

White Paper



Fünf Schritte zu hochwertigeren Unternehmensdaten



Datenqualität

Einführung

Datenqualität und Data Governance erfordern die Einbindung von Menschen, Prozessen und Technologie.

Jede gute Geschäftsentscheidung basiert auf einer guten Qualität der Daten. Gleich ob betriebliche, analytische oder strategische Entscheidungen, jede Abteilung in einem Unternehmen, von der Vorstandsetage bis zur Vertriebsabteilung, braucht zuverlässige, präzise Informationen, um solche Entscheidungen korrekt zu treffen. Die genaue Kenntnis darüber, welche Kunden welche Produkte haben, oder welche Kunden potenzielle Chancen für den Kauf eines höherwertigen Produkts bieten (Up-Selling), basiert auf korrekten Daten. Wahr ist aber auch, dass eine schlechte Datenqualität unabwendbar zu ungesicherten Geschäftsentscheidungen, falschen Strategien und mangelhaftem Kundenservice führt.

Wenn Unternehmen datenbezogene Initiativen wie Enterprise Resource Planning (ERP) oder Customer Relationship Management (CRM) ergreifen, oder ein Programm zu umfassender Data Governance implementieren wollen, sollten Sie daran denken, dass jedes Unternehmen grundsätzlich Datenqualitätsprobleme hat. Die erfolgreiche Lösung dieser Datenqualitätsprobleme – und die Etablierung von Kontrollen zur dauerhaften Sicherung der erreichten Datenqualität – sind für den Erfolg jeder dieser Initiativen entscheidende Voraussetzung.

Es gibt viele Gründe für die explosionsartige Zunahme inkonsistenter oder nutzloser Daten, angefangen von einfachen Fehlern, verursacht durch Menschen, bis hin zu ungenügenden oder gar nicht vorhandenen unternehmensweiten Datenstandards, die alle Systeme, Geschäftsbereiche und gegebenenfalls Konzerngesellschaften erfassen. Um die Datenprobleme zu überwinden, benötigen Unternehmen eine technologische Grundlage mit der sie die Datenqualitätsprobleme aufdecken und beheben können. Datenqualitätsprobleme können jedoch nicht durch Technologie allein gelöst werden.

Ebenso essentiell ist die Einbindung der richtigen Personen und Prozesse. Dieses Whitepaper erläutert anhand einer erprobten Datenqualitätsmethodik bestehend aus fünf aufeinander folgenden Abschnitten, wie Menschen, Prozesse und Technologie in ein effektives Datenmanagement für die Analyse, Verbesserung und Kontrolle der Geschäftsdaten eingebunden werden können.

Genauere Unternehmensdaten Datenqualität
 Elektronischer Datenaustausch Datenkonsistenz
 Entlastung von Datenprüfungen
 Internet Datenverkehr Datenfehler sichtbar machen
 Richtige Listen Eventprocessing
 Datenänderung melden
 Vertrauen in gespeicherte Daten Redundanzen
 Automatisierte Web-Kommunikation
 Datenfehler finden
 Routinetätigkeiten entfallen Datenkorrektur
 Fehlerfreier Datenverkehr mit Partnern
 Duplikate Prüfungsaufwand reduzieren
 Datenfehler werden nicht übermittelt
 Leere Datenfelder Falsche Daten Datenintegrität
 Korrekte Daten Kommunikationsfähigkeit
 Datenfehler vermeiden Dynamisch
 Daten effizient und anschaulich aufbereiten

Datenqualität (DQ) Probleme und Datenbereinigung

Genau und korrekte Daten sind die Grundlage für den Geschäftsdatenverkehr und Web-Service.

Unvollständige, doppelte oder fehlerhafte Datenbestände können in Ihrem Unternehmen enorme Kosten verursachen. So erschweren z. B. Dubletten, Fehler in Abrechnungs- und Adressdaten die Kommunikation. Leere Datenfelder führen zu falschen Ergebnissen in Listen und Statistiken. Die automatische Übermittlung fehlerhafter Daten beim elektronischen Datenaustausch führen zur Vermehrung der Fehler und zu erhöhtem Aufwand zur Berichtigung von Hand, ganz abgesehen vom Image-Schaden, der durch die Multiplikation von fehlerhaften Daten entstehen kann. Es gibt viele unterschiedliche Faktoren, die für eine schlechte Datenqualität verantwortlich sein können. Unnötige Kosten und ein gesteigener Bearbeitungsaufwand sind die Folgen. Siehe Abbildung 1: Datenbanktabelle.

Folgende exemplarische Faktoren tragen entscheidend zu einer verminderten Datenqualität bei:

- **Unvollständige Daten**
- **Uneinheitliche Daten**
- **Inkorrekte Daten**

Kernprobleme sind:

- Fehlende Werte (Leere Datenfelder)
- Duplikate
- Keine Eindeutigkeit
- Widersprüche
- Schreibfehler
- Repräsentation
- Referenzintegrität
- Falsche Werte

Datenbanktabelle mit Fehlerbeispielen

		Repräsentation	Widersprüche	Ref. Integrität			
KUNDE	Knr	Name	Geb.datum	Alter	Geschlecht	Telefon	PLZ
	1234	Jran, Leo	18.2.80	36	m	999-9999	68693
	1234	Ann Joy	32.2.70	32	f	368-4511	22222
	1235	Leo Jran	18.2.80	23	m	367-3211	68693
						Fehlende Werte	Duplikate
ADRESSE	PLZ	Ort					
	28693	Ilmenau					
	28684	Ilmenauh	Falsche Werte				
	28766	BRD	Schreibfehler				

Abbildung 1: Datenbanktabelle

Folgen mangelnder Datenqualität

- Listen bringen falsche Ergebnisse
- Datenfehler werden durch Datenübermittlung multipliziert
- Großer Arbeitsaufwand für Fehlersuche
- Großer Arbeitsaufwand für Datenkorrektur
- Vertrauen in die IT geht verloren
- Elektronischer Datenaustausch nicht möglich
- Falsche Abrechnungen (Berechnungen)
- vieles mehr

Aus den aufgeführten Problemen ergeben sich folgende Fragen:

Fragen zur Verbesserung der Datenqualität

- Wie lassen sich DQ-Probleme aufdecken?
- Wie kann man DQ-Probleme überwinden?
- Wie kann man DQ-Probleme vermeiden?
- Wie lassen sich Datenfehler sichtbar machen?

