HTW Chur
Hochschule für Technik und Wirtschaft
University of Applied Sciences

Master Thesis

MSc BA – Major Information & Data Management 2016

Herausforderung Datenqualität

Modellgestütztes Vorgehen für die systematische Verbesserung der Qualität von betrieblichen Daten

Autor: Markus Rüf
MSc BA – Major Information & Data Management 2016
markus.ruef@msc.htwchur.ch

Advisor: Prof. Dr. Ingo Barkow, Pulvermühlestrasse 57, 7000 Chur
Co-Advisor: Prof. Dr. Wolfgang Semar, Pulvermühlestrasse 57, 7000 Chur

Chur, August 2018

Formales

Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen in der vorliegenden Arbeit verzichtet. Sämtliche Personenbezeichnungen gelten für beide Geschlechter.

Kurzfassung - Abstract

Die Qualität der Daten avanciert in Zeiten von Big Data, Machine Learning und bei Business Intelligence und Data Mining Anwendungen zu **DEM** entscheidenden Faktor. Ohne die korrekten Grunddaten werden diese Vorhaben schnell zur Farce. Die Arbeit versucht die Ausprägungen, die Ursachen und die Auswirkungen der Qualität von betrieblichen Daten aufzuzeigen und entwickelt ausgehend vom aktuellen Forschungsstand ein Vorgehensmodell für die proaktive Steuerung von Datenqualität in einem Unternehmen.

Mit einer partizipativen Fallstudie wird das Vorgehensmodell auf seine Praxistauglichkeit in einem KMU untersucht. Die Datenanalyse wird mit Data Profiling durch die Validierung des Datenbestandes anhand von Metadaten wie Business Rules für die Qualitätsniveau Bestimmung praxisnah demonstriert.

Nachhaltig verankert wird Datenqualität erst mittels Datenqualitätsmanagement, mit dem Rollen und Zuständigkeiten für die Daten definiert werden. Über einen PDCA Regelkreis werden Datenqualitätsprobleme aktiv aufgedeckt und entsprechende Lösungen und Massnahmen mit den Datennutzern erarbeitet. In den Do und Check Phasen werden die Massnahmen mit einem abgestimmten DataQuality Monitoring überwacht und soweit iterativ verfeinert, dass sie in der Act Phase effizienzsteigernd und wertschöpfend in die Geschäftsprozesse integriert werden.

Die Arbeit schliesst mit Umsetzungsempfehlungen für ein Nutzen stiftendes Datenqualitätsmanagement in Unternehmen.

Schlagwörter: Datenqualität, Datenqualitätsmanagement, Datenqualitätmetriken, Datenqualitätsmanagementvorgehensmodell, DataProfiling, Datenqualitätskennzahlen, Stammdatenmanagement, DataCleansing, Datenbereinigung.

Inhalt

1. Einleitung 8

1.1. Motivation und Relevanz des Themas 9

1.2. Ziele und Forschungsfragen 10

1.3. Forschungsmethoden und Vorgehensweise 10

2. Terminologie und Grundlagen 12

2.1. Daten - Information - Wissen 12

2.2. Datenarten 14

2.3. Metadaten 16

2.4. Definition Datenqualität 16

3. Treiber, Ursachen und Wirkungen von Datenqualität 25

3.1. Treiber der Datenqualität 25

3.2. Herausforderungen und Ausprägungen schlechter Datenqualität 26

3.3. Ursachen schlechter Datenqualität 27

3.4. Auswirkungen schlechter Datenqualität 29

4. Datenqualitätsmanagement 31

4.1. Datenqualitätsmanagementansätze 31

4.2. Datenqualitätsmanagementmodell 33

4.3. Verbesserung der Datenqualität 34

4.4. Vereinfachendes generisches Vorgehensmodell 35

4.5. Vermeidung von neuen Datenqualitätsproblemen 35

4.6. Datenanalyse 36

4.7. Evaluation von Data Profiling Software Tools 38

5. Fallstudie Datenqualitätsmanagement in einem Produktionsunternehmen 44

5.1. Permapack – Einführung eines Datenqualitätsmanagement in einem heterogenen Handels- und Produktionsunternehmen 44

5.2. Ausgangssituation Datenlandschaft 45

5.3. Handlungsdruck 45

5.4. Ziele der Fallstudie 46

5.5. Realisierung von Datenqualitätsmanagement bei Permapack 46
5.6. Erkenntnisse aus der Fallstudie bei Permapack 59
6. Zusammenfassung und Fazit 61
6.1. Zusammenfassung 61
6.2. Einschränkungen 62
6.3. Handlungsempfehlung für ein systematisches Datenqualitätsmanagement 62
6.4. Resümee 63
Literaturverzeichnis 64

Tabellenverzeichnis

Tabelle 1: Übertragbarkeit der Ansätze von Garvin auf die Datenqualität (Schumacher, 2011, S.166) 20
Tabelle 2: Bewertungsmatrix der Data Profiling Software (eigene Darstellung) 43
Tabelle 3: Kennzahlendefinition (eigene Darstellung) 58
Tabelle 4: Antworten der Umfrage (eigene Darstellung) 68
Tabelle 5: Datenrelevanzfestlegung (eigene Darstellung) 70
Tabelle 6: Datentypfestlegung (eigene Darstellung) 70
Tabelle 7: Datenmanipulationsrechte (eigene Darstellung) 70
Tabelle 8: Metadatendokumentation (eigene Darstellung) 72
Tabelle 9: Formular Tabellen Feld Zuordnung (eigene Darstellung) 74
Tabelle 10: Valdidierungsregelen (eigene Darstellung) 75
Tabelle 11: neue Default-Werte (eigene Darstellung) 75

Abbildungsverzeichnis

Abbildung 1: Von Daten zur Information (Tiemeyer, 2015, S. 554) 13

Abbildung 2: Begriffslandkarte für das Datenqualitätsmanagement (Otto & Österle, 2016, S. 28) 13

Abbildung 3: Typologie von Daten (vgl. Schemm, 2009, S. 20) 14

Abbildung 4: Zusammenhang zwischen den Datenarten (Schmidt, 2010, S. 18) 16

Abbildung 5: typischer zeitlicher Datenqualitätsverlauf in Unternehmen (Otto & Österle, 2016, S. 32) 19

Abbildung 6: Datenqualitätspyramide (eigene Darstellung nach Gebauer & Windheuser, 2008, S. 89) 20

Abbildung 7: Taxonomie von Datenqualitätskriterien (Apel et al., 2010, S. 22) 22

Abbildung 8: Taxonomie von Datenqualitätskriterien (Apel et al., 2010, S. 23) 24

Abbildung 9: Geschäftstreiber von Datenqualität (Apel et al., 2010, S. 30) 25

Abbildung 10: Daten Qualitätsprobleme bei Stammdaten (Schuh & Stich, 2014, S. 19) 27

Abbildung 11 Ursachen von schlechter Datenqualität (Apel et al., 2010, S. 34) 27

Abbildung 12: Ursachen für schlechte Stammdatenqualität (Schuh & Stich, 2014, S. 21) . 28

Abbildung 13: Total Quality data Management Methodik nach English (Apel et al., 2010, S. 25) 31

Abbildung 14: Vergleich der Ansätze des Datenqualitätsmanagements (Ofner, 2013, S. 103) 32

Abbildung 15: Framework für Datenqualitätsmanagement (Otto & Österle, 2016, S. 23) ... 34

Abbildung 16: Vorgehensmodell mit Prozessschritten (eigene Darstellung) 35

Abbildung 17: Verbesserung der Prozesse in den Quellsystemen (Apel et al., 2010, S. 104) 36

Abbildung 18: Aufgaben der Datenanalyse (Apel et al., 2010, S. 110) 37

Abbildung 19: Talend Quality Data Open Studio Dataprofiing Konfiguration (eigene Darstellung) 39

Abbildung 20: Mustererkennung auf Attributebene FirstUser (eigene Darstellung) 39

Abbildung 21: Data Cleaner – Werteverteilung auf Attributebene (eigene Darstellung) 40

Abbildung 22: Microsoft SQL Data Quality Client (eigene Darstellung) 41

Abbildung 23: Microsoft SQL Data Quality Data Dictonary u Knowledge Base (eigene Darstellung) 41

Abbildung 24: Beispiel Projekt und die Datenprofil Ansicht für die Musterverteilung des Attributs Bezeichnung 1 der Artikel Tabelle (eigene Darstellung) 42

Abbildung 25: Datensystem-Landschaft von Permapack (eigene Darstellung) 45

Abbildung 26: Rollen und Gremien in der Aufbauorganisation (Apel et al., 2010, S. 54) 48

Abbildung 27: Umfrageergebnis der C3 Nutzer nach Datenqualitätskategorien (eigene Darstellung) 49

Abbildung 28: Umfrageergebnis der C3 Nutzer nach Datenqualitätsdimensionen (eigene Darstellung) 50

Abbildung 29: Umfrageergebnis Datenqualitätsdimensionen pro Abteilung (eigene Darstellung) 50

Abbildung 30: Metadatendokumentation (eigene Darstellung) 51

Abbildung 31: Metadatendokumentation (eigene Darstellung)Apel et al., 2010, S. 54) 52

Abbildung 32: Musterverteilung von Attribut ZNR (Artikelnummer) mit MS SSIS (eigene Darstellung) 53

Abbildung 33: Verfahren zur Validierung und Filterung von Daten (Apel et al., 2010, S. 150) 55

Abbildung 34: zyklische Verbesserung der Datenqualität mit PDCA (eigene Darstellung) . 56

Abbildung 35: Metadaten eines Formulares im C3: Pflichtfelder, Default-Werte (eigene Darstellung) 57

Abbildung 36: Produktivitätskennzahlen einer Fachabteilung bei Permapack (eigene Darstellung) 59

Abbildung 37: Online Umfrage (eigene Darstellung) 68

Abbildung 38: Eingabemaske Artikeldaten (eigene Darstellung) 71

Abbildung 39: Vorgehensmodell mit Prozessschritten (eigene Darstellung) 73

Abbildung 40: Validierungsregel auf Datenbank Ebene (eigene Darstellung) 75

Abbildung 41: C# RegEx Funktion (eigene Darstellung) 76

Abbildung 42: Verwendung der RegEx Funktion in SQL 2014 (eigene Darstellung) 77

Abkürzungsverzeichnis

API	Application Programming Interface
BE	Business Engineering
BI	Business Intelligence
CDQ	Corporate Data Quality (engl. für konzernweite Datenqualität)
CDQM	Corporate Data Quality Management (engl. für Qualitätsmanagement konzernweit genutzter Daten)
DB	Datenbank oder Database
DGIQ	Deutsche Gesellschaft für Informations- und Datenqualität
DQM	Data Quality Management (engl. für Datenqualitätsmanagement)
DRM	Data Relationship Management
DWH	Datawarehouse
etc.	et cetera
ERP	Enterprise Ressource Planning
GUI	Graphical User Interface (engl. für graphische Benutzerschnittstelle)
HTW	Hochschule für Technik und Wirtschaft Chur
ISO	Internationale Organisation für Normung
IT	Informationstechnologie
MDM	Master Data Management (engl. für Stammdatenmanagement)
RDBMS	Relational Database Management System
Small Data	strukturierte Daten, klein genug um vom Menschen verstanden zu werden
SOA	Service-Oriented Architecture (engl. für Serviceorientierte Architektur)
SQS	Schweizerischen Vereinigung für Qualitäts- und Managementsysteme.
TDQM	Total Data Quality
UWG	Gesetz gegen den unlauteren Wettbewerb
WMS	Wissensmanagementsystem
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformation
z. B.	zum Beispiel

1. Einleitung

Revival of Dataquality Management (DQM) oder warum wird Datenqualität in Unternehmen meist unsystematisch und selten proaktiv als vorbeugender Supportprozess betrieben?

Daten sind der Treibstoff für die Geschäftsprozesse von Unternehmen im 21 Jahrhundert. - Zusammen mit den Informationstechnologien ermöglichen sie gänzlich neue Geschäftsmodelle. Aber auch etablierte Unternehmen sind durch gewachsene und komplexere Systeme immer mehr auf Daten und deren Qualität angewiesen. Es zieht sich quer durch alle Branchen, vom Handel über Dienstleistungsunternehmen bis hin zu Produktionsbetrieben - kaum ein Arbeitsplatz bei der die Informationstechnologien und der Umgang mit Daten nicht die entscheidende Rolle spielen.

Bedingt durch den Wachstum der digitalen Systemlandschaften sind in den Unternehmen in den letzten Jahren und Jahrzehnten grosse Datenberge erzeugt worden. Im Handel sind das Berge von Abwicklungsdaten wie Aufträge, Lieferscheine und Rechnungen usw. In Produktionsfirmen entstehen hingegen viele Fertigungsdaten wie, Messwerte, Bewegungs- und Rückverfolgungsdaten bei der Produktfertigung. Besonders wichtig sind für den Handel die Kunden- und Lieferantendaten und in der Produktion die Maschinen- und Artikeldaten, die als Stammdaten bezeichnet werden.

Massen an Daten aus Einzelanwendungen wie Maschinensteuerungen und deren Sensoren, Kundendaten aus verschiedensten Unternehmensbereichen, Logistiktransaktionsdaten, Materialwirtschaftsdaten usw. werden heute schon über Enterprise Ressource Planning (ERP) Systeme zusammengeführt. Der Vorteil dabei ist, dass Teile dieser Daten, bspw. Transaktionsdaten auf gemeinsame Stammdaten zurückgreifen und so vor allem der Aufwand für die Mehrfacherfassung und die Redundanz, mit all seinen Folgen, verringert werden kann. Datawarehouse- (DWH) und Business-Intelligence- (BI) Lösungen versuchen dem Benutzer einen Mehrwert aus diesen Daten durch sinnvolle Umformung und Aggregierung in Form von „Information“ zu bieten.

Durch IOT und die Industrie 4.0 ist eine zunehmende Dezentralisierung der Datenhaltung feststellbar, da nun nicht mehr an festgelegten Kontrollpunkten wie z. B. beim Wareneingang der Informations- und Güterfluss zusammenkommen, sondern das Produkt selber seine Daten mittransportiert. Für diese vielen Kontaktpunkte braucht es einen standardisierten Datenaustausch und einen dynamischen Zugriff auf übergeordnete Datencenter. Durch diese Zunahme der Komplexität und des Datenaustausches ist die Qualität der Daten immer mehr der entscheidende Faktor für eine reibungslose Funktion der Systeme und Geschäftsprozesse.

1.1. Motivation und Relevanz des Themas

> *It's like when you see a skyscraper: You're impressed by the height, but nobody is impressed by the foundation. But make no mistake, it's important. You have to have good data management to take advantage of AI."*
>
> Akshay Tandon, Vice President/ Head of Strategy and Analytics, LendingTree Inc.

Den Grundstock aller Geschäftsprozesse bilden die sogenannten Stammdaten. Durch Firmenfusionen, neue Geschäftsfelder, engere Verzahnung in der Lieferkette (Supply Chain) und Just-in-time Produktion müssen heterogene Datenbestände zusammengeführt werden. Dafür ist eine gute Stammdatenqualität fundamental. Sie wird aber oft vernachlässigt. Juan (1988) schätzte schon 1988, dass ein Viertel des operativen Unternehmensgewinns aufgrund schlechter Datenqualität verloren ginge.

Dies macht ein Management der Daten notwendig. In der unternehmerischen Praxis hat es aber nicht immer den Stellenwert, den es bräuchte. Stocken Prozesse so liegt es oft an der Qualität der Stammdaten. Durch mangelnde Pflege und Sorgfalt (zu viele Zugriffsrollen, unklare Zuständigkeiten) nehmen Inkonsistenzen und Redundanzen weiter zu.

Vor allem zwei Hauptursachen sind der Grund für mangelnde Datenpflege:

Die konkreten Auswirkungen sind oft nicht unmittelbar bei der Erfassung der Daten sichtbar, sie sind vielmehr zeitlich verschoben. Die andere Ursache liegt in fehlenden Lösungsansätzen in den Betrieben, wie die Datenqualität in den Griff zu bekommen ist. Die Hersteller von Informationssystemen wie ERP und DWH Systemen helfen meist auch nur bedingt weiter oder haben oft nur inadäquate Lösungen.

Ganzheitlich betriebene Datenpflege ist schon lange nicht mehr nur ein Thema der Informatikabteilungen, sondern beinhaltet strategische, organisatorische und informationstechnische Massnahmen. Anforderungen an die Datenqualität kommen von gesetzlichen Regelungen, betrieblichen Regeln wie Business Rules und weitere Regeln von den Datennutzern. Um hohe Datenqualität langfristig zu sichern, reichen kurzfristige Datenbereinigungsmassnahmen der aufgetretenen Mängeln nicht aus. Deshalb sind passende Methoden zur Fehler-Analyse, Werkzeuge zur Bereinigung und Datenqualitätskennzahlen zur Überwachung der Qualität zu planen und zu definieren um die benötigte Datenqualität kontinuierlich sicherzustellen.

1.2. Ziele und Forschungsfragen

Das Ziel dieser Arbeit ist es Unternehmen Möglichkeiten und Massnahmen für die Optimierung der Qualität ihrer betrieblichen Daten aufzuzeigen. Die übergeordnete Fragestellung lautet dem gemäss:

Wie kann die Datenqualität in einem Unternehmen nachhaltig verbessert werden?

Daraus resultieren folgende Unterfragestellungen:

Was sind betriebliche Daten?

Was ist unter Datenqualität zu verstehen?

Was sind die Treiber für Datenqualität?

Warum soll die Datenqualität proaktiv gesteigert werden, wenn derzeit keine offensichtlichen Probleme auftreten und die Massnahmen dafür sehr teuer sind und ihr Nutzen unklar ist?

Welche Konzepte und Ansätze gibt es für die Steuerung von Datenqualität?

Wie gehen Unternehmen vor, um ihre Datenqualität in den Griff zu bekommen?

1.3. Forschungsmethoden und Vorgehensweise

Um die Forschungsfrage zu beantworten werden folgende Forschungsmethoden angewendet.

1.3.1. Literaturrecherche

Die Literaturrecherche ist die Grundlage jedes Forschungsprojektes (Garfield, 1997). Im Verlauf der Literaturrecherche sollen die für das Forschungsthema in Frage kommenden Quellen identifiziert werden, um schliesslich die Forschungsergebnisse auf einer fundierten Wissensbasis abstützen zu können (Rowley, 2004, S. 31). Die Literaturrecherche ist auch deshalb notwendig, um zu eruieren, ob es nicht schon ähnliche Problemstellungen gibt, die bereits erforscht wurden. Die Literaturrecherche wurde im Internet, auf Google Scholar und auf Edis (http://www1.unisg.ch/www/edis.nsf), der elektronischen Dissertationsdatenbank der Universität St. Gallen durchgeführt. Vor allem auf der Forschungsplattform Alexandria (https://www.alexandria.unisg.ch/index.html) der Universität St. Gallen sind viele wissenschaftliche Publikationen und Artikel zu Stammdatenmanagement vorhanden.

1.3.2. Partizipative Fallstudie

Mit einer partizipativen Fallstudie arbeiten Wissenschaftler und Fachexperten eines Unternehmens an einer konkreten Problemstellung und versuchen gemeinsam die

Anwendbarkeit eines Modells in der Praxis zu demonstrieren. Die dabei aus der Praxis gewonnenen Erkenntnisse sollen zurück in das Modell fliessen (Susman & Evered, 1978, S. 580). Aus der Literaturrecherche wird ein Modell für Datenqualitätsmanagement identifiziert und ein daraus entwickeltes Vorgehensmodell soll in Form einer Fallstudie in einem realen Produktionsunternehmen auf seine Tauglichkeit hin untersucht werden.

1.3.3. Umfrage Datenqualitätswahrnehmung der Datennutzer

Mit einer Umfrage können gezielt Meinungen und Eindrücke der Teilnehmer mit Hilfe eines Fragebogens erfasst werden (Siau & Rossi, 2007, S. 8). Eine Umfrage über die subjektive Wahrnehmung von Datenqualitätsdimensionen wurde im Rahmen dieser Arbeit genutzt, um einerseits Transparenz und Vertrauen über das Forschungsvorhaben bei den Fallstudienteilnehmern zu schaffen und andererseits um erste Anhaltspunkte über Datenqualitätsprobleme zu erhalten.

1.3.4. Summative Evaluation von Data Profiling Software

Eine summative Evaluation dient zur Ergebnisbewertung von Prozessen und Funktionen. Mit der Evaluation von Daten Profiling Software Tools sollen ihre Eignungen auf die Unterstützung und Wirksamkeit auf Datenqualitätsanalysen unter dem Fokus der Effizienz anhand eines Anforderungskataloges bewertet werden.

1.3.5. Aufbau und Gliederung

Damit die Fragestellungen mit den Forschungsmethoden und der Vorgehensweise beantwortet werden können, gliedert sich die Arbeit in folgende Kapitel:

In Kapitel 2 werden die Grundlagen für das Verständnis der Problemstellung mit der Begriffserklärung von Daten, Datenarten, Datenqualität und Datenqualitätsdimensionen geschaffen. Kapitel 3 widmet sich den Treibern, den Ursachen und Wirkungen von Datenqualität im Unternehmen. Kapitel 4 betrachtet den Stand der Forschung für Datenqualitätsmanagement Modelle und erläutert einen übergeordneten Rahmen für Datenqualitätsmanagement als Unternehmensprozess. Teilaspekte dieses Prozesses werden im Detail erläutert. Basierend auf diesem Framework wird ein Vorgehensmodell für die Umsetzung von Datenqualitätsmanagement mit einer Aufbauorganisation entwickelt. Der wichtige Prozessschritt Datenanalyse und die Softwareauswahl wird beschrieben

Kapitel 5 beschreibt die Fallstudie mit der praktischen Erprobung des Vorgehensmodells und deren Durchführung mit den jeweiligen Ergebnissen. In Kapitel 6 wird ein Fazit gezogen und es werden Handlungsempfehlungen für die Erfolgreiche Implementierung von Datenqualitätsmanagement in einem KMU gegeben.

2. Terminologie und Grundlagen

In diesem Kapitel werden die zentralen und wesentlichen Begriffe, die zum Verständnis der Arbeit beitragen dargestellt. Zunächst werden die oft synonym verwendeten Begriffe Daten Information und Wissen definiert. Daraus abgeleitet werden die Arten von betrieblichen Daten präzisiert. Es folgen Erläuterungen zu den Begriffen Datenqualität, Datenqualitätsdimensionen, die Ausprägungen von schlechter Datenqualität anhand Datenqualitätskriterien und Datenqualitätsmanagement.

2.1. Daten - Information - Wissen

Betriebliche Daten in Form von Zeichen beschreiben und speichern Merkmalswerte von Datenelementen, Datensätzen bzw. Objekten die in unternehmerischen Geschäftsprozessen verarbeitet werden.

Gemäss ISO/IEC 2382-1 (1993) sind Daten die formalisierte, d. h. für die weiterführende Verarbeitung, Interpretation und Kommunikation geeignete Repräsentation der Eigenschaften von Geschäftsobjekten.

Betriebliche Informationssysteme stellen die Daten in einem bestimmten Nutzungskontext dar. Durch z. B. Analyse, Interpretation und Organisation werden die Daten zu Information für den Nutzer. Die Transformation der Daten durch Hinzufügen von deren Bedeutung ist im Allgemeinen unabhängig vom Nutzer der Daten. Je nach Kontext und Sachlage kann die Information relevant oder irrelevant sein (Kuhlen, Semar, & Strauch, 2013, S. 4). Die Information (Datenobjekte) werden in den Geschäftsprozessen verarbeitet und sind der Ausgangspunkt für Entscheidungen im Unternehmen (English, 1999, S. 8).

Wird Information in einen pragmatischen Handlungskontext eingebunden, entsteht Wissen. Wissen resultiert somit aus der Verarbeitung von Information durch den eigenen Erfahrungskontext des Nutzers durch Qualifizierung und Quantifizierung, Verknüpfung und Verbreitung. Diese Transformation hängt vom Nutzer der Information und dessen Situation ab. Das als Ergebnis dieses Transformationsprozesses entstandene Wissen ermöglicht dem Informationskonsumenten auf Ereignisse zu reagieren (Kuhlen et al., 2013). Beispiel dafür ist eine Zahl (Datum), die den gewünschten Liefertermin einer Kundenbestellung darstellt (Information) und die eine Handlung in Form von Kommissionierung und Versand der Waren auslöst wenn der Liefertermin abzüglich der Versanddauer (vorhandenes Wissen) auf heute fällt (Ereignis). In der Praxis insbesondere in der Informationsverarbeitung wird der Begriff Daten aber weitergefasst. Siehe dazu auch Abbildung 1 in dem der Informationsbegriff aus der Perspektive der Informatik dargestellt ist. Für die weitere Arbeit sind mit betrieblichen

Daten die Daten, also ihre Werte und ebenso ihre Bedeutung durch ein Schema und ihre Bedeutung in einem ganz bestimmten Kontext gemeint. (Pipino, Lee, & Wang, 2002).

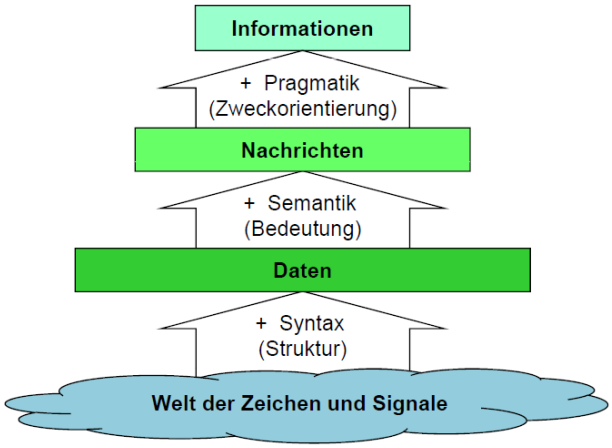


Abbildung 1: Von Daten zur Information (Tiemeyer, 2015, S. 554)

Somit beziehen sich die Begriffe Datenqualität und Datenqualitätsmanagement auf Daten und Information gleichermassen. Einen ersten guten Überblick über betriebliche Daten, ihre logische Struktur und die Datenarten, Datenqualitätsdimensionen sowie deren Steuerung und Management stellt Abbildung 2 dar.

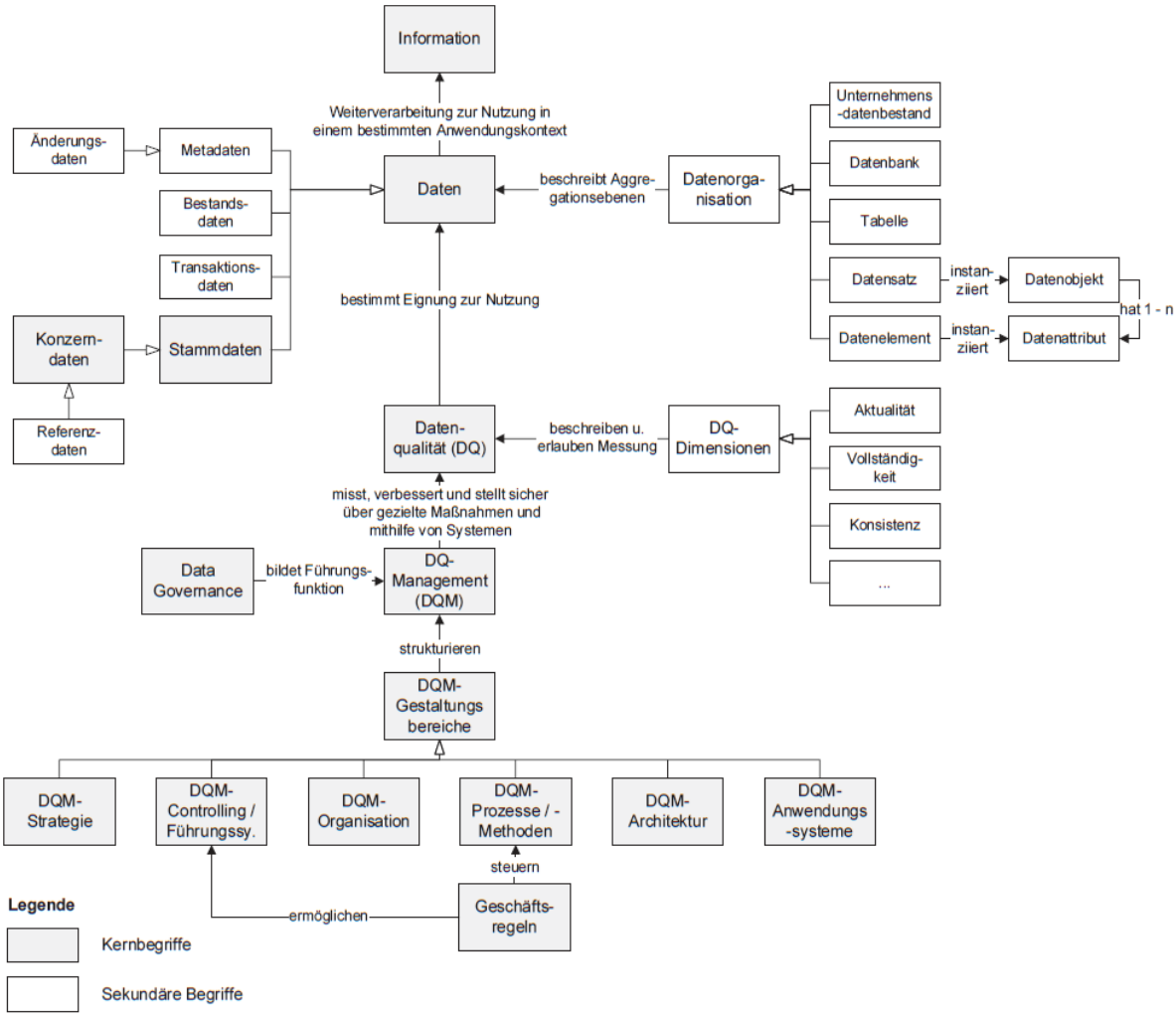


Abbildung 2: Begriffslandkarte für das Datenqualitätsmanagement (Otto & Österle, 2016, S. 28)

Die logische Datenorganisation wird über Datenstrukturen ermöglicht (Levitin & Redman, 1998), bei der die Datenelemente die unterste Menge der Datenattribute bilden und ein Datensatz die Instanzierung eines Datenobjektes ist, das aus ein oder mehreren Datenattributen bestehen kann. Datensätze wie die komplette Beschreibung eines Kunden mittels einzelnen Datenelementen werden in Tabellen zusammengefast. Die Datenorganisation ist vom zugrunde liegenden relationalen Datenmodell abhängig und im idealen, semantisch eindeutigen Fall wird ein reales Geschäftsobjekt in genau einem Datenobjekt abgebildet. Im Regelfall ist das nicht immer so eindeutig und das Datenmanagement ist gefordert, dass Richtlinien, Geschäftsregeln und Prozesse so implementiert werden, dass im jeweiligen Geschäftsprozess stets mit den korrekten Datenobjekten, das den realen Geschäftsobjekten zugeordnet ist gearbeitet wird (Otto & Österle, 2016, S. 29).

