

# **Spezifikation und Messung der Datenqualität in Data-Warehouse-Systemen**

## **-Vorstellung des Dissertationsprojektes-**

**Markus Helfert**

Universität St. Gallen (markus.helfert@unisg.ch)

### *Zusammenfassung:*

Die Sicherstellung der Datenqualität ist einer der entscheidenden Erfolgsfaktoren für die Nutzung von Data-Warehouse-Systemen, denn betriebliche Entscheidungen werden von der Qualität der zur Verfügung gestellten Daten beeinflusst. In der Praxis lassen sich zahlreiche Projekte finden, die gerade aufgrund von Problemen der Datenqualität scheitern. Bislang existieren sowohl in der Praxis als auch in der Wissenschaft nur unzulängliche Ansätze zur Sicherstellung einer angemessenen Datenqualität. Die Dissertation beschäftigt sich daher mit dieser Fragestellung. Insbesondere werden die für ein Datenqualitätsmanagement zentralen operativen Aufgabenbereiche der Qualitätsplanung und der Qualitätslenkung untersucht.

Ziel der Dissertation ist es, ein Datenqualitätsmodell zur Spezifikation und Messung der Datenqualität in Data-Warehouse-Systemen zu entwickeln und so ein Bewertungssystem für Qualitätsaussagen zu schaffen. Das Datenqualitätsmodell ermöglicht es, Qualitätsziele festzulegen und die sich entwickelnde Datenqualität abzuschätzen. Das Datenqualitätsmodell soll anhand eines im Mittelpunkt stehenden Fallbeispiels einer Schweizer Universalbank untersucht werden.

*Schlüsselworte: Datenqualitätsmanagement, Datenqualität, Datenqualitätsmessung, Data-Warehouse-Systeme*

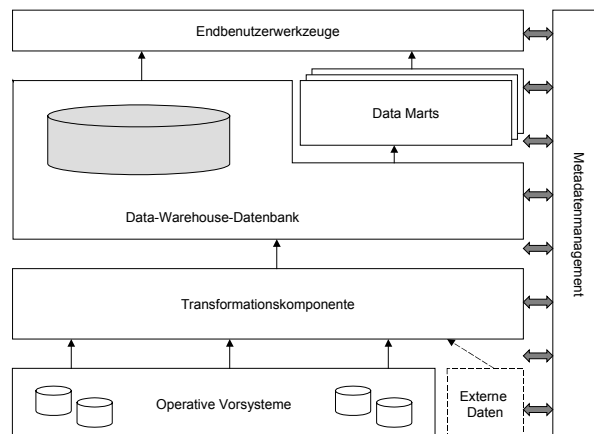
## **Konzeptionelle Grundlagen**

Im ersten Teil des Artikels werden konzeptionellen Grundlagen für die Arbeit dargestellt. Ziel ist es, ein für die Arbeit geeignetes Begriffsverständnis zu schaffen und den Ansatz des Datenqualitätsmanagements darzustellen. Hierzu wird zunächst das Konzept des Data-Warehouse-Systems und dessen Komponenten erläutert. Anschliessend wird auf die Begriffe Daten, Qualität und Datenqualität

eingegangen. Im letzten Abschnitt wird das Datenqualitätsmanagement beschrieben. Aufbauend auf diesen konzeptionellen Grundlagen beschäftigt sich der zweite Teil des Artikels mit dem Dissertationsprojekt und der Zielsetzung der Arbeit.

## Data-Warehouse-Systeme

In den letzten Jahrzehnten entwickelten sich zu mengenorientierten operativen Informationssystemen zahlreiche Ansätze und Systeme für analyseorientierte Informationssysteme.[Holt99, S. 36] Ein, in diesem Zusammenhang häufig diskutiertes Konzept, welches insbesondere die Datenversorgung des Managements bezweckt, kann sehr allgemein unter dem Begriff des "Data Warehousing" zusammengefasst werden. Zentrale Komponente des Konzeptes ist eine Data-Warehouse-Datenbank. Diese stellt eine von den operativen Systemen losgelöste, logisch zentralisierte, einheitliche und konsistente Datenbasis dar.[Dev197, S. 20] Die Gesamtheit der Anwendungen und Datenbanken, die diese Datenbank nutzbar macht, wird umfassend als Data-Warehouse-System bezeichnet.[Wint00, S. 128] Abbildung 1 gibt einen groben Überblick über die Komponenten eines Data-Warehouse-Systems, die im folgenden kurz beleuchtet werden sollen.



**Abbildung 1: Data-Warehouse-System [Müll00, S. 104]**

Ausgangspunkt sind die operativen Vorseiten und externe Informationssysteme, die als Datenlieferanten agieren. Mit Hilfe einer Transformationskomponente werden diese Daten extrahiert, transformiert und dann in die zentrale Data-Warehouse-Datenbank überführt. Teilweise werden auch, zur fachspezifischen Datenbereitstellung, kleinere, redundante Teilausschnitte aus der unternehmensweiten Datenbasis in sogenannten Data Marts vorgehalten. Auf diese Datenbestände können die Anwender mit vielfältigen Endbenutzerwerkzeugen zugreifen.

Sie reichen von Werkzeugen zur Erstellung von Berichten über Ad-Hoc-Abfragen bis hin zu vielschichtigen, mehrdimensionalen Auswertungen.

Diese Komponenten dienen der Sicherstellung des Datenflusses von der Datenquelle bis hin zur Datenverwendung und bilden das Basissystem des Data-Warehouse-Systems. Neben diesem existiert ein Koordinationssystem, welches im allgemeinen als Metadatenverwaltungssystem bezeichnet wird.[Holt99, S. 95 ff.] Im Kern umfasst es eine Metadatenbank und Verwaltungswerkzeuge zur Speicherung und Verwaltung der Systemkomponenten und der Datenflüsse.

## Datenqualität

### Datenbegriff

Da es bislang sowohl in der Betriebswirtschaft noch in der Informatik als auch in der Wirtschaftsinformatik noch nicht gelungen ist, die häufig verwendeten Begriffe Informationen, Daten und Wissen klar abzugrenzen und zu definieren,[MaLe95; Stre96, S. 17 ff.; Bode97] soll zur Einordnung des Begriffs der Datenqualität zunächst eine für die Arbeit geeignete Begriffsbestimmung<sup>1</sup> beschrieben werden.