Wie sich Datenqualität steigern und Datenintegrität sichern lässt

Dataquality Management - Erhöhen Sie die Datenqualität in Ihren IT-Lösungen

Vollständige, aktuelle und richtige Daten sind die Basis optimaler Geschäftsprozesse. Datenqualität ist die Basis für Ihren Geschäftserfolg. Daten und Informationen sind die Grundlage sämtlicher Geschäftsprozesse Ihres Unternehmens. Je höher die Datenqualität, desto flexibler und schlagkräftiger können Sie Ihre operativen Geschäftsprozesse gestalten und in Wettbewerbsvorteile verwandeln. Anders gesagt: Datenqualität ist die Basis für Ihren Geschäftserfolg – auf allen Anwendungsgebieten.

Richtige und genaue Daten sind die Grundlage.

Die Lösung

Effiziente und effektive Datenanalyse mit grafischen Geschäftsregeln - "Visual Rules"

Mit der Lösung "Visual Rules" für grafische Geschäftsregeln lassen sich mit Konsistenzregeln und Plausibilitätsregeln Prozesse zur Datenprüfung und Bereinigung erstellen. Die Methodik besteht aus den folgenden fünf Bausteinen:

5 Schritte zur Datenqualität

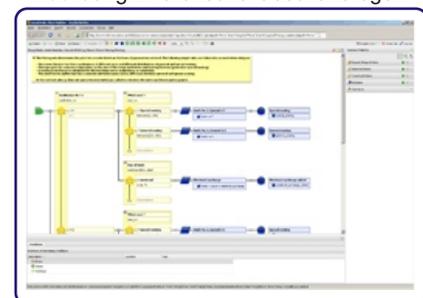
- **Daten-Profiling** - Überprüfen der Daten in Hinblick auf Fehler, Inkonsistenzen, Redundanzen oder Unvollständigkeit
- **Daten-Qualität** - Korrigieren, Standardisieren und Verifizieren der Daten
- **Daten-Integration** - Abgleichen, Zusammenführen oder Verknüpfen von Daten aus unterschiedlichen Quellen
- **Daten-Anreicherung** - Erweitern von Daten mittels interner oder externer Datenquellen
- **Daten-Kontrolle** - Kontinuierliche Überprüfung und Steuerung der Datenintegrität

Mit dem Business Rules Management System "Visual Rules" können beliebige Prüf-, Kontroll- und Bereinigungsverfahren für Datenquellen (Datenbanken) einfach erstellt werden. Auf einer speziell entwickelten grafischen Oberfläche erstellen Anwender in Fachabteilungen individuell auf Ihre Bedürfnisse zugeschnitten Prüfregeln (Wenn-Dann-Regeln und Entscheidungstabellen). Damit können alle Datenbanken, Tabellen und Datenfelder der installierten Softwaresysteme im Unternehmen gleichzeitig auf Fehler, Plausibilitäten und Änderungen geprüft und permanent überwacht werden.

Die Daten unterschiedlicher Anwendungsprogramme (CRM, ERP, PPS, Finanzbuchhaltung etc.) werden mit den grafischen Regeln verbunden und gleichzeitig gemeinsam geprüft und überwacht.

Grafische Geschäftsregeln sind leicht verständlich und einfach änderbar. Prüfprozesse sind mit "Visual Rules" übersichtlich dokumentiert und können zur dauerhaften Sicherung automatisiert ausgeführt werden.

Abbildung 2: Grafische Geschäftsregeln



Bildschirm: Grafische Geschäftsregeln mit "Visual Rules"

Eberle bietet Lösungen für jede Stufe professioneller Data Governance: Datenanalyse, Datenqualität, Datenintegration, Aufwertung und Kontrolle der Daten. Der "Visual Rules"-Lösungsansatz erlaubt den Aufbau einer Infrastruktur, die Unternehmensdaten kontinuierlich in konsistente, korrekte und verlässliche Informationen umwandelt.

Datenqualitäts-Management

Unternehmensweite Datenqualität - die Lösung "Visual Rules"

Etablieren Sie mit der regelorientierten Plattform "Visual Rules" die durchgängige Datenqualität, die genaue Sicht auf die Prozesse und die Organisation Ihres Unternehmens (Abbildung 3).

Grafische Geschäftsregeln als Kooperationsplattform für Konsistenz und Plausibilität

Über die web-orientierte Plattform "Visual Rules" können Sie Ihre Daten auf intelligente Weise prüfen, abgleichen, zusammenfassen und konsolidieren. Dies kann automatisiert in Echtzeit, ereignis- oder prozessabhängig und manuell erfolgen.

Bei der Messung der Datenqualität spielen aber auch Konsistenz- und Plausibilitätsregeln, die sich aus den aktuell implementierten Geschäftsprozessen ergeben, eine wichtige Rolle. Mit Hilfe von Konsistenzregeln wird im ersten Schritt geprüft, ob Daten innerhalb eines Systems oder systemübergreifend vorhanden und identisch sind. Wenn die Konsistenz gegeben ist, kann man mit Hilfe von Plausibilitätsregeln zusätzlich die inhaltliche Korrektheit der Daten prüfen.

In "Visuals Rules" prüfen "Wenn-Dann-Regeln" Datenfelder und Datenbanktabellen verschiedener Systeme gleichzeitig. Sie erkennen Datenfehler und Datenänderungen. Die Regeln vergleichen Einträge und prüfen nach vorgegebenen Plausibilitäten.

Neben dem Monitoring können Regeln automatisiert Entscheidungen treffen, Daten ändern und Ereignisse auslösen um Routinearbeiten zu automatisieren und von manuellen Tätigkeiten zu entlasten. So können beispielsweise Nachrichten und Dokumente über nicht konsistente Daten an Mitarbeiter verteilt und Datenfelder automatisiert berichtigt werden.

