

Oracle Enterprise Data Quality for Products

Ralf Kittel, ORACLE Deutschland B.V. & Co. KG

Eine unternehmensweite ERP-/MDM-Lösung erlaubt die gemeinsame Nutzung von Daten und ermöglicht eine globale Zusammenarbeit über die Firmengrenzen hinaus mit Kunden und Lieferanten. Jetzt gilt es sicherzustellen, dass die Daten immer konsistent sind und keine Dubletten enthalten.

Die Qualität des Master-Daten-Management (MDM) und speziell des Produkt-Informationen-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.

Mit der „Oracle Enterprise Data Quality for Products“-Lösung (EDQP) stehen jetzt Werkzeuge zur Verfügung, um die Daten automatisch zu reinigen, zu extrahieren, zu standardisieren und zu klassifizieren. Dubletten und Referenzen werden erkannt sowie die Übersetzung von nicht standardisierten Daten in mehrere Sprachen sichergestellt. EDQP nutzt die patentierte semantische DataLens-Technologie, um ein konsistentes MDM von strukturierten und unstrukturierten Produkt-Informationen zu erreichen. Die Semantik-Technik 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. Das sind die häufigsten Ursachen für schlechte Stammdaten in eine IT-Applikation:

- Verschiedene Mitarbeiter an verteilten, internationalen Standorten haben unterschiedliche Gewohnheiten im Umgang mit IT-Systemen
- Verschiedene Muttersprachen
- Verschiedene Mitarbeiter benutzen verschiedene Abkürzungen für die gleiche Sache
- Tippfehler

- Kontinuierliche Integration von Lieferantendaten (verschiedene Anbieter mit unterschiedlichen Produkt-Codes)
- Lieferanten, die Produkt-Informationen nicht im geforderten Format liefern können
- Fusionen und Übernahme von Firmen
- Verschiedene Systeme verwenden verschiedene Codes/Werte für ein und dieselbe Sache
- Unzureichende Recherche-Tools, kombiniert mit Zeitdruck, zwingt Benutzer dazu, doppelte Datensätze anzulegen
- Keine Einschränkungen, da keine Kontrolle vorhanden ist
- Mehrere ERP-System im Firmenverbund im Einsatz
- Fehlende Systemharmonisierung/-Konsolidierung

Stammdatenqualität in integrierten IT-Applikationen durch Semantik-Technologie

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

Schlechte Produkt-Datenqualität ist wie Schmutz auf der Windschutzscheibe. Man kann damit lange Zeit weiterfahren. Der Schmutz nimmt jedoch immer mehr zu und irgendwann wird ein Punkt erreicht, an dem es gefährlich wird, weiterzufahren. Deshalb muss die Qualität der Produktdaten zum Bestandteil der umfassenden Qualitätsstra-

tegie des Unternehmens werden. Die Stichwörter lauten Total Quality Management (TQM) und Total Data Quality Management (TDQM).

Darstellung der Ist-Situation und der Anforderungen

Aktuelle Data-Warehouse-Analysen in den Unternehmen zeigen, dass bis zu 80 Prozent der befragten Unternehmen unzufrieden mit ihrer Datenqualität sind. Mehr als 70 Prozent scheuen die Bereinigung und Standardisierung von Produktdaten, weil das Durchführen von Maßnahmen zur Verbesserung der Produktdaten-Qualität zu schwierig, zu zeitaufwändig, 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 von Metadaten, die in strukturierter oder auch unstrukturierter Form mit unterschiedlichen Schreibweisen, Synonymen und Gültigkeiten vorliegen, sowie daraus entstehende Folgerungen und Zusammenhänge (etwa Schraube, Ring, Leiter).

Ein leicht nachvollziehbares Beispiel für die Vielfalt von Produktdaten und Kategorien sind Varianten von Elektromotoren, die von vielen Firmen weltweit zum Kauf angeboten werden. Die Produkte sind teilweise austauschbar, sie besitzen die gleichen Kennwerte. Dennoch sind Vollständigkeit, Formatierungen und Produktbezeichnungen sehr unterschiedlich.

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 hin zu den vielen Lieferanten, zu den Standorten, zu Kunden – und das 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, jedoch ändern sich Kategorien, Strukturen, Bezeichnungen etc. durch die eng vernetzten, globalen Informationsflüsse in den vielfältigen Anwendungssystemen wie:

- E-Commerce, Vertrieb
- Product Design
- Lager und Logistik
- Kundendienst
- Business Intelligence

Produktdaten harmonisieren

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

Durch Homogenisierung und hohe Transparenz in den Produktdaten kann die Komplexität von Produkten, Varianten, Kategorien und Regelwerken reduziert werden. Die Pflege von Produkt-/Stammdaten ist die Grundlage für wichtige Geschäftsprozesse und -entscheidungen.

Schwierig ist die Antwort auf die Frage, wie der Grad der Produktdaten-Qualität 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 aufwändig. 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.