2.2. Datenarten

Nach ihren Charaktereigenschaften wie Zeitbezug, Änderungshäufigkeit, Volumenkonstanz und Grad der existenziellen Unabhängigkeit (Wedekind, 2001, S. 72), bzw. dem unterschiedlichen Verwendungszweck, können betriebliche Daten in die Kategorien Stammdaten, Bestandsdaten, Transaktionsdaten (Bewegungsdaten), Änderungsdaten und Metadaten unterschieden werden (Stahlknecht & Hasenkamp, 2005, S. 138).

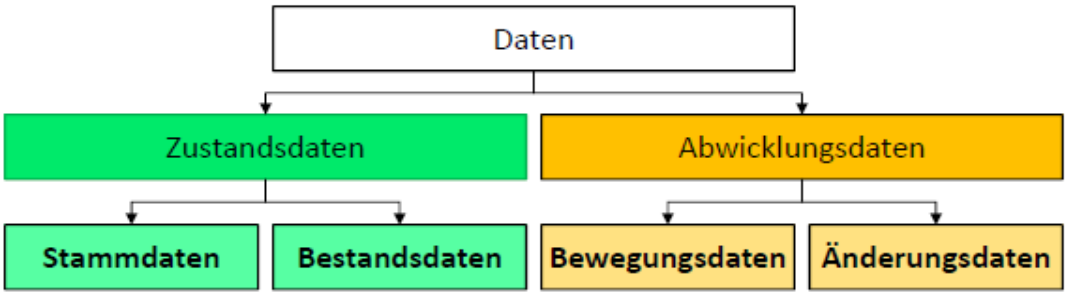


Abbildung 3: Typologie von Daten (vgl. Schemm, 2009, S. 20)

2.2.1. Stammdaten

Als Stammdaten werden jene zustandsorientierten Daten bezeichnet (vgl. Abbildung 2 und 3), die existenziell unabhängig von anderen betrieblichen Datenarten sind und die Grundinformationen über betrieblich relevante Objekte wie Kunden, Lieferanten, Produkte, Personal und Maschinen etc. enthalten. Diese Grundinformationen sind eher statischer Natur, haben also meistens keinen zeitlichen Bezug und sind für die laufende periodische Verarbeitung notwendig. Sie werden über entsprechendes Stammdatenmanagement gepflegt, wobei hier besonders ihre konzeptionelle Architektur (Vollständigkeit) und die

Datenqualität (Fehlerfreiheit) im Fokus stehen, da sie mehrfach und langfristig genutzt werden (Hansen & Neumann, 2009, S. 165). Im Stammdatenmanagement werden nur ausgesuchte Datenattribute eines Stammdatenobjektes unternehmensweit verwaltet, der volle Datenumfang wäre dafür zu komplex. Nach Otto & Österle (2016, S. 30) können für die Bestimmung der Attribute durch eine strategische Analyse im Unternehmen folgende Unterscheidungsmerkmale helfen:

- Organisatorische Reichweite: Unterscheidung zwischen unternehmensweit genutzten Konzerndaten und lokalen Daten
- Datentyp: Differenzierung zwischen „strukturierten“ Daten (Smart Data), meist in relationalen Datenbanken verwaltet und „unstrukturierten“ Daten wie Produktinformationen (z. B. Bilder, Werbetexte, Applikationsvideos)
- Ort der Metadatendefinition: Unterscheidung zwischen firmeninterner Definition von Bedeutung (Semantik), Formaten und Standardwerten und externer Definition (z. B. bei Länder- und Währungscodes der ISO sowie Klassifikationsstandards wie eCl@ss und UN/SPSC), sogenannte Referenzdaten.

Referenzdaten sind eine spezielle Form der Stammdaten. Sie dienen ausschliesslich der Klassifizierung und für den Abgleich von Stammdatenobjekten mittels extern verwalteter Codelisten. Referenzdaten werden bei firmen- bzw. länderübergreifenden Standardisierungsinitiativen durch Einigung definiert. Beispiele sind Codes für Länder, Währungen oder Sprachen der Internationalen Organisation für Normung (ISO) sowie Produktklassifikationen oder Gefahrgutklassen (Schemm, 2009, S. 20).

2.2.2. Bestandsdaten

Im Unterschied zu den Stammdaten beschreiben Bestandsdaten die betrieblichen Mengen und Wertestrukturen. Lagerbestände von Materialien und Kontostände sind Beispiele dafür. Bestandsdaten sind ebenfalls zustandsorientiert, weisen aber eine höhere Änderungshäufigkeit als Stammdaten auf und sie werden meist durch einen Prozess verändert (Hansen & Neumann, 2009, S. 9; Schemm, 2009, S. 21).

2.2.3. Bewegungsdaten

Bewegungsdaten sind Abwicklungsdaten und bilden betriebswirtschaftliche Vorgänge ab. Sie entstehen somit immer wieder neu. Typische Beispiele sind Lieferscheine, Bestellungen und Rechnungen. Sie verändern Bestandsdaten durch mengen- oder wertmässige Zu- und Abgänge (Hansen & Neumann, 2009, S. 9).

2.2.4. Änderungsdaten

Änderungsdaten sind ebenfalls abwicklungsorientierte Daten und lösen Änderungen von Stammdaten aus. Das Korrigieren, Ergänzen und Löschen von Stammdaten in Datenbeständen bezeichnet man als Änderungsdienst (Hansen & Neumann, 2009, S. 9).

Einen Überblick über die Zusammenhänge zwischen den verschiedenen Datenarten gibt Abbildung 4 an.

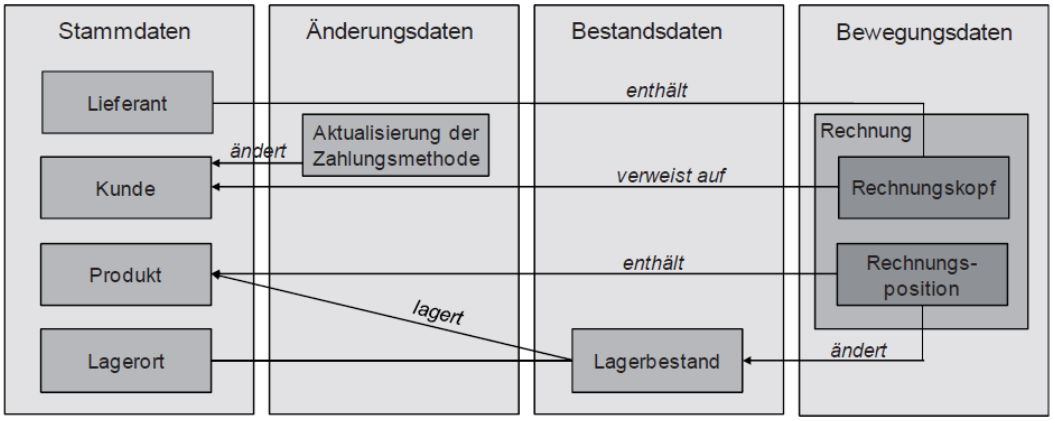


Abbildung 4: Zusammenhang zwischen den Datenarten (Schmidt, 2010, S. 18)

2.3. Metadaten

Für das Stammdaten- und Datenqualitätsmanagement sind die Metadaten sehr entscheidend, werden doch mit ihnen die Eigenschaften, die Bedeutung, die Struktur und die korrekte Verwendung von Daten (Otto & Hühner, 2009, S. 25) festgelegt. In operativen Systemen sind Metadaten meist nicht abgebildet. Fachliche Metadaten sind z. B. in der Informationssystem Benutzerdokumentationen enthalten, die die betriebswirtschaftlichen Zusammenhänge näher beschreiben.

Zu beachten gilt, dass die Einteilung in diese Datenarten nicht starr ist, sondern die Übergänge zwischen Stamm- und Bewegungsdaten fliessend sind.

2.4. Definition Datenqualität

Bei dem Begriff Datenqualität handelt es sich um einen parametrisierten Begriff - Daten und Qualität. Somit soll im Folgenden zuerst auf den Begriff Qualität eingegangen werden. (Apel, Behme, Eberlein, & Merighi, 2010, S. 16).

2.4.1. Begriff Qualität

Das Lateinische Wort „qualitas“ bedeutet „Eigenschaft“ oder „Beschaffenheit“. Ursprünglich war der Begriff weder positiv noch negativ besetzt. Heute wird er aber übelicherweise

automatisch als positiv wahrgenommen. Es gibt eine grosse Anzahl von Definitions- und Interpretationsversuchen. Eine allgemein akzeptierte Begriffsbeschreibung ist die DIN Norm 55 350: Danach ist die „Qualität die Gesamtheit von Eigenschaften und Merkmalen eines Produktes oder einer Tätigkeit, die sich auf deren Eignung zur Erfüllung festgelegter oder vorausgesetzter Erfordernisse beziehen.“ (Apel et al., 2010, S. 17).

Demgegenüber steht die aktuelle Norm nach ISO 9000:2015, in der die Qualität, als „Grad, in dem ein Satz inhärenter Merkmale Anforderungen erfüllt“ definiert wird.

Inhärent bedeutet, der Einheit innewohnend, ein ständiges Merkmal.

Diese Definition indiziert ein anderes Qualitätsverständnis, als dem Mass für die bestehenden Anforderungen an ein Produkt (z. B. Daten).

Garvin hat einen der ersten Systematisierungsansätze entworfen, der fünf generelle Qualitätsvorstellungen unterscheidet (Garvin, 1984, S. 40; Apel et al., 2010, S. 17):

- produktorientierter Ansatz
- anwenderorientierter Ansatz
- prozessorientierter Ansatz
- wertbezogener Ansatz
- transzendenter Ansatz

Die **produktbezogene** Sicht entspricht einem objektiven Qualitätsbegriff, weil Qualität als eine messbare, genau spezifizierbare Grösse, die das Produkt beschreibt, gesehen wird. Sie ist unabhängig von der subjektiven Wahrnehmung und vom Kunden und wird nur über die Produkteigenschaften bestimmt.

Der **kunden- oder anwenderbezogene** Ansatz hingegen definiert die Qualität eines Produkts über die Wahrnehmung durch den Produktnutzer. Die subjektive Beurteilung des Kunden entscheidet, inwieweit das Produkt der geforderten Qualität entspricht. Diese Sichtweise wird als „fitness for purpose“ oder „fit for use“ genannt. Zu beachten gilt, verschiedene Endbenutzer können unterschiedliche Bedürfnisse haben, sodass die Qualität desselben Produkts unterschiedlich bewertet werden kann.

Der **prozessorientierter** Ansatz nimmt an, dass Qualität dann entsteht, wenn der Herstellungsprozess optimal und kontrolliert verläuft und alle Vorgaben und Normen (Produktspezifikationen) eingehalten werden. Abweichungen von dem definierten Prozess werden als Qualitätsverlust angesehen.

Der **wertbezogene** Ansatz versucht einen Bezug zwischen Preis und Qualität im Sinne von Nutzen herzustellen. Ein Produkt ist dann von hoher Qualität, wenn die Kosten und die empfangene Leistung in einem positiven Verhältnis stehen. Somit ist die eigentliche Qualität des Produktes nur indirekt angegeben.

Der **transzendente** Ansatz entspricht in etwa der umgangssprachlichen Sicht auf Qualität als vorgegebene Vortrefflichkeit, Einzigartigkeit oder Superlativ. Qualität wird als Synonym für hohe Standards und Ansprüche angesehen. Nach dieser Definition kann Qualität nur schwer quantifiziert werden, weil er davon ausgeht, dass Qualität nicht messbar, sondern nur erfahrbar ist (Apel et al., 2010, S. 18).

2.4.2. Datenqualität

Die zuvor genannten Definitionen zeigen, dass Qualität je nach Betrachtungsart unterschiedlich wahrgenommen werden kann und somit die Beurteilung der Qualität eines materiellen oder immateriellen (digitalen) Produktes (Daten) unterschiedlich ausfallen kann. Aus diesen allgemeinen Definitionen von Daten und Qualität und unter Berücksichtigung der von Garvin (1984) gefundenen Qualitätsvorstellungen wird der Begriff Datenqualität näher beschrieben.

Von Helfert (2002) wurden die in der Literatur vorhandenen Ansätze (Wang & Strong, 1996, S. 14) zur Definition von Datenqualität untersucht und einander gegenübergestellt.

Das Untersuchungsergebnis zeigt, dass der Benutzer das Qualitätsniveau festlegt und damit im Sinne der Datenverarbeitung ausschliesslich der anwenderorientierte Ansatz sinnvoll ist (Helfert, 2002, S. 69). Es macht Sinn, Qualität von Daten nur im Kontext ihrer Nutzung zu beurteilen und dadurch kann auch der Begriff der Informationsqualität synonym verwendet werden (Hildebrand & Otto, 2011, S. 87).

Daraus definiert Würthele (2003, S. 21) Datenqualität als

„mehrdimensionales Maß für die Eignung von Daten, den an ihre Erfassung/Generierung gebundenen Zweck zu erfüllen. Diese Eignung kann sich über die Zeit ändern, wenn sich die Bedürfnisse ändern."

Damit wird deutlich, dass die Qualität von Daten vom Zeitpunkt der Betrachtung sowie von dem zu diesem Zeitpunkt an die Daten gestellten Anspruchsniveau abhängt. Dieser weiche Faktor Anspruchsniveau ist vom Können und Willen der an den Prozessen beteiligten Nutzer abhängig, also immer Benutzerkontext bezogen, was später folgend im Datenqualitätsmanagement mit der Rolle der Datenverantwortlichen (Data Owner) näher erläutert wird.

Kontextabhängig bedeutet, dass die Datenqualität für einen Geschäftsvorfall ausreichend sein kann, für einen anderen hingegen nicht. Deshalb ist Datenqualität insgesamt definiert als ein Mass für die Eignung der Daten für bestimmte Anforderungen in den Geschäftsprozessen, in denen sie verwendet werden (Hildebrand & Otto, 2011, S. 21).

Auch ändert sich die Datenqualität über die Zeit, weil die Daten lediglich eine Momentaufnahme der Wirklichkeit darstellen, sich diese Wirklichkeit aber verändert. Zum Beispiel ändern Kunden ihre Telefonnummern oder die Ansprechpartner beim Kunden ändern sich.

Viele Unternehmen lösen ihre Datenqualitätsprobleme oft nur ad hoc, wenn die Auswirkungen von schlechter Datenqualität in mehreren Geschäftsprozessen sichtbar werden. Migrationsprobleme, hoher manueller Korrekturaufwand in Geschäftsprozessen, Management Reports mit unterschiedlichen Werten zur gleichen Kennzahl sind Beispiele für die Datenqualitätsprobleme, die plötzlich sichtbar werden, nachdem die Datenqualität zuvor schleichend abgefallen ist (Otto & Hühner, 2009, S. 32).

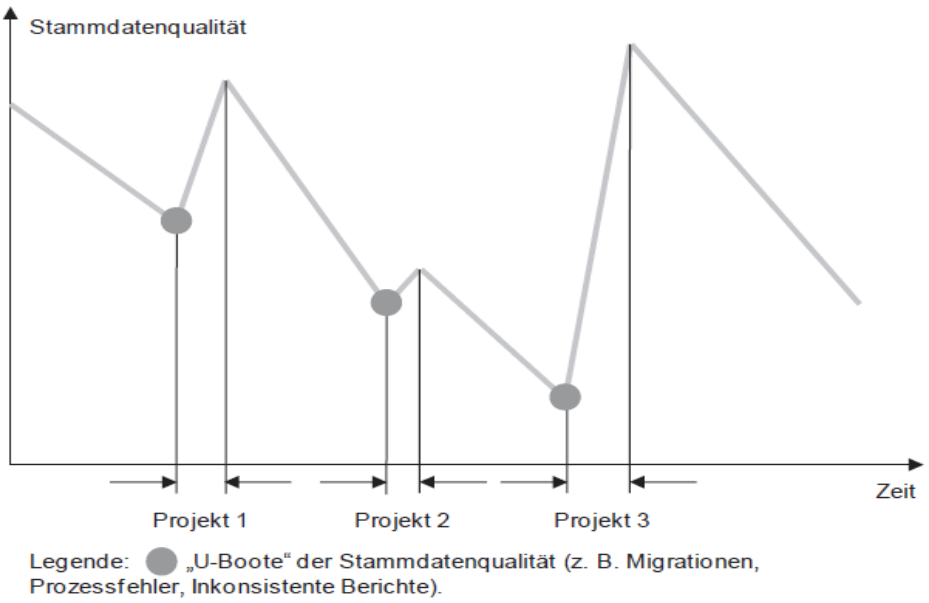


Abbildung 5: typischer zeitlicher Datenqualitätsverlauf in Unternehmen (Otto & Österle, 2016, S. 32)

Damit die Zusammenhänge mit dem Messen der Datenqualität in weiterer Folge in dieser Arbeit begreiflicher wird, wird nun in Anlehnung an die Definition von Würthele (2003, S. 21) sowie den Definitionen von Daten, Information und Qualität der Begriff Datenqualität folgendermassen definiert:

Die Datenqualität ist der Grad der Übereinstimmung der Ansprüche eines Nutzers und den Eigenschaften der Daten. Der Anspruch kann sich über die Zeit ändern und kann aufgrund unterschiedlicher Verwendungszwecke eine andere Bedeutung erhalten. Zudem sollen Daten subjektiv als qualitativ hochwertig wahrgenommen werden.

Tabelle 1 zeigt zusammengefasst die Übertragung der Systematisierungsansätze von Garvin (1984) auf die Datenqualität (Schumacher, 2011, S.166).

Systematisierung von Qualität	Produktqualität	Datenqualität
Produktorientierter Ansatz	Über spezifische Merkmale messbar	Merkmals als eine Datenprodukteigenschaft
Anwenderorientierter Ansatz	Anwender kann sie beurteilen	Verwendungszweck des Anwenders erfüllt
Prozessorientierter Ansatz	Einhaltung der Spezifikation und Fertigungsnormen	harmonisierte und durchgängige Daten in den Prozessen
Wertbezogener Ansatz	Produkt Preis-/Leistungsverhältnis	Kosten-/ Nutzen Betrachtung der Datenqualität
Transzendenter Ansatz	nur subjektiv erfahrbar	nicht geeignet – da nicht exakt fassbar

Tabelle 1: Übertragbarkeit der Ansätze von Garvin auf die Datenqualität (Schumacher, 2011, S.166)

Es gibt kein einzelnes Merkmal, was die Datenqualität vollständig beschreibt, sondern vielmehr die Summe der einzelnen objektiven Datenqualitätsmerkmale bilden die Datenqualität ab. Datenqualitätsmerkmale können erst über zu definierende Datenmetriken gemessen werden, damit kann die Datenqualität operationalisiert werden. Sie bilden das Qualitätsmass, mit dem eine quantitative Aussage über den Erfüllungsgrad der Datenqualität für den Datenbenutzer möglich wird. Es müssen die Anspruchsniveaus jeweils abhängig vom Reifegrad des Datenmanagements individuell festgelegt werden, dass bedeutet, dass sich je nach erreichter Qualitätsstufe die Sollwerte wieder ändern können. Diese Qualitätsmetriken bilden die operationelle Basis zur Bestimmung der Datenqualität. Abbildung 6 zeigt die Operationalisierung durch die Datenqualitätspyramide (Gebauer & Windheuser, 2008, S. 89).

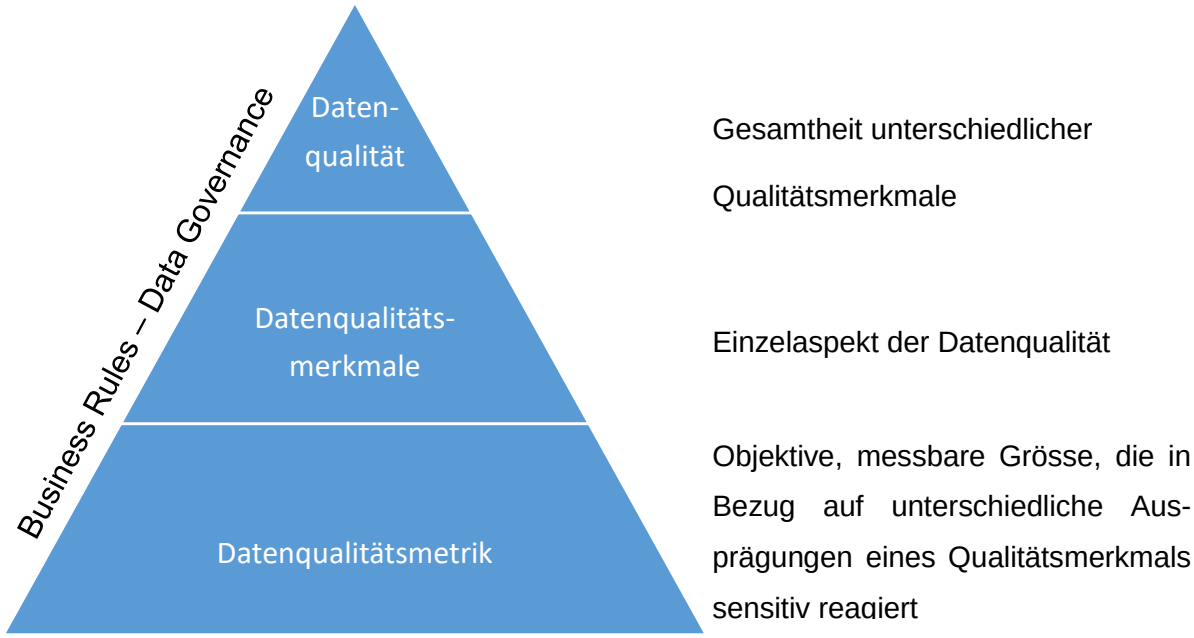


Abbildung 6: Datenqualitätspyramide (eigene Darstellung nach Gebauer & Windheuser, 2008, S. 89)

2.4.3. Datenqualitätsdimensionen - Datenqualitätskriterien

Die Qualität eines Datenobjektes kann durch die Überprüfung bestimmter Eigenschaften der Werte der enthaltenen Datenelemente gemessen werden (Loshin, 2001, S. 280-300). Weisen die Datenelemente eines Datenobjektes alle geforderten Eigenschaften auf, ist seine Qualität einwandfrei. Eigenschaften von Daten (z. B. letzter Änderungszeitpunkt eines Datenelements, Vollständigkeit eines Datenelements) können z. B. als Geschäftsregeln formuliert und dadurch überprüfbar gemacht werden. Datenqualitätsmerkmale werden häufig in übergeordneten Kategorien gruppiert. Es entsteht ein Datenqualitätsmodell. Eine wesentliche Eigenschaft eines solchen Modells ist seine Zerlegungssystematik (Apel et al., 2010, S. 22).

In der Literatur gibt es dazu viele verschiedene Ansätze, um in Kombination der Datenqualitätsmerkmale eine Aussage über die Datenqualität zu machen. Als Ansatz kann ein intuitiver, empirischer, dienstleistungsorientierter, semiotischer oder ein systemorientierter Ansatz gewählt werden. In der Literatur werden allgemein vier Fehlertypen identifiziert, aus denen innere Datenqualitäts-Merkmale abgeleitet werden. Zu diesen typischen Qualitätskriterien, zählen die Korrektheit, Vollständigkeit, Konsistenz und Aktualität (Apel et al., 2010, S. 20).

- Korrektheit: Stimmen die Daten der betrieblichen Datenobjekte mit den Eigenschaften des realen Objektes überein.
- Vollständigkeit: Sind alle Datenattribute eines Datensatzes wie definiert komplett vorhanden?
- Konsistenz: Falls es mehrere Datenobjekt Versionen desselben realen Objektes im System vorhanden sind, sind sie kongruent?
- Aktualität: Stimmen die Daten zu jedem Zeitpunkt mit dem aktuellen Status des realen Objektes überein und werden die Daten angepasst, wenn es sich ändert?

Ein systemorientierter Ansatz ist der von Larry Englisch (Englisch 1999) der drei Hauptmerkmale mit ihren insgesamt 10 Untermerkmalen unterscheidet:

- Qualität der Datendefinition
  - Datenspezifikation
  - Geschäftsregeln
  - Integritätsbedingungen
- Inhaltliche Datenqualität (Korrektheit der Datenwerte)
  - Vollständigkeit
  - Eindeutigkeit
  - Einhaltung der Geschäftsregeln
  - Genauigkeit und Fehlerfreiheit

- Qualität der Datenpräsentation
  - Rechtzeitige Bereitstellung
  - Angemessenheit des Formats
  - Verständlichkeit des Formats

Dieser Ansatz ist recht pragmatisch, weil er sich an den Erfordernissen eines zu implementierenden Datenmodells orientiert. Grundsätzlich scheint er sich für die vorliegende Arbeit gut zu eignen, da im später erklärten Data Profiling und der Daten Analyse besonders auf die Datendefinitionen und inhaltliche Datenqualität wie Einhaltung der Geschäftsregeln zurückgegriffen wird (Heinrich & Klier, 2008).

Der am weitesten untersuchte und verbreitete Ansatz ist aber der von Wang & Strong (1996) bzw. Otto & Hühner (2009), der mehr den Datennutzer und damit das Thema Informationsqualität in den Vordergrund stellt.

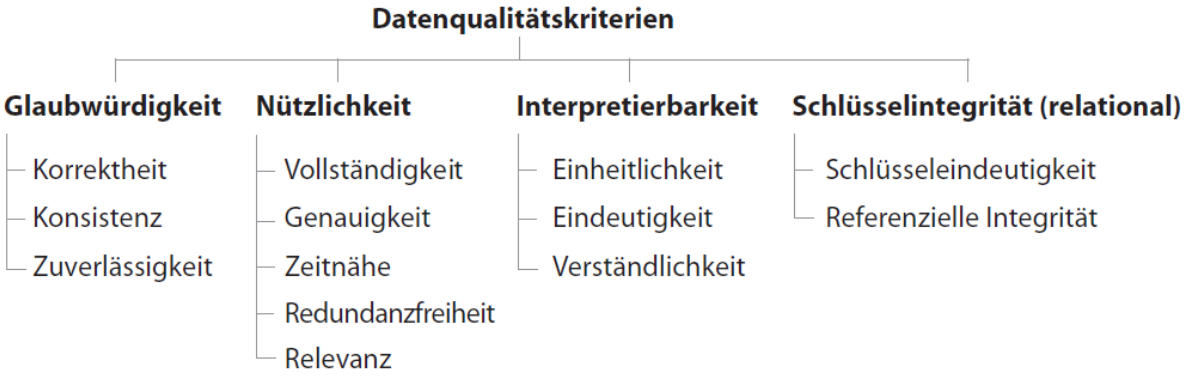


Abbildung 7: Taxonomie von Datenqualitätskriterien (Apel et al., 2010, S. 22)

Diesem Ansatz stellt die Deutsche Gesellschaft für Informations- und Datenqualität (DGIQ 18) eine Kategorisierung gegenüber, die aus einer Befragung von IT Anwendern hervorgegangen ist (Wang & Strong, 1996).