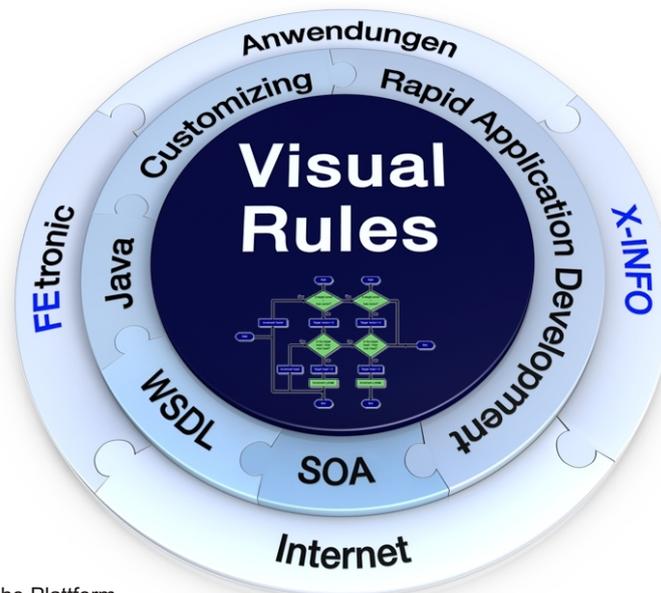


Abbildung 3: Einheitliche Plattform

Wir ermöglichen echte Zusammenarbeit von Fachabteilung und IT

Die Erstellung grafischer Geschäftsregeln mit "Visual Rules" ist einfach. Mit grafischen Geschäftsregeln können Fachabteilung und IT-Seite eines Unternehmens in praxisbezogenen Datenmanagementprojekten sehr einfach und verständlich zusammen arbeiten. Routinearbeiten wie das Prüfen von gedruckten Listen kann vielfach automatisiert werden.

Fachanwender können selbst die Logik in Regeln zur Datenprüfung und zur automatisierten Korrektur erstellen und nutzen, ohne aufwendig zu programmieren.

Fachabteilungen und IT arbeiten einfach zusammen.

Event Processing - Datenfehler automatisch anzeigen.

Grafische Regeln als zentrale Geschäftslogik für alle Programme.

Entlastung von Routinetätigkeiten.

Visual Rules die Business Rules Management Plattform.

Visual Rules, die marktführende Lösung für regelbasierte Web-Lösungen.

Die Plattform ermöglicht ein effizientes Management von Geschäftsregeln, angefangen von der Regelerstellung, über Simulation, Test, Integration, bis hin zum Monitoring und der anschließenden Pflege. Die Plattform verfügt zudem über eine intuitiv benutzbare Oberfläche, die sowohl professionelle IT-Anwender, als auch die Mitarbeiter in der Fachabteilung bei der Verwaltung komplexer Daten unterstützt.

Quick Facts

- Leistungsfähige Kollaborationsplattform für Business und IT
- Grafische Benutzeroberfläche zur Modellierung und Pflege von Geschäftsregeln
- Integriertes Test- und Qualitätsmanagement
- Transparente Regelausführung und Revisionsfähigkeit
- Durchgängige Prozessunterstützung
- Unternehmensweites Regel-Repository
- Servicefähigkeit für den Einsatz von SOA
- Hohe Performance in der Regelverarbeitung

Einfache und selbstverständliche Geschäftslogik

Die Herausforderung ist, die Fachbereiche in die Lage zu versetzen, ihre Regeln selbst zu erstellen und dabei die große Zahl an operativen Entscheidungen zu automatisieren - ohne Programmierung. Die Regelerstellung in Visual Rules ist denkbar einfach (Abbildung 4).

Geschäftslogik wird in grafische Symbole übersetzt.

Beispiele für Prüfregelein (Wenn-Dann-Regeln):

Konsistenzregel für das Attribut Zahlungsart könnten also lauten:

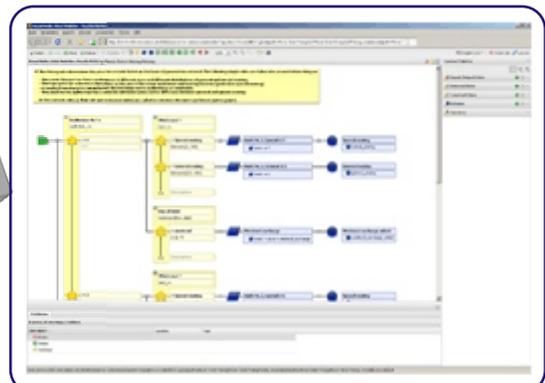
1. Wenn (Zahlungsart = Rechnung): Dann: Fehler, wenn Rechnungsadresse nicht gefüllt
2. Wenn (Zahlungsart = Lastschrift): Dann: Fehler, wenn Bankverbindung nicht gefüllt
3. Wenn (Zahlungsart = nicht gefüllt oder nicht Rechnung oder nicht Lastschrift): Dann: Fehler

Einfache Plausibilitätsregel für die Attribute Rechnungsadresse und Bankverbindung könnten lauten:

1. Rechnungsadresse: Fehler, wenn Postleitzahl keine fünfstellige Zahl ist
2. Bankverbindung: Fehler, wenn Bankleitzahl nicht zwischen 10000000 und 99999999

Aus Text wird Grafik

Abbildung der Prüfregelein in Form von grafischen Geschäftsregeln.



Text wird zum Symbol!(*)

“If - Then” = 
“Else”

Abbildung 4: Aufbau von Prüfregelein mit “Visual Rules”

Übersichtliche Repräsentation der Prüfregelein als grafischer Baum.

Einfach für jeden verständliche Symbolsprache.

Agilität und Flexibilität sind die entscheidenden Faktoren für Erfolg.



Die Zukunft des Datenmanagements

Datenqualität, Datenintegration und Stammdatenmanagement für alle Softwareanwendungen und Unternehmensbereiche mit einer einheitlichen Lösung "Visual Rules" als zentraler Plattform.

Data Quality Monitoring

Mit Visual Rules erhält man einen Datenqualitäts-Monitor, der die Datenqualität nach von Ihnen festgelegten Business Regeln überwacht. Datenänderungen werden anhand bestimmter Schwellwerte, Gültigkeitsbedingungen, Plausibilitätsregeln oder Ereignissen usw. permanent überwacht und angezeigt. Werden vorgegebene Werte überschritten oder Regeln verletzt, benachrichtigt das System den Anwender. Data Quality Monitoring wird zum Beispiel für die Analyse von Finanz-, Transaktions-, Statistik- oder Auftragsdaten eingesetzt.

Typische Einsatzszenarien

- Ad-hoc: Überprüfung nach Datenübernahme.
- Projektbezogen: Überprüfung im Rahmen eines Datenqualitätsprojektes.
- Dauerhaft: Regelmäßige Überprüfung nach festgelegten Regeln.

Funktionalität des Problem Managements

Die Anwendung von Visual Rules deckt den gesamten Datenqualitätsmanagement (DQM) - Prozess zur Überprüfung, Transformation und Integration von Unternehmensdaten ab. Grundlage für alle Abläufe sind frei definierbare, austauschbare und wieder verwendbare grafische Geschäftsregeln (Business Rules).