Zunächst soll das der Arbeit zugrundegelegte Wissenverständnis eingeführt werden. "**Wissen** ist jede Form der Repräsentation von Teilen der realen oder gedachten (d. h. vorgestellten) Welt in einem materiellen Trägermedium." [Bode97, S. 458] Kennzeichnend für diesen Wissensbegriff ist die Repräsentation von Ausschnitten der realen Welt. Wissen als Abbildung ist mit den Realitätsausschnitten nicht identisch, steht jedoch stets in Relation zu diesen und besitzt somit eine Bedeutung (Semantik). Aufbauend auf diesem Begriffsverständnis kann Information als echte Teilmenge von Wissen verstanden werden. "**Informationen** sind Wissensbestandteile, die in Form menschlicher Sprache repräsentiert sind." [Bode97, S. 459] Menschliche Sprache wird dahingehend eingeschränkt, dass diese die Übermittlung von Wissen zwischen Menschen erlaubt. Zur Unterscheidung zwischen maschinenverarbeitbaren und nicht maschinenverarbeitbaren Informationen soll der Begriff Daten als echte Teilmenge von Informationen definiert werden. **Daten** sind Informationen, "deren sprachliche Repräsentationsform und materielle Träger auf eine maschinelle Verarbeitung gerichtet sind" [Bode97, S. 460].

---

<sup>1</sup> Eine Diskussion hierzu findet sich in [Helf01, S. 55 f.]. Der Definitionsansatz bezieht sich im Wesentlichen auf den von [Bode97].

## Qualitätsbegriff

Als Folge der Entwicklung des Qualitätsbegriffs und der damit verbundenen Diskussion existiert eine Vielfalt an Definitions- und Interpretationsansätzen des Begriffs der Qualität. Ziel der Begriffsbestimmung ist es, die Komplexität des Qualitätsphänomens zu reduzieren und zu operationalen Aussagen zu gelangen.

Nach dem Systematisierungsansatz von [Garv84, S. 26] lassen sich fünf Qualitätsdefinitionen unterscheiden. Neben diesen eher beschreibenden Ansätzen gibt es eine Reihe von Versuchen, insbesondere durch Normierungsinstitutionen, den Begriff der Qualität im Sinne einer Definition zu erfassen.[Wall90, S. 8] Diese können jedoch entsprechend in die fünf Grundtypen eingeordnet werden. Der **transzendente** Ansatz kennzeichnet Qualität als angeborene Einzigartigkeit oder Superlative, als ein Synonym für hohe Standards und Ansprüche. Die diesem, eher abstrakt philosophischen Verständnis folgende Auffassung, Qualität könne nicht exakt definiert werden, sondern sei nur durch Erfahrung erkennbar,[Garv84, S. 25] ist für die weitere Betrachtung unzulänglich. Materielle Produkteigenschaften bestimmen bei **produktbasierten** Qualitätsdefinitionen die Qualität eines Produktes. Qualität ist so präzise messbar und eine inhärente Eigenschaft des Produktes selbst. Beim **anwenderbezogenen** Ansatz liegt die Auffassung vor, dass Qualität durch den Produktbenutzer und nicht durch das Produkt selbst festgelegt wird. Ein Produkt wird dann als qualitativ hochstehend angesehen, wenn es dem Zweck der Benutzung durch den Kunden während des Gebrauchs dient. Die individuellen Bedürfnissen des Kunden sind dabei bestimmend. Nach dem **prozessbezogenen** Ansatz bedeutet Qualität die Einhaltung von Spezifikationen und die Abwesenheit von Fehlern. Im Mittelpunkt dieses Ansatzes stehen die auf Einhaltung der Spezifikation kontrollierten Produktionsprozesse. Der **wertbezogene** Ansatz stellt Qualität im Sinne von Nutzen in Bezug zu den hierfür aufgewendeten Kosten. Ein Produkt ist dann von hoher Qualität, wenn die zu entrichtenden Kosten und die empfangene Leistung in einem akzeptablen Verhältnis stehen.

## Begriff der Datenqualität

Ähnlich wie der Qualitätsbegriff, wird der Begriff der Daten- und Informationsqualität in der Literatur in sehr unterschiedlichen Sichtweisen und mit verschiedenen operationalisierenden Merkmalen beschrieben.[WaZL01, S. 2; Müll00, S. 15; NaRo99; Wolf99, S.85; Eng199; TaBa98; JaVa97; WaSt96; Redm96; Hauk84] So haben sich bereits zahlreiche Autoren insbesondere mit der Auflistung von Qualitätskriterien beschäftigt. Ergebnis dieser Arbeiten sind eine Vielzahl von Kriterienlisten und Einordnungsrahmen für unterschiedliche Anwendungsgebiete. Im Rahmen des Dissertationsprojektes sollen diese weiter untersucht und als Grundlage zur Entwicklung eines Datenqualitätsmodells dienen.

Wenngleich die Begriffe Daten- und Informationsqualität uneinheitlich verwendet werden, lässt die grobe Untersuchung der Ansätze eine weitgehende Überein-

stimmung dahingehend erkennen, dass Daten- bzw. Informationsqualität hinsichtlich des Beitrags zur Erreichung der Ziele des Datenempfängers bestimmt wird. Damit folgen die Ansätze weitgehend dem anwenderorientierten Qualitätsbegriff. Dieser Begriff wird durch einer Fülle von Qualitätsmerkmalen konkretisiert, die in ihrer Bedeutung und Intensität erheblich vom Betrachtungszusammenhang abhängen. Eine Übereinstimmung bzgl. der relevanten Qualitätsmerkmale, deren Definition und Beziehungen lässt sich allerdings nicht erkennen. Im allgemeinen werden die Qualitätskriterien intuitiv anhand von Erfahrungen aufgestellt [BaPa85; Laud86; More82], auf Basis der in der Literatur genannten Kriterien zusammengestellt [NaRo99] oder durch empirische Untersuchungen erfasst [WaSt96].

Grundlage weiterer Arbeiten bilden die im Rahmen einer empirischen Untersuchung erfassten Qualitätskriterien.[WaSt96, S. 20] Ausgehend von diesen Kriterien schlagen Jarke und Vassiliou Qualitätsfaktoren für Data-Warehouse-Systeme vor,[JaVa97, S. 303] wobei spezifische Anforderungen hinsichtlich aggregierter und historisierter Daten stärker berücksichtigt werden. Die Qualitätskriterien sind in Abbildung 2 dargestellt und bilden die Grundlage für die weitere Arbeit. Nachdem der Begriff der Datenqualität beschrieben wurde, soll im nächsten Abschnitt das Konzept eines Datenqualitätsmanagements skizziert werden.