2.4.4. DGIQ Datenqualitätsmodell

Mit den 15 Dimensionen die DGIQ vorschlägt können alle denkbaren Qualitätsaspekte von Daten aus Anwendersicht beschrieben werden. Der Datennutzer kann die Qualitätsaspekte aus seiner Sicht aber erst priorisieren, nachdem er alle 15 Dimensionen beachtet hat. Die Dimensionen werden aufgrund inhaltlicher Zusammenhänge in Bezug auf den Untersuchungsgegenstand in Ordnungskategorien unterteilt. Z. B. können die systemunterstützten Dimensionen Zugänglichkeit und Bearbeitbarkeit nur untersucht werden, indem man sich das datenverarbeitende System und seine Benutzeroberfläche anschaut.

Die 15 Datenqualitäts-Dimensionen im Überblick (Apel et al., 2010, S. 33):

Kategorie System:

1. Zugänglichkeit (accessibility): Daten sind dann zugänglich, wenn sie auf einfachen Weg für den Nutzer abrufbar sind.
2. Bearbeitbarkeit (ease of manipulation): Daten sind leicht bearbeitbar, wenn sie leicht zu ändern und für unterschiedliche Zwecke zu verwenden sind.

Kategorie Nutzen/Zweckabhängig:

3. Angemessener Umfang (appropriate amount of data): Daten sind von angemessenem Umfang, wenn die Abbildung realer Objekte den gestellten Anforderungen genügt.
4. Vollständigkeit (completeness): Daten sind vollständig, wenn sie nicht fehlen und zu den festgelegten Zeitpunkten in den jeweiligen Prozess-Schritten zur Verfügung stehen
5. Relevanz (relevancy): Information ist dann relevant, wenn sie für den Nutzer notwendige Information liefert.
6. Aktualität (timeliness): Daten sind aktuell, wenn sie die tatsächliche Eigenschaften des beschriebenen Objektes zeitnah abbilden.
7. Wertschöpfung (value-added): Daten sind wertschöpfend, wenn ihre Nutzung zu einer quantifizierbaren Steigerung einer monetären Zielfunktion führen kann.

Kategorie inhärent/Inhalt:

8. Glaubwürdigkeit (believability): Daten sind glaubwürdig, wenn Daten als wahr und zuverlässig angesehen werden. Allgemeine Indikatoren für hohe Glaubwürdigkeit sind Zertifikate, die einen hohen Qualitätsstandard ausweisen sowie der Aufwand, der für die Informationsgewinnung, -verbreitung und -überprüfung betrieben werden.
9. Fehlerfreiheit (free of error): Daten sind fehlerfrei, wenn sie mit der Realität übereinstimmen.
10. Objektivität (objectivity): Information ist objektiv, wenn sie streng sachlich und wertfrei ist.
11. Hohes Ansehen (reputation): Information ist hoch angesehen, wenn die Informationsquelle, das Transportmedium und das verarbeitende System im Ruf einer hohen Vertrauenswürdigkeit und Kompetenz steht.

Kategorie Darstellung:

12. Übersichtlichkeit (concise representation, clearness): Daten sind übersichtlich, wenn genau die benötigten Daten in einem passenden und leicht fassbaren Format dargestellt sind.
13. Einheitliche Darstellung (consistent representation): Daten sind einheitlich dargestellt, wenn sie fortlaufend auf dieselbe Art und Weise abgebildet werden.

14. Eindeutige Auslegbarkeit (interpretability): Information ist eindeutig auslegbar, wenn sie in gleicher, fachlich korrekter Art und Weise begriffen wird.
15. Verständlichkeit (understandability): Information ist verständlich, wenn sie unmittelbar von den Nutzern verstanden und für deren Zwecke eingesetzt werden kann.

Ein Informationssystem kann erst dann ausreichend funktionieren, wenn alle diese 15 kritischen Datenqualitätsdimensionen eine hohe oder zumindest ausreichende Qualität aufweisen. In der Abbildung 8 wird dies durch die Positionierung der Informationsqualität (IQ) im zentralen Kreis dargestellt: Daten müssen den vom Nutzer verlangten Zweck erfüllen („fit for use“-Konzept), erst dann sind sie „rund“.

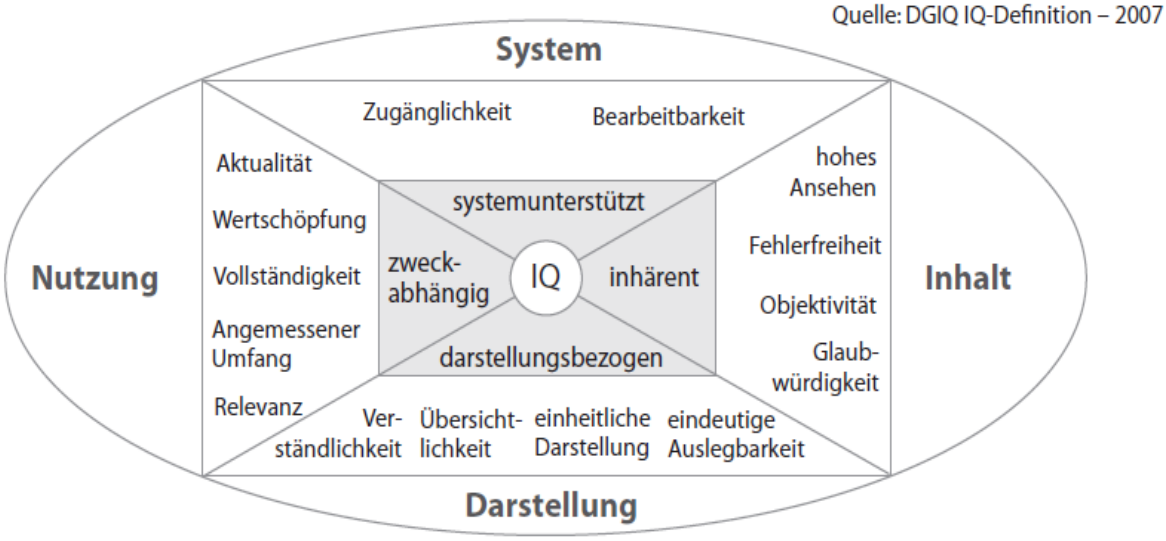


Abbildung 8: Taxonomie von Datenqualitätskriterien (Apel et al., 2010, S. 23)

Die vier Untersuchungsgegenstände Inhalt, System, Nutzung und Darstellung können gemeinsam als die organische Struktur der Informationsverarbeitung eines Unternehmens bezeichnet werden. Kern eines jeden Informationssystems sind die Inhalte, die verarbeitet werden. Die Inhalte werden in Systemen verarbeitet und innerhalb des Unternehmens weitergeleitet. Die Inhalte werden mithilfe der Systemfunktionalitäten für die Nutzer in eine Darstellung überführt. Alle Prozesse der Informationsverarbeitung und Bereitstellung laufen im Kontext der Unternehmensprozesse unter Berücksichtigung der Nutzung ab. Die Unternehmensprozesse umfassen damit Systeme, Inhalte und Darstellung (Apel et al., 2010, S. 23).

3. Treiber, Ursachen und Wirkungen von Datenqualität

Wie zuvor schon definiert, hängt die Datenqualität von der Anwendung und den subjektiven Anforderungen der Benutzer ab. Die Ursachen schlechter Datenqualität sind oft vielfältig und lassen sich auch unter grössten Anstrengungen nicht ganz vermeiden. Die Benutzer müssen auf die Ursachen und Ausprägungen von schlechter Datenqualität und deren Umgang damit sensibilisiert werden (Apel et al., 2010, S. 29).

3.1. Treiber der Datenqualität

Gerade die schlechte Datenqualität von Stammdaten kann auf viele unterschiedliche Geschäftsprozesse Einfluss haben, die die Benutzer so nie vermuten würden. Eine Verbesserung der Datenqualität ist deshalb im Sinne Aller. Verbesserungen sind oftmals schwierig und aufwendig durchzuführen und aus diesem Grund sollten auf jeden Fall Kosten-/Nutzen Überlegungen angestellt werden.

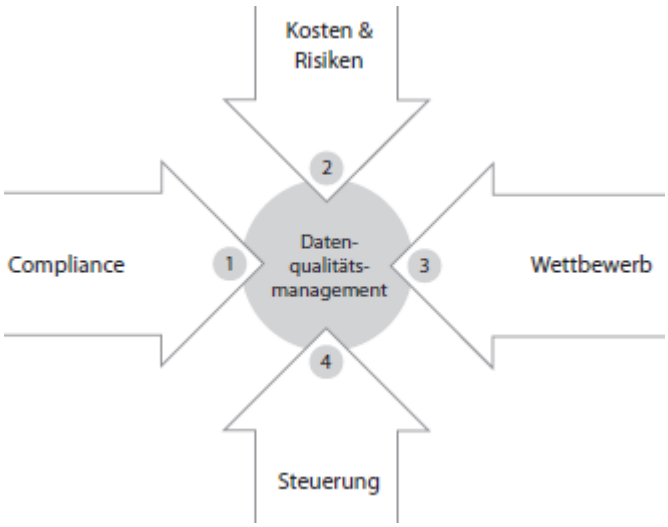


Abbildung 9: Geschäftstreiber von Datenqualität (Apel et al., 2010, S. 30)

Abbildung 9 zeigt die vier wesentlichen Geschäftstreiber, welche die Datenqualität besonders einfordern. Als Konsequenz aus diesen Abhängigkeiten entsteht in weiterer Folge die Forderung nach einem nachhaltigen Datenqualitätsmanagement.

Compliance

Aufgrund gesetzlicher und behördlicher Regelungen und firmeninternen Richtlinien wie dem Data Governance legen die interne und externe Revision immer mehr regulatorische Vorgaben fest, die das Unternehmen erfüllen muss. Diese verlangen die Definition und Umsetzung firmenweiter einheitlicher Standards und Mechanismen. Im Data Governance als übergeordnete Steuerung des Datenmanagements und des Datenqualitätsmanagements werden z. B. festgelegt wer im Unternehmen, für welche Daten welche Rechte zur Definition

und zur Nutzung hat und wer die Verantwortung für welche Daten trägt (Data Owner) (Apel et al., 2010, S. 31).

Kosten und Risiken

Durch das frühzeitige Erkennen von Datenqualitätsproblemen kann das operative Risiko im Unternehmen reduziert werden. Besonders der operative Geschäftsbereich ist auf Grund von immer schnelleren Prozessen auf qualitativ hochwertige Daten angewiesen. Das aktive Beheben von schlechter Datenqualität mindert generell die Risiken von Schnittstellen- und Datenmigrations-Projekten, da immer weniger Zeit für projektinterne Datenqualitätsmassnahmen bleibt (Apel et al., 2010, S. 31).

Wettbewerb

Für die langfristige Verbesserung der Kundenbeziehung, dem Erfolg von Kampagnen und der Entwicklung neuer Produkte ist eine gute Datenqualität Grundvoraussetzung (Apel et al., 2010, S. 31).

Steuerung

Die Datenqualität hat unmittelbare Auswirkungen auf Entscheidungen für die Unternehmenssteuerung. Sind die Daten falsch oder unvollständig, beeinträchtigt das die getroffenen Entscheidungen erheblich. Die Akzeptanz von Informationssystemen durch die Anwender ist in hohem Ausmass von der Datenqualität abhängig (Apel et al., 2010, S. 31).

3.2. Herausforderungen und Ausprägungen schlechter Datenqualität

Diese dargestellten Geschäftstreiber erfordern jeweils eine bestimmte Qualität der Daten. Die Ausprägungen von schlechter Datenqualität sind grundsätzlich sehr vielfältig und werden subjektiv vom Nutzer wahrgenommen. Sie lassen sich anhand der bereits genannten Datenqualitätskriterien festmachen. In der Studie „Stammdatenmanagement in der produzierenden Industrie“ (Schuh & Stich, 2014) wurde folgendes Ergebnis der Häufigkeit von Datenqualitätsproblemen festgestellt. Daraus ist ersichtlich, dass die grössten Probleme in der **Vollständigkeit**, **Aktualität** und **Strukturiertheit** von Daten liegen. Am wenigsten Probleme gab es mit der Verfügbarkeit und Zugänglichkeit der Daten.

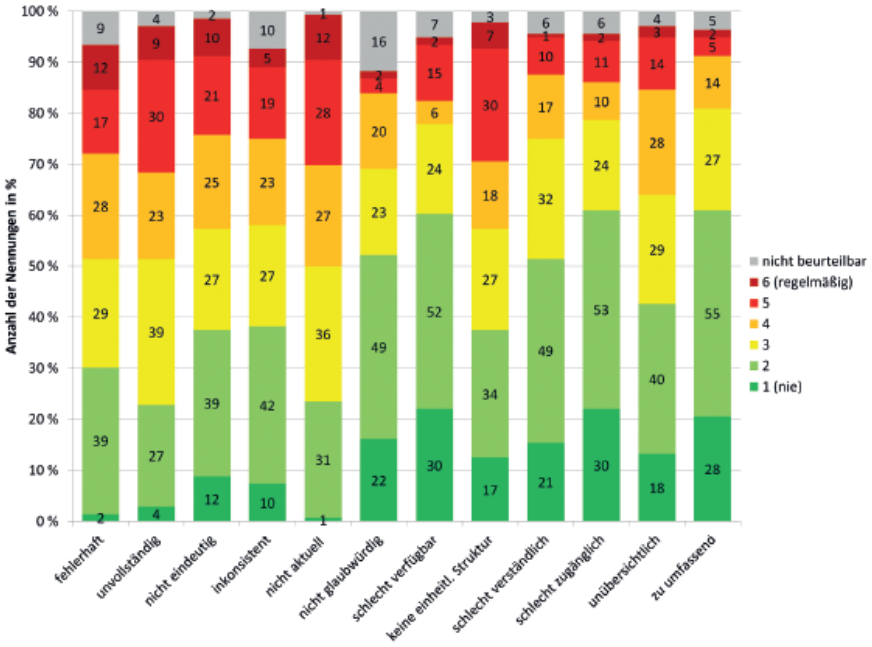


Abbildung 10: Daten Qualitätsprobleme bei Stammdaten (Schuh & Stich, 2014, S. 19)

3.3. Ursachen schlechter Datenqualität

Ursachen für schlechte Datenqualität können wie die Ausprägungen unterschiedlich sein. Nur wenn die Ursache bekannt ist kann die Datenqualität dauerhaft verbessert werden.

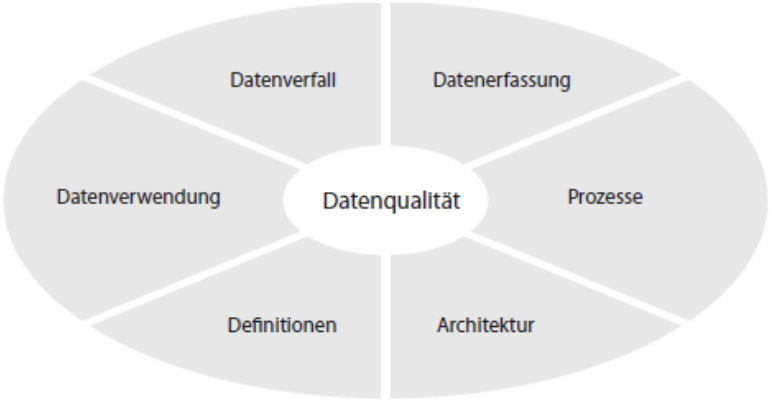


Abbildung 11 Ursachen von schlechter Datenqualität (Apel et al., 2010, S. 34)

Abbildung 11 stellt die grundlegenden Datenqualitätskategorien der Ursachen für schlechte Datenqualität dar.

Datenerfassung

Die Datenerfassung ist eine der wesentlichsten Fehlerquellen. Es beginnt bei einfachen Benutzerfehlern wie Tippfehler, fehlendes Knowhow für die korrekte Bedienung oder Verständnisfehler der zu erfassenden Daten oder fehlende Information und geht bis zu fehlenden Ressourcen für eine korrekte Datenerfassung.

Eine der grundlegenden Ursachen ist das fehlende Problembewusstsein für die Auswirkungen mangelhafter Datenqualität sowie die fehlende Motivation für eine korrekte

Erfassung von Daten. Auch die Eingabemasken sind mögliche Ursachen, weil die Möglichkeiten zur Validierung bzw. Einschränkung der Datenwerte und Mussfelder nicht genügend genutzt wird (Apel et al., 2010, S. 35).

In der Studie „Stammdatenmanagement in der produzierenden Industrie“ (Schuh & Stich, 2014) wird das bestätigt und menschliche Fehler gefolgt von mangelndes Verständnis, als die häufigste Ursache von Datenqualitätsproblemen festgestellt.

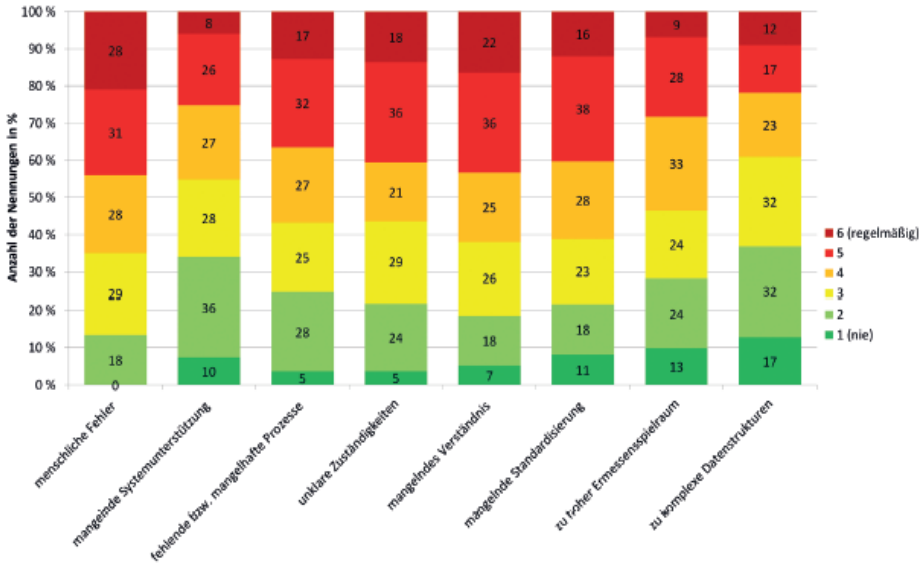


Abbildung 12: Ursachen für schlechte Stammdatenqualität (Schuh & Stich, 2014, S. 21)

Prozesse

Ist die Komplexität der Geschäftsprozesse von der Datenerfassung bis hin zu den Auswertungen hoch bzw. sind sie generell mangelhaft hat das Auswirkungen auf die Datenqualität. Fehlende Prüf- und Validierungsprozesse verstärken das zusätzlich. In einem Prozess sind oft mehrere Unternehmensbereiche, welche sich miteinander abstimmen müssen, involviert. Aufgrund fehlender Abstimmung werden potenzielle Fehler erst spät erkannt (Apel et al., 2015, S. 30).

Architektur

Das Design der Datenarchitektur, der verwendeten Technologien und die Zugänglichkeit durch den Anwender hat ebenso Einfluss auf die Datenqualität. Durch Modifikationen und Erweiterungen bestehender Applikationen über den geplanten Einsatzbereich hinaus und weitere Systeme kann es zu redundanter Datenhaltung kommen.

Durch Schnittstellen und Migrationen zwischen den Applikationen müssen Daten wiederholt transformiert werden. Die Nachvollziehbarkeit der Transformationen in den Datenflüssen ist eine wesentliche Voraussetzung für die nachhaltige Sicherstellung der Datenqualität (Apel et al., 2015, S. 31).

Fachliche Definitionen

Inkonsistente fachliche Beschreibungen sind weitere Ursachen für schlechte Datenqualität. Übergreifende identische Felder oder Datenobjekte sind unterschiedlich definiert oder verwendet. Durch die immer grösseren Anwendungslandschaften und die dadurch bedingte erhöhte Abstimmung während der Konzeption verstärken sich die Auswirkungen von unterschiedlichen Definitionen (Apel et al., 2015, S. 32).

Datenverwendung

Empfänger von Daten leiden bei inkorrekter Nutzung von Anwendungen durch z. B. falsche Datenkorrekturen in verschiedenen Applikationen, bzw. wenn sie in operativen Systemen nicht nachgeführt werden. Solche Korrekturen in nachgelagerten Systemen sind immer problematisch sogar bei entsprechender Dokumentation und Abstimmung mit der Revision z. B. bei Bilanzauswertungen (Apel et al., 2015, S. 32).

Datenverfall

Daten haben oftmals nur eine bestimmte Gültigkeit. Besonders bei Stammdaten sind solche Änderungen oft bedenklich und es muss vordefinierte Regeln geben wie damit umzugehen ist. Z. B. mag das inaktiv Setzen eines Kunden unproblematisch sein, aber wenn sich z. B. die Ansprechpartner ändern kann das weitreichende Folgen haben. Abhilfe kann eine Historisierung bei Stammdatenänderungen schaffen (Apel et al., 2015, S. 32).

Fazit

Das Fehlen der Definition von Verantwortung und Regeln für die Behebung von Datenproblemen sowie inkonsistente Datenarchitekturen sind die wesentlichsten Ursachen für schlechte Datenqualität. Oftmals sind Datenqualitätsprobleme bekannt, aber es werden keine Massnahmen zu deren Beseitigung ergriffen. In vielen Fällen ist das auf ein fehlendes Bewusstsein für die Auswirkungen schlechter Datenqualität bei den Entscheidungsträgern zurückzuführen. Das kann auch daran liegen, dass es nicht immer einfach ist die Wirtschaftlichkeit von Datenqualitätssicherungsmassnahmen transparent aufzuzeigen.

3.4. Auswirkungen schlechter Datenqualität

„Elwis rockt bei Lidl nicht mehr“: Mit dem System „Elvis“ wollte der weltweit tätige Lebensmitteldiscounter Lidl seine Stammdaten zentralisiert verwalten und an allen Standorten mehr oder weniger einheitliche Prozesse ausrollen. Ob die unterschiedlichen Ausprägungen der Stammdaten, die unterschiedlichen und komplexen Daten-Schnittstellen

in den verschiedenen Ländern und Standorten, oder ob die Entscheidung zum Projektstart im Jahr 2011 schlicht auf Daten mit nicht ausreichender Datenqualität beruhte, ist unklar (itb, 2018).

Bei Schlagzeilen wie diesen rückt die Datenqualität und Datenmanagement wieder in den Fokus der Unternehmen und es werden hektische Überlegungen angestellt, ob die eigene Datenqualität ausreichend ist.

Eine Studie von 2015 belegt, dass Unternehmen mit korrekten Daten ein um 35 % höheres Wachstum aufweisen (eoda, 2018) als Unternehmen mit Daten schlechterer Datenqualität. Laut einer Studie von The Data Warehousing Institute (2015) (TDWI) betragen in den USA die Kosten wegen schlechter Datenqualität ca. 600 Milliarden US Dollar pro Jahr.

Schlechte Datenqualität verursacht für ein Unternehmen Kosten und macht sich auf verschiedenen Ebenen bemerkbar:

- Schlechte Rücklaufquoten für Werbeaktionen wegen nicht korrekter oder veralteter Adressen.
- Schlechtes Image wegen mehrfach erhaltener Werbepost, aufgrund Adressdaten Dubletten, falschen Anreden usw.
- Aufwände für die Pflege von Datensätzen bei redundanter Datenhaltung ohne automatischer Synchronisierung.
- Treffen von falschen Entscheidungen aufgrund von falschen Daten.
- Inkonsistenzen zwischen System und realem Lagerbestand.
- Vermehrte Rückfragen von Nutzern nachgelagerter Prozesse
- Höhere Prozess Durchlaufzeiten

Fazit

Trotz dieser Argumente ist es für Unternehmen oft schwierig, die Wirtschaftlichkeit von Massnahmen zur Verbesserung der Datenqualität nachzuweisen. In der Literatur und im Internet (Data Cleaning Software Anbieter) werden viele beeindruckende Zahlen von Beispielen angeführt. Dennoch bleibt es eine Herausforderung, diese Beispiele auf das eigene Unternehmen zu übertragen und damit die Skepsis beim Management für die Durchführung prozessgestützter Datenqualitätsmassnahmen auszuräumen. Das Aufzeigen von Einsparungen durch die Steigerung der Datenqualität ist nicht zuletzt deshalb auch schwierig, da Datenqualitätsmassnahmen in den Fachabteilungen nicht unbedingt als sehr attraktiv angesehen sind. Wer mag schon gerne seinen Adressstamm bereinigen, weil z. B. ein neues System eingeführt wird, oder eine nachgelagerte Abteilung zusätzliche oder andere Daten benötigt?

4. Datenqualitätsmanagement

Kapitel 3 hat die Ursachen und Folgen von schlechter Datenqualität aufgezeigt. Wie in den vorigen Kapiteln gezeigt, werden Unternehmen zunehmend von externen Faktoren wie gesetzliche Auflagen etc. gezwungen ihre Daten akkurat zu halten (Apel et al., 2010, S. 18). Nachträgliche Datenbereinigung ist im Vergleich zu qualitätssichernden Massnahmen um den Faktor 5-10 teurer (Hankins, 1999).

Für die langfristige Sicherstellung von guter Datenqualität bedarf es einer nachhaltigen Verankerung von Datenqualitätssicherungs-Prozessen im Unternehmen.

4.1. Datenqualitätsmanagementansätze

Datenqualitätsmanagement (DQM) als Unternehmensfunktion ist der proaktive Ansatz um Datenqualität nachhaltig zu verbessern. Dazu sind Tätigkeiten in strategischer, organisatorischer, technischer, aber auch kultureller Hinsicht erforderlich (Hildebrand, Gebauer, Hinrichs, & Mielke, 2018, S. 276).

Datenqualitätsmanagementmodelle entwickelten sich aus dem Qualitätsmanagement und dem Datenmanagement (Hildebrand et al., 2018, S. 69).

Die etablierten Qualitätssicherungsmethoden aus der Fertigungsindustrie sind auch auf Daten und Information anzuwenden (Hildebrand et al., 2018, S. 68). Der Ansatz dazu verfolgt die Total Data Quality (TDQM) Methodik. Im TDQM erfolgt eine ganzheitliche Betrachtungsweise. Datenqualität ist bei TDQM ganzheitlich im Sinne von über alle Aktivitäten des gesamten Unternehmens zu sehen. Wang (1996) analysierte und adaptierte in den 90er Jahren am Massachusetts Institute of Technology (MIT) Qualitätsmanagement-Konzepte und entwickelte TDQM (Apel et al., 2010, S. 25). Die Aufgaben von TDQM werden in einem Regelkreislauf von Definition, Messung, Analyse und Verbesserung der Datenqualität beschrieben (Hildebrand et al., 2018, S. 70).

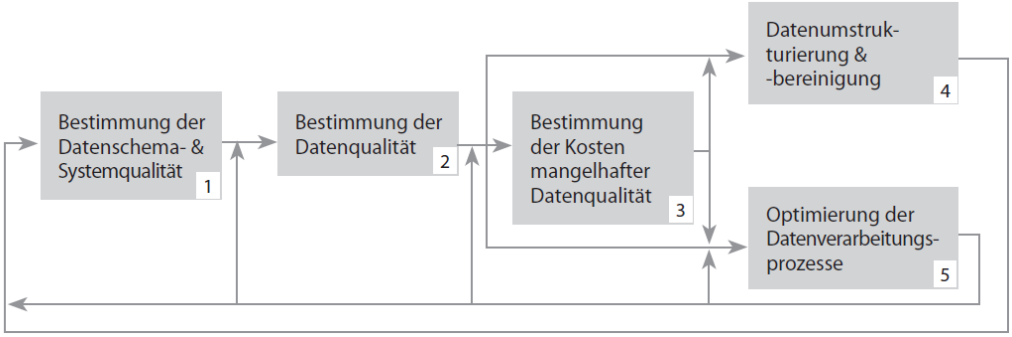


Abbildung 13: Total Quality data Management Methodik nach English (Apel et al., 2010, S. 25)

Zur Erfüllung der Hauptaufgabe der Hebung der Datenqualität ist Datenqualitätsmanagement eine Querschnittsfunktion über viele Unternehmensbereiche. Laut Weber (Hildebrand et al., 2018, S. 269) bezeichnet Datenqualitätsmanagement das qualitätsorientierte Management der Unternehmensressource Daten.

Technische Datenqualitätssicherungsmassnahmen müssen über geeignete Methoden und Verfahren zur Verfügung gestellt werden. Nur im Zusammenhang mit der Definition von Prozessen, Richtlinien, Plänen und Verantwortungsrollen sind sie fortlaufend auszuführen (Hildebrand et al., 2018, S. 273).

Datenqualitätsmassnahmen können in datenorientierte und prozessorientierte Massnahmen unterteilt werden (Hildebrand et al., 2018, S. 50). Prozessorientierte Ansätze versuchen eine Verbesserung der Datenqualität durch eine Neu- oder Umgestaltung der Geschäftsprozesse zu erreichen. Im Gegensatz dazu zielen datenorientierte Ansätze auf eine Bereinigung vorhandener Daten, um allfällige Datendefekte zu eliminieren. In dieser Arbeit wird für die Fallstudie der nachhaltigere prozessorientierte Ansatz weiterverfolgt.