Unser Angebot umfasst ebenfalls vorgefertigte, branchenspezifische Rules. Beispielsweise um die Fakturierung mit der Finanzbuchhaltung dynamisch abzugleichen und zu verbinden. Individuelle DQM-Prozesse können damit ohne großen Anpassungsaufwand sofort gestartet werden.

Beispielfunktionen bei der Fehlererkennung

- E-Mail-Benachrichtigung bei Nichteinhalten festgelegter Regeln bzw. Grenzwerte.
- Erkennung der Ursachen für Datenfehler.
- Statistiken zu Trends im zeitlichen Verlauf.
- Ermittlung statistischer Kennzahlen.
- Erstellen eigener Reports über Zugriff auf Repository.
- mehr ...

Vorteile & Nutzen die Kommunikationsfähigkeit und Kollaboration optimieren.

Elektronischer Datenaustausch fehlerfrei, sicher und automatisiert möglich.

Vorteile eines professionellen Datenqualitätsmanagements (DQM)



Eberle verhilft seinen Kunden systematisch zu einer raschen Einschätzung der Art und Ursachen sowie Verbesserung problematischer Daten ihrer Organisation und damit zu einer Grundlage für unternehmensweite Initiativen rund um **Data Governance, Compliance und Master Data Management (MDM)**.

Vorteile sind:

- Reduzierung von Kosten und Risiken
- Optimierung von Kundenbeziehungen
- Verbesserung von Entscheidungsgrundlagen
- Datenaktualität permanent
- Erfüllung von gesetzlichen Vorgaben
- Optimierung der Kommunikationsfähigkeit
- Marktforschungsdaten effizient und anschaulich aufbereiten
- **mehr unter www.FE-tronic.de**

Der Nutzen korrekter und konsistenter Daten

Da Stammdaten im gesamten Wertschöpfungsprozess des Unternehmens eingesetzt werden, bieten sie ein sehr hohes Einspar- und Optimierungspotential an verschiedenen Stellen:

- Keine redundanten Stammdaten in den Datenbanken
- Verringerung der Kosten für aufwendige Bereinigung
- Effektivere, treffsichere Marketingaktionen
- Richtige Listen
- Korrekter elektronischer Datenverkehr
- Optimierung der Lagerhaltung durch Vermeidung von Redundanzen
- Einkaufspotentiale werden durch identifizierte Gleichteile erhöht
- Datenaustausch zwischen Lieferanten wird effizienter
- Mehraufwand durch erfolglose Suche wird reduziert
- Transparentere Planungsgrundlage für die Produktionsplanung
- Identifikation „unbekannter“ Gleichteile/Artikel
- **und vieles mehr!**

Workflow Datenqualität

Systematische Schritte zur dauerhaften Datenqualität.

Der systematische Ablauf der Datenüberprüfung mit grafischen Geschäftsregeln kann in folgende Schritte (Workflow) gegliedert werden:

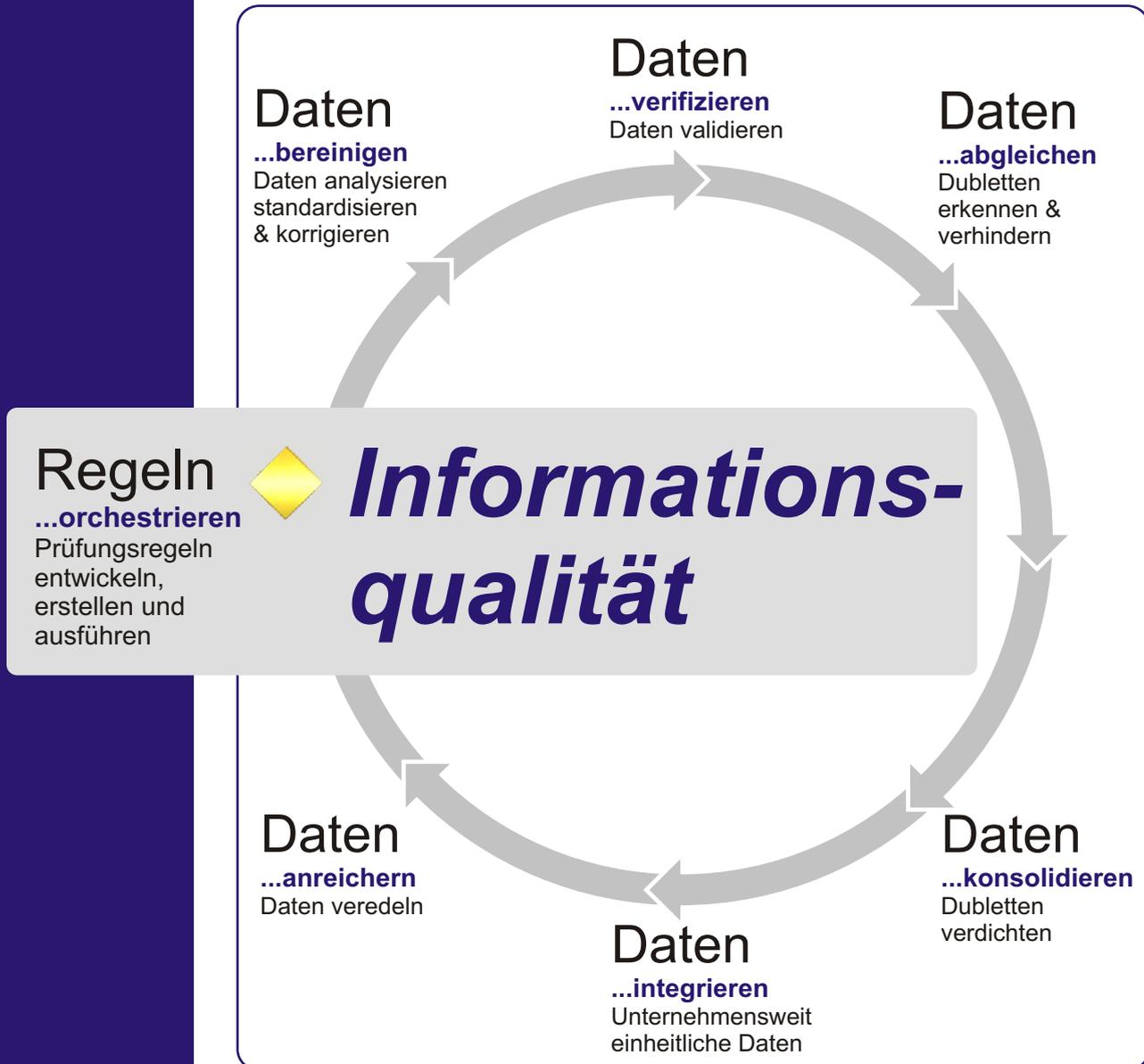


Abbildung 5: Workflow Datenqualität

Datenqualität und Datenmanagement führen zum Erfolg!