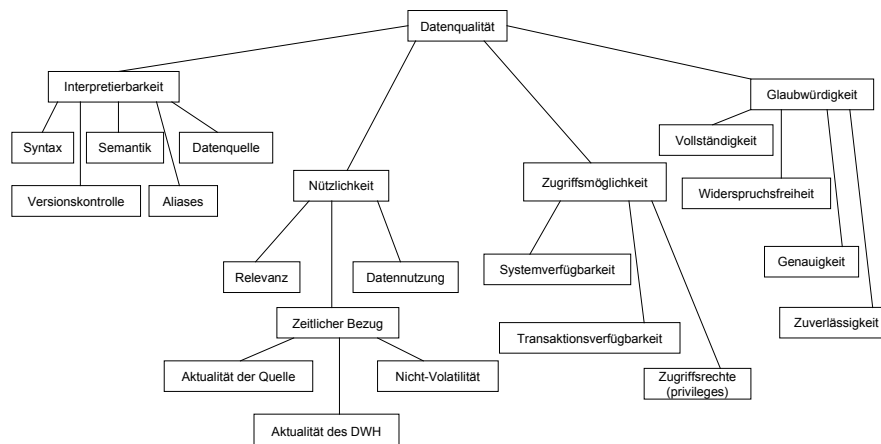


Abbildung 2: Qualitätskriterien nach Jarke und Vassiliou [JaVa97, S. 303]

## Konzept eines Datenqualitätsmanagements

In Anlehnung an allgemeine Konzepte des Qualitätsmanagements wird im folgenden ein Konzept zur Sicherstellung einer möglichst angemessenen Datenqualität in Data-Warehouse-Systemen aufgezeigt. Überträgt man DIN ISO 8402 sinngemäss, so umfasst Datenqualitätsmanagement alle Tätigkeiten der Gesamtfüh-

rungsaufgabe, welche die Datenqualitätspolitik, die Datenqualitätsziele und die Verantwortungen für die Datenqualität festlegt.[DIN95] Ausgehend von diesem Ansatz und aufbauend auf den Grundsätzen des Total Quality Management (TQM) lässt sich ein Konzept für ein Datenqualitätsmanagement für Data-Warehouse-Systeme entwickeln, welches sich grundsätzlich auf drei zentrale Bereiche stützt:[Helf00a, S. 67 f.; Engl99, S. 70 ff.; Wolf99, S. 74]

- Die Verpflichtung des Managements, Datenqualität als Philosophie und Unternehmenskultur vorzuleben. Auf Basis formulierter Unternehmensgrundsätze und -ziele ist eine Datenqualitätspolitik und eine Datenqualitätsstrategie abzuleiten. (in Anlehnung an [Segh96, S. 51])
- Ein Qualitätsmanagementsystem, welches den organisatorischen Rahmen verkörpert, ist zu etablieren. Dieses umfasst nach DIN ISO 8402 die Aufbau- und Ablauforganisation, die Zuständigkeiten, Prozesse und Mittel für die Qualitätssicherung.[DIN95] Es stellt sicher, dass in allen Bereichen geeignete Prozesse, Richtlinien, Pläne sowie Test- und Prüfverfahren etabliert sind, die die geforderte Datenqualität gewährleisten. Hierzu ist eine ständige Überprüfung, Analyse und Verbesserung der gewählten Massnahmen und durchzuführenden Prozesse erforderlich.
- Zur Unterstützung der Mitarbeiter bei der Ausübung der Qualitätsprozesse sind in allen Phasen geeignete Methoden, Verfahren und Werkzeuge zur Verfügung zu stellen.

Auf der operativen Ebene können als Teil des Datenqualitätsmanagements vier zentrale Aufgabenbereiche identifiziert werden. Aufgabe der **Qualitätsplanung** ist es, Bedürfnisse und Erwartungen zu erfassen, diese in Vorgaben zu transformieren und Leistungen sowie Prozesse zu gestalten.[Segh96, S. 72] Im Rahmen der Qualitätsplanung werden Anforderungen an die Daten (Produktqualität) und deren Bereitstellungsprozesse (Prozessqualität) festgelegt. Es sind Qualitätsmerkmale auszuwählen, zu klassifizieren und zu gewichten.[Wall90, S. 19] Die **Qualitätslenkung** zielt auf die Einhaltung von Spezifikationen und der Beherrschung der Prozesse ab.[Segh96, S. 76] Hierfür sind zunächst geeignete Prozesse zu identifizieren und Massnahmen zur Erreichung der Prozesskonformität zu ergreifen. Produkt- und Prozessqualität müssen im Rahmen der Qualitätslenkung gemessen und in quantitativen Kennziffern ausgedrückt werden. Wichtigstes Hilfsmittel für die Qualitätslenkung sind Qualitätsprüfungen.[Wall90, S. 19] Schlussendlich sind Verantwortlichkeiten für die Qualitätslenkung festzulegen und die Messergebnisse als Rückkopplung in Form von Regelkreisen zurückzuführen.

Die **Qualitätssicherung** ist als strukturelle Unterstützung der Qualitätsplanung und Qualitätslenkung zu verstehen, die darauf abzielt, Risiken systematisch zu erkennen, aufzudecken und ihre Wirkung zu bekämpfen.[Segh96, S. 108] Voraussetzung der Qualitätssicherung sind Risikoanalysen, wie beispielsweise die der Fehlermöglichkeits- und -einflussanalyse (FMEA).[Segh96, S. 99] Der vierte Aufgabenbereich des operativen Qualitätsmanagements ist die **kontinuierliche**

**Verbesserung.**[Segh96, S. 111] Während Qualitätslenkung und Qualitätssicherung stabilisieren und veränderungshemmend wirken, fördert die kontinuierliche Verbesserung die dynamische Steigerung des Qualitätsniveaus. Als wichtigstes Instrumentarium der Qualitätsverbesserung sind Verbesserungsprojekte zu nennen.

## **Überblick über das Dissertationsvorhaben**

Die Dissertation beschäftigt sich mit der Spezifikation und der Messung der Datenqualität in Data-Warehouse-Systemen. Nachdem im ersten Teil des Artikels die grundlegenden Begriffe dargestellt wurden, beschäftigt sich dieser Teil mit der Forschungsfrage und der Zielsetzung der Arbeit, den bisherherigen Forschungsansätzen in der Literatur sowie der beabsichtigten Vorgehensweise und dem zugrundeliegenden Forschungsansatz.

### **Ableitung der zentralen Forschungsfrage**

#### **Problematik der Datenqualität in Data-Warehouse-Systemen**

Die Sicherstellung der Datenqualität ist einer der entscheidenden Erfolgsfaktoren für die Nutzung von Data-Warehouse-Systemen.[Häus98, S. 75] In der Praxis lassen sich zahlreiche Projekte finden, die gerade aufgrund von Problemen der Datenqualität scheitern.[Helf00a, S. 65] Gestellte Anforderungen an das Data-Warehouse-System, insbesondere an die Aussagekraft der Daten, können nicht oder nur teilweise erfüllt werden. Nach einer 1999 durchgeführten Umfrage wird die Sicherstellung der Datenqualität bei nahezu allen Unternehmen als problematisch eingeschätzt.[Helf00b, S. 13] Insbesondere die Durchsetzung von Massnahmen zur Datenqualitätssicherung in den operativen Systemen sowie semantische Aspekte sind problematisch und schwer zu lösen.

