and in-memory processing allow easy scaling to handle of many millions of items across thousands of categories.

• **Integrated governance** – Dedicated interface for data stewards and product specialists to review process effectiveness (quality metrics and KPIs) as well as interact directly with exception data for rapid remediation.

• **Built for business users** – Allowing the business users who really understand the data to create and maintain rules allows for faster and more precise operation than if IT is forced to be the middle-man in rules creation.

Oracle Product Data Quality – Capabilities

The purpose of *Oracle Product Data Quality* is to provide an integrated capability to recognize, cleanse, match, govern, validate, correct and repurpose product data from any source to any target requirement.
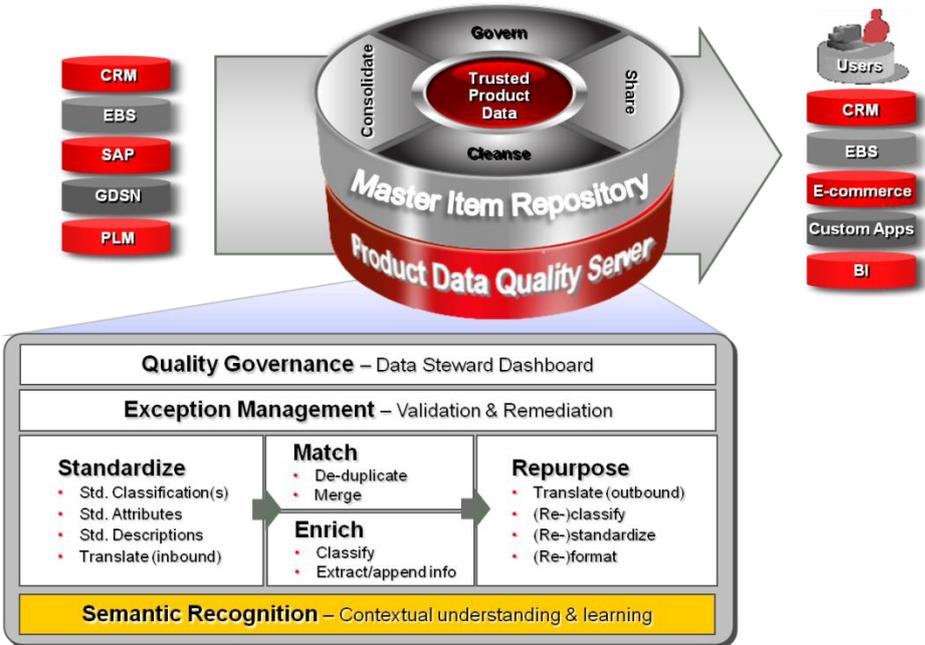

Illustration. 1: Oracle Product Data Quality Functionality Scope

The broad functional capabilities of the system are as follows:

• **Semantic Recognition** – Each product record is compared to a semantic model that can recognize and extract required information through reference to the semantic context. This is done irrespective of variations in input format, punctuation, terminology etc. and can be applied to both structured and unstructured data. This is the core requirement for handling product data and without it the variability of typical product data will cause subsequent processes to fail. In the case where the Oracle Product Hub is used, semantic metadata from the PIM system will be used – ensuring that all data loaded to the PIM already conforms to its defined standards.

• **Standardization** – Enforce any required PIM standards on product data:

- *Classification* – Classify an item to the PIM master item class and any number of alternate catalogues which can be based on industry or custom schemas. Supports multiple classifications per item.
- *Attributes* – Enforce PIM standards on attribute data including unit of measure conversion, range standardization, including color ranges, terminology and abbreviations.
- *Descriptions* – Enforce PIM standards on descriptions including word order, punctuation, terminology etc. For descriptive information that has to fit a target character length without truncation an AutoAbbreviate function can progressively abbreviate the line until it meets the length requirement while maintaining maximum readability.
- *Translation* – Attributes and descriptions can be translated into multiple languages with full support for double-byte character sets and parsing left-to-right and right-to-left.

• **Matching** – Support for many individual and hybrid matching methods from simple key matching (with or without cleanup) to semantic matching with full control over mandatory, optional and weighted attributes. A semantic key is created to ensure high performance when de-duplicating an input file, or when matching against PIM production tables. Matched records can optionally be merged and cross references and associations maintained.

• **Enrichment** – Extract or append information to enrich the item record:

- *Internal* – extract information embedded in descriptive fields
- *External* – append information from legacy systems, subscription sources etc.

• **Repurposing** – Transform product data into any form for storage in the PIM or for external publication. Capabilities include all those listed under standardization above and can be called as required to avoid redundant storage in the PIM production tables.

• **Governance** – Data Stewards can access a graphical dashboard of quality metrics and KPIs to give insight on where and how to drive process improvement. The dashboard is fully configurable to monitor transactional data within the system as well as any other source that can be accessed through a SQL query.

Illustration. 2: Oracle Product Data Quality, Configurable Quality Metrics Dashboard

• **Exception management** – Single console for exception management by data stewards or product specialists. Exceptions can be routed to different work queues for manual or automated remediation and validation.

Illustration. 3: Product Data Quality Console allows product specialists to verify exception data

Integrated to Oracle Product Hub

Oracle Product Data Quality is designed to integrate easily with any other system or process in real-time or batch mode and is not dependent on any specific system of record, however, since many companies are adopting PIM, or MDM for product data systems as a core component of their ability to manage product data, ***Oracle Product Data Quality*** is pre-integrated into the Oracle Product Hub at a semantic level, that is, semantic metadata from the Product Hub is used to seamlessly create the required semantic recognition models. For operational processes such as loading and matching, the Product Data Quality Server has access to both interface and production data tables for maximum operational simplicity and ease of use. A semantic key is created to enable high performance semantic matching against PIM production tables. This level of integration represents unprecedented co-ordination between a PIM system and a dedicated product data quality system.

Contact Address:

Alec Towlson
Business Development Leader for Oracle Product Data Quality EMEA
ORACLE Corporation UK Limited Company
Building 1, Blythe Valley Park, Shirley
Solihull GB B90 8AD

Phone:	+44 (0)7785 577 735
Email	alec.towlson@oracle.com
Internet:	www.oracle.com