Unternehmensweite Datenqualität Ziele, To Do's, Vorteile und Nutzen

Daten Qualitäts
Monitoring

Datenintegration

Datenüberprüfung - Ziele, Aspekte zur Datenqualität

Nachfolgend sind Argumente für die Sicherung der Datenqualität aufgelistet. Der sichere, fehlerfreie elektronische Datenaustausch, korrekte Listen und das Vertrauen aller in das Datenfundament des Unternehmens sind grundsätzlichen Ziele der Maßnahmen des Datenqualitätsmanagements.

- Datenvollständigkeit überprüfen
- Datenfehler automatisch sofort erkennen und beheben
- Analyse der Daten unternehmensweit
- Vollständigkeit von Sachbearbeitungen prüfen
- Ereigniserkennung mit Anzeige
- Datenänderungen automatisch erkennen
- Abteilungen verbinden
- Ursachen und Fehlererkennung
- Zeitabhängige Datenprüfung
- Konsistenz gewährleisten
- Daten automatisch berichtigen
- Fehler, Irrläufer und Fehlaussendungen vermeiden
- Analyse der Daten bereichsübergreifend
- Softwareprogramme verbinden
- Datengenauigkeit sichern
- Arbeitsprozesse gewährleisten
- Überschneidungen entfernen
- Daten effizient und anschaulich aufbereiten
- Elektronischen Datenaustausch sichern
- Datenbestand für Listen und Statistiken prüfen
- mehr ...

Wir erarbeiten individuelle Konzepte für Datenqualität

Die aufgeführten Argumente sind eine Auswahl. Jedes Unternehmen benötigt eine eigene Konzeption bezogen auf die individuelle Datenstruktur und die Geschäftsprozesse für das Datenqualitätsmanagement.

Die fünf Schritte zu wertvolleren Unternehmensdaten

Data Profiling, Datenqualität, Datenintegration, Datenanreicherung und Datenkontrolle sind die fünf Prozesskomponenten eines Datenmanagementprogramms.

Technologie ist nur ein Teil der Lösung, die zu konsolidierten, hochwertigen Unternehmensinformationen führt. Eine robuste Technologie, die jede Datenqualitätsinitiative unterstützt, ist von zentraler Bedeutung. Noch wichtiger für den Erfolg der Initiative ist jedoch die richtige Methodik für den Einsatz dieser Technologie.

Eberle hat eine Methodik für Datenqualität und Data Governance in Unternehmen entwickelt, die aus fünf Komponenten oder Teilverfahren besteht. Sie resultiert aus der Erfahrung mit Implementierungen, die in vielen Projekten gesammelt wurden.

Die fünf Verfahren erfassen Data Profiling, Datenqualität, Datenintegration, Datenanreicherung und Datenkontrolle. Zusammen bilden sie einen bewährten, praktischen Ansatz für Data Governance – hin zu umfassendem, kontrolliertem Datenmanagement, das Unternehmen ermöglicht, ihre Daten zu analysieren, zu verbessern und zu kontrollieren.

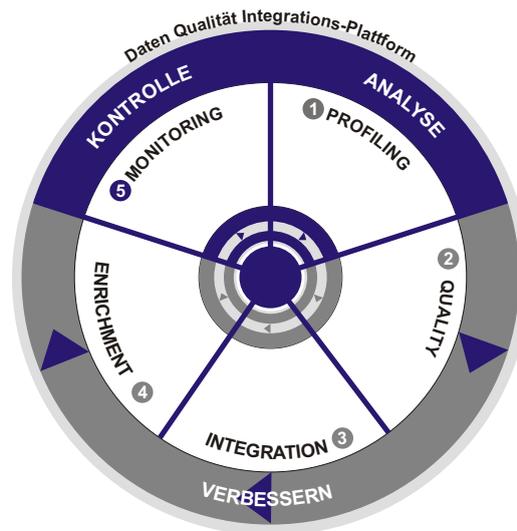


Abbildung 6 – Die fünf Komponenten eines Datenmanagementprogramms.

Diese Komponenten bieten eine Systematik, mit der sich der gesamte Prozess für eine verbesserte Datenqualität erfolgreich managen lässt. Nutzt man sie konsequent, resultiert diese Methodik in einer einheitlichen Sicht jeder Art von Daten im gesamten Unternehmen, darunter die zentralen Daten zu Kunden, Produkten und Lieferanten.

1. Datenanalyse

Datenanalyse bietet ein genaues Bild des gegenwärtigen Zustands der Unternehmensdaten.

Der erste Schritt in diesem Ansatz, das Data Profiling, untersucht die Struktur, die Beziehungen und den Inhalt vorhandener Datenquellen, um ein genaues Bild vom Zustand der Daten zu ermöglichen. Die Einschätzung des gegenwärtigen Datenzustands als Ergebnis der Datenanalyse bildet die Grundlage für die Planung der sinnvollsten Wege zur Korrektur und Harmonisierung der Informationsbestände.

Das Data Profiling erfasst drei verschiedene Aspekte der Daten und analysiert sie jeweils auf eigene Weise. Zusammen ergeben sie ein klares Bild von der Art und dem Ausmaß der Datenqualitätsprobleme.

Data Profiling besteht im Wesentlichen aus dem Prozess, Kernfragen zu den Daten für jeden der folgenden Aspekte zu beantworten:

- Datenstruktur: Entsprechen die Datenmuster den vorgesehenen Mustern? Entsprechen die Daten den korrespondierenden Metadaten?
- Dateneinträge: Sind die Datenwerte vollständig, genau und eindeutig? Sind die Daten entsprechend der geltenden Regeln standardisiert?
- Datenbeziehungen: Weisen die Daten in allen Spalten und Tabellen die erforderliche Zuordnung der spezifizierten Schlüsselbeziehung auf?
- Bestehen abgeleitete Relationen über Spalten, Tabellen und Datenbanken hinweg? Gibt es redundante Daten?