Hühner unterteilt Datenqualitätsmassnahmen in proaktiv und reaktiv, um eine hohe Datenqualität zu gewährleisten (Schumacher, 2011, S. 28). Unter proaktiven Massnahmen versteht er vorbeugende Massnahmen wie Validierungen bei der Dateneingabe etc.

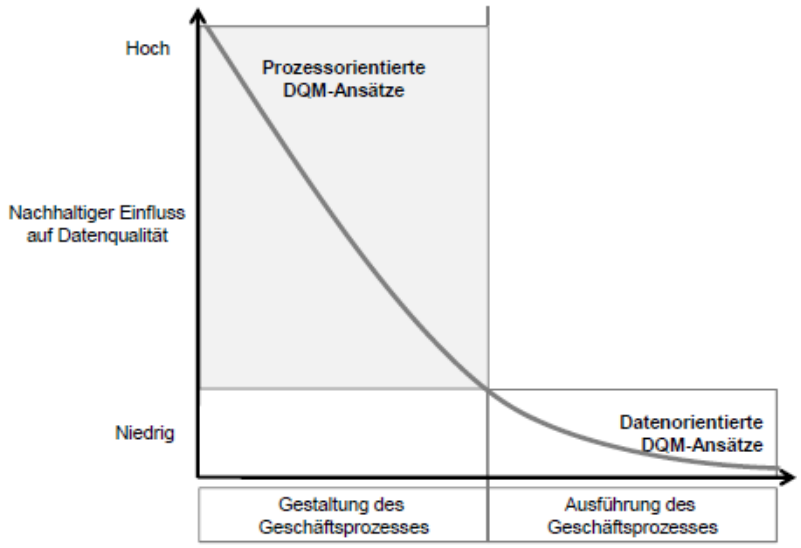


Abbildung 14: Vergleich der Ansätze des Datenqualitätsmanagements (Ofner, 2013, S. 103)

Die aktuelle Forschung definiert Datenqualitätsmanagement als die Gesamtheit aller aufeinander abgestimmten Tätigkeiten, zum Leiten und Lenken einer Organisation, die zu einer umfassenden Datenqualität führen. Die Erfassung, Sicherstellung und Verbesserung der Datenqualität kann über Data Profiling, Data Cleansing und Data Quality Monitoring gewährleistet werden.

Datenqualitätsmanagement beinhaltet weiter die Ermittlung von Anforderungen an die Daten, die Definition von Metriken und die Berechnung der Kennzahlen der ausgesuchten Datenqualitätsdimensionen für die Kontrolle der Qualität. Mit Qualitätszielen soll die Soll-Qualität der definierten Datenqualitätsdimensionen unter Berücksichtigung von Kosten-/ Nutzen Aspekten festgelegt werden. Diese Schritte sind nach der Einführung von DQM periodisch iterativ durchzuführen, bis das gewünschte Datenqualitätsniveau erreicht wird.

4.2. Datenqualitätsmanagementmodell

Um Datenqualität unternehmensweit zu betreiben sind in der aktuellen Forschung verschiedene Frameworks und Vorgehensmodelle entwickelt worden. Das in der Literatur am häufigsten zitierte und adaptiere Modell ist der Ordnungsrahmen bzw. Framework von Otto & Österle (2016, S. 23). Abbildung 15 zeigt diesen Ordnungsrahmen von Otto & Österle (2016).

Dieses Modell folgt dem Ansatz des Business Engineering und berücksichtigt neben organisatorisch-betriebswirtschaftlichen Anforderungen auch die unternehmensweite Anwendbarkeit von Datenqualitätsmanagement.

Der Ordnungsrahmen ordnet sechs Gestaltungsbereiche in die drei Gestaltungsebenen des Business Engineering ein (Otto & Österle, 2016, S. 23):

Ebene Strategie: Entwicklung einer Datenqualitätsstrategie

> Mit der Datenqualitätsstrategie wird das Datenqualitätsmanagement an den Unternehmenszielen ausgerichtet.

Ebene Organisation:

> Einführung eines Führungssystems.
>
> Entwurf einer DQM Organisation (Data Governance), dass die Zuordnung der Aufgaben, Rollen und Verantwortlichkeiten definiert.
>
> Definition von DQM-Prozessen: Datenqualitätsmanagement kann nur dann wirksam sein, wenn die Soll-Qualität der definierten Daten festgelegt ist und über geeignete Metriken entsprechende Datenqualitätskennzahlen gemessen werden.

Ebene Systeme:

> Spezifikation der Datenarchitektur und Datenhaltung.
>
> Systemunterstützung für DQM.

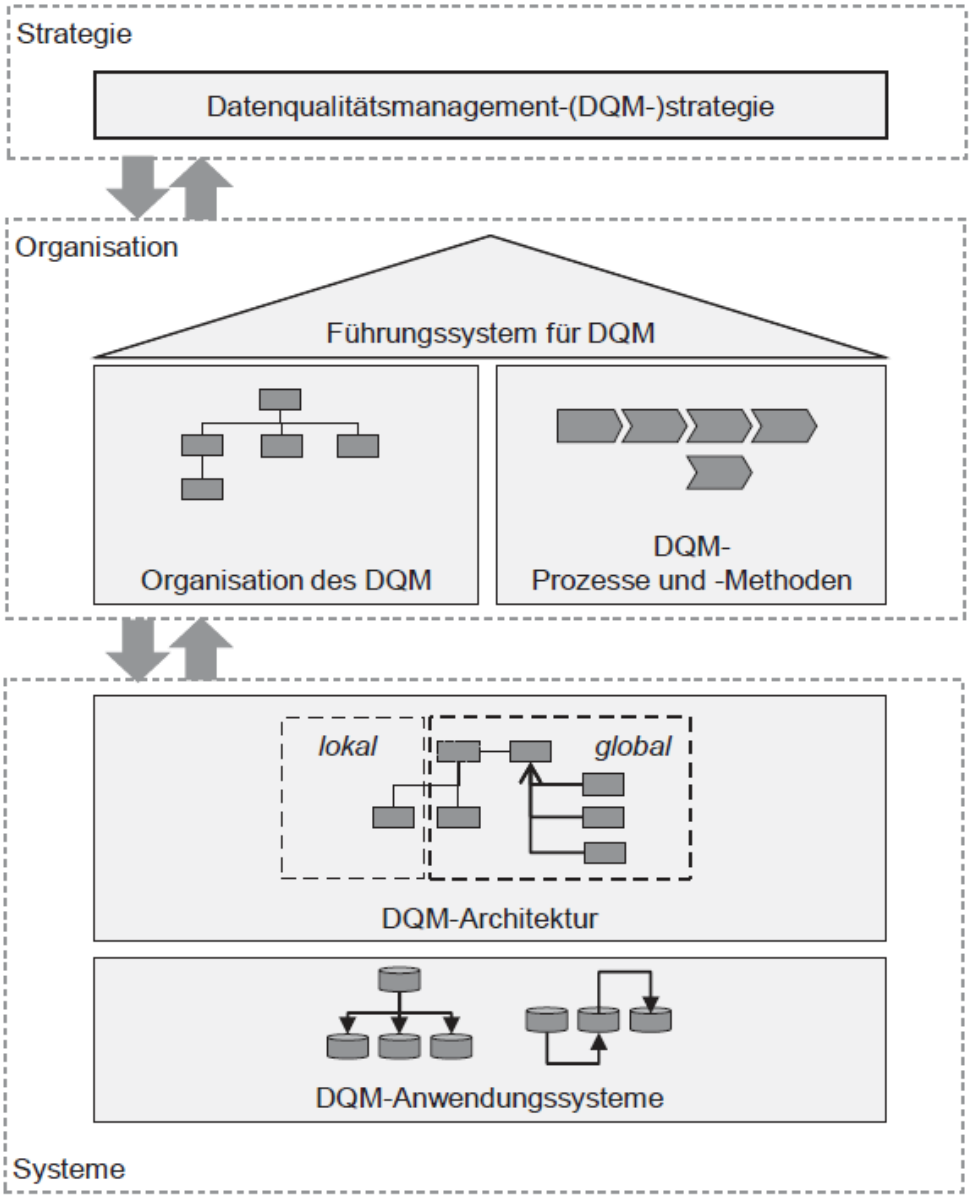


Abbildung 15: Framework für Datenqualitätsmanagement (Otto & Österle, 2016, S. 23)

4.3. Verbesserung der Datenqualität

Nach Apel et al. (2010) kann die Datenqualität in einem System mit zwei Schritten nachhaltig verbessert werden:

- Korrektur der fehlerhaften Daten
- Massnahmen zur Verhinderung neuer Datenqualitätsprobleme

Dazu bedarf es umfangreicher Vorarbeiten und Analyseschritten welche nachfolgend erläutert werden sollen.

4.4. Vereinfachendes generisches Vorgehensmodell

Um Datenqualitätsmassnahmen für KMUs fassbar und effizient anwendbar zu machen muss ein abgestimmtes Vorgehensmodell entwickelt werden. In der Literatur werden viele verschiedene Funktionenmodelle und Vorgehensmodelle beschrieben. In Anlehnung an das Funktionenmodell von Otto & Österle (2016) und Weigel (Hildebrand et al., 2018) wird folgendes Vorgehensmodell mit den Prozessschritten für die nachhaltige Umsetzung von Datenqualitätsmanagement in KMUs vorgeschlagen (Abbildung 16).

Es folgt initial einem zeitlich, sequenziellen Ablauf von ereignisgesteuerten Prozessschritten mit a) definiertem Input, b) Aufgaben und Funktionen und c) Output-Ereignissen. Einzelne Prozesse werden iterativ ausgeführt.

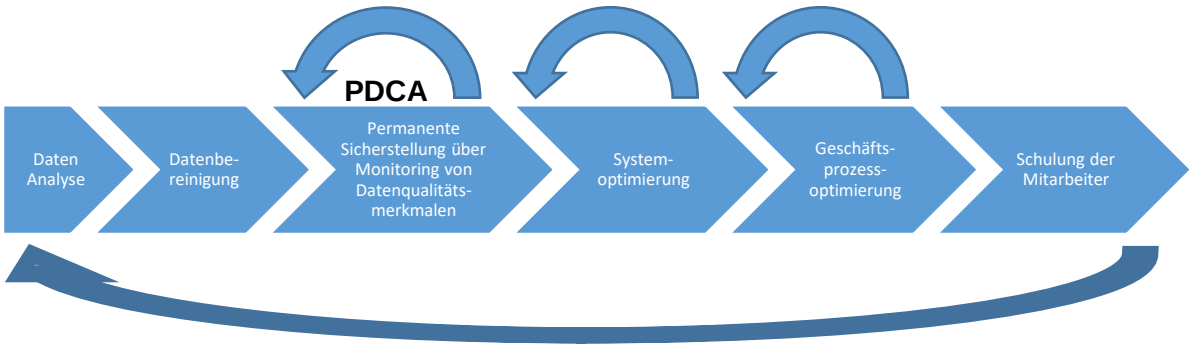


Abbildung 16: Vorgehensmodell mit Prozessschritten (eigene Darstellung)

Alle Prozessschritte sind auf die vorgelagerten Prozesse aufbauend. Die Datenanalyse dient dazu den Datenbestand zu definieren und dessen Ist-Zustand zu bestimmen. Vor der Datenbereinigung muss ein Soll-Zustand, ausgehend von den dokumentierten Compliance und Geschäftsregeln definiert werden. Die Datenbereinigung erfolgt durch Validierung des Datenbestandes mit den definierten Bedingungen und Regeln des Soll-Zustandes. Nach der erstmaligen Datenbereinigung werden die Problemursachen erkundet und entsprechende Massnahmen definiert, dazu werden Kennzahlen erhoben um das Qualitätsniveau permanent zu beurteilen. Ein wichtiger Aspekt ist die Definition einer Strategie und das definieren von Zuständigkeiten Die entwickelten Massnahmen werden in den betroffenen Systemen umgesetzt. Danach werden Datenobjekte der Geschäftsprozesse nach der Datenmanagementstrategie ausgerichtet und optimiert. Die Mitarbeiter bekommen Schulungen über die angepassten Abläufe und umgesetzten Massnahmen. Weitere Details der Prozessschritte werden in den nachfolgenden Kapiteln und in der Fallstudie erklärt.

4.5. Vermeidung von neuen Datenqualitätsproblemen

Die meisten Datenfehler treten durch fehlerhafte Prozesse auf. Damit Korrekturen nicht ständig wiederholt werden müssen, müssen diese Prozessfehler gefunden und korrigiert

werden und durch entsprechende permanent wirksame Massnahmen wie Dokumentation oder automatisierte Korrekturen sichergestellt werden. In weiterer Folge muss das System nach Kosten-/Nutzen Abwägungen so weit optimiert werden, dass diese Fehlerart möglichst vermieden wird.

Apel et al. (2010) schlägt vor die Prozesse in fünf Schritten zu verbessern:

1. Datenqualitätsprobleme eindeutig identifizieren und auf Grundlage einer Wirtschaftlichkeitsbetrachtung priorisieren.
2. Datenmetriken mit Soll-Zielwerten definieren.
3. Gründe der Datenqualitätsprobleme analysieren und die verursachenden Prozesse identifizieren.
4. Prozessänderungen, um die Datenqualitätsprobleme zu beheben.
5. Erneute Messung und Vergleich mit den Zielwerten.

Abbildung 17 zeigt diese fünf Schritte im Überblick.

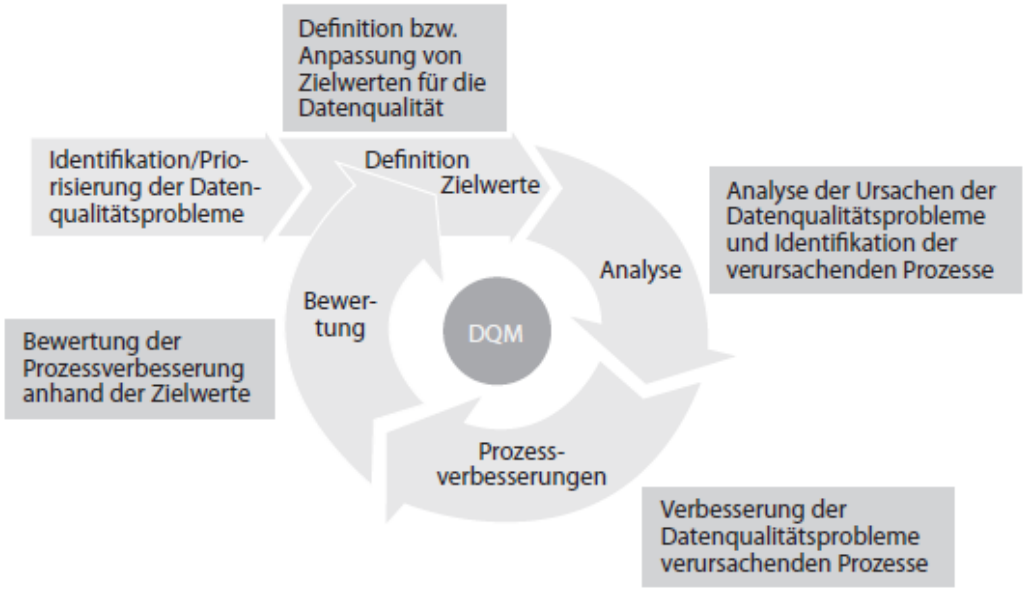


Abbildung 17: Verbesserung der Prozesse in den Quellsystemen (Apel et al., 2010, S. 104)

Ist die Abweichung zwischen Messung nach der Prozessverbesserung und dem ursprünglichen Zielwerten des Datenqualitätsmerkmals zu gross, wird diese Vorgehensweise iterativ angewandet oder die Zielwerte entsprechend korrigiert.

4.6. Datenanalyse

Einer der zentralen und zu gleich komplexesten Schritte im Datenqualitätsmanagement ist die Analyse der Probleme bzw. der Daten. Die Datenanalyse soll Aussagen über die Qualität der Daten liefern. Dazu werden verschiedene Verfahren des Data Profilings angewendet, bei

der die bekannten Metadaten Informationen wie Geschäftsregeln, Wertebereiche etc. auf die existierenden Daten angewendet werden. Dadurch erhält man neue Metadateninformationen und bereits bestehende Datenqualitätsprobleme können hoffentlich erkannt werden. Dieser Prozess sollte nur so weit iterativ betrieben werden, bis ein ausreichendes Verständnis über den Datenbestand vorliegt. Nicht so gut wie möglich sondern so gut wie nötig. Das Ziel ist es die bestehenden Metadaten zu verfeinern, dass ein gutes Verständnis der Daten und Geschäftslogik vorhanden ist.

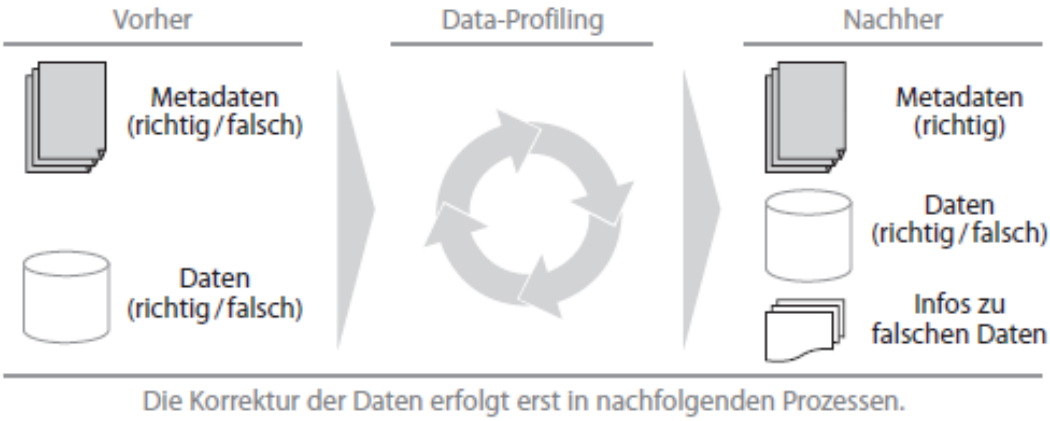


Abbildung 18: Aufgaben der Datenanalyse (Apel et al., 2010, S. 110)

Data Profiling ist meist komplex. Deshalb sollten in dieser Phase entsprechende Tools zur Unterstützung eingesetzt werden. Dadurch können diese Analyseschritte zeitsparend automatisiert wiederholt werden.

Folgende Faktoren identifiziert Apel et al. (2010) um ein effizientes und erfolgreiches Data Profiling durchzuführen (Apel et al., 2010, S. 147):

- fundierte Kenntnisse über die Fachlichkeit und die zugehörigen Prozesse,
- methodisches Vorgehen,
- richtige Kombination und Einsatz der verschiedenen Verfahren und
- optimale Zusammensetzung und Erfahrung des Teams.

4.7. Evaluation von Data Profiling Software Tools

In diesem Kapitel werden vier Software Tools für die relationale Datenbank Microsoft SQL Server (ab Version 2012) vorgestellt, die das Data Profiling massgeblich unterstützen sollen. Microsoft (MS) SQL Server gehört zu den weltweit am weitesten verbreiteten RDBMS für ERP Systeme in KMUs und ist auch das Datenhaltungssystem (DB) aus der Fallstudie.

4.7.1. Funktionale Kriterien für die Auswahl

Die Software Tools wurden dahingehend evaluiert, wie weit und einfach sie bei der

- Attributanalyse
- Datensatzanalyse
- Tabellenanalyse von relationalen Daten helfen.

In den nächsten Absätzen werden die evaluierten Software Tools vorgestellt.

Talend Data Quality Open Studio

Die auf Java basierte Software Talend Data Quality ermöglicht Profiling, Bereinigung und Maskierung von Daten sowie die langfristige Überwachung der Datenqualität unabhängig von Format oder Grösse. Es unterstützt De-Duplizierung, Validierung und Standardisierung der Daten um saubere Bestände für Zugriffs-, Reporting- und Analysezwecke zu erhalten. Auch können externe Referenzdaten z. B. für die Adressvalidierung, Unternehmensidentifizierung, Beurteilung der Kreditwürdigkeit etc. eingebunden werden.

Ein kurzer Test hinterlässt einen zwiespältigen Eindruck, da die Benutzerführung gerade in der Auswertephase nicht übersichtlich genug ist, sowie die Software nur bedingt mit Microsoft SQL Server umgehen kann und eine Regex (Muster) Analyse z. B. für die Pattern Erkennung nur erschwert umgesetzt werden kann (deshalb ist nachträglich im Rahmen dieser Arbeit eine SQL Regex Funktion im Microsoft SQL Server (bzw. DB Instanz) integriert worden. Siehe Anhang C). Die Firma bietet viele weitere Software für Datenmanagement an, wie z. B. „Talend Open Studio for MDM“ (Masterdatamanagement). Die Oberfläche und die Bedienung ähneln sich, daher ist die Einarbeitungszeit entsprechend verkürzt, was wiederum für deren Verwendung spricht.

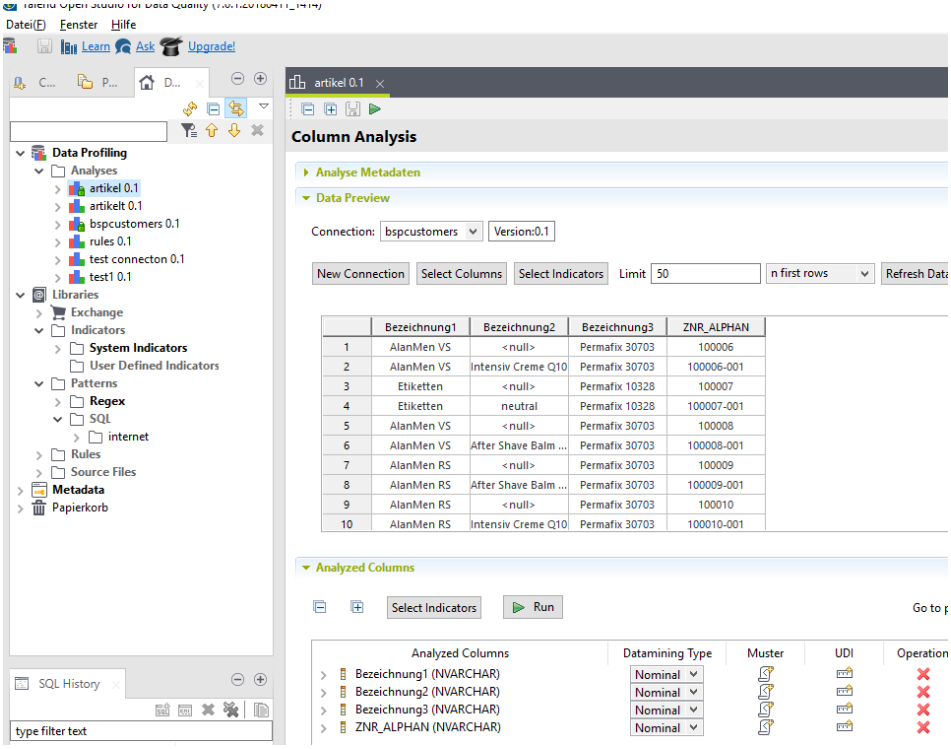


Abbildung 19: Talend Quality Data Open Studio Dataprofiing Konfiguration (eigene Darstellung)

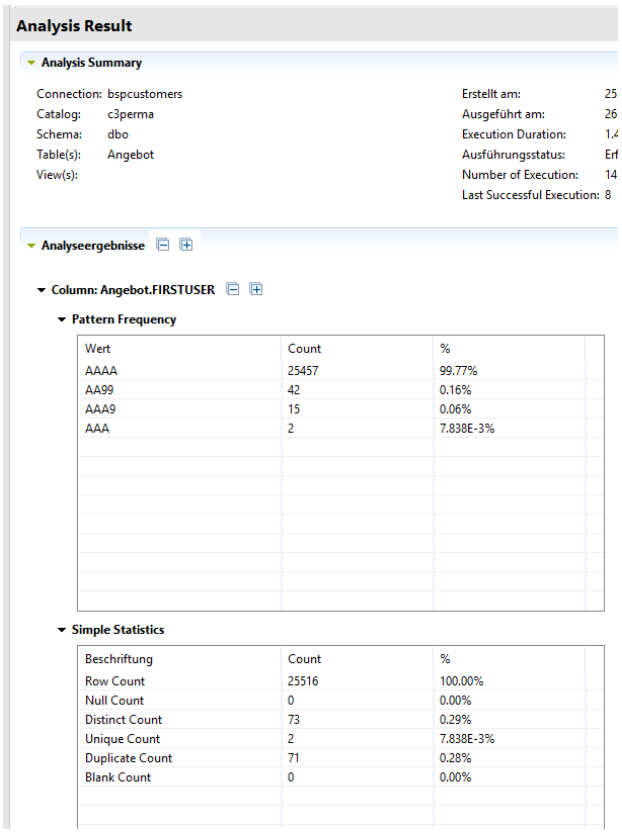


Abbildung 20: Mustererkennung auf Attributebene FirstUser (eigene Darstellung)

Die Patternanalyse (Mustererkennung) folgt nicht bekannter Regular Expression Schreibweise und wird als problematisch bei der Weiterverwendung der Ergebnisse angesehen.

DataCleaner

DataCleaner ist eine Opensource Software und bietet sehr viele Funktionen für die Analyse, dem Data Profiling und der Daten Bereinigung an. Es funktioniert mit den gängigsten Datenbanksystemen. Die verschiedenen Analyse Funktionen sind sehr übersichtlich und klar in der Bedienung und Konfiguration. Die Resultate werden in Form von Reports wiedergeben. Dazu lassen sich Daten vor der Analyse über so genannte Transformatoren aufbereiten oder je nach Bedarf miteinander kombinieren.

Nach der kurzen Einarbeitung verblüfft der grosse Funktionsumfang, der sich nur mit meist teuren Analysewerkzeugen vergleichen lässt. DataCleaner setzt für die Bedienung allerdings fortgeschrittene Datenbank Kenntnisse voraus. Diese sind für ein effektives, effizientes und zielführendes Data Profiling aber Grundvoraussetzung.

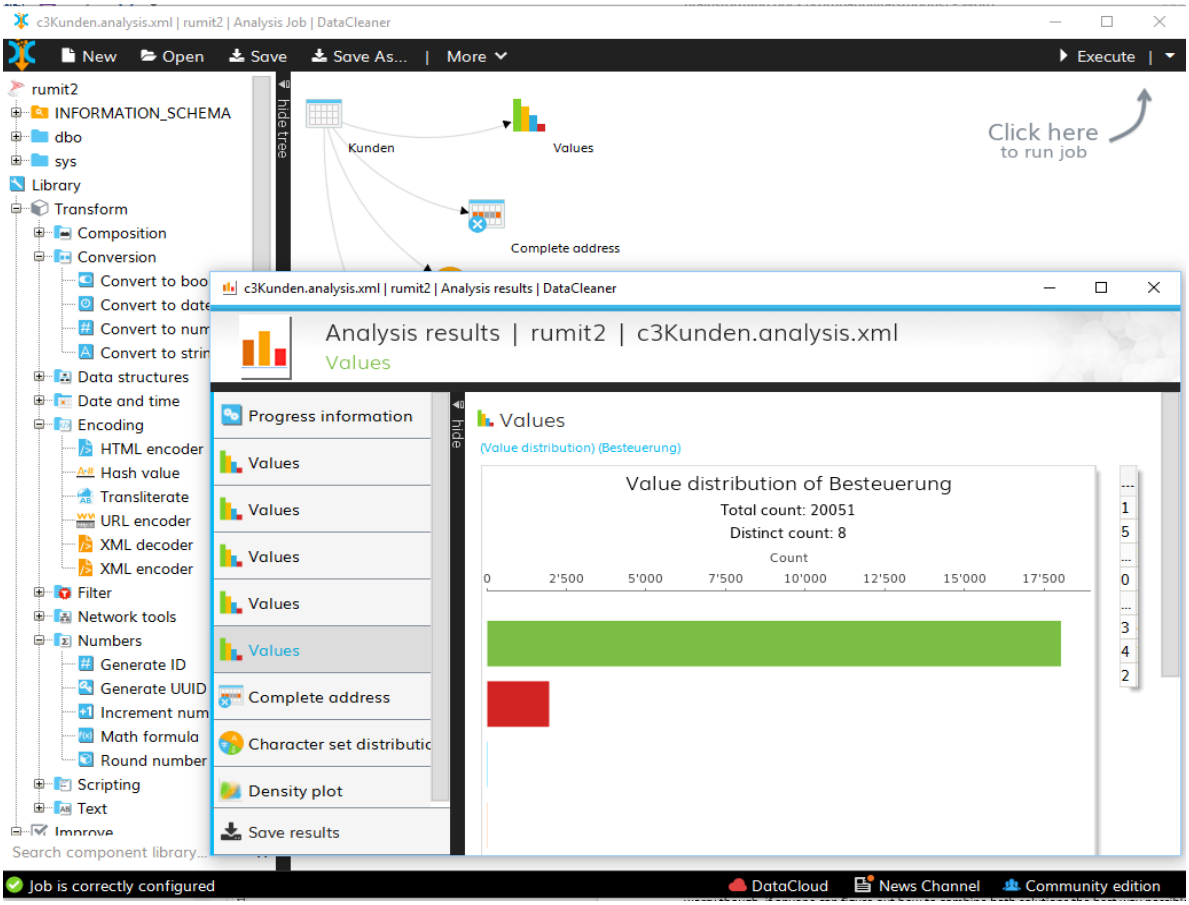


Abbildung 21: Data Cleaner – Werteverteilung auf Attributebene (eigene Darstellung)

Microsoft Data Quality Client

Mit der im Installationsumfang von Microsoft SQL Server (Developer Edition) mitgelieferten Software Data Quality Client können DataDictonaries, also Metadaten Repositories erstellt und Datenprüfungen auf Grundlage dieser Metadaten durchgeführt werden. Das Tool ist sehr übersichtlich aufgebaut und intuitiv in der Bedienung, bietet aber über die grundlegenden Analysen hinaus wenig Unterstützung beim Erforschen der Daten Strukturen. Das Tool kann zum Verwaltung der Regeln und Abhängigkeiten genutzt werden, welche für die nächste evaluierte Software verwendet werden kann. Somit können auch Nutzer aus den Fachbereichen die Geschäftsregeln erstellen und modifizieren und brauchen keine tiefer gehenden Informatikkenntnisse.