Lösungsstrategien für eine bessere Produktdaten-Qualität

Produktdaten sind etwas Besonderes im Unternehmen und unterscheiden sich deutlich von anderen Datenbeständen. Personal-, Kontakt- und Adress-Daten 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, Einkauf, Logistik, Produktion oder Kundendienst. Auch 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:

- Zwischen Anwendern und Abteilungen
- Zwischen Unternehmen
- Zwischen Anwendungssystemen

Daten und Information haben als immaterielles Produkt denselben Anspruch an Produktqualität wie die materiellen Erzeugnisse des Unternehmens. Produktdaten-Qualität muss Teil der TQM-Strategie des Unternehmens werden als Vision für TDQM. Die sogenannten „Master- oder Produktda-

ten“ 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 aufwändiger ist die Sicherstellung der Synchronisation und der Konsistenz von unstrukturierten Daten, Daten-Inhalten oder -Kategorien in den unterschiedlichen Anwendungssystemen. Hier kommen die Vorteile des EDQP-Managements mit semantischer Suche gegenüber bisherigen Methoden klar zum Vorschein:

- Manuelle Arbeitsweise zum Bereinigen von Produktdaten ist
 - zu zeitintensiv
 - zu langsam
 - zu teuer
- Programmierbare Tools sind
 - wenig flexibel wegen der Programmierung
 - aufwändig für die vielen Ausnahmeregelungen
 - selten skalierbar
 - kaum selbstlernend
 - dediziert
 - beschränkt auf wenige Anwendungskategorien beziehungsweise Anwendungssysteme

Die 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
- Daten-Repository, Kombinatorik und Ablaufregeln sind grafisch interaktiv definiert

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

- Wann kam der erste Mensch auf den Mond?
21. Juli 1969 02:56:20 (UTC)
- 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 EDQP-Tool werden Regeln und Filter zur Bereinigung und Zusammenführung von Datenbeständen grafisch interaktiv erstellt und können leicht erweitert werden. In Verbindung mit dem EDQP-Tool kann PIM im laufenden Betrieb die Produktdaten-Qualität 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 EDQP-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 Wörterbücher aufzubauen und mit den Begrifflichkeiten auszustatten, die die Anwender täglich benutzen.

Oracle hat kürzlich erst die Firma Silver Creek übernommen und seine PIM/MDM-Architektur um deren Data-Lensing-/Cleansing-Lösung ergänzt. Diese Data-Lensing-Funktion ist unter Oracle Product Data Quality Solution (EDQP) verfügbar. Definitionen der unternehmensspezifischen, semantischen Modelle können mit EDQP in kurzer Zeit erstellt werden und passen sich an (self-tuning). Der Einsatzbereich von EDQP geht über Suchen, Standardisieren, Zusammenführen, Übersetzen bis hin zu Umformatieren. Dialoge, Regelwerke und Abläufe für die Datenkonsolidierung erstellt der EDQP-Manager ohne Programmierung mithilfe der grafischen Technik. Ein grafisches Dashboard bringt Übersicht, zeigt die Statistik über den Prozess der Datenbereinigung und unterstützt bei der Regelung von Ausnahmen.

EDQP in eine bestehende IT-Infrastruktur integrieren

Oracle EDQP kann Daten aus vielen verschiedenen Datenquellen extrahieren, etwa Text-

te, Microsoft-Excel-Tabellen, Web-Services, XML und Datenbanken. Dabei kann EDQP sogar auf implementierte DB-Funktionen/-Prozeduren zugreifen.

Oracle EDQP kann die bearbeiteten Daten in verschiedenen Zielsysteme speichern – auch gleichzeitig in mehreren Zielsysteme, etwa im XML-, Excel- oder Text-Format. Die Daten können auch automatisch an einen FTP-Server versendet werden.

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 „Product Information Management“ (PIM). 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 in CRM, ERP, PLM, LVS, E-Commerce etc. Dieser Ansatz bildet die Basis für ein konsistentes, unternehmensweites Datenmanagement und befreit Abteilungen, Standorte oder Zulieferer davon, das gleiche System oder einheitliche Datenformate nutzen zu müssen.

Nutzenpotenzial

Eine Analyse der Datenbestände mit der semantischen Abfragetechnik (Data-Cleansing) von Oracle EDQP zeigt oft überraschend hohe Trefferquoten und legt somit den Handlungsbedarf bei der Verbesserung der Produktdaten-Qualität offen. Es ist zu empfehlen, über ein Analyseprojekt lohnende Kategorien für Data Cleansing und beispielsweise Dubletten-Bereinigung zu finden. Geeignet für Prüfung, Konsolidierung und Transformierung von Produktdaten sind große Warengruppen und Produktkategorien unterschiedlicher Lieferanten und Hersteller aus den verschiedensten Industrien wie:

- Handel
- Elektronik
- Medizintechnik
- Nahrungsmittel
- Maschinen und Anlagen

- Konsumgüter
- Ersatzteile für Anlagenbetreiber

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

- ERP
- CRM
- PDM/PLM (Product Lifecycle Management)
- E-Commerce, Web-Kataloge

Fazit

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 Produktdaten-Qualität ist die schnellere und bessere Auskunftsfähigkeit durch höhere Transparenz der Produktdaten. Die Reduktion von Teilleistungen 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 EDQP und dem PIM/MDM-Lösungsansatz kann diese Bremse gelöst und neuer Schwung in das wertvollste Gut des Unternehmens – die Produktinformationen und den Informationsfluss – gebracht werden. Die Qualität der Produktdaten muss Bestandteil der Qualitätsstrategie des Unternehmens sein.

Ralf Kittel
ralf.kittel@oracle.com