Abbildung 7 zeigt die Art der Berichte, die während der Datenanalyse erzeugt werden. Diese Analyse ist ein Beispiel für die Ermittlung von Daten. Sie zeigt für einige Felder in der Datenbank Nullwerte an. Es fällt auf, dass die Gesamtzahl von 5.000 Einträgen konstant bleibt, während einige Felder eine hohe Zahl von Nullwerten aufweisen. Aus der Information, dass hier in einigen Schlüsselfeldern ein hoher Anteil an Werten fehlt – Adresse, Stadt, Staat und besonders Geschlecht – können Geschäftsprozessanalytiker sofort das Ausmaß und die Art der Datenqualitätsprobleme erkennen.

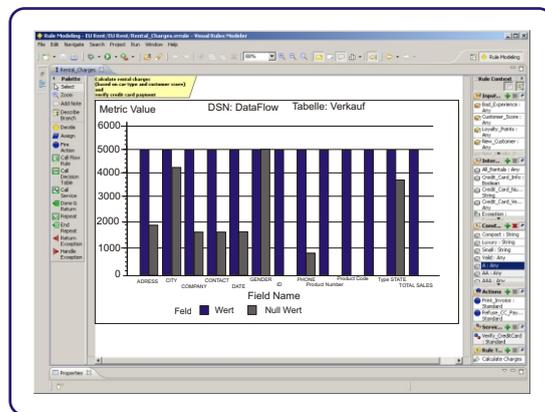


Abbildung 7: Monitoring der Datenfehlerstatistik.

Die Grafik zeigt eine Vertriebsdatentabelle mit einem hohen Anteil an Nullwerten.

Data Profiling liefert quantifizierbare, detaillierte Informationen über die Stärken und Schwächen der Unternehmensdaten. Aus diesen Kenntnissen lassen sich die weiteren Maßnahmen für eine Verbesserung der Datenqualität ableiten. Die durch die Datenanalyse gewonnenen Informationen bilden zudem den Ausgangspunkt für das nächste Verfahren, die Datenqualität.

Datenfehler werden sichtbar.

Stärken und Schwächen der Unternehmensdaten.

2. Datenqualität

Das Datenqualitätsverfahren nutzt die Erkenntnisse aus der Datenanalyse zur Lösung der Datenprobleme und beginnt mit dem Aufbau einer besseren Datenqualität.

Das Datenqualitätsverfahren nutzt die Resultate der Datenanalyse, um mit dem Aufbau hochwertigerer Daten zu beginnen. In diesem Schritt werden Fehler korrigiert, Informationen standardisiert und Daten unternehmensweit validiert. Es können verschiedene Taktiken verfolgt werden. Die am häufigsten verwendeten Methoden sind:

- Die Datenstandardisierung, um vielfältige Permutationen von Daten herauszufiltern und zu korrigieren. Ein Beispiel: „ACME Manufacturing Corporation“ kann in derselben Datenquelle als „Acme Mftg Corp“, „ACME“ und „ACME Manufacturing“ repräsentiert sein. Intelligentes Fuzzy Matching dient dazu, diese Varianten zu standardisieren und eine einheitliche Nomenklatur zu schaffen.
- Die Musterstandardisierung, um für alle Tabellen und Spalten geltende Datenmuster zu erstellen. Manche Datenbestandteile wie Telefonnummern oder Sozialversicherungsnummern sind in Musterabweichungen leicht zu finden.

Andere, wie Produkt- oder Teiledaten können in Unternehmen oder Branchen unterschiedlichen Standards folgen. Die Musterstandardisierung kann Informationen aus nicht standardisierten Formaten in ein gewünschtes Standardformat umwandeln.

- Die Adressverifizierung, um sicher zu stellen, dass Adressen gültig und nutzbar sind. Beispielsweise kann eine Regel (Visual Rules) erkennen, wenn die Stadt und der Staat nicht mit der Postleitzahl übereinstimmen. Die Software sucht nach postalischen Daten, findet die richtige Postleitzahl und ändert den Eintrag in eine korrekte Adresse.

Abbildung 8: Die Technologie führt durch die Schritte zum Aufbau von Geschäftsregeln für hochwertige Daten. In diesem Beispiel beginnt der Ablauf mit der Quelltable, setzt sich über die Standardisierung fort und führt die Verifizierung und weitere Routinen aus. Die vorausgegangene Data Profiling-Routine bringt drei Probleme ans Licht, für deren Lösung der Anwender die jeweiligen Schritte einleitet.

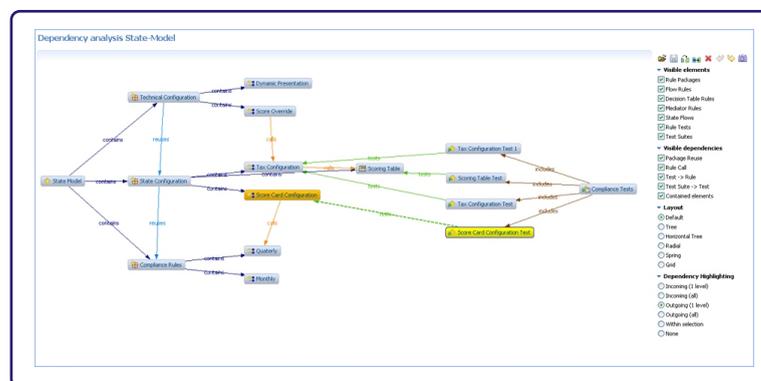


Abbildung 8: Beispiel für Regelbeziehungen für ein Datenqualitätsverfahren.

Der Ansatz ist variabel und kann jeweils an den Datensatztyp angepasst werden. Die Entscheidung für die Wahl des Ansatzes obliegt dem Fachbereich, der für die jeweiligen Daten verantwortlich ist. Zu berücksichtigen ist, dass das Datenqualitätsverfahren die Qualität vorhandener Daten verbessert, nicht jedoch die Wurzel des Datenproblems löst. Für eine nachhaltige Verbesserung der Datenqualität muss ein Unternehmen auch die Gründe ermitteln, die zu den Fehlern in den Daten führten, und geeignete Maßnahmen ergreifen, um die Daten kontinuierlich auf einem höheren Qualitätsniveau zu halten. Dies können Veränderungen in den Geschäftsprozessen, aber auch das Honorieren des korrekten Umgangs mit den Daten sein.