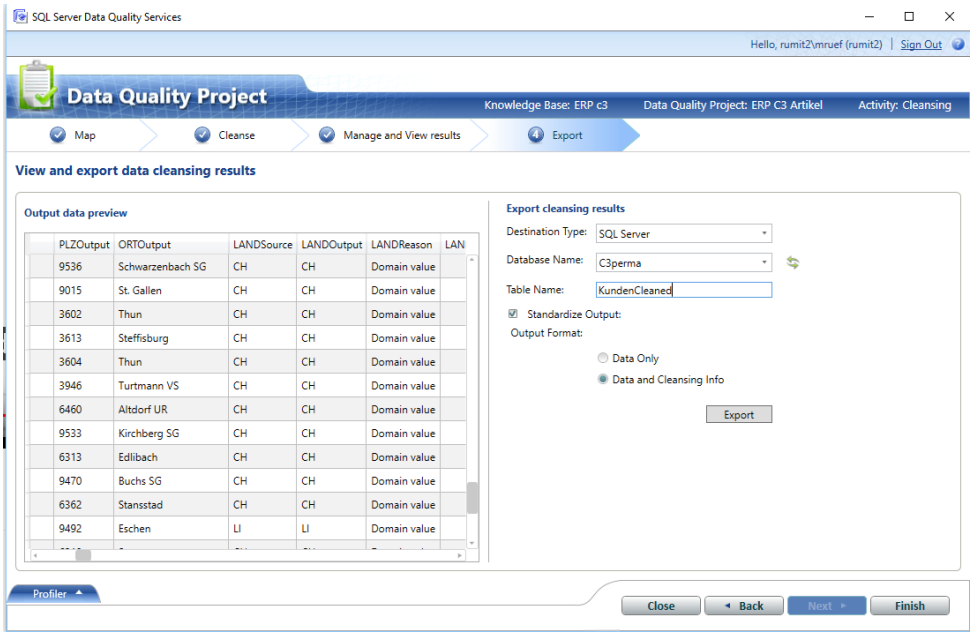


Abbildung 22: Microsoft SQL Data Quality Client (eigene Darstellung)

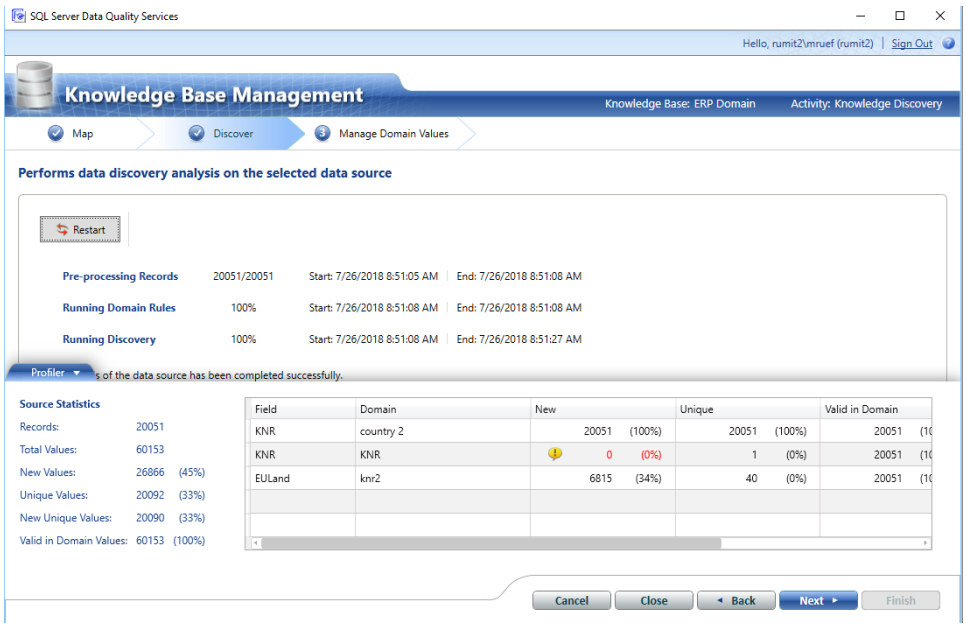


Abbildung 23: Microsoft SQL Data Quality Data Dictonary u Knowledge Base (eigene Darstellung)

Microsoft SQL Server Integration Services (SSIS) für Microsoft Visual Studio

SSIS folgt der Software „Data Transformation Services“ nach und das Tool kann vor allem bei Datentransformationen und Datenmigrationen (ETL) helfen. Es bietet viele verschiedene Funktionen die grafisch bedient und konfiguriert werden. Die erstellten Projekte in Microsoft Visual Studio können dann automatisch vom SQL Server Agent ausgeführt werden. Der Vorteil dieses Produktes ist, dass viele Microsoft (MS) SQL Datenbankadministratoren mit dem SSIS Produkt vertraut sind und das Tool somit bedienen können um Datenanalysen zielführend durchführen zu können.

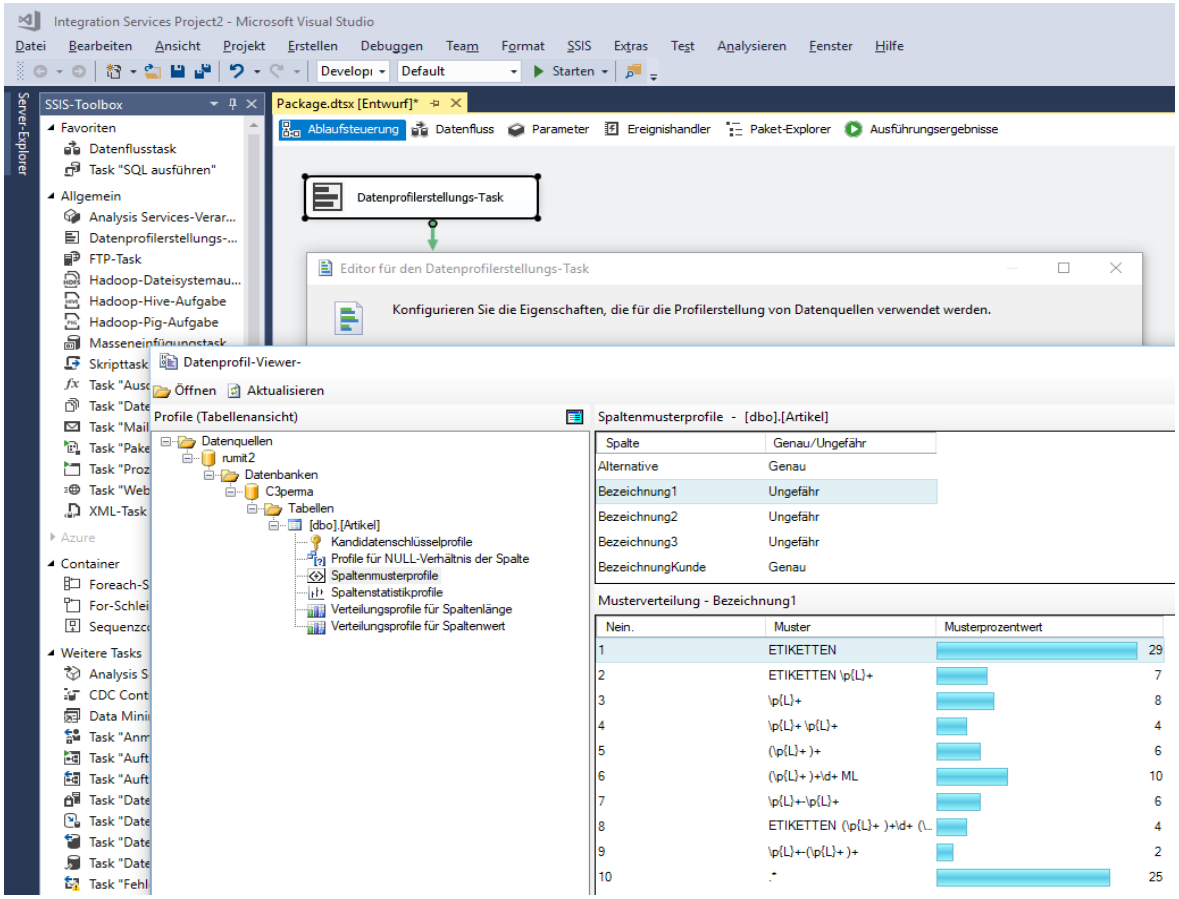


Abbildung 24: Beispiel Projekt und die Datenprofil Ansicht für die Musterverteilung des Attributs Bezeichnung 1 der Artikel Tabelle (eigene Darstellung)

4.7.2. Fazit und Bewertung der Software

Grundsätzlich erfüllen alle getesteten Software Tools die geforderten Funktionen um Data Profiling durchzuführen (siehe Tabelle 2). MS Data Quality Client bietet einen zu geringen Funktionsumfang, eignet sich aber für die automatisierte Validierung der Daten gegenüber Geschäftsregeln und anderen definierten Regeln und kann auch ohne umfangreiche Informatikkenntnisse bedient werden. DataCleaner als OpenSource Software wird für die Daten Analyse unbedingt empfohlen. Microsoft SSIS eignet sich aufgrund der logischen Bedienung und vielen Funktionen ebenso gut für das Data Profiling. Diese Toolevaluierung wurde durchgeführt, weil eine passende Software viel Arbeit bei der quantitativen Analyse des Datenbestandes abnehmen kann.

Daten Profiling Software	**Talend Data Quality**	**DataCleaner**	**MS Data Quality Client**	**MS SSIS**
Attributanalyse	+	++	+	+
Datensatzanalyse	++	+	+	++
Tabellenanalyse	++	++	+	++
Metadaten Verwaltung	+	+	++	++
Geschäftsregeln	+	+	+	+
Plausibiliätsprüfungen	+	+	+	+
Abgleich mit Referenzdaten	+	+	+	+
Dublettenerkennung	0	++	0	+
Mustererkennung	+	++	+	+
Massenbearbeitung	+	+	+	+

Tabelle 2: Bewertungsmatrix der Data Profiling Software (eigene Darstellung)

5. Fallstudie Datenqualitätsmanagement in einem Produktionsunternehmen

Durch die Literaturrecherche konnte ein passendes Datenmanagementmodell und ein Vorgehensmodell gefunden und adaptiert werden, welches helfen soll, die Forschungsfrage zu beantworten, wie Datenqualität nachhaltig in einem Unternehmen verbessert werden kann.

Mit der Praxis Fallstudie soll aufgezeigt werden welche Herausforderungen dabei auftreten können und mögliche Lösungsansätze für die Umsetzung von Datenqualitätsmanagement beschrieben werden. Das Ergebnisdokument zur Fallstudie ist im Anhang B gelistet.

5.1. Permapack – Einführung eines Datenqualitätsmanagement in einem heterogenen Handels- und Produktionsunternehmen

Die Permapack AG (Permapack) ist eine 1959 gegründete erfolgreiche schweizerische KMU in Familienbesitz. Sie beschäftigt derzeit ca. 380 Mitarbeiter an zwei Standorten in Rorschach am Bodensee. Permapack ist organisatorisch in die zwei Vertriebs-Divisionen „Flexible Verpackung“ und „Technik“, sowie in die Dienstleistungs-Division „Finanzen Services & IT“ (FSI) unterteilt. Sie vertreiben als Handelsunternehmen Dichtstoffe und Klebebänder für die Bauindustrie. Für den Kosmetik, Food- und NonFood Markt produziert Permapack Laminattuben und bedruckt selbstklebende Etiketten und Verpackungsfolien.

Permapack operiert als Handelsunternehmen in einem stark umkämpften Markt. In diesem Segment bedient Permapack schweizweit Retail-Kunden und vereinzelt Endkunden wie Bauunternehmen oder Malergeschäfte. Für die gesamte Geschäftsabwicklung ist die 2017 eingeführte ERP Software Microsoft AX (AX) im Einsatz mit den Hauptmodulen Beschaffung, Dispo, Artikel-, Lieferanten- und Kundenstamm und CRM.

In der Division Flexible Verpackung werden Artikel auf Kundenwunsch wie Tuben-Etiketten oder Sachet für die Kosmetikindustrie hergestellt. Dieser Markt ist stark preisgetrieben, bzw. die schnelle Ermittlung der korrekten Herstellkosten bzw. Verkaufspreises für die Angebotserstellung ist eine der Herausforderungen für die Division. Das ERP System C3 (C3) von Theuerer.com stellt Funktionen für die gesamte Produktionsabwicklung vom Einkauf, Lager, Produktentwicklung, Arbeitsvorbereitung, Planung über die eigentliche Produktion mit der Betriebsdatenerfassung bis hin zum Verkauf zur Verfügung.

5.2. Ausgangssituation Datenlandschaft

Da die Finanzbuchhaltung und Verrechnung im AX stattfindet, sind beide ERP Systeme über Schnittstellen miteinander verbunden. Im AX werden die Debitoren- und Kreditorenstammdaten gepflegt, im C3 wird der Artikel- und Materialstamm für die Eigenproduktion gepflegt. Rechnungen für Kundenaufträge werden im C3 automatisch aus erledigten Lieferscheinen generiert. Die Schnittstelle zwischen den beiden ERP Systemen dient für den Export des Artikelstammes und der Rechnungen aus C3 ins AX und für den Export der Debitoren und Kreditoren aus AX ins C3. Weiter sind Business Intelligence Anwendungen wie QlikView und digitale Schnittstellen zu Kunden z. B. über EDI und Behörden wie z. B. Zoll über E-Trans angebunden.

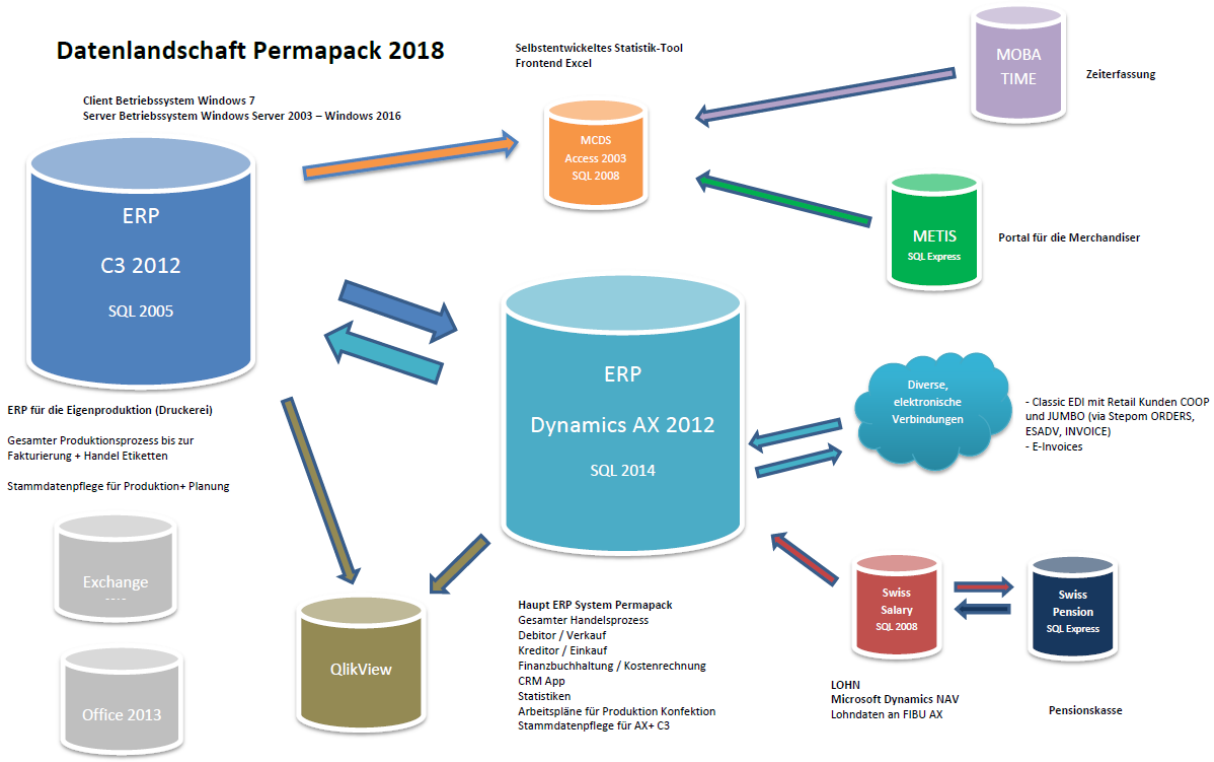


Abbildung 25: Datensystem-Landschaft von Permapack (eigene Darstellung)

Damit diese heterogenen Systeme möglichst gut miteinander operieren und kommunizieren sind Daten mit hoher Qualität ausschlaggebend.

5.3. Handlungsdruck

Die Eigenproduktion unterliegt hohen gesetzlichen Anforderungen hinsichtlich Exportdokumentation, Konformitätserklärungen für den Lebensmittelbereich wie REACH, RoHS und weiteren. Das ISO 9001:2015 und ISO 14001:2015 zertifizierte Unternehmen ist also gefordert jederzeit den Nachweis über die Unbedenklichkeit seiner Eigenartikel und die

lückenlose Rückverfolgbarkeit aller seiner Artikel für die Kunden aber auch Behörden zu erbringen und zu gewährleisten. Alle drei Jahre erfolgt eine Re-Zertifizierung aller Geschäftsprozesse durch die SQS, der Schweizerischen Vereinigung für Qualitäts- und Managementsysteme.

Aufgrund neuer externer Anforderungen für Konformitätserklärungen der Artikel gegenüber den Kunden und internen Anforderungen ein neues Berichtswesen mit Business Intelligence Anwendungen aufzubauen, soll die Datenqualität der Stammdaten im C3 erhöht werden.

Bisher war Datenqualität bei Permapack noch nicht wichtig genug um es strukturiert anzugehen. **Datenqualitätsprobleme** in den Stammdaten wurden wenn überhaupt auf Zuruf manuell durch die Informatikabteilung oder den Fachabteilungen gelöst. Es werden auch noch keine systematischen **Datenqualitätsmessungen** durchgeführt. Datendefekte fallen somit erst bei Problemen in Geschäftsprozessen wie der Erstellung von Konformitäten **(Datennutzung)** auf.

5.4. Ziele der Fallstudie

In dieser Fallstudie soll dargestellt werden, wie Datenqualitätsmanagement als Prozess bei Permapack umgesetzt werden kann.

Darüber hinaus wird der Teilaspekt Datenanalyse unter Zuhilfenahme von Data Profiling Software anhand der Problemstellung die Datenqualität der Artikel- und Materialstammdaten zu erhöhen praxisnah dargestellt.

Ein Beispiel bei der die mangelnde Datenqualität der Stammdaten offensichtlich wird, ist bei der Erstellung von Konformitäten, bei denen verschiedene Daten aus dem Artikel- und Materialstamm genutzt werden. Weiter werden digital abgelegte Materialspezifikationen der Lieferanten benötigt. Diese sind unsystematisch abgelegt und nicht im Materialstamm verlinkt bzw. vermerkt. Die Folgen davon sind jeweils zeitintensive Nachforschungen über die Dokumente, wo sie abgelegt sind und welche gültig sind, oft bis zurück zum Lieferanten.

5.5. Realisierung von Datenqualitätsmanagement bei Permapack

Unter Verwendung des erarbeiteten Vorgehensmodelles aus Kapitel 4 soll bei Permapack ein Datenqualitätsmanagement realisiert werden. Dazu definiert die Unternehmensleitung eine Datenqualitätsmanagement Strategie anhand von Zielen, die an der Unternehmensstrategie ausgerichtet sind.

5.5.1. Datenqualitätsstrategie

In Anlehung an das Framework für Datenqualitätsmanagement von Otto & Österle (2016, S. 23) wird eine unternehmensweite Datenqualitätsstrategie entwickelt, die besagt,

„dass proaktiv Datenfehler schon bei der Erfassung von Daten minimiert werden sollen."

Grund ist, dass die Stammdaten aufgrund ihrer geringen Erstellungshäufigkeit durchwegs manuell erfasst werden, aber Fehler durch die Eingabe oft zeitnahe Auswirkungen auf andere operative Geschäftsprozesse haben, z. B. die falsche Zuordnung eines Materials zu einer Warengruppe, die Geschäftsprozesse Einkauf und Lager empfindlich stört.

5.5.2. Ziele der Unternehmensführung

Aufgrund der geschilderten Ausgangssituation mit den Anforderungen erhofft sich die Division „Flexible Verpackungen" folgende Punkte von einem Datenqualitätsmanagement:

- höhere Kundenzufriedenheit und verbesserte Kundenbindung durch aktive Erstellung von Konformitäten nach deren zeitlichem Gültigkeitsverfall
- kürzere Durchlaufzeiten bei der Erstellung von Konformitäten
- Reduktion von Datenpflegeaufwänden und bessere zeitliche Planung der Datenpflege
- validere und umfassendere, sowie einheitlichere Konformitäten
- daraus resultierende mögliche Teil-Automatisierung der Konformitätserstellung
- teilweise automatisierte und systemgestützte Konformitätserstellung
- Auswertemöglichkeiten und Kennzahlen zu Konformitäten
- höhere Qualität des Artikel- und Materialstammdaten die durch ein Datenqualitäts Monitoring regelmässig geprüft wird
- höheres Bewusstsein bei den Mitarbeitern für Datenqualität

Um diese Zielvorgaben zu erreichen, soll das Projekt „Stammdatenqualität im C3" über das in Kapitel 4 vorgestellte Vorgehensmodell als Aufbauorganisation initiiert werden, um damit die bestehenden Datenqualitätsprobleme zu mindern und Datenqualitätsmanagement als unternehmensweiten Prozess zu verankern. Das Wesen einer Aufbauorganisation ist die Definition der Rollen und der Zuständigkeit/Verantwortung für Aufgaben.

5.5.3. Definition von Rollen und Zuständigkeiten

In der Literatur gibt es verschiedene Modelle, um in einer Aufbauorganisation die Rollen und Zuständigkeiten für das Datenqualitätsmanagement zu regeln. Den meisten gemein ist eine Dreiteilung in einen **Datenqualitätslenkungsausschuss** (Entscheidungsgremium, das übergreifende Datenqualitätstandards definiert), einem **zentralen Datenqualitätsmanager** (Zuständig für die Koordinierung und Verfolgung von Datenqualitätsproblemen) und

dezentrale Data Stewards, die Aufgaben in der Datenqualitätssicherung, Datendefinitionen sowie Freigabe von Daten zum Zugriff übernehmen (Apel et al., 2010, S. 53). Die Systemverantwortlichen in der Informatik stellen den operativen technischen Betrieb der Informations- und Datenbereitstellungssysteme sicher und unterstützen die Verantwortlichen der Datenerfassung. Die Datenerfassung ist für die korrekte und vollständige Dateneingabe sowie deren Korrektur, wenn Datenprobleme erkannt werden zuständig. Damit die Rollen und Verantwortung gut benannt werden können, ist ein Regelwerk über die Quellsysteme, Schnittstellen, die genutzten Datenfelder und die Geschäftsverantwortung eines Datenbereiches zu erstellen (Apel et al., 2010, S. 57). Die Daten Stewarts definieren in der Ablauforganisation dann selbstständig Kennzahlen und Reports, identifizieren Datenqualitätsprobleme und leiten Massnahmen für deren Behebung ein. Siehe auch Abbildung 26 die die Aufbauorganisation für Datenqualitätsmanagement mit den Ebenen und den Rollen darstellt.

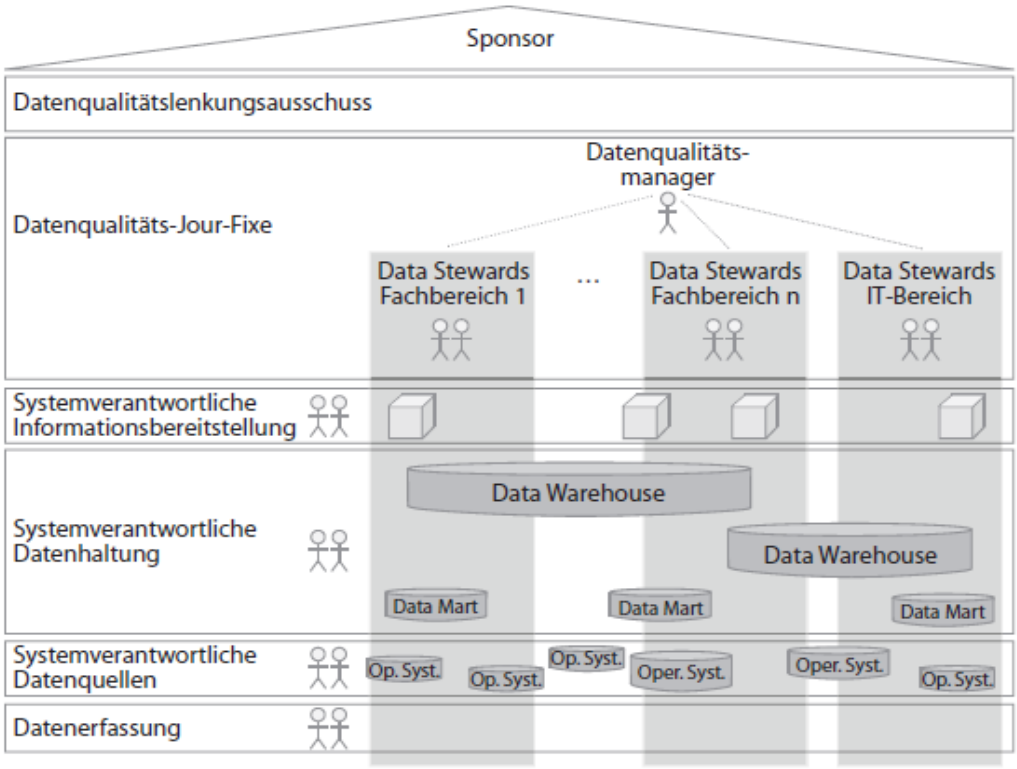


Abbildung 26: Rollen und Gremien in der Aufbauorganisation (Apel et al., 2010, S. 54)

Die Unternehmensleitung definierte als erste proaktive Massnahme neue Rollen und Verantwortlichkeiten für die Datenmanagementprozesse **Datenbereinigung** und **Datenqualitätsverbesserungsmassnahmen**.

5.5.4. Projektteam

Die Unternehmensleitung stellte für die Realisierung der Aufbauorganisation ein Projektteam bestehend aus dem C3 ERP Verantwortlichen und je einem Fachverantwortlichen von Einkauf und Verkauf zusammen.

5.5.5. Befragung der Datennutzer

Als erstes führten sie eine Umfrage durch, mit der alle C3 Benutzer in den Bereichen Einkauf, Verkauf und Arbeitsvorbereitung befragt wurden, welche Datenqualitätsprobleme sie kennen oder subjektiv wahrnehmen. Insgesamt wurden ca. 70 C3 Nutzer eingeladen an der Umfrage teilzunehmen, davon haben 31 eine gültige Antwort zurückgesendet. Folgende Datenqualitätsdimensionen wurden nach Diskussion im Projektteam für die Beurteilung der Datenqualität in den Artikel- und Materialstammdaten durch die C3 Nutzer ausgewählt:

1) Zugänglichkeit, 2) Bearbeitbarkeit, 3) angemessener Umfang, 4) Vollständigkeit,

5) Relevanz, 6) Glaubwürdigkeit, 7) Fehlerfreiheit, 8) Übersichtlichkeit, 9) Verständlichkeit

Diese Dimensionen sind den Kategorien Darstellung, Inhalt, Nutzung und System zugeordnet.