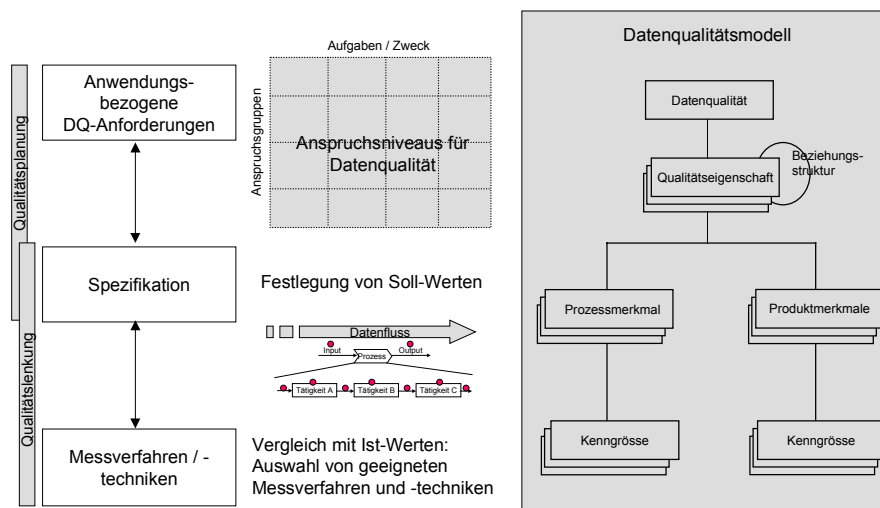
Mangelnde Datenqualität führt nicht nur zur aufwendigen Suche nach korrekten Werten und nachträglichem Aufwand bei der Datenbereinigung sondern hat darüber hinaus weitere Auswirkungen. So führt mangelnde Datenqualität nicht selten zur Verringerung der internen Akzeptanz des Data-Warehouse-Systems, zu nicht optimalen Entscheidungen sowie unzureichender Unterstützung operativer Geschäftsprozesse.[Helf01, S. 55]

Neben der Möglichkeit die Datenqualität durch aufwendige Datenbereinigungsprozesse (siehe hierzu [Müll00]) zu verbessern, erscheint die Sicherstellung und Steigerung der Datenqualität anhand des oben dargestellten Konzepts sinnvoll. Zentrale Aufgabenbereiche des Datenqualitätsmanagements sind insbesondere die Qualitätsplanung und Qualitätslenkung.[Engl99, S. 70 ff.; HuLW99, S. 16] Kön-

nen diese Aufgabenbereiche nicht realisiert werden, ist ein Datenqualitätsmanagement im obigen Sinne nur schwer oder nicht möglich. Daher richtet sich der Fokus des Dissertationsprojektes auf die operativen Aufgaben des Qualitätsmanagements und insbesondere der Qualitätsplanung und Qualitätslenkung.

### Zielsetzung der Arbeit

Aufgabe der Qualitätsplanung ist es, Qualitätsbedürfnisse und Qualitätserwartungen zu erfassen, in Vorgaben zu transformieren und diese Vorgaben im Rahmen der Qualitätslenkung einzuhalten. Ziel der Dissertation ist es, ein Datenqualitätsmodell zur Spezifikation und Messung der Datenqualität in Data-Warehouse-Systemen zu entwickeln und so die Qualitätsplanung und Qualitätslenkung zu ermöglichen. Im Überblick ist der Fokus der Dissertation in Abbildung 3 dargestellt.



**Abbildung 3: Fokus der Dissertation (Eigene Darstellung)**

Das Qualitätsmodell soll ein Bewertungssystem für Qualitätsaussagen schaffen, welches es ermöglicht, Qualitätsziele festzulegen und die sich entwickelnde Datenqualität abzuschätzen. Ausgehend von Qualitätsanforderungen, die von Anspruchsgruppen sowie dem Einsatzzweck des Data-Warehouse-Systems abhängen, soll ein Rahmen für eine Spezifikation abgeleitet werden und als Qualitätsvorgaben in das Datenqualitätsmodell einfließen. Die konkrete Instanzierung des Modells kann aufgrund der hohen Anwendungsabhängigkeit jedoch ausschliesslich im Einzelfall vorgenommen werden.

Ein wesentlicher Aspekt des Datenqualitätsmodells ist die Zerlegungssystematik von Qualitätseigenschaften.[Wall90, S. 46] Der allgemeine Qualitätsbegriff, der durch Qualitätseigenschaften charakterisiert ist, wird durch Ableiten von Quali-

tätsmerkmalen weiter operationalisiert. Anschliessend werden die jeweiligen Prozess- und Produktmerkmale durch Festlegung von Indikatoren, den sogenannten Kenngrössen, messbar gemacht. Für die Qualitätsindikatoren sind Messtechniken festzulegen sowie geeignete Messpunkte und Messzeitpunkte zu definieren.

Zentraler Ausgangspunkt zur Identifikation geeigneter Messgrössen im Data-Warehouse-System ist das Basissystem zur Datenbereitstellung, von der Datenerfassung über die operativen Systeme bis hin zur Datennutzung. Ziel ist es, einerseits Möglichkeiten zur Qualitätsmessung aufgrund von Datenbeständen und andererseits auf Grundlage typischer Datentransformationsprozesse zu identifizieren. Lösungsansätze werden vor allem im Bereich der Statistik gesehen, mit Hilfe derer dynamische Veränderungen und Widersprüche in den Datenbeständen und Transformationsprozessen erkannt werden können.