Ist der Datenqualitätsprozess abgeschlossen, verfügt eine Organisation über die Mittel, die Daten jeder Quelle hinsichtlich ihrer Vollständigkeit und Präzision auf das angestrebte Niveau zu bringen. Der nächste Schritt besteht in der Integration vielfältiger Quellen.

3. Datenintegration

Intelligente Regel-
technologie kann
Datenintegrations-
projekte deutlich
vereinfachen.
Die Anreicherung
von Unterneh-
mensdaten durch
Ergänzung um
Daten aus ver-
trauenswürdigen
externen Quellen
kann den Wert der
Daten steigern.

Für jeden Datentyp existieren häufig verschiedene Datensätze in mehreren Datenbanken, die dasselbe Objekt repräsentieren. Die Datenintegration dient dazu, die Daten aus verschiedenen Quellen zu optimieren. Datenintegration ermöglicht die Konsolidierung dieser voneinander abweichenden Einträge zu einem einzigen Stammdatensatz.

Beispielsweise pflegt eine Firma zwei Datenquellen für ihre Produkte:

Die Originalprodukt Datenbank, extrahiert aus ihrer Enterprise Resource Planning (ERS) Lösung in China, und eine Produkt Datenbank in Europa. Die Firma verkauft das gleiche Produkt in beiden Regionen, die Produktbezeichnungen in den Datenbanken sind jedoch unterschiedlich. Zusätzlich wurden die Produktnamen, Fabrikatsbezeichnungen und Produktbeschreibungen in jeder der Datenquellen von unterschiedlichem Personal eingegeben.

Die erste Aufgabe der Datenintegration besteht darin, ein Produkt zu finden das in beiden Quellen repräsentiert ist. Diesen Vorgang nennt man Linking. Die zweite Aufgabe besteht darin, die Daten in eine einzige Sicht des Produkts zu konsolidieren.

In Kundendaten befindet sich häufig ein in allen Datensätzen verwendetes Feld, beispielsweise die Steueridentifikationsnummer, das beim Auffinden entsprechender Einträge hilft. In diesem Fall lassen sich Vielfacheinträge für denselben Kunden schnell und leicht finden.

Auf Produktdaten trifft dies oft nicht zu. Im obigen Beispiel verhalf die Datenintegrationstechnologie dazu, die Produktinformationen zu finden, sie zu verbinden und die Daten zu konsolidieren.

Intelligentes Fuzzy Matching identifizierte eine Produktbeschreibung, die in beiden Quellen übereinstimmte. Die Datenbank in China enthielt die Beschreibung, das Fabrikat und den Produktnamen zusammen in einem Feld mit verschiedenen Mustern. Die Datenbank in Europa enthielt nur die Produktbeschreibungen, in unterschiedlichen Mustern und Abkürzungen. Zur Konsolidierung der Einträge nutzte die Integrationstechnologie die folgenden Abläufe:

- Identifikation und Isolierung (Parsen) der produktspezifischen Attribute und der Fabrikatsbezeichnung aus der Beschreibung in der Datenbank in China
- Abgleich der Unterschiede in den Fabrikatsbezeichnungen
- Abgleich der Unterschiede in den Produktattributen (Kurzformen, Abkürzungen etc.)
- Phonetische Anpassung der verglichenen Daten
- Erstellung eines Berichts zum Matching der Produkte

Nach der Datenintegration besitzt das Unternehmen jetzt die komplette Sicht über alle Daten zu einem Objekt. Die gleichen Prinzipien und Techniken können auf jeden Datentyp angewendet werden, um integrierte Stammdatensätze zu erstellen. Im nächsten Schritt soll der Wert der Daten durch Ergänzung um zusätzliche Informationen erhöht werden.

4. Datenanreicherung

Datenanreicherung kann zur Ergänzung um internationale Warenwirtschaftsstandards wie eCl@ss oder UNSPSC dienen. Durch die Integration sogenannter „Watch Lists“ und „Do-not-Call“-Verzeichnisse kann Datenanreicherung nützlich für Compliance Initiativen sein.

Bei der Datenanreicherung werden die Datensätze ergänzt, um den Informationswert und die Nützlichkeit der Daten zu erhöhen. Dies kann auf verschiedene Art und Weise erfolgen. So können beispielsweise die eigenen Daten durch Daten von Partnern oder anderen Firmen ergänzt und so genauere Informationen über Kunden aufgebaut werden – etwa zu ihrem Einkaufspotenzial oder ihrer Loyalität. Durch sogenannte Geokodierung können inzwischen geografische Informationen integriert werden. Auf diese Weise lässt sich etwa feststellen, dass alle Häuser einer bestimmten Postleitzahl in Kombination mit einer bestimmten geografischen Region nach 1980 gebaut wurden und einen bestimmten Immobilienwert übersteigen. Daraus lassen sich sehr spezifische Produktangebote für potenzielle Kunden in diesem Areal ableiten.

Eine weitere Möglichkeit der Datenanreicherung ist, die vorhandenen Kundendaten mit den Daten über das Kundenverhalten zu kombinieren. Letztere lassen sich durch die Analyse bestimmter Attribute ermitteln. Dies ermöglicht effektivere Kundensegmentierung und das Herausfiltern spezieller Vertriebschancen.

Abbildung 9 zeigt ein Beispiel für die Ergänzung einer einfachen Hausanschrift um wichtige Daten aus einer externen Quelle – mittels Matching im Rahmen der Technologie zur Verbesserung der Datenqualität.

Bestätigte Daten	Standardisierte Daten	
FETronic (Rules)	Eberle GmbH	
940 Adresse	Hausnummer 123	
Gemeinde	Elztal	
Neue Daten angefügt		
	Liefer Adresse ID	35
	Liefer Ort ID	2
	Land Nummer	049
	Land Name	Deutschland
	Bundesland	BW
	Latitude	36.7841
	Longitude	-87.784
	Block Gruppe	3457898221

Abbildung 9: Anreicherung der Daten einer Adresse.

Datenanreicherung lässt sich auf mehr als nur Adressdaten anwenden. Viele Arten von Daten, darunter Produkt-, Inventar- und Finanzdaten, können durch die Verbindung zu externen Quellen ergänzt und dadurch aufgewertet werden. So können internationale Warenwirtschaftsregister wie UNSPSC oder eCl@ss dazu dienen, die Produkte oder Lieferantendaten zu standardisieren und damit Einkauf und Verkauf zu rationalisieren. Die Integration dieser universellen Klassifizierung unterstützt die Genauigkeit der Lieferantendaten, was wiederum die Effizienz der Bestellungen erhöht und den Weg für reibungslose internationale Geschäfte ebnet.