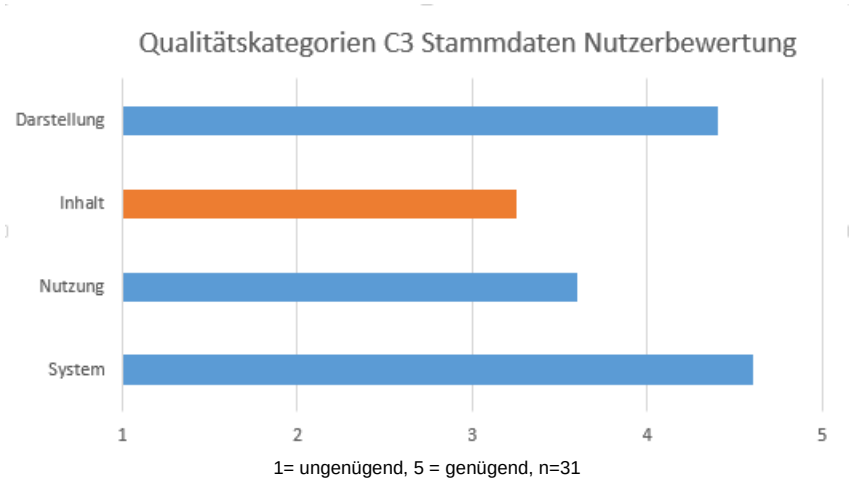


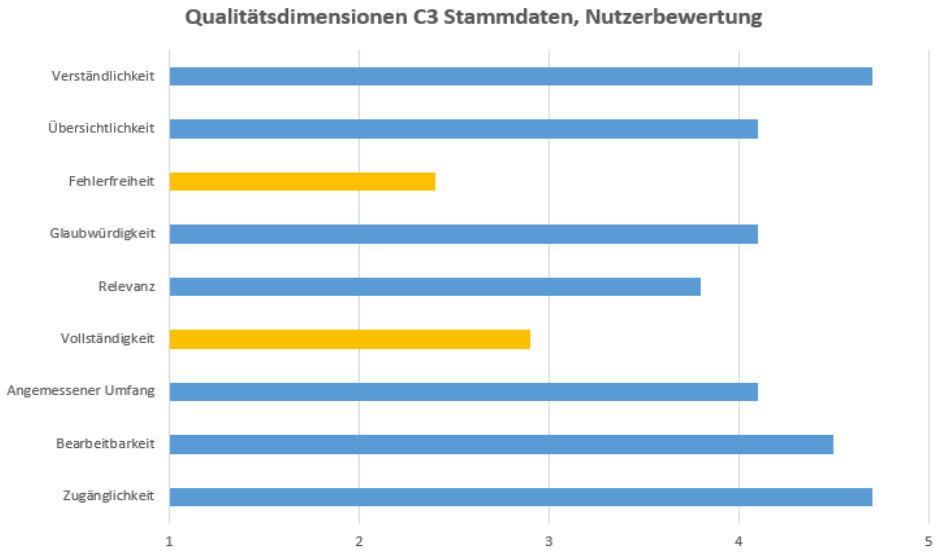
Abbildung 27: Umfrageergebnis der C3 Nutzer nach Datenqualitätskategorien (eigene Darstellung)

Die Auswertung der Umfrage ergibt folgende Ergebnisse (siehe Abbildung 27):

a) Mindere Datenqualität in der Kategorie Inhalt

Aus dieser Umfrage geht hervor, dass die Nutzer vor allem Datenqualitätsprobleme in der Kategorie Inhalt (Mittelwert 3.3) und Nutzung (3.6) sehen. Insgesamt wird die Datenqualität aber in keiner Kategorie als ungenügend angesehen.

Die Dimensionen mit der vermeintlich geringsten Qualität sind Fehlerfreiheit (Mittelwert 2.4) in der Kategorie Inhalt und Vollständigkeit (2.9) in der Kategorie Nutzung. Diese Dimensionen werden im Schritt Datenanalyse deshalb besonders untersucht.

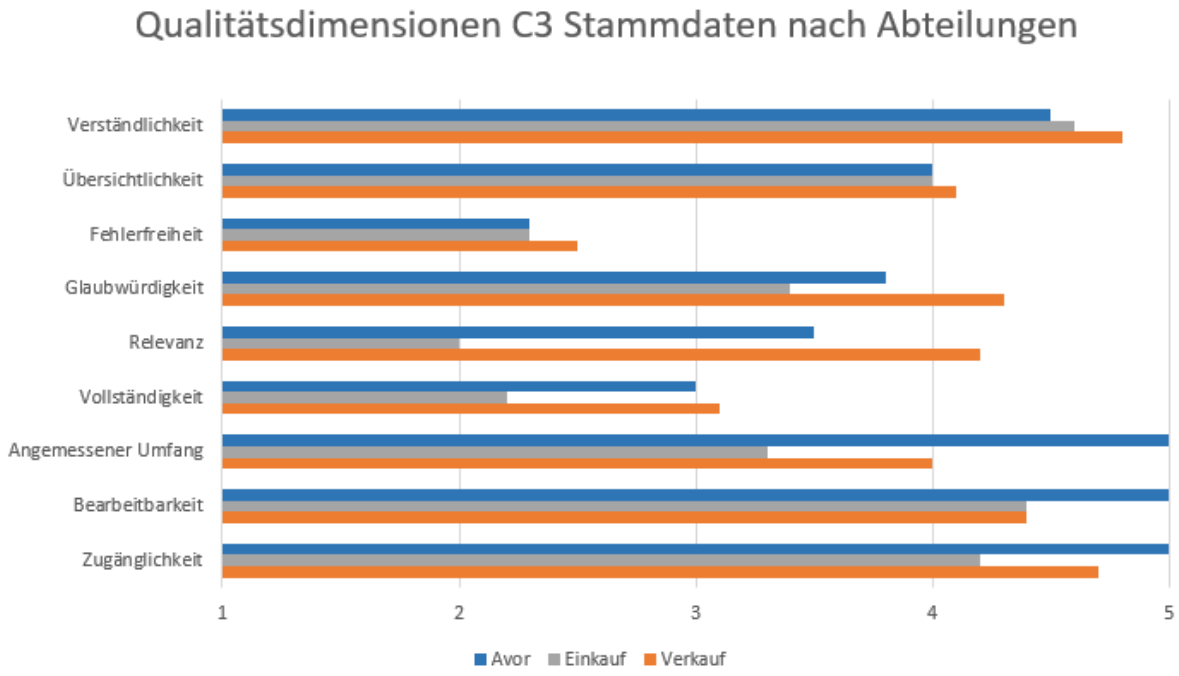


1= ungenügend, 5 = genügend, n=31

Abbildung 28: Umfrageergebnis der C3 Nutzer nach Datenqualitätsdimensionen (eigene Darstellung)

b) Unterschiedliche Wahrnehmung der Datenqualität zwischen den Fachabteilungen

Nach Abbildung 28 ist auffallend, dass es bei zwei Dimensionen Relevanz und Vollständigkeit vor allem die Nutzer aus der Abteilung Einkauf die Datenqualität von Relevanz bzw. Vollständigkeit schlechter als die anderen Abteilungen beurteilen. Während bei der Dimension „Angemessener Umfang“ die Abteilung Verkauf die Qualität als sehr gut wahrnimmt, die anderen Abteilungen abfallen. Das lässt darauf schliessen, da die Material- und Artikelstammdaten in der C3 Applikation über die gleichen Masken gepflegt werden, die Benutzerführung eventuell zu sehr auf den Verkauf ausgerichtet ist, bzw. Datenfelder in der Eingabemaske den Nutzern im Einkauf inhaltlich unklarer erscheinen.



1= ungenügend, 5 = genügend, n=31

Abbildung 29: Umfrageergebnis Datenqualitätsdimensionen pro Abteilung (eigene Darstellung)

5.5.6. Projekt Ziele

Ausgehend von dieser Umfrage definierte das Projektteam folgende Ziele für das Stammdatenqualitätsmanagement-Projekt:

- Schaffung eines Bewusstseins für die Notwendigkeit von Datenqualitätsmanagement bei der Unternehmensführung mit Hilfe einfacher Datenqualitätskennzahlen
- Schaffung von Transparenz über die Stammdatenqualität
- Installieren von systematischen Kontroll- und Überwachungsfunktionen, um Trends und Entwicklungen bei der Stammdatenqualität zu verfolgen
- Initiierung angemessener Datenbereinigungs- und Qualitätsverbesserungsaktivitäten für die identifizierten Handlungsfelder
- Schaffung von kontinuierlichen Prozessen für die Stammdatenpflege
- Definieren von Verantwortlichkeiten (Data Owner) für die Stammdaten

5.5.7. Datenanalyse des Ist-Zustandes

Um auf organisatorischer Ebene die Datenqualität erfolgreich zu steuern, ist die nächste Aufgabe den Ist-Zustand der Qualität vom Datenbestand aufzunehmen.

Eine Ist-Analyse beginnt laut Apel et al.(2010) mit der **Identifikation** der Systeme, die die Daten bereithalten. Ausgehend von Abbildung 25 liegen die Daten für den Artikel- und Materialstamm ausschliesslich in der C3 Datenbank.

5.5.7.1. Qualitative Analyse

Um eine **Qualitative Analyse** der Systeme durchzuführen gilt es die **Datenverantwortlichen** zu identifizieren und zu Interviewen. Es gibt meist nicht nur den technischen Datenverantwortlichen. Der fachliche Datenverantwortliche hat oft detailiertes Wissen über die Semantik der Daten, ihre Herkunft und ihre Verwendung. Der technische Datenverantwortliche hat dagegen eher systemseitige Information über die Daten wie Datenvolumen oder über welche Schnittstellen sie importiert und exportiert werden etc. Dieses Wissen über die Systeme und Datenobjekte der Datenverantwortlichen soll in einer Dokumentation festgehalten werden, da es für das zukünftige Datenqualitätsmanagement von grosser Bedeutung ist. Abbildung 30 zeigt eine Beispieldokumentation.

Pos	System	Datenobjekt	Standort	Art, Name	Verantworlich fachlich	Verantwortlich techn	Bemerkung
1	C3	Materialstamm	C3perma	permac3.Artikel	EK	IT-C3	Unterscheidung Material/Artikel nach Warenart 2 = Materialien. 1 = Artikel Produktklasse --> Sachmerkmal Stammmaterial + Breiten
2	C3	Produktionsartikel	C3perma	permac3.Artikel	30/40/50	IT-C3	Warenart 1 = Sortenvorlage + Varianten + Artikel
3	C3	Farben	C3perma	permac3.Artikel	FM	IT-C3	Warenart 3 = Farben, Farbdefinition in -->Farbe

Abbildung 30: Metadatendokumentation (eigene Darstellung)

Ein Ergebnis aus dieser ersten Analyse ist, dass der Einkaufsleiter monatlich die im letzten Monat angelegten Materialien auf ihre Verwendung hin überprüft. Eine erfreuliche Information ist, dass alle Datenobjekte klare fachliche (Data Owner) und technische Verantwortliche haben. Bei den Produktionsartikeln sind mehrere Verkaufsabteilungen zuständig, dies ist abhängig von der im Artikel festgelegten Warengruppe. Im Verkauf erfolgt im Gegensatz zum Einkauf keine Prüfung der neu oder bestehenden Artikel (= Sorte und Sortenvorlagen). Des Weiteren kam durch ein persönliches Interview zum Vorschein, auch wenn ein Mitarbeiter einen Artikel selbst angelegt hat, er sich nach dessen Erstellung für allfällige Korrekturen am Datenobjekt Artikel nicht unbedingt zuständig fühlt. Als übergeordneten Sponsor für die Stammdatenqualität konnten die Verkaufsabteilungsleiter gewonnen werden, die somit Datenkorrekturarbeiten als Vorgesetzte veranlassen können.

In der Qualitativen Analysephase werden vor allem einfache **Metadaten** erkundet. Wie in Kapitel 3 geschrieben sind sie für das Verständnis der Daten und für die Steuerung der Datenqualität essentiell. Quellen für Metadaten sind die zuvor identifizierten Datenverantwortlichen, Anwendungs- und bereits existierende firmeninterne Benutzerdokumenationen, technische Spezifikationen, Entity-Relationship Diagramme (Objektbeziehungsmodell des Datenbestandes) und weiteres mehr. Die gefundenen Metadaten werden inhaltich geprüft, ob sie mit der Wirklichkeit übereinstimmen, oder ob es Diskrepanzen gibt. Diese Metadaten und weitere Hinweise werden in einem Metadaten Repository oder Data Dictonary dokumentiert. In dieser Phase reicht es aus einen überblicksmässigen Eindruck von den Metadaten und ihrem Detailierungsgrad zu bekommen. Sie werden in den nächsten Analyseschritten weiter vervollständigt.

Pos	System	Datenobjekt	Version/Datum	Ablageort/Link	Art	Wer	Bemerkung
1	C3/Artikel	Material Datensatz	14_632_0176_001	permac3.dbo. Artikel	Datenobjekt	IT-C3	Theurer Datenmodell, neues Feld: LastQuantityUpdat
2	C3/Waren gruppen	Warengruppe	14_632_0176_001	permac3.dbo. SysWarengruppen	Datenobjekt	IT-C3	FK: [ProductionClasses] u WGProduktgruppe: def = 2
3	C3/WarenU ntergruppen	Warenunterg...	14_632_0176_001	permac3.dbo. SysWarenUntergru ppen	Datenobjekt	IT-C3	FK: [SysWarengruppen] ([WarenGruppe]) u WGProduktUnterGruppe: def ...

Abbildung 31: Metadatendokumentation (eigene Darstellung)Apel et al., 2010, S. 54)

Abbildung 31 zeigt die Dokumentation der Regeln bzw. Validierungs- und Auswahlfunktionen auf den Benutzermasken mit denen die Daten erfasst, manipuliert oder angezeigt werden. Es ist offensichtlich, dass bei einem beabsichtigten Einsatz von Data Profiling Software diese Dokumentation in ein Metadaten Repository zu übertragen ist.

5.5.7.2. Aufzeichnen von bestehenden Datenqualitätsproblemen

Die Datenanalyse soll nicht zu lange dauern und soll deshalb zeitlich limitiert sein. Deshalb ist es sinnvoll die Datennutzer über ihnen bekannte Datenprobleme zu befragen. Dann liegen

erste Anhaltspunkte vor, nach denen spätere Datenqualitätsmassnahmen besser geplant werden können und worauf der Fokus schon bei der quantitativen Analyse liegen kann. Persönliche Gespräche mit gleichzeitiger Demonstration sind am effizientesten, da dann z. B. Verständnis oder Benutzungsprobleme direkt sichbar werden.

5.5.7.3. Quantitative Datenanalyse

Bei der quantitativen Analyse der Datenhaltungssysteme in der Ist-Zustandsaufnahme geht es darum, möglichst detailiert alle Bedingungen, Geschäftsregeln, Abhängigkeiten zu dokumentieren, sowie den Datenbestand hinsichtlich der bekannten Datenqualitätsprobleme aus technischer Sicht zu untersuchen (Apel et al.,2010, S. 254). Dies geschieht durch ein detailiertes Data Profiling. Die Praxis hat gezeigt, dass zu Beginn eines Profilings, das das Bedienen der Applikation z. B. Erfassung, Manipulation und Weiterverarbeitung der Daten und die gleichzeitige Analyse was mit den dazugehörenden Daten auf Datenhaltungssystemebene passiert, für den Verständnisaufbau effizient ist. Fast alle Hersteller von relationalen Datenbanken wie Oracle, Microsoft etc. bieten für ihre Datenbanksysteme Zusatzsoftware für das Loggen der Datenbankaktivitäten. Bei richtiger Anwendung können die daraus gewonnenen Erkenntnisse aufschlussreich sein.

Da zu diesem Zeitpunkt schon einige Metadaten wie Geschäftsregeln, Abhängigkeiten usw. dokumentiert sind, kommen nun die Kapitel 4 beschriebenen Data Profiling Tools zum Einsatz. Diese Tools liefern verschiedenste Erkenntnisse über die Daten.

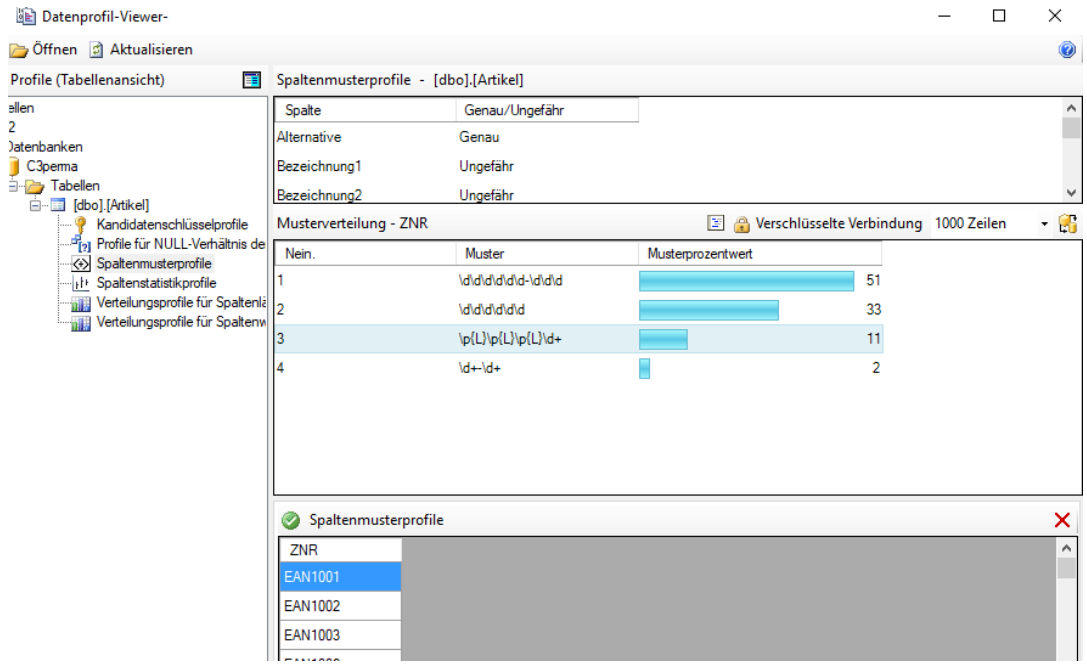


Abbildung 32: Musterverteilung von Attribut ZNR (Artikelnummer) mit MS SSIS (eigene Darstellung)

Wie im Kapitel 4.6 beschrieben werden die Grundinformationen über die Daten mit einer Attribut-, Datensatz und Tabellenanalyse erhoben. Diese Metadaten werden dem Metadaten Repository hinzugefügt. Abbildung 32 zeigt beispielhaft, was MS SSIS auf Attributebene (Mustererkennung in allen Werten einer Spalte) liefert. Die Tabellenanalyse dient dazu, funktionale Abhängigkeiten und Redundanzen zwischen den Tabellen zu erkennen. Wenn die Werkzeuge die schon dokumentierten Metadaten mitverarbeiten, kann neben der formalen Prüfung z. B. auf Vollständigkeit auch eine inhaltliche Prüfung z. B. auf Korrektheit der Daten durch deren Validierung an fachlichen Geschäftsregeln erfolgen. Beispielsweise kann eine Domänen Analyse aufgrund der in den Metadaten zulässigen Werte die Attributwerte validieren (Zulässige Werte innerhalb des Datentyp-Wertebereiches können aus einer Auswahlliste (Dropdown-Liste) aus den Eingabemasken stammen). Aber auch der umgekehrte Fall ist möglich, dass aus der Muster- und Wertehäufigkeit der Attributwerte neue Geschäftsregeln gefunden werden können.

Beispiel für eine Geschäftsregel (Business Rule): Artikel Nummer ZNR (char(10) constraint NOT NULL) hat immer die Länge 6, wenn das Attribut Variantenvorlage auf WAHR steht und die Warenart = SORTE ist. Nur numerische Werte und ein Bindestrich sind zulässig.

Im Gegensatz zu der technischen Validierungsregel: Datentyp Char 10: Das Datenattrbut speichert genau 10 alphanumerische Zeichen, egal ob nur 2 Zeichen eingegeben wurden.

Alle Analysearten aufzuzählen (z. B. Dublettenerkennung) würde den Rahmen dieser Arbeit sprengen.

Diese allgemeinen Prüfungen sollen durchgeführt werden um potenzielle Datenqualitätsprobleme zu entdecken, darüber hinaus sollen ganz gezielt die bereits von den Datennutzern aufgezeigten Datenqualitätsprobleme mit den Geschäftsregeln untersucht werden und Massnahmen für deren Minimierung entwickelt werden.

Schon länger bieten spezialisierte Unternehmen Data Profiling Dienste auch online an. Es lässt sich aber selber abschätzen wie gut die Prüfungsergebnisse ausfallen, wenn ihnen für die Datenprüfung die internen fachlichen Geschäftsregeln und das Datennutzerwissen nicht zur Verfügung stehen. Sie beschränken sich oft auf formale Validierungen. Trotzdem können schon diese Messergebnisse für andere technische Anforderungen wie der korrekten Indexierung und somit dem Performancetuning der Datenbank dienen.

5.5.8. Datenbereinigung

Dem Projektteam lag nun ein Bericht über die Qualität des Datenbestandes vor (siehe Anhang B). Für die einmalige, bzw. erstmalige Bereinigung müssen die Daten validiert und gefiltert werden. Die so identifizierten fehlerhaften Daten können in Tabellen oder Berichten möglichst in Kategorien zusammengefasst werden. Nur offensichtliche und kritische Fehler

können im Echtsystem bereinigt werden, ansonsten sollte ein Data Cleaning (Datenbereinigung) nie auf der Produktiv Datenbank erfolgen. Über uneindeutige Fehler wird im Team entschieden, wie mit ihnen verfahren wird (Löschen, bereinigen, belassen, kennzeichnen). Der betroffene Datennutzer sollte maximal um eine Einschätzung gefragt werden, weil er die Auswirkungen auf andere Geschäftsbereiche in der Regel nicht kennt. Die Bereinigung kann wo möglich automatisiert, oder durch die Mitarbeiter erfolgen. Vorgang und Verfahren der Validierung zeigt Abbildung 33.

Abgeleitete Massnahme des Projekt-Teams:

- Invalide Datensätze auflisten und den Abteilungen zur Korrektur übergeben.

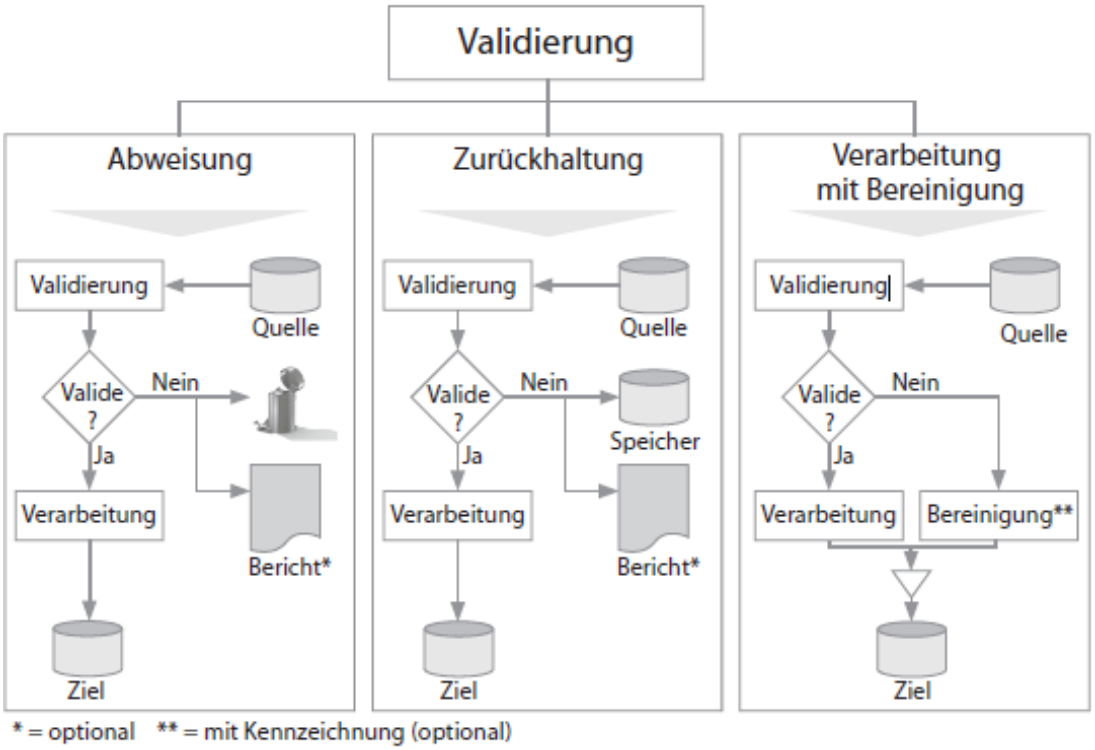


Abbildung 33: Verfahren zur Validierung und Filterung von Daten (Apel et al., 2010, S. 150)

5.5.9. Regelkreis für die nachhaltige Sicherstellung der Datenqualität - PDCA Zyklus

Ist die Datenbereinigung erfolgreich durchgeführt worden, ist der nächste entscheidene Schritt das erreichte Datenqualitätsniveau zu halten, bzw. nach den Anforderungen der Datennutzer auszurichten. Mit einem Prozess der den Schritten des Demingkreises oder PDCA Zykus folgt, kann durch entsprechende Planung, Durchführung und Überprüfung von datenqualitätshebenden Aktivitäten eine stetige Verbesserung der Datenqualität erreicht werden.

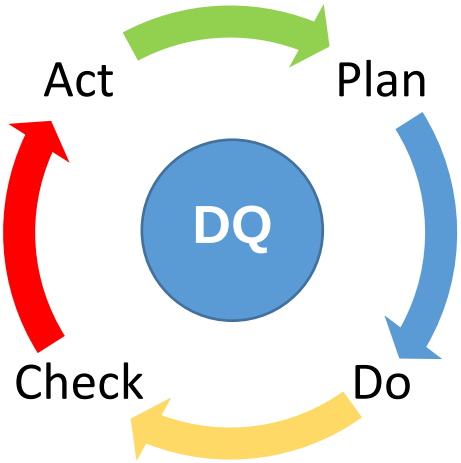


Abbildung 34: zyklische Verbesserung der Datenqualität mit PDCA (eigene Darstellung)

5.5.10. Geplante Massnahmen (Plan)

Um die Datenqualität zu verbessern, müssen für die gefundenen Datenprobleme entsprechende Lösungen erarbeitet werden und davon umsetzbare Massnahmen abgeleitet werden. Alle Massnahmen müssen unter Kosten-/Nutzen Aspekten beurteilt werden. Grössere oder fachbereichsübergreifende Massnahmen werden vom Datenqualitätsmanager koordiniert und überwacht. Massnahmen, die nur einen Fachbereich tangieren, werden von den jeweiligen Data Stewards des Fachbereiches initiiert und überwacht und auf den Datenqualitäts-Jour-Fixe Meetings mit den anderen Data Stewards abgestimmt.

Einmalige Datenqualitäts Massnahmen für Permapack:

Das Projekt Team entschloss sich die IT-Abteilung mit der Umsetzung der Pflichtfeldervalidierung der Artikelmasken zu beauftragen (siehe Abbildung 35). Die neu definierten Geschäftsregeln (Anhang B) müssen davor noch auf die folgenden drei Kriterien hin geprüft werden:

Geschäftskonform: Die Regel ist technisch umsetzbar und bewirkt im operativen Tagesgeschäft eine Verbesserung.

Geschäftsrelevant: Die Regel ist an den Unternehmenszielen ausgerichtet und fördert langfristige Ziele.

Messbarkeit: Die Regel ist möglichst formal-technisch messbar.

Das Team legte weiterfest, dass die nach dieser Kriterienprüfung noch verbliebenen Geschäftsregeln so weit möglich auf Datenbankebene vom C3 Verantwortlichen implementiert werden sollten, sofern die vom Hersteller umgesetzte Applikationslogik das zulässt.

Regelmässige Datenqualitäts Aufgaben und Systemoptimierungen für Permapack:

Das Projekt Team legte fest, dass mindestens halbjährlich die Geschäftsbereichsregeln auf neue Anforderungen aus den Fachbereichen hin untersucht werden.

Für die Ursachenanalyse wurde ein PDCA eingerichtet und die Verantwortlichen für die Durchführung definiert.