Zusammenfassend stellt sich die im Rahmen des Dissertationsprojektes zu bearbeitende Forschungsfrage wie folgt:

**Wie kann Datenqualität in Data-Warehouse-Systemen geplant, spezifiziert und gemessen werden?**

Die geplante Arbeit soll, unter Berücksichtigung von in der Literatur genannten Forschungsergebnissen einen Beitrag zur Planung und Messung der Datenqualität in Data-Warehouse-Systemen leisten. Fokus der Dissertation ist die Qualitätsplanung und Qualitätslenkung im Rahmen des Datenqualitätsmanagements für Data-Warehouse-Systeme. Insbesondere sollen die häufig isoliert betrachteten Forschungsansätze der Datenqualitätsplanung, der Qualitätsspezifikation und der Qualitätsmessung durch das zu entwickelnde Datenqualitätsmodell zusammengeführt und Möglichkeiten zur Integration in die Metadatenverwaltung aufgezeigt werden. Hierzu sind folgende Detailfragen zu lösen:

- Welche Anforderungen ergeben sich an ein Datenqualitätsmodell für Data-Warehouse-Systeme?
- Welche Defizite haben bisherige Ansätze im Vergleich zu den Anforderungen?
- Welche Anforderungen haben Endbenutzer an die Daten (anwenderbezogener Qualitätsbegriff, anwendergruppen- und aufgabenbezogene Anspruchsniveaus)?
- Wie können die Qualitätsanforderungen in Qualitätsvorgaben transformiert werden (Spezifikation für ein Anwendungsszenario)?
- Wie kann die Einhaltung von Datenqualitätsvorgaben überprüft werden (Messpunkte, Messzeitpunkte sowie Messtechniken und -verfahren)?
- Wie kann das entwickelte Datenqualitätsmodell für weitere Anwendungen angepasst werden (Generalisierbarkeit des Qualitätsmodells)?

- Wie kann die Qualitätsplanung und Qualitätslenkung innerhalb der herkömmlichen Entwicklung und des Betriebs eines Data-Warehouse-Systems berücksichtigt werden (Integration in die Metadatenverwaltung)?

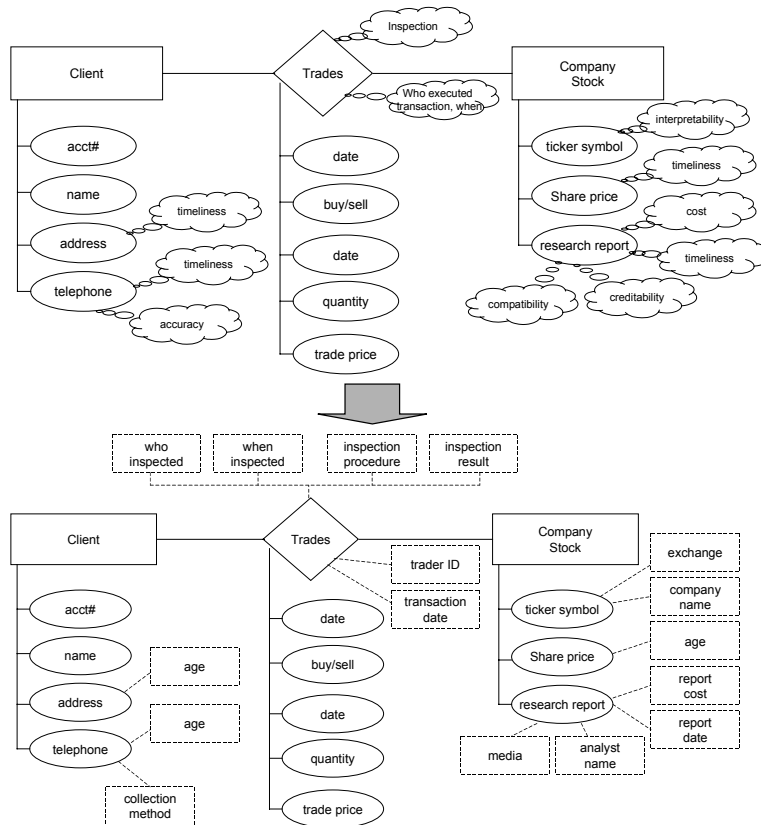
## **Mängel bisheriger Forschungsansätze**

Zahlreiche Forschungsarbeiten beschäftigen sich bereits mit Aspekten im Bereich der Datenqualität [WaSF95] und insbesondere mit der Fragestellung der Qualitätsplanung, deren Messung sowie der Integration in die Metadatenverwaltung. Im folgenden sollen ausgewählte Forschungsansätze kurz beschrieben und deren Defizite in bezug auf die oben genannte Forschungsfrage aufgezeigt werden. Zunächst werden Ansätze zur Datenqualitätsplanung und -spezifikation dargestellt. Anschliessend werden produkt- und prozessbezogene Messansätze beschrieben. Abschliessend wird ein Ansatz vorgestellt, der Qualitätsinformationen in die Metadatenverwaltung integriert. Im Verlauf der weiteren Arbeit gilt es, diese und weitere Ansätze in bezug auf die Forschungsfrage zu untersuchen und entsprechend einzuordnen.

### **Datenqualitätsplanung und -spezifikation**

Bisher wurde bei der Entwicklung von Informationssystemen die Erfassung und Beschreibung von Datenqualitätsanforderungen wenig beachtet. [WaKM93] Hervorzuheben sind die Arbeiten von Wang et al. [WaKM93; TuWa93; WaRK95; StWa98] und die des Forschungsprojektes „Foundations of Data Warehouse Quality“ (DWQ) [JaVa97; JLVV00]. Das Vorgehen des DWQ stützt sich weitgehend auf dem aus der Softwareentwicklung bekannten „Goal-Question-Metric“ (GQM) Ansatz [BaWe84] und untersucht dabei die Möglichkeiten der Integration dieser Qualitätsinformationen in die Metadatenverwaltung. Der Ansatz wird weiter unten näher beschreiben.

Die Arbeiten von Wang et al. beziehen sich auf Erweiterungen des Entity-Relationship-Modells (ERM) um Qualitätsattribute [WaKM93; TuWa93; WaRK95] sowie der Modellierung eigenständiger Qualitätsentitätstypen [StWa98]. Sie schlagen dabei ein Vorgehen anhand von vier Phasen vor. [WaKM93, S. 673] Ausgangspunkt ist ein anwendungsspezifisches ER-Datenmodell. Zu den Entitäts- und Beziehungstypen sowie den Attributen werden anschliessend subjektive Qualitätsparameter zugeordnet. Diese Qualitätsparameter werden dann in objektive Qualitätsindikatoren transformiert (vgl. hierzu Abbildung 4). In einem abschliessenden Schritt sollen die so erstellten anwendungsspezifischen Qualitätssichten in ein einheitliches Qualitätsschema zusammengeführt werden.



**Abbildung 4: Qualitätsplanung nach Wang et al. (Beispiel aus [WaKM93])**

Wenngleich das vorgeschlagene Vorgehen grundsätzlich geeignet erscheint, ist es bislang wenig konkretisiert. Zunächst stellt sich die Erfassung der subjektiven Qualitätskriterien und deren Überführung in objektive Qualitätsindikatoren bei komplexen, realistischen Anwendungssituationen als problematisch dar. Insbesondere ist die Aussagekraft des erweiterten Datenmodells fraglich. Zwar werden die Qualitätsindikatoren im Modell erfasst, jedoch wird keine Aussage über deren Zielgröße getroffen. Weiter ist die Beziehung zwischen den Qualitätsanforderungen, den Qualitätsindikatoren sowie deren Operationalisierung nicht berücksichtigt.