Datenanreicherung unterstützt auch das von häufigen Neuerungen betroffene Compliance Management. Es ermöglicht den Vergleich von Kundenlisten oder Transaktionen mit externen Listen auf Übereinstimmung mit den gesetzlichen und industriellen Vorschriften. Diese Listen können beispielsweise dazu dienen, der Gefahr von Geldwäsche oder Korruption oder Verstößen gegen „Do-not-Call“-Verzeichnisse vorzubeugen.

5. Datenkontrolle

Die Geschäftsprozessregeln, die im Verlauf des gesamten Verfahrens entwickelt wurden, können anschließend in der Datenkontrolle zur Sicherung vor dem Absinken der Datenqualität verwendet werden.

Geschäftsprozessregeln als Web-Service.

Zu einem umfassenden Data Governance Programm gehört eine kontinuierliche Datenkontrolle. Nur so ist gewährleistet, dass der Datenqualitätserfolg aus den aufwändigen Korrektur- und Aufwertungsmaßnahmen nicht durch die allmähliche Rückkehr von Datenfehlern zunichte gemacht wird. Mit aktiver Datenkontrolle haben Unternehmen Kenntnis vom Zustand ihrer Daten und können die Ursachen für Datenqualitätsprobleme isolieren und korrigieren. Datenkontrolle kann auf verschiedene Weise erfolgen. Die einfachste Datenkontrolle automatisiert Data Profiling Reports, die in regelmäßigen Abständen Analysen der Daten in Bezug auf Ausnahmen und nichtstandardisierten Daten liefern. Sie identifizieren beispielsweise Ereignisse wie einen zu großen Prozentsatz an außergewöhnlichen Daten, und können gegebenenfalls eine eMail oder eine Systemmeldung an zuvor bestimmte Adressaten oder Verantwortliche eines Bereichs auslösen.

Web-Service - dynamische Datenkontrolle

Eine dynamischere und gleichzeitig effektivere Methode der kontinuierlichen Datenkontrolle ist die Durchsetzung von Geschäftsprozessregeln in Form von Web Services. Mit dieser Methode können die gleichen Regeln, wie sie für die Bereinigung und Anreicherung der Daten benutzt wurden, dazu dienen, Daten in Echtzeit während der Eingabe oder auf ihrem Weg durch das Unternehmen zu überprüfen.

Die Vorteile dieses Ansatzes beinhalten die folgenden Möglichkeiten:

- Es können Data Governance Regeln aufgestellt und integriert werden. Damit ist gewährleistet, dass die unternehmensspezifischen Regeln für das Datenmanagement befolgt werden.
- Aufgabenkritische Prozesse werden erkennbar, indem Verstöße gegen oder Ausnahmen von Geschäftsregeln in einem Repository protokolliert werden. So lassen sich Trends verfolgen und die Prozesse entsprechend anpassen.
- Es können definierte Ereignisse ausgelöst werden, um die Daten automatisch zu korrigieren und entweder festgelegte Personen durch eMails darüber zu informieren, die Daten in einem Repository zu protokollieren oder einfach die korrigierten Daten in eine Datenbanktabelle einzutragen.

Unternehmensdaten müssen analysiert, optimiert und kontrolliert werden.

Zusammenfassung

Die Qualität jeder Geschäftsanalyse ist nur so gut wie die Daten, die ihre Grundlage bilden. Ohne Daten, die unternehmensweit konsistent, akkurat und zuverlässig sind, kann ein Unternehmen leicht irreführende, falsche und potentiell geschäftsschädigende Schlussfolgerungen ziehen. Für eine erfolgreiche Verbesserung der Datenqualität müssen Unternehmen zunächst eine Methodik und Lösung finden, die alle fünf wesentlichen Bestandteile umfasst: Data Profiling, Datenqualität, Datenintegration, Datenanreicherung und Datenkontrolle.

Visual Rules - Geschäftsregeln

Visual Rules ist die Technologie für Datenintegration, die es Unternehmen ermöglicht, Datenqualitätsmaßnahmen auf einem hohen Niveau zu realisieren.

Das Ziel ist eine einheitliche Sicht der Datenbestände zu Kunden, Produkten, Lieferanten und weiteren Datenarten – in allen möglichen Applikationen im gesamten Unternehmen.

Durch die Kombination einer ausgereiften Methodik mit geeigneter Technologie kann die Datenqualität erheblich verbessert werden.

Dies wird zu soliden Entscheidungen führen und mehr Wettbewerbsvorteile schaffen. Denn beide beruhen auf der Verfügbarkeit hochwertiger Daten.

Profitieren Sie von unserer Erfahrung mit regelorientierten wissensbasierenden Systemen

Seit über zehn Jahren beraten wir und entwickeln regelbasierende Lösungen für unsere Kunden. Profitieren Sie von unseren Erfahrungen für schlanke und durchdachte Prozesse.

Die Eberle GmbH ist auf Datenmanagement für die Sicherung hoher Datenqualität spezialisiert. Wir begleiten unsere Kunden von kurzfristigen Maßnahmen der Datenbereinigung bis hin zur Entwicklung einer langfristigen und unternehmensweiten Strategie.



Eberle GmbH

Wiesenweg 1
74834 Elztal Auerbach

Tel.: 06293/9278-0
Mail: s.eberle@x-info.de

www.FE-tronic.de

© Eberle GmbH, 2013. Alle Rechte vorbehalten.

Weitergabe und Vervielfältigung dieser Publikation oder von Teilen daraus sind, zu welchem Zweck und in welcher Form auch immer, ohne die ausdrückliche schriftliche Genehmigung durch die Eberle GmbH nicht gestattet. In dieser Publikation enthaltene Informationen können ohne vorherige Ankündigung geändert werden.

FEtronic ist eine eingetragene Marke von Eberle. MLDS, Visual Rules und Work Frame Relations sind eingetragene Marken der Bosch Software Innovations. Die innovations-Rule-Technologie ist patentamtlich geschützt (GBM 20014430).

Verwendete Produkt- und Firmenbezeichnungen sind eingetragene Marken und unabhängig von ihrer Kennzeichnung - Eigentum ihrer jeweiligen Inhaber
Textauszüge: Data Flux, Bilder, Innovations