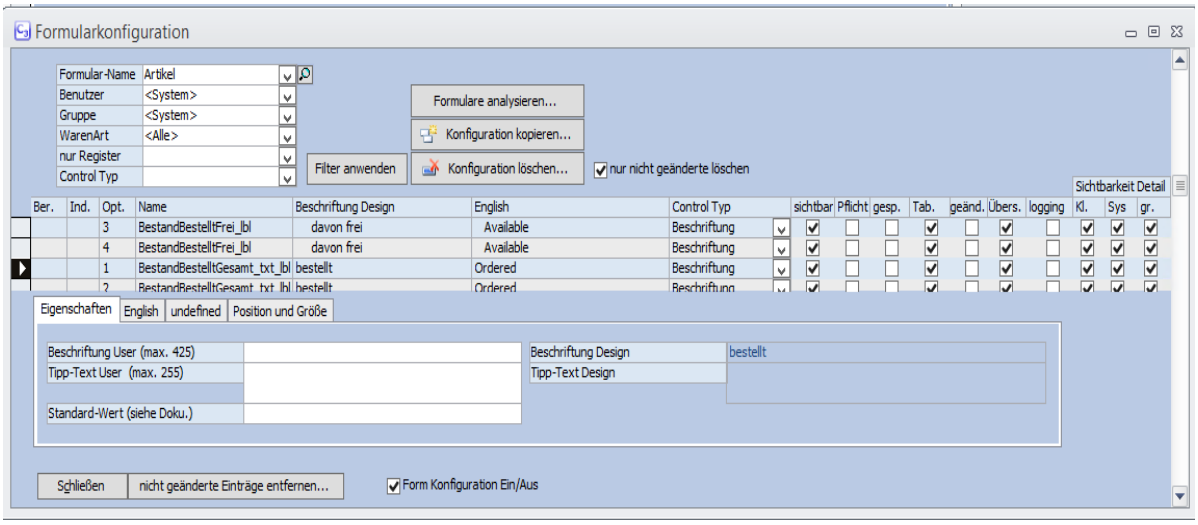


Abbildung 35: Metadaten eines Formulares im C3: Pflichtfelder, Default-Werte (eigene Darstellung)

Abbildung 35 zeigt wie bei Formularen, hier beispielsweise vom C3, mit Metadaten die Datenfelder einer Maske beeinflusst und konfiguriert werden können (Sichtbarkeit, Pflichtfeld, Defaultwerte, Bezeichnung etc.). Die meisten ERP Systeme bieten diese Möglichkeiten um die Masken auf den jeweiligen Kunden anzupassen.

5.5.10.1. Data Quality Monitoring in der Act Phase

Über definierte Kennzahlen oder erneutes Data Profiling muss die Datenqualität laufend überwacht werden um frühzeitig anhand Trends der Kennzahlenwerte neue Datenqualitätsprobleme zu erkennen und entsprechende Steuerungsmassnahmen (Plan) beim Abweichen von den definierten Soll-Zielwerten einzuleiten. Andererseits sollen sie die Auswirkungen (Check) von geplanten Datenqualitätsmassnahmen sichtbar machen. Auch hier gilt, es lassen sich nur jene Dinge verbessern, die auch gemessen werden können. Die Korrektheit der Messverfahren ist deshalb entscheidend. In der Literatur existieren viele Ansätze zur Datenqualitätsmessung, die sich vor allem in den zugrundeliegenden Messverfahren unterscheiden (Heinrich & Klier, 2008, S. 51) Es wird zwischen quantitativen und qualitativen Messverfahren unterschieden. Die quantitativen Messverfahren beruhen auf meist automatisiert berechneten Kennzahlen.

5.5.10.2. Data Quality Metriken

Data Quality (DQ) Metriken (Datenqualitätsmasse) sind die Grundlage von Kennzahlen mit denen die Datenqualitätsdimensionen geprüft werden sollen. Sie lassen sich in die drei Arten

Formal-technische in Form von einfachen Abfragen über Datensatzanzahl, Fehleranzahl

in **Inhaltliche** bei denen Geschäftsregeln überprüft werden und als Ergebnisse Häufigkeitsverteilungen der Werte liefern und

Qualitative bei denen z. B. eine visuelle Prüfungen durch Anwender erfolgt, unterteilen.

Damit Metriken eine Aussagekraft besitzen müssen sie gewissen Ansprüchen genügen (vgl. Apel et al., 2010, S. 87).

- Damit die Ergebnisse vergleichbar bleiben, müssen sie normiert werden (z. B. - absolute auf prozentuale Werte umrechnen).
- Die Werte sollten für eine zeitliche Vergleichbarkeit zahlenmässig gut erfassbar sein
- Sie sollten auf die jeweilige Anwendung hin sensibilisiert sein und eine Aussagekraft besitzen.
- Sie müssen fachlich interpretier- und reproduzierbar sein.

5.5.10.3. Data Quality Kennzahlen

Datenqualitätsmanagement kann nur dann wirksam sein, wenn der Soll-Zielwert der zu steuernden Datenqualitätskennzahl der Daten festgelegt ist und über geeignete Metriken entsprechende Datenqualitätskennzahlen gemessen werden.

Das Team entschied beispielhaft die Datenqualitätsdimension **Vollständigkeit** mit einer Kennzahl zu messen. Diese Kennzahl soll über die automatisierte Erfassung folgender Metriken berechnet werden.

Dimension	Kennzahl	Metrik
Vollständigkeit	Attributvollständigkeit	$1 - \frac{Anzahl\ Nullwerte\ eines\ Attributs}{Gesamtanzahl\ Attributwerte}$
	Datenobjekt-vollständigkeit	$\frac{\sum(Attributvollständigkeit\ pro\ Datenobjekt)}{Anzahl\ Datenobjekte}$

Tabelle 3: Kennzahlendefinition (eigene Darstellung)

Das Projektteam weiss, dass es weitere Kennzahlen aufbauend auf den umgesetzten Geschäftsregeln definieren muss, um daraus einen einheitlichen Datenqualitätsindex zu bilden. Ein normierter Datenqualitätsindex verdichtet alle berechneten Kennzahlen auf einen

Wert, damit ist die Entwicklung der Datenqualität im zeitlichen Verlauf besser vergleichbar. Die algorithmische Berechnung des Datenqualitätsindex muss mit der Unternehmensleitung abgestimmt sein. Das Projektteam beauftragte den C3 Verantwortlichen die definierte Kennzahl in einem Dashboard im ERP System zu implementieren. Für die Darstellung der Kennzahlen gibt es darauf spezialisierte Software, die ähnlich der Abbildung 36 die Kennzahlen in einem Dashboard im ERP System oder als Bericht aufbereitet zur Verfügung stellen.

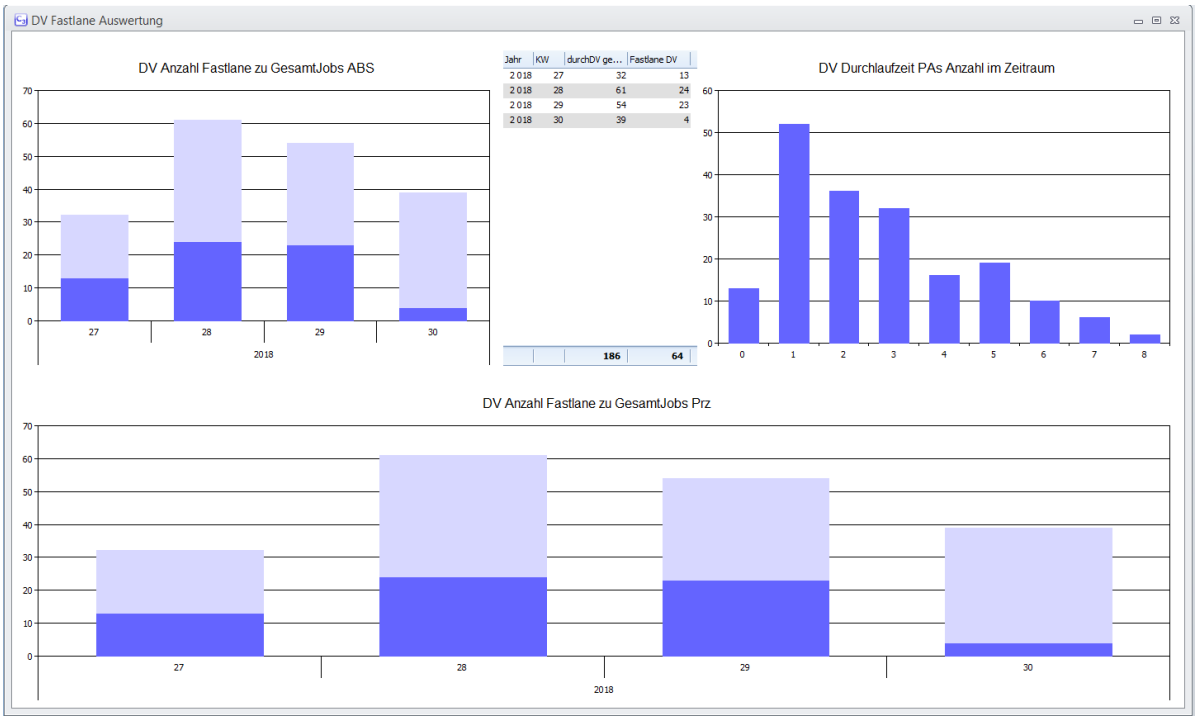


Abbildung 36: Produktivitätskennzahlen einer Fachabteilung bei Permapack (eigene Darstellung)

Dieser Regelkreis ist wesentlich damit das Datenqualitätsmanagement in einem Unternehmen als Prozess ausgeführt wird und gerade das Data Quality Monitoring gewährleistet eine grössere Transparenz und eine bessere Kosten-/Nutzen Abschätzung über die durchgeführten Massnahmen. Dazu bedarf es klarer Ziele (Plan) und konsistenter nachvollziehbarer periodisch erfasster Kennzahlen.

5.6. Erkenntnisse aus der Fallstudie bei Permapack

Die systematische Auseinandersetzung mit der Datenqualität hat bei den Datenverantwortlichen und bei den Projektmitgliedern neues Bewusstsein über die Wichtigkeit von Datenqualitätsmanagement als unternehmensweiten Prozess geschaffen. Datenqualität ist Grundvoraussetzung, dass Geschäftsprozesse die Compliance und Geschäftsregeln erfüllen können. Die Definition von Kennzahl-Sollzielwerten für die Datenqualität muss mit den betroffenen Fachabteilungsleitern abgestimmt sein, da sie die personelle Kompetenz haben Datenqualität bei ihren Mitarbeitern über Zielvereinbarungen

einzufordern. Aufkommende Datenqualitätsprobleme müssen aktiv angegangen werden und dürfen nicht wieder vertuscht werden, bis schlussendlich ganze Geschäftsprozesse stocken.

Die wichtigste Erkenntnis für Permapack ist die, dass Datenqualität kein Selbstzweck ist und das Bewusstmachen wie wichtig Datenqualität für das reibungslose funktionieren aller Geschäftsprozesse ist, bei den Mitarbeitern durch individuelle Zielvereinbarungen geschaffen wurde.

6. Zusammenfassung und Fazit

Dieses Kapitel fasst die Ergebnisse der Arbeit zusammen, diskutiert die Einschränkungen, und gibt Handlungsempfehlungen für die Umsetzung von Datenqualitätsmanagement in einem KMU. Es schliesst mit einem Resümee

6.1. Zusammenfassung

Mit dieser Arbeit wurden drei Ziele verfolgt:

- aktueller Forschungsstand zu Datenqualitätsmanagement in der Literatur
- Ausarbeitung eines Vorgehensmodells zur Umsetzung eines Datenqualitätsmanagement in einer KMU
- Auswahl und Verwendung von Software für den Prozessschritt Data Profiling.

Die Literaturrecherche hat gezeigt, dass Datenqualität weit mehr als die Bereinigung von Daten durch die Informatik Abteilung ist. Erst durch Sicherungs- und Vermeidungsmassnahmen in den fehlerhaften Prozessen kann die Datenqualität nachhaltig gesteuert werden. Nur durch die Bewusstmachung was Datenqualität zu einem bestimmten Zeitpunkt und genau in einem Geschäftsprozessschritt bedeutet, macht die Auswirkungen von Fehlern transparenter. Massnahmen um die Datenqualität zu heben müssen durch die Komplexität vieler Geschäftsprozesse wohl durchdacht und mit den Datenverantwortlichen abgestimmt werden. Datenqualität ist von vielen verschiedenen Bausteinen und Prozessschritten abhängig, und mit jeder Schnittstelle steigt das Risiko Fehler zu provozieren. Dazu braucht es Strukturen einer Aufbau- und Ablauforganisation, sowie passende Vorgehensmodelle um die Massnahmen zu koordinieren.

Ein entworfenes Vorgehensmodell konnte in einer Fallstudie mit der Permapack AG auf seine Praxistauglichkeit hin evaluiert werden. Die Verankerung eines Datenqualitätsmanagements über einen Regelkreis befindet sich bei Permapack zum Zeitpunkt der Erstellung dieser Arbeit, erst in der Aufbauphase und es liegen bis dato noch keine verwertbaren Ergebnisse vor. Die Erstbereinigung der fehlerhaften Daten wurde durch die meisten Fachabteilungen bereits umgesetzt. Die ersten manuellen Messungen haben gezeigt, dass die Datenqualität in den definierten Datenqualitätsdimensionen Vollständigkeit und Fehlerfreiheit von vormals jeweils unter 90 % auf jeweils mindestens 97.1 % angehoben werden konnte. Das Vorgehensmodell hat sich in der Durchführung wie Datenqualitätsmanagement in einer KMU umgesetzt werden kann bewährt. Der Regelkreis sollte durch flankierende Massnahmen wie der Dokumentaton z.B. der Prozessschrite, durch Analyseprotokolle und Checklisten verfeinert werden. Zielbeschreibungen eines Prozessschrittes helfen das Verständnis für die Schritte zu heben.

Ein weiteres Ergebnis der Arbeit ist die Evaluierung von Data Quality Software, die in der Data Profiling Phase eingesetzt werden kann. Es hat sich gezeigt, dass vor allem Data Quality Software, die vom Hersteller der Datenbanksoftware stammt, durch die meist kürzere Einarbeitungszeit und nahtlosere Integration in die Datenbanksoftwae, schnelle befriedigende Analyseergebnisse liefert.

6.2. Einschränkungen

Das entworfene Vorgehensmodell berücksichtigt nicht alle Aspekte, die in der Literatur für ein Datenqualitätsmanagementsystem behandelt werden. Immerhin konnte es sich in der Praxis einmal bewähren. Zwischenbetriebliches oder standortübergreifendes Datenqualitätsmanagement in grösseren Unternehmungen benötigt mehr Datenverantwortliche und einen grösseren Umsetzungsrahmen und mehr Koordination.

6.3. Handlungsempfehlung für ein systematisches Datenqualitätsmanagement

Als wichtigste Massnahme ist das Bewusstsein für Datenqualität bei den Mitarbeitern zu schaffen und dass jeder einzelne eine aktive Rolle zur Verbesserung der Datenqualität übernehmen muss. Diese Sensibilisierung kann auch durch konkrete Ziele in Bezug auf Datenqualität erfolgen. Denn ohne entsprechende Datenqualität werden die Geschäftsprozesse nicht so reibungslos laufen, wie sie könnten.

Das Verständnis fördern, dass die Information über Datenfehler an die Datenverantwortlichen weitergeleitet werden. Gerade bei kleineren Betrieben die Durchsetzung der Datenqualitäts-Regelung verlangen, dass auch wenn im Geschäftsprozess fehlerhafte Daten aktuell vorliegen, nicht am „System“ vorbeigearbeitet wird.

Damit Kosten-/Nutzen von Massnahmen abgeschätzt werden können, muss zuvor der Ist-Zustand der Datenqualität erhoben werden. Und die Massnahmen sind auf den Ist-Zustand abzustimmen und zu priorisieren. Mit den Massnahmen beginnen, die den grössten Benefit versprechen.

Datenverantwortliche für den Datenbestand bestimmen und mit ihnen die Relevanz und Regeln für die Qualität der Daten definieren. Erst dann Datenmodelle analysieren, um weitere Regeln und Bedingungen abzuleiten und Redundanzen aufzudecken.

Mit den gefundenen Regeln und Bedingungen als Validierungsgrundlage eine Datenbereinigung durchführen und gleich im Anschluss Massnahmen wie Validierungsfunktionen auf Eingabemasken definieren, um bei der Datenerfassung die

Fehler zu minimieren. Einfach abrufbare Dokumentation z. B. durch Hinweise auf den Datenfeldern in der Maske erstellen und die Mitarbeiter immer wieder schulen.

6.4. Resümee

Die Forschungsfrage zu entwickeln war die grösste Herausforderung für die vorliegende Arbeit. Der Fokus zu Beginn der Arbeit lag auf Qualität im Stammdatenmanagement und der Verbesserung der Qualität durch automatische Datenanalyse über Machine Learning. Dieses Thema ist weg vom Inhalt dieses Masterstudienganges.

Der Schwerpunkt der Recherche verlagerte sich darauf auf das Thema Stammdatenmanagement. Die Literaturrecherche hat gezeigt, dass dazu bereits viele wissenschaftliche Konzepte ausgearbeitet wurden. Leider konnten wenige konkrete Umsetzungen der Konzepte in die Praxis bei KMUs recherchiert werden.

Ziel war aber eine machbare und praxisrelevante Vorgehensweise zu entwickeln mit dem Datenqualitätsmanagement auch tatsächlich in KMUs realisiert werden können.

Literaturverzeichnis

Anderegg, B. (2000). *IT-Prozessmanagement effizient und verständlich: Projekterfolg und Qualitätsverbesserung in 24 Schritten.* Braunschweig: Vieweg.

Apel, D., Behme, W., Eberlein, R., & Merighi, C. (2010). *Datenqualität erfolgreich steuern. Praxislösungen für Business Intelligenceprojekte.* München: Carl Hanser Verlag.

ap-verlag. (20. 03 2018). Von neue einsatzgebiete fuer business intelligence und datenmanagement: http://ap-verlag.de/neue-einsatzgebiete-fuer-business-intelligence-und-datenmanagement/17257/ abgerufen

barc. (21. 03 2018). Von bi trend monitor 16: http://barc.de/bi-trend-monitor-16 abgerufen

Beims, M. (2012). *IT-Service Management mit ITIL - ITIL Edition 2011, ISO 20000:2011.* München: Carl Hanser.

Berson, A., & Dubov, L. (2007). *Master Data Management and Customer Data Integration for a Global Enterprise.* New York : McGraw-Hill.

Blessing, L. T., & Chakrabarti, A. (2009). *DRM, a design research methodology.* Berlin: Springer.

commvault. (21. 04 2018). Von using a data management platform: https://www.commvault.com/de-de/solutions/by-topic/using-a-data-management-platform abgerufen

Dippold, R., Meier, A., & Schnider, W. (2005). *Unternehmensweites Datenmanagement.* Wiesbaden: Vieweg,.

Eisenhardt, K. M. (4 1989). Building Theories form Case Study Research. *Academy Of Management Review*, S. 532-550.

English, L. P. (1999). *Improving data warehouse and business information quality: Methods for reducing costs and increasing profits.* New York: Wiley.

eoda. (20. 7 2018). *Whitepaper Datenqualität*. Von https://www.eoda.de/files/Use_Case_Seiten/Whitepaper/Whitepaper_Datenqualitaet.pdf abgerufen

Garfield, E. (3 1977). Proposal for a new profession: scientific reviewer. *Essays of an Information Scientist*, S. 84-87.

Garvin, D. A. (1984). What does Product Quality Really Mean. *Sloan Management Review*(Vol. 26), 25-43.

Gebauer, M., & Windheuser, U. (2008). Strukturierte Datenanalyse, Profiling und Geschäftsregeln. In K. Hildebrand, M. Gebauer, & H. Hinrichs, *Daten- und*

Informationsqualität - Auf dem Weg zur Information Excellence (S. 88-101). Wiesbaden: Vieweg+Teubner.

Hankins, L. (2 1999). Cleansing Looms Important in Data Warehouse Efforts. *SIGNAL AFCEA's International Journal*.

Hansen, H. R., & Neumann, G. (2009). *Wirtschaftsinformatik 1. Grundlagen und Anwendungen.* Stuttgart: Lucius & Lucius.

Heinrich, B., & Klier, M. (2008). Datenqualitätsmetriken für ein ökonomisch orientiertes Qualitätsmanagement. In K. Hildebrand, M. Gebauer, & H. e. Hinrichs, *Daten- und Informationsqualität. Auf dem Weg zur Information Excellence* (S. 49-67). Wiesbaden: Vieweg+Teubner.

Helfert, M. (2002). *Proaktives Datenqualitätsmanagement in Data Warehouse Systemen: Qualitätsplanung und Qualitätslenkung.* Berlin: Dissertation.

Hildebrand, K., & Otto, B. (2011). *Stammdatenmanagement, HMD 279.* Heidelberg : dpunkt Verlag.

Hildebrand, K., Gebauer, M., Hinrichs, H., & Mielke, M. (2018). *Daten- und Informationsqualität - Auf dem Weg zur Information Excellence.* Berling Heidelberg: Springer.

ISO/IEC. (1993). *Information technology: Vocabulary Part 1: Fundamental terms.* Genf: ISO/IEC 2382-1.

it bei lidl rockt das projekt elwis. (23. 7 2018). Von dice.com: http://de.dice.com/nachrichten/it-bei-lidl-rockt-das-projekt-elwis/ abgerufen

Juran, J. M. (1988). *Juran on Quality by Design: The New Steps for Planning Quality into Goods and Services.* New York.

Kuhlen, R., Semar, W., & Strauch, D. (2013). *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis.* Berlin, Boston: De Gruyter Saur.

Legner, C., & Otto, B. (2009). *Stammdatenmanagement Whitepaper.* St. Gallen: HSG.

Levitin, A. V., & Redman, T. C. (1 1998). Data as a Resource: Properties, Implications, and Prescriptions 40. *Sloan Management Review* , S. 89-101.

Loshin, D. (2001). *Enterprise knowledge management: the data quality approach.* San Diego, California: Morgan Kaufmann.

Ofner, M. H. (2013). *Datenqualitätsmanagement aus Prozessperspektive - Methoden und Modelle.* St. Gallen: Logos Verlag Berlin GmbH, Dissertation.

Ortner, E. (1991). *Informationsmanagement.* Berlin: Springer-Verlag.

Otto, B., & Hüner, K. (2009). *Funktionsarchitektur für unternehmensweites Stammdatenmanagement.* St. Gallen: Universität St. Gallen.

Otto, B., & Österle, H. (2016). *Corporate Data Quality - Voraussetzung erfolgreicher Geschäftsmodelle.* Berlin: Springer.

Paschke-Hess, A. (20. 05 2018). *Stammdatenmanagement: Eine Frage der Organisation, nicht alleine der Technologie*. Von SAP News: https://news.sap.com/germany/stammdatenmanagement-eine-frage-der-organisation-nicht-alleine-der-technologie/ abgerufen

Pipino, L., Lee, Y., & Wang, R. (4 2002). Data quality assessment. *Communications of the ACM 45* , S. 211–218.

Rautenstrauch, C., & Schulze, T. (2003). *Informatik für Wirtschaftswissenschaftler und Wirtschaftsinformatiker: Mit 40 Tabellen.* Berlin: Springer.

Redman, T. C. (1996). *Data Quality for the Information Age.* Boston: Artech House.

Scheibmayer, M., Naß, E., & Birkmeier, M. (30. 05 2018). *Stammdatenmanagement - White Paper*. Von RWTH Aachen: https://www.fir.rwth-aachen.de/publikationen/whitepaper abgerufen

Schemm, J. W. (2009). *Zwischenbetriebliches Stammdatenmanagement. Lösungen für die Datensynchronisation zwischen Handel und Konsumgüterindustrie.* Berlin Heidelberg: Springer.

Schmidt, A. (2010). *Entwicklung einer Methode zur Stammdatenintegration.* St. Gallen: Dissertation.

Schuh, G., & Stich, V. (2014). *Stammdatenmanagement in der produzierenden Industrie.* RWTH Aachen: FIR e. V.; knapp:consult.

Schumacher, J. C. (2011). *Prozess- und Data Governance im industriellen Anlagenmanagement.* Karlsruhe: Dissertation.

Siau, K. R. (2007). Evaluation techniques for systems analysis and design modelling methods - a review and comparative analysis. *Information Systems Journal*, S. 455-474.

Spath, D., Weisbecker, A. (., & Kokemüller, J. (2009). *Stammdaten-Managementsysteme.* Stuttgart: Fraunhofer Verlag. Von Spath, D.; Weisbecker, A. (Hrsg.); Kokemüller, J.: Stammdaten-Managementsysteme 2009. Stuttgart: Fraunhofer Verlag. 2009 abgerufen

Stahlknecht, P., & Hasenkamp, U. (1999). *Einführung in die Wirtschaftsinformatik.* Berlin: Springer.

Stöckler. (22. 04 2018). *sol4bus*. Von datenmanagement strategie und sap hana: http://www.sol4bus.de/themen/enterprise-resource-planning/artikel/datenmanagement-strategie-und-sap-hana/ abgerufen

Suchanek, P. D. (02. 07 2018). *Wirtschaftslexikon Gabler*. Von https://wirtschaftslexikon.gabler.de/definition/vertrauen-50461 abgerufen

Susman, G. I., & Evered, R. D. (1987). An Assessment of the Scientific Merits of Action Research. *Administrative Science Quarterly*, 582-603.

The Data Warehousing Institute. (1 2015). Von Data quality: does poor data quality significantly impact the effectiveness of data analysis and other business intelligence activities?: https://www.researchgate.net/publication/272565867_Data_quality_does_poor_data_quality_significantly_impact_the_effectiveness_of_data_analysis_and_other_business_intelligence_activities abgerufen

Tiemeyer, E. (2015). *Handbuch IT-Management - Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis.* München: Carl Hanser Verlag GmbH & Co. KG.

Wang, R. Y., & Strong, D. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers . *Journal of Management Information Systems*, 5-33.

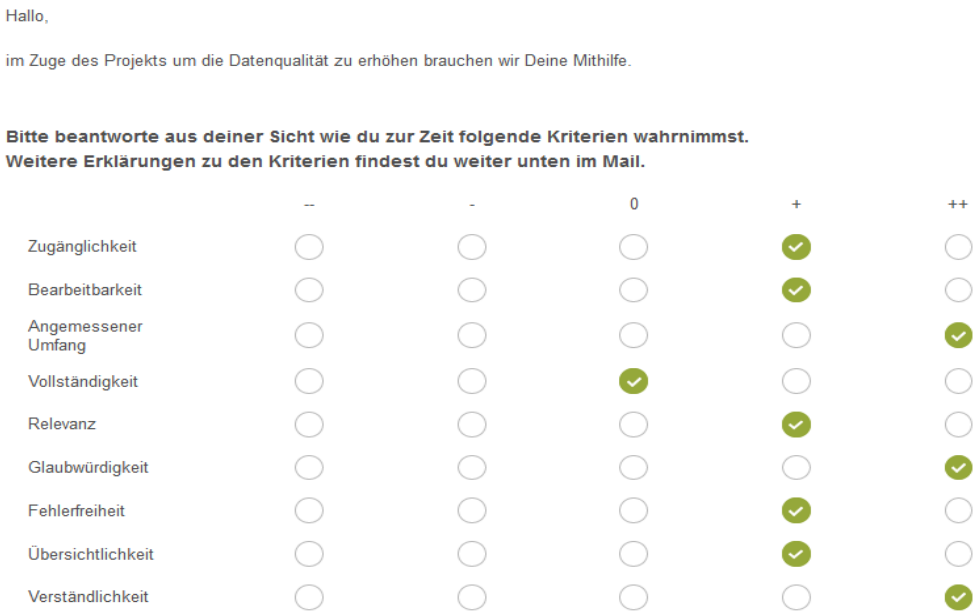
Wedekind, H. (2001). Bestandsdaten, Bewegungsdaten, Stammdaten Lexikon der Wirtschaftsinformatik,. In P. Mertens, A. Back, J. Becker, W. König, H. Krallmann, B. Rieger, . . . H. Wedekind, *Lexikon der Wirtschaftsinformatik* (S. 72-80). Berlin: Springer.

Würthele, V. G. (2003). *Datenqualitätsmetrik für Informationsprozesse.* Zürich: ETH Disseration.

Yin, R. K. (2002). *Case study research: design and methods.* Thousand Oaks: Sage Publications.

Anhang A

Umfrage Datenqualität C3 Nutzer


Hallo,

im Zuge des Projekts um die Datenqualität zu erhöhen brauchen wir Deine Mithilfe.

Bitte beantworte aus deiner Sicht wie du zur Zeit folgende Kriterien wahrnimmst.
Weitere Erklärungen zu den Kriterien findest du weiter unten im Mail.