### Produkt- und Prozessbezogene Messverfahren

Zahlreiche Ansätze zur Datenqualitätsmessung beruhen auf der Zuordnung von Werten zu bestimmten, auf Daten bezogene Kriterien und sollen daher als produktbezogene Messverfahren bezeichnet werden. Die hierbei angewendeten Messverfahren (siehe Beispielsweise [Redm96, S. 185 ff.; HuLW99, S. 60 ff.;

WSKL99; BoMY99; NaRo00, S. 149 ff.] unterscheiden sich hinsichtlich der Ebene der Semiotik, der Durchführung, des Umfangs und des Zeitbezugs.

Anhand der **Semiotik** können die Messverfahren in syntaktische, semantische und pragmatische Verfahren untergliedert werden. Verfahren auf der syntaktischen Ebene vergleichen Daten anhand syntaktischer Regeln. Semantische Verfahren betrachten die inhaltliche Bedeutung der Daten. Verfahren auf der pragmatischen Ebene beziehen sich auf die Datenanwendung und deren Zweck. Die **Messungsdurchführung** kann subjektiv durch den Anwender selbst oder objektiv durch einen Experten oder eine entwickelte Technik vorgenommen werden.<sup>2</sup> In bezug auf den **Messumfang** kann die Auswertungsmenge entweder alle Daten oder lediglich eine ausgewählte Stichprobe berücksichtigen. Hinsichtlich des **Zeitbezugs** kann die Messung dynamisch oder statisch erfolgen. Während dynamische Messungen die Historie, wie beispielsweise Erfahrungen des Anwenders, berücksichtigen, beziehen sich statische Verfahren lediglich auf einen zeitpunktbezogenen Zustand.

Neben den produktbezogenen Messverfahren existieren Ansätze die Aussagen über die Datenqualität in Abhängigkeit von durchgeführten Datentransformationen ermitteln.[BWPT98; More82] Im Vordergrund steht der Datenfluss und die Datentransformationen. Daher sollen diese Verfahren als prozessbezogene Messverfahren bezeichnet werden. Daten werden entlang des Datenflusses durch bestimmte Datentransformationen verändert. In Abhängigkeit derer Prozesseigenschaften beeinflussen diese wiederum die Datenqualität der Ausgabedaten. Mit Kenntnis der Qualität der Eingangsdaten und den Prozesseigenschaften kann so die sich resultierende Datenqualität abgeleitet werden. Hierbei ist die Bestimmung der Qualität der Eingangsdaten sowie der relevanten Prozesseigenschaften schwierig.

Vergleicht man die unterschiedlichen Ansätze zur Datenqualitätsmessung so kann zusammenfassend festgestellt werden, dass die Messverfahren sich meist nur auf wenige Messmöglichkeiten beziehen. (siehe Beispielsweise [HuLW99, S. 60 ff; WSKL99; BoMY99] Häufig wird zur Bestimmung der Datenqualität lediglich ein Messansatz berücksichtigt. Prozessbezogenen Messverfahren sind bislang wenig betrachtet. Insbesondere werden bislang produkt- und prozessbezogene Messverfahren selten zur Bestimmung der Datenqualität kombiniert. Die beabsichtigte Dissertation verfolgt dagegen den Ansatz, verschiedene Messverfahren geeignet zu kombinieren, so dass eine für den Datenverwender ausreichende Bestimmung der Datenqualität möglich wird.

---

<sup>2</sup> Anzumerken ist, dass obgleich die Bewertung durch Experten subjektive Einflüsse beinhaltet, von einer objektiven Messung gesprochen werden kann.[NaRo00, S. 156] Die Objektivität bezieht sich lediglich auf den Anwender und dessen Anwendungssituation.



## **Forschungsansatz und Vorgehensweise**

Bislang existiert keine Einigkeit über die in der Wirtschaftsinformatik anzuwendenden Forschungsansätze.[FKKT99, S. 71] Welche Forschungsmethoden anzuwenden sind, hängt in erster Linie von den verfolgten Forschungszielen, der Forschungsthematik sowie der Forschungsfrage ab. Forschungsziele der Wirtschaftsinformatik sind sowohl erkenntnis- als auch handlungsorientiert, wenngleich i. d. R. ein hoher Anwendungsbezug befürwortet wird.[Schü98, S. 11; MüK196, S. 5] Während beim theoretischen Wissenschaftsziel Erkenntnisse über den Gegenstand mit Hilfe wissenschaftlicher Theorien gewonnen werden, zeichnet sich die handlungsorientierte Wissenschaft durch einen hohen Pragmatismus aus.[Schü98, S. 10]

Die Dissertation verfolgt in erster Linie ein pragmatisches Gestaltungsziel im Rahmen anwendungsbezogener Forschung, die sich durch folgende Eigenschaften auszeichnet:[Ulri84, S. 202 f.]

- Die untersuchten Probleme entstehen in der Praxis,
- angewandte Forschung ist interdisziplinär,
- das Forschungsziel besteht in der Gestaltung der betrieblichen Realität durch Entwicklung von Modellen, Handlungsanleitungen und Instrumenten (siehe auch [Schü98, S. 10 f.]),
- die Aussagen sind normativ und wertend sowie
- das Fortschrittskriterium ist die praktische Problemlösungskraft der Modelle und Handlungsanweisungen (siehe auch [ÖsBH91, S. 35]).

Wird ein angewandtes Forschungsziel angestrebt, so eignen sich besonders qualitative Ansätze, die sich durch einen hohen Grad an Interdisziplinarität, durch eine verstärkte Verbindung der Wissenschaft mit der Praxis, durch eine Problem- und Handlungsorientierung, durch zyklisches Vorgehen und durch die Partizipation seitens der Praxis auszeichnen.[PrRa95, S. 8 ff.] Im Mittelpunkt des Dissertationsprojektes steht daher ein Fallbeispiel einer Schweizer Universalbank, mit dessen die Entwicklung und Umsetzung des Datenqualitätsmodells ermöglicht werden soll (explorativer Forschungscharakter [Krom94, S. 426]).

