

# **Produkt Daten Qualität in IT Applikationen durch Semantiktechnologie**

**Dr.-Ing. Helmut Maier  
Dr. Maier CSS GmbH & Co.KG  
Stutensee, Technologieregion Karlsruhe**

## **Schlüsselworte:**

Produktdaten, Produkt Daten Qualität, Semantiktechnologie, Data-Lensing, Data-Cleansing, TQM, TDQM, PDQ, PIM, Informationsfluss, Business Intelligence,

## **Einleitung**

Die Qualität des Master-Daten-Management (MDM) und speziell des Produkt-Informations-Management (PIM) über integrierte IT-Applikationen hinweg ist meist unbefriedigend. Daten sind inkonsistent. Kategorien, Bezeichnungen und Attribute stimmen nicht überein. Vorlagen für MDM-Abgleich, Konvertierung, Umrechnung und Formatierung sind starr. Datenpflege bleibt wegen der Kosten auf ein Minimum begrenzt.

Die Product Data Quality Suite (PDQ) von Oracle bietet eine flexible Lösung für mehr Qualität der Produktdaten. PDQ nutzt Semantik Technologie, um ein konsistentes MDM von strukturierten und unstrukturierten Produktinformationen zu erreichen. Die Semantiktechnik bringt Daten zusätzlich noch auf der Ebene ihrer Bedeutung zueinander in Beziehung (semantische Relation). Das führt schneller zu konsistenten und validierten Daten von und zu beliebigen Datenquellen. Hier liegen hohe Nutzenpotenziale für ein Unternehmen.

## **Produkt Daten Qualität in integrierten IT Applikationen durch Semantik-Technologie**

Die Informationsflut im Unternehmen wächst täglich. Viele datenführende Systeme wie ERP, CRM, PLM, aber auch webbasierende Kataloge enthalten Tausende von Produktdaten, die die selbe Bedeutung haben, aber unterschiedlichste Formate und Inhalte führen. Dies macht dem Anwender das Suchen und Verarbeiten wichtiger Informationen schwer, kostet Zeit und führt zu Fehlern.

Schlechte Produktdatenqualität ist wie Dreck auf der Windschutzscheibe. Man kann damit lange Zeit weiterfahren. Der Dreck nimmt aber immer mehr zu und irgendwann wird ein Punkt erreicht, an dem es lebensgefährlich wird weiterzufahren.

Qualität der Produktdaten muss zum Bestandteil der umfassenden Qualitätsstrategie des Unternehmens werden (TQM - Total Quality Management und TDQM - Total Data Quality Management).

## **Wie stellen sich Istsituation und Anforderungen dar?**


Aktuelle Data Warehouse Situationen der Datenqualität in den Unternehmen zeigt, dass bis zu 80% der befragten Unternehmen unzufrieden mit ihrer Datenqualität sind. Über 70% scheuen die

Bereinigung und Standardisierung von Produktdaten, weil das Durchführen von Maßnahmen zur Verbesserung der Produktdatenqualität zu schwierig, zu zeitaufwendig, zu kostenintensiv und somit fast undurchführbar sei.


Heterogene inkonsistente Datenbestände entstehen durch jahrelang mangelnde Datenpflege, durch Fusion und Übernahmen, durch mehrere Altsysteme, durch verschiedene Applikationen, die jede für sich eigene Metadaten führen. Darin gibt es unterschiedliche Typen und viele Kategorien der Metadaten, die in strukturierter oder auch unstrukturierter Form vorliegen mit unterschiedlichen Schreibweisen, Bedeutungen, Synonymen und Gültigkeiten sowie daraus entstehende Folgerungen und Zusammenhänge (bspw. Skt-Schraube, Ring, Leiter).