	--	-	0	+	++
Zugänglichkeit	○	○	○	✓	○
Bearbeitbarkeit	○	○	○	✓	○
Angemessener Umfang	○	○	○	○	✓
Vollständigkeit	○	○	✓	○	○
Relevanz	○	○	○	✓	○
Glaubwürdigkeit	○	○	○	○	✓
Fehlerfreiheit	○	○	○	✓	○
Übersichtlichkeit	○	○	○	✓	○
Verständlichkeit	○	○	○	○	✓

Abbildung 37: Online Umfrage (eigene Darstellung)

Auswertung der C3 Nutzer Umfrage (Stichprobenumfang n=31):

	Legende: 1=ungenügend, 5=sehrgut			A=Verkauf,B=Einkauf,C=Arbeitsvorbereitung					
Kategorie	**System**	**System**	**Nutzung**	**Nutzung**	**Nutzung**	**Inhalt**	**Inhalt**	**Darstellung**	**Darstellung**
Datenqualitäts-dimension	**Zugänglichkeit**	**Bearbeitbarkeit**	**Angemessener Umfang**	**Vollständigkeit**	**Relevanz**	**Glaubwürdigkeit**	**Fehlerfreiheit**	**Übersichtlich**	**Verständlich**
Teilnehmer	4.7	4.5	4.1	2.9	3.8	4.1	2.4	4.1	4.7
A1	5	4	3	3	4	4	2	4	5
A2	4	4	3	3	4	4	2	4	5
A3	4	4	3	3	4	4	2	4	5
A4	4	4	3	3	4	4	4	4	5
A5	4	4	3	3	4	4	2	4	5
A6	5	5	4	4	5	3	3	5	5
A7	5	4	3	3	4	4	2	4	3
A8	5	5	4	3	4	4	2	4	5
A9	5	5	5	3	5	5	3	4	5
A10	5	5	5	5	4	5	4	4	5
A11	5	5	5	5	4	5	2	4	4
A12	5	5	5	3	5	5	2	5	5
A13	4	4	5	4	4	5	4	4	5
A14	5	4	5	3	4	4	2	5	5
A15	5	3	3	3	5	5	2	4	5
A16	5	4	5	3	4	5	2	4	5
A17	5	4	5	2	5	4	2	4	3
A18	4	5	3	2	-	3	4	4	5
A19	5	5	4	3	4	4	2	4	5
A20	5	4	4	2	4	5	2	4	5
A21	5	5	4	2	4	4	2	4	5
A22	5	5	5	3	4	4	2	4	5
B1	5	5	4	2	2	4	2	4	5
B2	2	3	2	1	1	2	2	4	3
B3	4	4	-	3	3	4	-	4	5
B4	5	5	3	2	2	4	3	4	5
B5	5	5	4	3	2	3	2	4	5
C1	5	5	5	3	2	4	2	4	3
C2	5	5	5	3	4	3	3	4	5
C3	5	5	5	3	4	4	2	4	5
C4	5	5	5	3	4	4	2	4	5

Tabelle 4: Antworten der Umfrage (eigene Darstellung)

Anhang B

Stammdatenqualität C3 Permapack DQM Initial

Artikel und Materialstamm

1) Ausgangslage

Nach der Einführung der neuen Business Intelligence Anwendung QlikView mit der Auswertung C3 Daten wurden vermehrt Datenqualitätsprobleme wie doppelte Artikel bei gleichen Kunden, fehlende Angaben bei Artikeln und Material etc. sichtbar. Weiter treten immer wiederkehrend Probleme bei der Dateneingabe von Artikeln und bei der Materialanlage im C3 auf. Aus diesem Grund soll die Datenqualität bei den Stammdaten im C3 geprüft und allenfalls korrigiert werden. Die Anwendungsmasken im C3 sollen auf Schwachstellen hin untersucht werden und Verbesserungsvorschläge mit dem Hersteller Theurer.com besprochen und umgesetzt werden. Einerseits werden Datenbankseitig zu implementierende Validierungen und Restriktionen wie Geschäftsregeln gesucht, andererseits in der C3 Applikation bei der Dateneingabe Validierungen durchzuführen. Um Validierungen zu ermöglichen, muss der Datenbestand auf Fehler analysiert werden. Diese müssen vorgängig bereinigt werden, ansonsten kann es zu Laufzeitfehler kommen. Um eine Validierung zu ermöglichen müssen folgende Punkte analysiert werden.

1. Sind die Daten für den Prozess relevant (Eingabe/Erfassung notwendig)?
2. In welchen Geschäftsprozessen werden die Daten definiert oder gelesen?
3. Welche Datenwerte und Datenvarianten sind möglich/erlaubt?

2) Datenanalyse

a) Relevanz und Vollständigkeit

Sämtliche Artikel und Materialstammdatenfelder werden von den Fachabteilungen Einkauf, Verkauf, Arbeitsvorbereitung und Produktion auf deren Relevanz für die eigenen Prozesse analysiert und in einer Metadatenbeschreibung dokumentiert. Ebenso sollen alle Abhängigkeiten, Werteinschränkungen und Geschäftsregeln erfasst werden. Mit dieser Analyse lässt sich bestimmen, wo im Prozess welche Daten erzeugt werden und welche Geschäftsprozesse die Daten nutzen. Weiter werden sämtliche Artikelstammdatenfelder anhand der persistierten Werte analysiert, dies hilft um Muster und Eingabefehler in den Felder zu erkennen und ob und wie sie bewirtschaftet werden. Die Relevanz kann in folgende neun Kategorien (Tabelle 5) unterteilt werden:

	Typ	Beschreibung	Massnahme
1	Pflicht	Die Daten werden im weiteren Prozess zwingend benötigt	Dateneingabe muss im Prozess beschrieben werden
2	Optional	Die Daten haben einen Informationswert	Informationen, welche im Prozess nicht zwingend benötigt werden
3	Fix Wert	Die Daten haben immer den gleichen Inhalt	Der Wert wird hinterlegt, das Feld wird disabled oder ausgeblendet.
4	Leer	Die Daten werden nie gefüllt	Irrelevante Daten. Das Feld kann ausgeblendet werden
5	Bedingung	Die Daten nehmen den Typ 1 – 7 anhand einer Bedingung an.	Validierung erfolgt durch Definition von Geschäftsregeln
6	Systemdaten	Die Daten werden vom System generiert	Der Wert wird vom System vergeben, das Feld wird disabled.
7	Read-only	Die Daten können nicht direkt bearbeitet werden	Der Wert wird zur Information angezeigt
8	Validierung	Die Dateneingabe ist Pflicht und wird vom System überprüft	Dateneingabe wird validiert. Speicherung ohne Werte nicht möglich.
9	Sollzustand	Die Standardwerte werden noch hintelegt	Programmierung erforderlich

Tabelle 5: Datenrelevanzfestlegung (eigene Darstellung)

Die Daten werden auf ihre Vollständigkeit überprüft, indem analysiert wird ob die Attribute mit Werten (aus den Metadatenregeln) belegt sind. Die Datenfelder werden mit ihrem jeweiligen Datentyp (Tabelle 6) ausgezeichnet, dies ist hilfreich für allfällige Validierungen.

	Typ	Beschreibung	Feldinhalt
D	Dropdown	Ein Wert kann aus einer Liste selektiert werden	Leer oder Wert aus Liste
T	Freitext	Freitext	Leer oder Alphanumerischer Text
C	Checkbox	Entscheidungsfeld	Ja oder nein
P	Bild	Bilddatei	Alle gängigen Bildformate
Dt	Datum	Datumseingabe	Leer oder Datum
N	Nummer	Numerischer Wert	0 – n

Tabelle 6: Datentypfestlegung (eigene Darstellung)

Die Daten werden auf ihre Vollständigkeit überprüft, indem analysiert wird ob die Attribute mit Werten (aus den Metadatenregeln) belegt sind. Die Datenfelder werden mit ihrem jeweiligen Datentyp ausgezeichnet, dies ist hilfreich für allfällige Validierungen.

In den Metadaten werden auch die Datenmanipulations-Rechte pro Benutzergruppe (Einkauf, Verkauf, Arbeitsvorbereitung (AVOR) und Produktion) festgehalten. Die Ausprägungen sind in Tabelle 7 dargestellt.

Lesen	Erstellen	Ändern		Stufe	Beschreibung
			N	Keine	Benutzer dürfen Datensätze weder erstellen, noch lesen oder ändern
			R	Lesen	Benutzer können Datensätze nur lesen
			C	Erstellen	Benutzer können Datensätze lesen und erstellen
			E	Ändern	Benutzer können Datensätze erstellen und/oder bearbeiten sowie löschen

Tabelle 7: Datenmanipulationsrechte (eigene Darstellung)

In der Spalte Std-Wert ist der Wert erfasst, welcher beim Anlegen eines Datensatzes im Feld vordefiniert ist. Ist das Feld rot markiert, entspricht der Wert dem Soll-Zustand.

b) Artikelstammdaten Metadatendokumentation

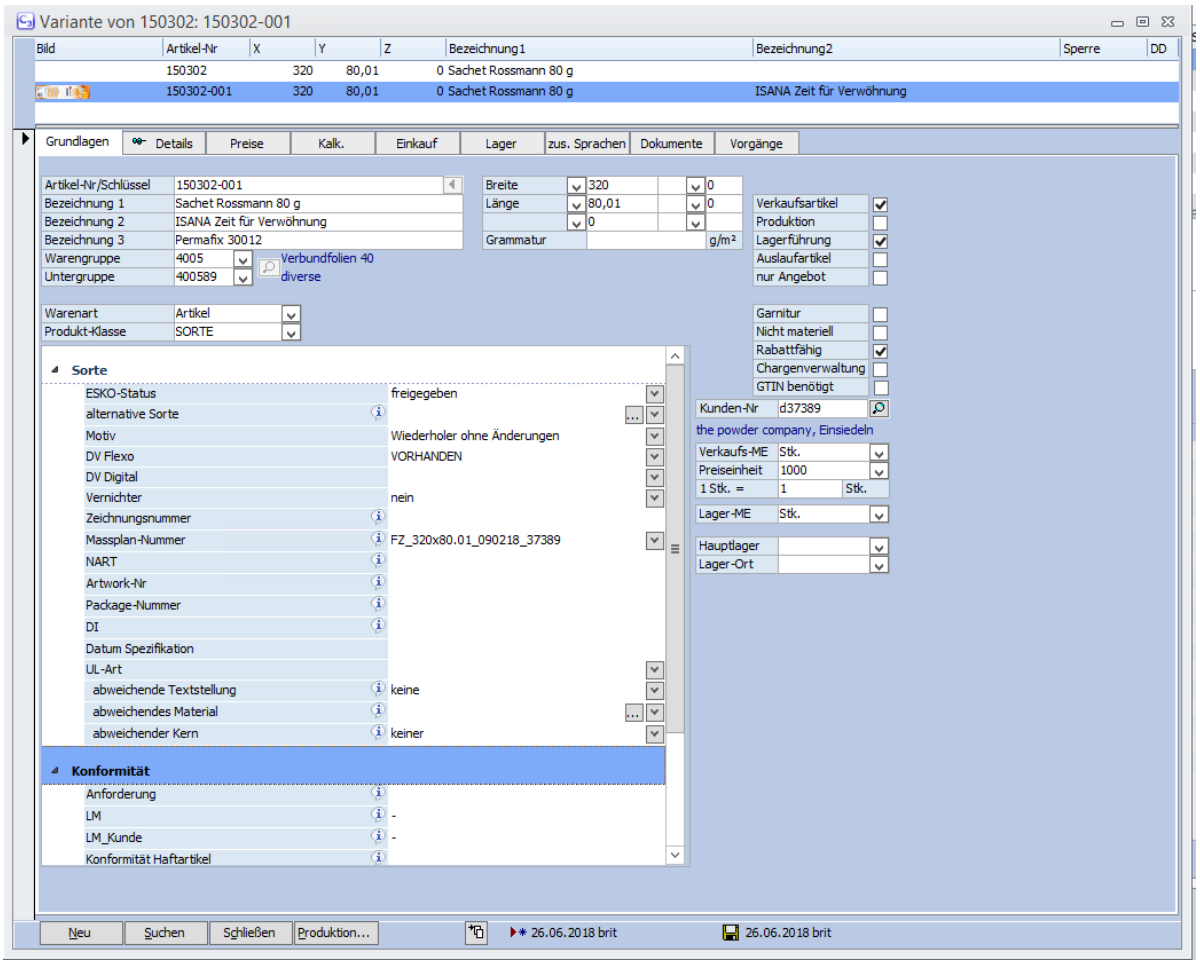


Abbildung 38: Eingabemaske Artikeldaten (eigene Darstellung)

Mit den Fachabteilungen zusammen wurde folgende Metadatendokumentation erstellt.

	Feldname	Vollständigkeit	Beschreibung	Std. Wert	EK	VK	AVOR	PR
D	Warenart	93.87 % Artikel, 3.1% Material	Material: siehe Geschäftsregel MAT-001	Artikel	C	C	E	R
D	Produkt-Klasse	85 % = Sorte,7% Sortenvorlage, 3.1% Material, 2.4% Werkzeug	nähere Klassifizierung ob SV oder Sorte Werkzeug: Einschränkungen	Sorte	C	C	E	R
T	Artikel-Nr/Schlüssel	Manuell nach Muster oder vom System erzeugt 51% ######-###,33% ######,11% \p{L}\p{L}\p{L}\d+ 2% \d+,\d+	ZNR nach Q-Dokument K-21-003-021	Nummer-001	R	R	R	R
T	Bezeichnung 1	3% NULL, 25% LEER, max. 50 Zeichen	Text nach Q-Dokument K-21-003-021		C	C	E	R
T	Bezeichnung 2	Vorgaben spezifizieren, 34.3% NULL, 12% LEER, max. 50 Zeichen	externe Bezeichnung für Kundenberichte, in vielen Listen Kombi aus Bezeichnung1 + Bezeichnung2		E	E	E	R
T	Bezeichnung 3	53% NULL, 32% LEER, max. 50 Zeichen	externe Bezeichnung für Kundenberichte, in vielen KundenListen Kombi aus Bezeichnung2 + Bezeichnung3		E	E	E	N
D	Warengruppe	1.3% leer	Beim Erzeugen eines Artikels verlangt System zwingend Warengruppe	von Vorlage	E	E	E	N
D	Untergruppe	5.3% leer	gleich wie Warengruppe	von Vorlage	E	E	E	N
T	Zeich. Datum	100% leer	nicht verwendet		N	N	N	N

T	Änd./ind.	100% leer	nicht verwendet		N	N	N	N
N	Breite	86.3%<=9999	abhängig vom Label. 98.3% Breite	Breite	E	E	E	N
N	Länge	94.3%<=9999	abhängig vom Label. 87.3% Länge	Länge	E	E	E	N
N	Dicke (ym)	53% ##,14.3% leer	abhängig vom Label. 54.3% Dicke	Dicke	E	E	E	N
N	Grammatur	87.4% leer	abhängig vom Label. 84.3% Grammatur	Gram	E	E	E	N
C	Vorlage	21% wahr	Wenn Vorlage = Sortenvorlage, Variante kann dann nicht 1 sein	1	R	R	R	R
C	Verkaufsartikel	67% wahr	Artikel kann in Angebot/Auftragsmasken ausgewählt werden	WAHR	E	E	E	N
C	Produktion	37% wahr	nur wenn wahr kann im Betriebsauftrag der Artikel ausgewählt werden. Wird in Produktionslisten verwendet	WAHR	E	E	E	R
C	Lagerführung	87.3% wahr	Lagerstände werden berücksichigt. Dispolisten	FALSCH	E	R	E	N
C	Auslaufartikel	3% wahr	Artikel kann in Angebot/Auftragsmasken nicht mehr angewählt werden	FALSCH	R	E	E	N
D	Verkaufs-ME	45% lfm,34% Stk, 2% Rollen	hat grossen Einfluss auf Umrechnung in Kalkulation und Einkauf	aus Kalk	R	R	E	N
D	Preiseinheit	32% 1, 35% 1000, 12.5% anderes	Bezugsgrösse für Verrechnung	aus Kalk	R	R	E	N
D	Lager-Me	43% lfm,34% Stk, 2% Rollen	sollte gleich wie Verkaufs-ME	aus Kalk	R	R	E	N
D	Hauptlager	5.1% leer	Lagerplatz von oder wohin es Bestandsänderungen gebucht wird		R	R	E	E
D	Lager-Ort	4% leer wenn Hauptlager <> leer	gleiche Bedeutung		R	R	E	E

Tabelle 8: Metadatendokumentation (eigene Darstellung)

c) Geschäftsregeln

Mit den Geschäftsregeln können Regeln und Datenkonstellationen beschrieben werden, welche zu jedem Zeitpunkt gültig sind. Diese Regeln können unter anderem als Validierung wie auch zur Datenqualitätsanalyse verwendet werden.

- Der Produkttyp ist immer „Artikel“ ausser bei Betragsartikel und CU 79 Produkte *1
- Der Produkttyp bei Betragsartikel ist immer Service *1
- Der Produkttyp bei CU 79 Produkte ist Service100 %
- Die Produktnummer und die Artikelnummer sind immer identisch 100 %
- Wenn der Produkttyp Artikel ist, dann ist die Artikelgruppe Handel 99.512 %
- Wenn der Produkttyp Artikel ist, dann ist die Lagerdimensionsgruppe WHS 99.512 %
- Wenn der Produkttyp Artikel ist, dann ist die Reservierungshierarchie STD 99.446 %
- Produktgruppen welche mit 15**** beginnen, sind ungültig 99.990 %
- Der Produktname darf nicht leer sein 99.736 %
- Die Rückverfolgungsgruppe ist immer „Keine“ 100 %
- Externe Charge kann nur selektierbar, wenn der Standardauftragstyp „Produktion“ ist 96.437 %
- Alle Artikel vom DIV (MU10, MU30, MU40) haben die Kostenstelle 6010 100 %

3) Verbesserung der Datenqualität

Um eine langfristige Verbesserung der Datenqualität zu garantieren, sollten alle Aspekte, welche sich auf die Datenqualität auswirken berücksichtigt werden. Alle Massnahmen zur Erhöhung der Datenqualität müssen mit den Fachverantwortlichen auf ihre Wirtschaftlichkeit und Sinnhaftigkeit hin geprüft werden.

Abbildung 39: Vorgehensmodell mit Prozessschritten (eigene Darstellung)

a) Datenbereinigung

Die eigentliche Bereinigung der identifizierten Datenfehler, kann als einmalige Aktivität implementiert werden. Die Bereinigung kann wo möglich automatisiert, oder durch die Mitarbeiter erfolgen.

- Invalide Datensätze in Listen auflisten und den Abteilungen zur Korrektur übergeben.

b) Permanente Sicherstellung

Eine permanente Sicherstellung der Datenqualität erfordert eine Integration von Datenqualitätsmessungen in Form von Validierungen direkt im C3. Das heisst die Qualitätsmängel bereits an der Quelle der Entstehung zu beseitigen. Ebenfalls sollte der Datenbestand in regelmässigen Abständen anhand der definierten Geschäftsregeln überprüft werden um allfällige Fehler zu detektieren, bevor diese im operativen Betrieb sichtbar werden.

- Abklärungen und Prüfung des Aufwands für Validierungen im C3.
    - Pflichtfeldervalidierung
    - Geschäftsregelvalidierungen
- Regelmässige Datenanalyse
    - Regelformulierung, Erweiterung und Anpassung der Geschäftsregeln
    - Ursachenanalyse (PDCA)

c) Systemoptimierung

Neben der reinen Bereinigung der Daten sollten auch die Informationssysteme betrachtet werden um die Datenqualität zu erhöhen. Die Weitergabe von Daten über Systemgrenzen stellen häufig Fehlerquellen dar. Ein besonderes Augenmerk sollte deshalb auf die Schnittstellenprogramme gelegt werden. Beispiel C3 zu AX, AX zu Stepcom, etc.

- Kontinuierliche Optimierung durch die IT

d) Geschäftsprozessoptimierung

Die Geschäftsprozesse bestimmen wie die Daten verarbeitet werden. Die unterschiedlichen Geschäftseinheiten haben unterschiedliche Prozesse entwickelt um die gleichen Daten zu pflegen. Hier könnte es ebenfalls sinnvoll sein, die Prozesse wo möglich zu vereinheitlichen.

- Prozesse durch die Abteilungen beschreiben, Synergien erkennen und gegebenenfalls vereinheitlichen.
- Sicherstellen, dass die gleichen Felder in den Geschäftseinheiten gleich interpretiert und genutzt werden.
- Data Owner pro Geschäftseinheit

e) Schulung der Mitarbeitern

Aufgrund der unterschiedlichen Organisationen der Geschäftseinheiten haben die Mitarbeiter bei der Datenerfassung unterschiedliche Wissensstände und auch unterschiedliche Anforderungen an die Daten. Ein pragmatischer Ansatz die Datenqualität zu erhöhen könnte das Erstellen von Checklisten für unterschiedliche Produktausprägungen (Material, Artikel, Produktionsartikel, etc.) sein. Ebenfalls sollten bei Entscheidungsfelder die Auswirkungen aufgezeigt werden. Dies könnte vor allem bei Mitarbeiter, welche selten Artikel anlegen oder bei Spezialfällen hilfreich sein.

- Schulungsunterlagen erweitern oder anpassen
- Mitarbeiter regelmässig schulen
- Checklisten für unterschiedliche Produktausprägungen erstellen

f) Formular – Tabellen – Feld –Zuordnung für Artikel

Typ	Feldname	Tabelle	Feld
Produkt erstellen			
T	Artikel Nr	Artikel	Znr_alphan
D	Warenart	Artikel	Warenart
T	Produkt-Klasse	Artikel	DisplayProductNumber
T	Artikelnummer	InventTable	ItemId
T	Produktname	Artikel	Name
D	Artikelgruppe	Artikel	ItemGroupId

Tabelle 9: Formular Tabellen Feld Zuordnung (eigene Darstellung)

g) Validierungen

Bei den Produktstammdaten sollen zusätzliche Validierungen eingeführt werden, die prüfen, ob das beschriebene Datenfeld einen Inhalt hat. Es muss sichergestellt werden, dass Produkte welche die folgenden Felder nicht erfasst haben im Prozess nicht verwendet werden können.

Anzeigename	Feldname
Produktname	EcoResProductTranslation.Name
Einzelhandelskategorie	EcoResCategory.Name
Produktgruppe	AXPERMProductGroup.ProductGroupId
Einkäufergruppe	InventBuyerGroup.Group
BusinessUnit	Finanzdimensionen
Kostenstelle	Finanzdimensionen

Tabelle 10: Valdidierungsregelen (eigene Darstellung)

Auf Datenbankebene sollen Validierungen nur zuletzt durchgeführt werden, ausser sie sind ein-eindeutig. Abbildung 40 zeigt einen Update Trigger auf Artikel Tabelle

```
ALTER TRIGGER [dbo].[PP_Artikel_UTrig] ON [dbo].[Artikel] FOR Update AS

Declare @errCount int
SET @errCount = 0

    Declare @WarenArt as nvarchar(50)
    Declare @Ursprungsland as nvarchar(50)
    Declare @knr as nvarchar(50)
    declare @msgstr nvarchar(150)
    set @msgstr =  nchar(13) + nchar(10) + nchar(13) + nchar(10) + N'Permapack-Geschäftsregel: ' + nchar(13) + nchar(10) + 'Hauptkundennummer (F
    Select @WarenArt = inserted.WarenArt from inserted

    select @Ursprungsland = inserted.IntrastatUrsprungsland from inserted where warenart = 1 and IntrastatUrsprungsland is not null
--    select @knr=inserted.knr from Inserted
    if ((@WarenArt = 1) and (@Ursprungsland is null))
    begin
        Declare @ZNR as nvarchar(50)
        Select @ZNR = inserted.znr_alphan from inserted

        Update artikel set IntrastatUrsprungsland = 'CH' where znr_alphan = @ZNR
    end

    if (@WarenArt = 2 and (update(meldebestand) or update(mindestbestand)))
    begin
        if ((select isnull(inserted.meldebestand,0) from inserted) < (select isnull(inserted.mindestbestand,0) from inserted))
        begin
            Set @msgstr = 'Meldebestand muss grösser als Mindestbestand sein'
            RAISERROR(@msgstr, 16, 1)
            set @errCount  =  @errCount  + 1
        end
    end


if @errCount > 0
    ROLLBACK TRANSACTION
```

Abbildung 40: Validierungsregel auf Datenbank Ebene (eigene Darstellung)

h) Default-Werte

Bei den folgenden Datenfeldern soll ein Default-Wert gesetzt werden.

Anzeigename	Feldname	Default Wert
Lagersteuerungsgruppe	InventModelGroupItem. ModelGroupId	D
Artikelgruppe	InventItemGroupItem.ItemGroupId	Handel
Lagerdimensionsgruppe	EcoResStorageDimensionGroupItem. StorageDimensionGroup	WHS
Rückverfolgungsgruppe		Keine
Reservierungshierarchie	WHSRerervationHiererchyItem. RerervationHiererchy	STD
Artikel MwSt. EK	InventTableModule.TaxItemGroupId; Modultype=1	AV
Artikel MwSt. VK	InventTableModule.TaxItemGroupId; Modultype=2	AV
Neuster Einkaufspreis	InventTable. PurchModel	checked

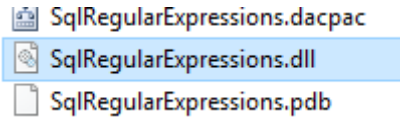
Tabelle 11: neue Default-Werte (eigene Darstellung)

Anhang C

SQL RegEx Implementierung

MS SQL bringt leider nur begrenzte String Match Funktionen mit, deshalb hier eine kurze Implementierungsadaption. SQL kann von Haus aus nur ‚^', '[ ]', $', ‚%' mit dem Like Operator verarbeiten.

```
/// <summary>
/// die erzeugte DLL als Assembly registriere und eine benutzerdef Funktion im SQL
/// ersetllen. Danach kann die Funktion als bekannte RegEx verwendet werden ;-)
/// </summary>
```



MS VC Projekt Source Anpassungen

SqlRegularExpressions — SQLSrvRegex

```
1   using System;
2   using System.Data.SqlTypes;
3   using System.Text.RegularExpressions;   //Match, Regex
4   using Microsoft.SqlServer.Server;       //SqlFunctionAttribute
5
6   /// <summary>
7   /// RegEX Match for SQL CLR for MasterThesis MS DataQuality Client 20180801
8   /// </summary>
9   public partial class SQLSrvRegex
10  {
11      /// <param name="input">The string to be tested for match.</param>
12      /// <param name="pattern">The regular expression pattern to match.</param>
13      /// <param name="options">A bitwise OR combination of RegexOption (-> see doc) enumeration values.</param>
14      /// <returns>true - if inputted string matches to pattern, else - false</returns>
15      /// <exception cref="System.ArgumentException">Regular expression parsing error.</exception>
16      [SqlFunction(Name="RegexMatch", IsDeterministic=true, IsPrecise=true)]
17      public static SqlBoolean RegexMatch(SqlString input, SqlString pattern, SqlInt32 options)
18      {
19          if (input.IsNull) {
20              return SqlBoolean.Null;
21          }
22          if (pattern.IsNull) {
23              pattern = String.Empty;     //""
24          }
25          if (options.IsNull) {
26              options = 0;  //RegexOptions.None
27          }
28
29          try {
30              Match match = Regex.Match((string)input, (string)pattern, (RegexOptions)(int)options);
31              if (match.Value != String.Empty) {
32                  return SqlBoolean.True;
33              }
34          }
35          catch {
36              throw;
37          }
38          return SqlBoolean.False;
39      }
40  };
```

Abbildung 41: C# RegEx Funktion (eigene Darstellung)

SQL Code

```
-- 1 begin
sp_configure 'clr enabled', 1
GO
RECONFIGURE
GO

-- 2
CREATE ASSEMBLY SqlRegularExpressions
-- assembly name and full path to assembly dll
from 'C:\Program Files (x86)\Microsoft SQL Server\120\SDK\Assemblies\SqlRegularExpressions.dll'
WITH PERMISSION_SET = SAFE

-- 3 function signature
CREATE FUNCTION RegexMatch(@Text nvarchar(max),
      @Pattern nvarchar(255), @Options INT) RETURNS BIT
--function external name
AS EXTERNAL NAME SqlRegularExpressions.SQLSrvRegex.RegexMatch

-- end

-- 4 using c3perma.. 20180801
--count(207)
select count(*) from kunden where 1 = dbo.RegexMatch(MatchCode, 'Coop',1)
--178 beginning
select count(*) from kunden where 1 = dbo.RegexMatch(MatchCode, '^Migro',1)
--0 end
select count(*) from kunden where 1 = dbo.RegexMatch(MatchCode, '$C',1)
--STARTING WITH C 2 times end
select * from kunden where 1 = dbo.RegexMatch(MatchCode, '^[C]{2}',1)

-- RegExpMatches sample
DECLARE @Text nvarchar(max);
DECLARE @Pattern nvarchar(255);

SET @Text = 'This is comprehensive compendium provides broad and thorough investigation of all '
+ 'aspects of programming with ASP.Net. Entirely revised and anna updated for the 2.0 '
+ 'Release of .Net, this book will give you the information you need to master ASP.Net '
+ 'and build a dynamic, successful, enterprise Web application.';
SET @Pattern = '\b(a\S+)';   --get all words that start from 'a'
SET @Pattern = '\b(c\S+)';   --get all words that start from 'a'

select C3perma.dbo.regexstrMatch(@Text, @Pattern,2)
```

Abbildung 42: Verwendung der RegEx Funktion in SQL 2014 (eigene Darstellung)

Declaration

I herewith declare that this is my independent work written by me and using only admissible aides and no other sources than those given. I have marked as such, all passages which have been taken literally or analogously from another source. I am aware that if this is not the case, the executive board of the university of applied sciences is entitled to rescind any qualifications awarded or any title bestowed based on this work.

Erklärung

Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und erlaubten Hilfsmittel benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen worden sind, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls die Hochschulleitung zum Entzug der aufgrund meiner Arbeit verliehenen Qualifikation oder des für meine Arbeit verliehenen Titels berechtigt ist."

Diepoldsau, 10.08.2018	
(Ort, Datum)	Markus Rüf