Häufige Kritik, die sich Forschung anhand von Fallstudien (insbesondere Einzelfallstudien) aussetzt, ist die der mangelnden Eignung generalisierbare Ergebnisse zu liefern. Fallstudien stellen weder repräsentative Beispiele einer bestimmten Grundgesamtheit dar, noch sagen sie etwas über die Häufigkeitsverteilung der beschriebenen Phänomene in der Realität aus. Sie erscheinen ebenfalls ungeeignet zur Plausibilisierung bestimmter Theorien oder Modelle. Ihre Berechtigung haben sie aber bei der Neu- und Weiterentwicklung von Modellen und Theorien.[Yin94, S. 10] Der Informationsgehalt von Fallbeispielen ist im Hinblick auf pragmatische

Aussagen extrem hoch, wengleich sich der Allgemeinheitsgrad i. d. R. geringer darstellt.[MüK196, S. 87]

Aus diesem Grunde erhebt die Dissertation nicht den Anspruch auf „statistische Generalisierbarkeit“ sondern bezweckt insbesondere die Entwicklung eines Datenqualitätsmodells im Sinne einer analytischen Generalisierbarkeit“ [Yin94, S. 10]. Als Instrumentarien stehen hierzu vor allem die Durchführung von Interviews, die direkte Beobachtung, die teilnehmende Beobachtung als auch die Untersuchung von Werkzeugen zur Verfügung.[Yin94, S. 85 ff.] Anhand des Fallbeispiels sollen mit Hilfe dieser Instrumentarien Qualitätsanforderungen erfasst, Messmöglichkeiten hinsichtlich produkt- und prozessbezogener Messverfahren identifiziert sowie die Intergrationsmöglichkeiten in die Metadatenverwaltung des Data-Warehouse-Systems untersucht werden. Das so entwickelte Datenqualitätsmodell für Data-Warehouse-Systeme soll anschliessend hinsichtlich möglicher Verallgemeinerungen überprüft werden.

Neben den zu identifizierenden Messverfahren ist die Auswahl relevanter Qualitätskriterien für die Dissertation zentral. Daher soll, ausgehend von den in der Literatur genannten Datenqualitätskriterien, zu Beginn der Dissertation deren Relevanz und Priorität sowie deren Beziehungstruktur durch eine empirische Studie ermittelt werden. Aufgrund von Erfahrungen im Rahmen der Projektarbeit sowie der in der Literatur belegten Anspruchsgruppen mit individuellen Zielen [JLVV00, S. 137; Meye00, S. 87], wird von aufgaben- und anspruchsspezifischen Qualitätsanforderungen ausgegangen. Diese Annahme soll im Rahmen der empirischen Studie zumindest gestützt (bzw. erweitert) werden.

## **Zusammenfassung und Ausblick**

Im Rahmen des Artikels wurde die Relevanz des Datenqualitätsmanagements für Data-Warehouse-Systeme betrachtet. Insbesondere zeigen sich noch offene Fragestellungen bzgl. der Datenqualitätsplanung und -lenkung. Hierzu sind relevante Datenqualitätsmerkmale auszuwählen und ein Datenqualitätsmodell zu entwickeln, so dass die Qualitätsvorgaben durch quantitativen Kennziffern überprüft werden können. Weiter ist die Integration der Qualitätsinformationen in die Metadatenverwaltung noch nicht umfassend gelöst.

Ziel der Dissertation ist es daher, ein Datenqualitätsmodell zur Spezifikation und Messung der Datenqualität in Data-Warehouse-Systemen zu entwickeln und Möglichkeiten für die Integration in die Metadatenverwaltung aufzuzeigen. Hierzu soll in einem ersten Schritt der allgemeine Datenqualitätsbegriff im Rahmen einer empirischen Untersuchung weiter operationalisiert und relevante Datenqualitätskriterien ausgewählt werden. Diese bilden die Grundlage zur Entwicklung eines Datenqualitätsmodells anhand eines Fallbeispiels einer Schweizer Universalbank.

Anhand dieses Anwendungsszenarios sollen mögliche Messverfahren, Messpunkte und Messzeitpunkte untersucht und geeignete ausgewählt werden. Abschliessend soll die Übertragung des erstellte Datenqualitätsmodells auf mögliche weitere Anwendungsfälle betrachtet werden.

## Literatur

[BaPa85] Ballou, D. P.; Pazer, H. L.: Modeling Data Process Quality in Multi-input, Multi-output Information Systems. In: Management Science 31 (1985) 4, S. 150-162.

[BaWe84] Basili, V. R.; Weiss, D. M.: A methodology for Collecting Valid Software Engineering Data. In: IEEE Transaction on Software Engineering 10 (1984) 6, S. 728-738.

[Bode97] Bode, J.: Der Informationsbegriff in der Betriebswirtschaftslehre. In: Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung 49 (1997) 5, S. 449-468.

[BoMY99] Bobrowski, M.; Marre, M.; Yankelevich, D.: A homogeneous framework to measure data quality. In: Proceedings of the 1999 International Conference on Information Quality, Cambridge, MA 1999, S. 115-124.

[BWPT98] Ballou, D. P.; Wang R. Y.; Pazer, H.; Tayi, G. K.: Modelling information manufacturing systems to determine information product quality. In: Management Science 44 (1998) 4, S. 462-484.

[Dev197] Devlin, B.: Data Warehouse: From architecture to implementation. Addison-Wesley Longman, Reading, MA et al. 1997.

[DIN95] o. V.: ISO 8042: Qualitätsmanagement – Begriffe. Beuth, Berlin 1995.

[Eng199] English, L.: Improving Data Warehouse and Business Information Quality. Wiley, New York 1999.

[FKKT99] Frank, U.; Klein S.; Krcmar H.; Teubner, A.: Aktionsforschung in der WI - Einsatzpotentiale und Einsatzprobleme. In: Schütte R.; Siedentopf, J.; Zelewski, S. (Hrsg.): Wirtschaftsinformatik und Wissenschaftstheorie. Arbeitsbericht des Instituts für Produktion und Industrielles Informationsmanagement Nr. 4, Universität GH Essen, Essen 1999, S. 71-90.

[Garv84] Garvin, D. A.: What does "Product Quality" really mean?. In: Sloan Management Review 26 (1984) 1, S. 25-43.

[Hauk84] Hauke, P.: Informationsverarbeitungsprozesse und Informationsbewertung. GBI, München 1984.