Ein leicht nachvollziehbares Beispiel der Vielfalt von Produktdaten und Kategorien sind Varianten von Elektromotoren als Kaufteil, die von vielen Firmen weltweit angeboten werden. Die Produkte

Energieklasse:	EFF1
Leistung:	0,18 kW
Spannung:	380 V
Drehzahl:	2700 - 3000 U/min
Bauform:	B3
Montageart:	Gewindeflansch



Hersteller	Beschreibung
eccDrives	ACA 63 A-2
VEM	K21R 63 K2
HEGA	2T63 A2
SBN	Sb_39141
SEW Eurodrive	DV112M-F4/TF/ASB15



ABC123	AA Inc.	10hp motor 115V Yoke mount	Motor, AC/DC	10hp	115V	Yoke
abc-123	A.A.	mtr. ac(115) 10 horsepower 115volts	AC/DC Motor	10	115 AC.	
ABC/123/Q	AA/Craft	10 Caballos, Motor, 115 Voltios	Mot-AC	10 H-pow	115	
QA-ST5	Craft	TEAO HP = 10.0 1725RPM 115V 48YZ YOKE MTR	28101804			
Z99	Z99	MOT-10,115V,48YZ,YOKE	Z99			

sind teilweise austauschbar. Es sind die gleichen Kennwerte. Dennoch sind Vollständigkeit, Formatierungen und Produktbezeichnungen sehr unterschiedlich.

Abb. 1: Varianten von Elektromotordaten, je nach Hersteller

Informationen und Produktdaten sind lebensnotwendig im Unternehmen. Der Informationsfluss läuft in der Regel dem Materialfluss voraus. Er ist global. Neben strukturierten Daten begleiten ebenso viele unstrukturierte und sogar grafische Daten die Geschäftsprozesse hinzu den vielen Lieferanten, zu den Standorten, zu Kunden und in den verschiedensten Sprachen. Produktdaten sollten daher unternehmensweit eine genaue, eindeutige und gültige Sicht auf die Produkte, ihre Technologien, Regeln und Prozesse abbilden. Dennoch hat jeder Anwenderbereich andere Sichten auf Daten, Informationen und Abläufe.

Produktdaten haben häufig eine hohe Dynamik, zwar völlig anders als die Bewegungsdaten, durch die engen Zusammenhänge in den globalen Informationsflüssen jedoch ändern sich Kategorien, Strukturen, Bezeichnungen usw. in den Anwendungssystem ständig wie bspw.:

- o E-commerce, Vertrieb
- o Product Design
- o Lager & Logistik
- o Kundendienst
- o Business Intelligence

### Warum sollte man Produktdaten harmonisieren?

Selbst das beste Informationssystem wird durch schlechte Datenqualität ausgehebelt. Das zieht ich durch alle Anwendungssysteme hindurch. Schlechte Produktdatenqualität führt zu hohem Suchaufwand, unscharfen Trefferquoten, geringer Transparenz und Fehlern im Prozess. Unzureichende Qualität der Produktdaten beeinflusst somit Produktqualität und Kundenzufriedenheit.

Durch hohe Transparenz in den Produktdaten kann die Komplexität von Produkten, Varianten, Kategorien und Regelwerken reduziert werden.

Schwierig ist die Antwort auf die Frage, wie der Grad der Produktdatenqualität gemessen und monetär bewertet werden kann? Es fehlen Methoden, um die Vorteile konsistenter Daten und durchgängiger Informationsflüsse bei unterschiedlichen Anwendungen zu quantifizieren. Eine Analyse des Bestands an Produktdaten auf Dubletten, ähnliche Teile, abweichende Beschreibungen usw. erscheint aufwendig. Sie setzt voraus, dass Abfragetools und Auswertungen flexibel formuliert werden können. Darüber hinaus müssen die Daten nicht nur in ihren Strukturen, sondern auch in unstrukturiertem Kontext analysiert werden. Das setzt lernfähige, leicht definierbare und semantische Suchwerkzeuge voraus, ähnlich wie man sie von der Internetsuche her kennt. Die Qualität der Suchergebnisse selbst sollte bei der semantischen Suche in den Produktdaten aber wesentlich besser sein als im Internet.