- [Häus98] Häussler, C.: Datenqualität. In: Martin, W. (Hrsg.): Data Warehousing. ITP, Bonn 1998, S. 75-89.
- [Helf00a] Helfert, M.: Massnahmen und Konzepte zur Sicherung der Datenqualität. In: Jung, R.; Winter, R. (Hrsg.): Data Warehousing Strategie – Erfahrungen, Methoden, Visionen. Springer, Berlin et al. 2000.
- [Helf00b] Helfert, M.: Eine empirische Untersuchung von Forschungsfragen beim Data Warehousing aus Sicht der Unternehmenspraxis. Arbeitsbericht des Instituts für Wirtschaftsinformatik der Universität St. Gallen BE HSG/CC DWS/05, St. Gallen 2000.
- [Helf01] Helfert, M.: Managing and Measuring Data Quality in Data Warehousing. In: Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics, Orlando, FL 2001, S. 55-65.
- [Holt99] Holthuis, J.: Der Aufbau von Data Warehouse-Systemen: Konzeption – Datenmodellierung – Vorgehen. Dt. Univ.-Verlag; Gabler, Wiesbaden 1999.
- [HuLW99] Huang, J., Lee, Y. W., Wang, R. Y.: Quality Information and Knowledge. Prentice Hall, Upper Saddle River, NJ 1999.
- [JaVa97] Jarke, M.; Vassiliou, Y.: Foundations of Data Warehouse Quality – A Review of the DWQ-Project. In: Proceedings of the 2nd International Conference on Information Quality, Cambridge, MA 1997, S. 299-313.
- [JLVV00] Jarke, M.; Lenzerini, M.; Vassiliou, Y.; Vassiliadis, P.: Fundamentals of Data Warehouses. Springer, Berlin et al. 2000.
- [Krom94] Kromrey, H.: Empirische Sozialforschung: Modelle und Methoden der Datenerhebung und Datenauswertung. 6. Aufl., Leske und Budrich, Opladen 1994.
- [Laud86] Laudon, K. C.: Data quality and due process in large interorganizational record systems. In: Communication of the ACM 29 (1986) 1, S. 4-11.
- [MaLe95] Maier, R.; Lehner, F.: Daten, Informationen, Wissen. In: Lehner, F.; Maier, R.; Hildebrand, K. (Hrsg.): Wirtschaftsinformatik. Hanser, München; Wien 1995, S. 165-272.
- [Meye00] Meyer, M.: Organisatorische Gestaltung des unternehmensweiten Data Warehousing. Difo-Druck, Bamberg 2000.
- [More82] Morey, R. C.: Estimating and improving the quality of information in the MIS. In: Communication of the ACM 25 (1982) 5, S. 337-342.
- [MüKl96] Müller-Boling, D.; Klandt, H.: Methoden empirischer Wirtschafts- und Sozialforschung: eine Einführung mit wirtschaftswissenschaftlichem Schwerpunkt. 3. Aufl., Förderkreis Gründungs-Forschung, Köln; Dortmund 1996.

- [Müll00] Müller, J.: Transformation operativer Daten zur Nutzung im Data Warehouse. Dt. Univ.-Verlag; Gabler, Wiesbaden 2000.
- [NaRo99] Naumann, F.; Rolker, C.: Assessment Methods for Information Quality Criteria. In: Proceedings of the 2000 Conference on Information Quality, Cambridge, MA 2000, S. 148-162.
- [ÖsBH91] Österle, H.; Brenner, W.; Hilbers, K.: Unternehmensführung und Informationssysteme: Der Ansatz des St. Galler Informationssystem-Managements. 2. Aufl., Teubner, Stuttgart 1991.
- [PrRa95] Probst, G. J. B.; Raub, S. P.: Action Research - Ein Konzept angewandter Managementforschung. In: Die Unternehmung 49 (1995) 1, S. 3-19.
- [Redm96] Redman, T. C.: Data Quality for the Information Age. Artech House, Norwood 1996.
- [Schü98] Schütte, R.: Grundsätze ordnungsmässiger Referenzmodellierung: Konstruktion konfigurations- und anpassungsorientierter Modelle. Dt. Univ. Verlag; Gabler, Wiesbaden 1998.
- [Segh96] Seghezzi, H. D.: Integriertes Qualitätsmanagement: das St. Galler Konzept. Hanser, München; Wien 1996.
- [Stre96] Streubel, F.: Theoretische Fundierung eines ganzheitlichen Informationsmanagements. Arbeitsbericht des Lehrstuhls für Wirtschaftsinformatik der Ruhr-Universität Bochum, Nr. 96-21, Bochum 1996.
- [StWa98] Storey, V. C.; Wang, R., Y.: Modeling Quality Requirements in Conceptual Database Design. In: Proceedings of the 1998 Conference on Information Quality, Cambridge, MA 1998, S. 64-87.
- [TaBa98] Tayi, G. K., Ballou D.: Examining Data Quality. In: Communications of the ACM, 41 (1998) 2, S. 54-57.
- [TuWa93] Tu, S.; Wang, R. Y.: Modeling Data Quality and Context through Extension of the ER Model. In Hevner, A; Kamel, N. (Hrsg.): Third Annual Workshop on Information Technologies and Systems (WITS-93), Orlando, FL 1993, S. 40-47.
- [Ulri84] Ulrich, H.: Managment. Haupt, Bern 1984.
- [WaKM93] Wang, R. Y.; Kon, H. B.; Madnick, S. E.: Data Quality requirements analysis and modeling. In: Proceedings of the 9th international conference on data engineering (ICDE), IEEE Computer Society, Wien 1993, S. 670-677.
- [Wall90] Wallmüller, E.: Software-Qualitätssicherung in der Praxis. Hanser, München; Wien 1990.

[WaRK95] Wang, R. Y.; Reddy, M. P.; Kon, H. B.: Towards quality data: An attribute-based approach. In: Journal of Decision Support Systems 13 (1995) 3/4, S. 349-372.

[WaSF95] Wang, R. Y.; Storey, V. C.; Firth, C. P.: A Framework for analysis of data quality research. In: IEEE Transactions on Knowledge and Data Engineering 7 (1995) 4, S. 623-640.

[WaSt96] Wang, R. Y.; Strong, D. M.: Beyond Accuracy: What Data Quality Means to Data Consumers. In: Journal of Management Information Systems 12 (1996) 4, S. 5-33.

[WaZL01] Wang, R. Y.; Ziad, M.; Lee, Y. W.: Data Quality. Kluwer Academic Publishers, Boston; Dordrecht; London 2001.

[Wint00] Winter, R.: Zur Positionierung und Weiterentwicklung des Data Warehousing in der betrieblichen Applikationsarchitektur. In: Jung, R.; Winter, R. (Hrsg.): Data Warehousing Strategie: Erfahrungen, Methoden, Visionen. Springer, Berlin et al. 2000, S. 127-139.

[Wolf99] Wolf, P.: Konzept eines TQM-basierten Regelkreismodells für ein „Information Quality Management“ (IQM). Praxiswissen, Dortmund 1999.

[WSKL99] Wang, R. Y.; Strong, D. M.; Kahn, B. K.; Lee, Y. W.: An information quality assessment methodology. In: Proceedings of the 1999 international Conference of Information Quality, Cambridge, MA 1999, S. 258-265.

[Yin94] Yin, R. K.: Case study research: Design and Methods. 2. Aufl., Sage, Thousand Oaks; London; New Delhi 1994.