### **Welche Lösungsstrategie für bessere Produktdatenqualität ist praktikabel?**

Produktdaten sind etwas besonders im Unternehmen und unterscheiden sich deutlich von anderen Datenbeständen. Personal-, Kontakt- und Adressdaten bspw. sind einigermaßen systematisiert. Aber trotz STEP und ISO 10303 gibt es für die umfassende Beschreibung von Produktdaten keinen hinreichenden Standard. Anwender von Produktdaten haben jeweils unterschiedliche Sichten auf diese Daten - sei es im Bereich e-commerce, Produktentwicklung oder Einkauf, Logistik, Produktion oder Kundendienst. Die Zusammenhänge in den Anwendersichten auf die Produktdaten sind unterschiedlich, ebenso Bezeichnungen, Gültigkeit und Regeln.

Produktdaten haben einen sehr hohen, zentralen, häufig unterschätzten materiellen Wert für das Unternehmen. Deswegen ist es zwingend, dass Produktdaten und die Informationsflüsse im Unternehmen in gleicher Weise effizient behandelt werden wie die Materialien und deren physikalische Materialflüsse. Wie beim Materialfluss wird der gesamte Geschäftsprozess gestört, sobald der Informationsfluss unterbrochen ist. Es kommt sehr auf effiziente und konsistente (Informations-)Prozesse an

- o Zwischen Anwendern und Abteilungen,
- o Zwischen Unternehmen und
- o Zwischen Anwendungssystemen .

Daten und Information haben als "Immaterielles Produkt" denselben Anspruch an Produkt-Qualität wie die materiellen Erzeugnisse des Unternehmens. Produktdatenqualität muss Teil der TQM-Strategie des Unternehmens werden (TQM – Total Quality Management) als Vision für TDQM (Total Data Quality Management).

Die sogenannten Masterdaten oder Produktdaten sollten möglichst zentral gepflegt und verwaltet werden. Hierfür gibt es bekannte Lösungsansätze in Form von Master-Daten-Management (MDM) oder bezogen auf Produktdaten das Produkt-Information-Management (PIM).

Die Frage bleibt, wie erreicht man eine bessere Qualität in den vorhandenen Produktdaten, die noch dazu oft in mehreren Anwendungssystemen entstehen und gepflegt werden. Bereits der Austausch von strukturierten Produktdaten zwischen verschiedenen Anwendungssystemen stellt sich oft schwierig dar. Noch aufwendiger ist die Sicherstellung der Synchronisation und der Konsistenz von unstrukturierten Daten, Dateninhalten oder Datenkategorien in den unterschiedlichen Anwendungssystemen. Hier kommen die Vorteile des PDQ-Management mit Semantischer Suche gegenüber bisherigen Methoden klar zum Vorschein:

- o Manuelle Arbeitsweise zum Bereinigen von Produktdaten ist
  - zu zeitintensiv,
  - zu langsam und
  - zu teuer.
- o Programmierte Tools sind
  - wenig flexibel wegen der Programmierung,
  - aufwendig für die vielen Ausnahmeregeln, selten skalierbar, kaum selbstlernend,
  - dediziert und beschränkt auf wenige Anwendungskategorien, bzw. Anwendungssysteme.
- o Vorteile eines semantisch arbeitenden Data-Quality-Tools (Semantic Based Data Lensing, Data Cleansing) sind
  - Semantik benutzt die Bedeutung von natürlichen Worten, Synonymen und deren Kontext.
  - Datenrepository, Kombinatorik und Ablaufregeln werden grafisch interaktiv definiert,

Bei der Suche im Web beispielsweise wird semantisch gesucht. Je nach Zusammenhang in der Fragestellung können sich dabei andere Antworten ergeben. Beispiel:

- o Wann kam der erste Mensch auf den Mond? = 21. Juli 1969 02:56:20 (UTC)
- o Wer war der erste Mensch auf dem Mond? = Neil Alden Armstrong

Auf Produktdaten bezogen können mit semantischer Suchtechnik relativ schnell Analysen von Datenbeständen durchgeführt werden. Mit dem PDQ-Tool werden Regeln und Filter zur Bereinigung und Zusammenführung von Datenbeständen grafisch interaktiv erstellt und können leicht erweitert werden. In Verbindung mit dem PDQ-Tool kann PIM im laufenden Betrieb die Produktdatenqualität innerhalb der Informationsflüsse sicherstellen. Das Einrichten der Rolle eines Produktdaten-Qualitätsmanagers ist sehr hilfreich, eigentlich unumgänglich.

Semantische Suche hilft ebenso bei der Migration von Daten aus verschiedenen Anwendungssystemen. Das PDQ-Tool kann für die Transformation der Produktdaten aus einem ins andere Anwendungssystem genutzt werden.

Für den unternehmensspezifischen Einsatz von Semantischer Suche kommt es im wesentlichen darauf an, die technischen Dictionaries aufzubauen und mit den Begrifflichkeiten auszustatten, die die Anwender täglich benutzen. Oracle hat kürzlich erst die Firma Silver Creek übernommen und deren Data-Lensing/Cleansing Lösung in seine PIM/MDM Architektur ergänzt. Diese Data-Lensing-Funktion ist unter Oracle PDQ - Product Data Quality Solution - verfügbar. Definitionen der unternehmensspezifischen, semantischen Modelle können mit PDQ in kurzer Zeit erstellt werden und passen sich an (self-tuning).

Der Einsatzbereich von PDQ geht über Suchen, Standardisieren, Zusammenführen, Übersetzen bis hin zu Umformatieren. Dialoge, Regelwerke und Abläufe für die Datenkonsolidierung erstellt der PDQ -

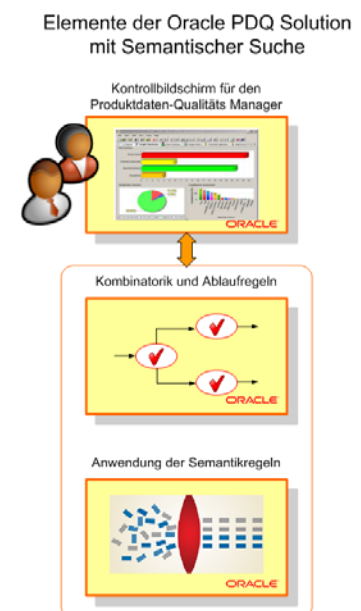


Abb. 2: PDQ mit semantischer Suchfunktion

Manager ohne Programmierung mit Hilfe der grafischen Technik zusammen. Ein grafisches Dashboard bringt Übersicht, zeigt die Statistik über den Prozess der Datenbereinigung und unterstützt bei der Regelung von Ausnahmen.

### Wie fügt sich PDQ in die Architektur des Oracle Data Hub?

Die Strategie des Master Data Management (MDM) trennt die Verwaltung der Produktdaten von der Nutzung dieser Masterdaten in den Anwendungssystemen. Das Master Data Management stellt die Zusammenfassung der Stammdaten, der Strukturen, der Verfahren und Data Repositories unabhängig von den Anwendungssystemen dar. Bezogen auf das Produktdaten Management spricht man auch von PIM Product Information Management. Dadurch wird erreicht, dass das MDM/PIM entkoppelt wird und in den jeweiligen Anwendungen spezifische Sichten auf die Daten, Datenstrukturen und Datenformate bestehen bleiben können, wie bspw in CRM, ERP, PLM, LVS, e-commerce usw..

Dieser Ansatz des Oracle Data Hub mit MDM/PIM bildet die Basis für ein konsistentes, unternehmensweites Datenmanagement und befreit Abteilungen, Standorte oder Zulieferer davon, das gleiche System oder einheitliche Datenformate nutzen zu müssen.

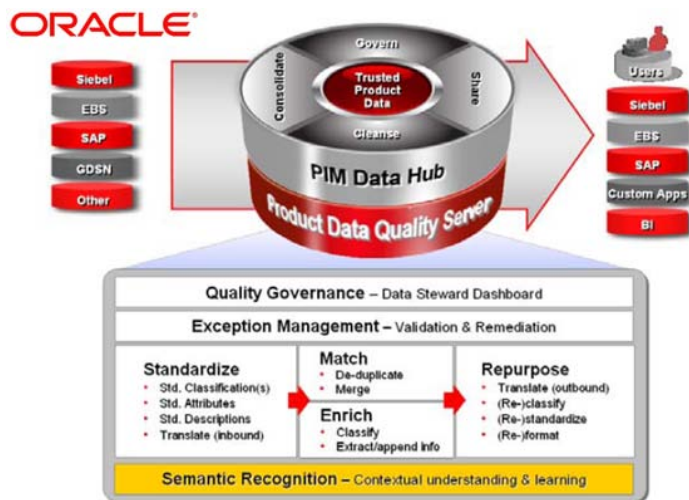


Figure - Product Data Quality Server 'footprint'  
Quelle: Oracle Product Data Quality Data Sheet 2010

Mit dem kürzlich vollzogenen Kauf von Silver Creek Systems ergänzt Oracle den PIM Data Hub mit dem Modul PDQ (Product Data Quality) um einen wertvollen Baustein, damit die MDM/PIM Datenqualität im täglichen Informationsfluss des Unternehmens umfassend gewährleistet wird. Für die Integrationen zwischen dem PIM Data Hub mit den sicheren und zuverlässigen Daten von und zu den internen und externen Anwendungssystemen des Unternehmens stehen die Techniken der SOA Middleware zur Verfügung.

Abb. 3: Product Data Quality Server

### Welches Nutzenpotenzial kann erschlossen werden?

Eine Analyse der Datenbestände mit der semantischen Abfragetechnik (Data-Cleansing) von Oracle PDQ zeigt oft überraschend hohe Trefferquoten und legt somit den Handlungsbedarf bei der Verbesserung der Produktdatenqualität offen. Es ist zu empfehlen über ein Analyseprojekt, lohnende Kategorien für Data Cleansing und bspw. Dublettenbereinigung zu finden.

Geeignet für Prüfung, Konsolidierung und Transformierung von Produktdaten sind große Warengruppen und Produktkategorien unterschiedlicher Lieferanten und Hersteller aus den verschiedensten Industrien wie

- o Elektronik
- o Maschinen und Anlagen
- o Medizintechnik
- o Konsumgüter
- o Ersatzteile f. Anlagenbetreiber

Ebenso läßt die Verwaltung von Produktdaten in mehreren Anwendungssystemen an unterschiedlichen Standorten mit verschiedenen Anforderungen an die Sichten auf Daten, Beschreibungen und Sprachen eine hohe Trefferrate vermuten, wie sie typischerweise anzufinden sind in:

- o ERP
- o CRM
- o PDM/PLM
- o E-commerce, Web-Kataloge

Durch hochwertige, geprüfte und konsistente Daten und Informationen wird Datenbruch vermieden. Das führt zur Verringerung von Mehraufwänden der Mitarbeiter durch schlechte Datenqualität und verbessert die Effizienz in allen Unternehmensbereichen, bei Zulieferern und bei Kunden.

Wesentlich geringere Kosten und Zeitaufwände entstehen sowohl für einmalige Datenbereinigung, als auch für laufende Prüfung und Datenkonsolidierung. Ergebnis einer gleichbleibend guten Produktdatenqualität ist schnellere und bessere Auskunftsfähigkeit durch höhere Transparenz der Produktdaten.

Reduktion von Teilebeständen durch Vermeiden von Dubletten erzielt nachweislich eine spürbare monetäre Verbesserung der Bevorratung und Lieferbereitschaft und dadurch eine deutliche Reduktion von Gemeinkosten.

Ungenauigkeit und schlechte Qualität der Produktdaten im Informationsfluss sind eine stille, meist unentdeckte Wachstumsbremse. Mit Oracle PDQ und dem PIM/MDM Lösungsansatz kann diese Bremse gelöst werden und bringt neuen Schwung in das wertvollste Gut des Unternehmens - die Produktinformationen und den Informationsfluss. Qualität der Produktdaten muss Bestandteil der Qualitätsstrategie des Unternehmens sein.

**Kontaktadresse:**

**Dr.-Ing. Helmut Maier**

Dr. Maier CSS GmbH & Co.KG  
 Am Sohlweg 6  
 DE 76297 Stutensee

Telefon: +49 (0) 7244-947100, mobil +49 (0) 171 400 1516  
 Fax: +49 (0) 7244-947103  
 E-Mail: [helmut.maier@maiercss.de](mailto:helmut.maier@maiercss.de)  
 Internet: [www.maiercss.de](http://www.maiercss.de)