

Institut für Wirtschaftsinformatik



Universität St.Gallen

## A Framework for Planing and Controlling Data Quality in Data-Warehouse-Systems

Markus Helfert  
markus.helfert@unisg.ch

Software Trends.ch  
Markus Helfert  
Slide 2

### Überblick

Data-Warehouse-Systeme und Datenqualität

Datenqualitätsmanagement

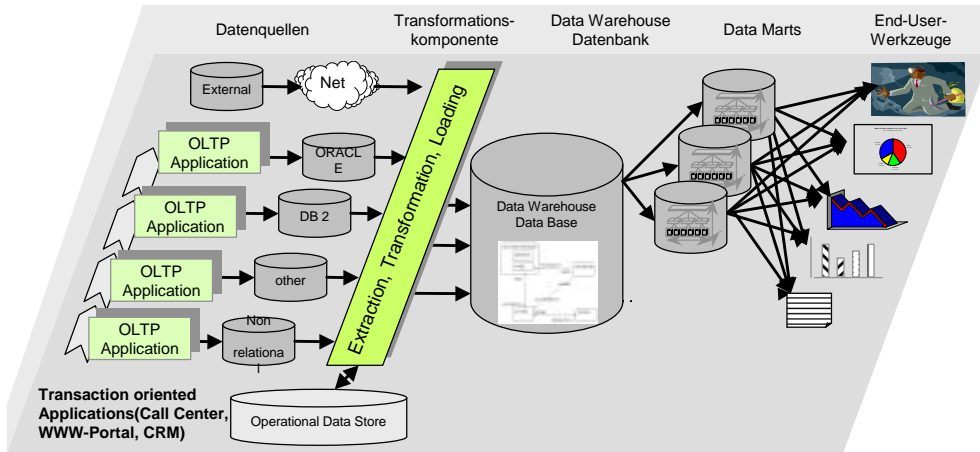
*Datenqualität*

*Modell zur Qualitätsplanung und -lenkung*

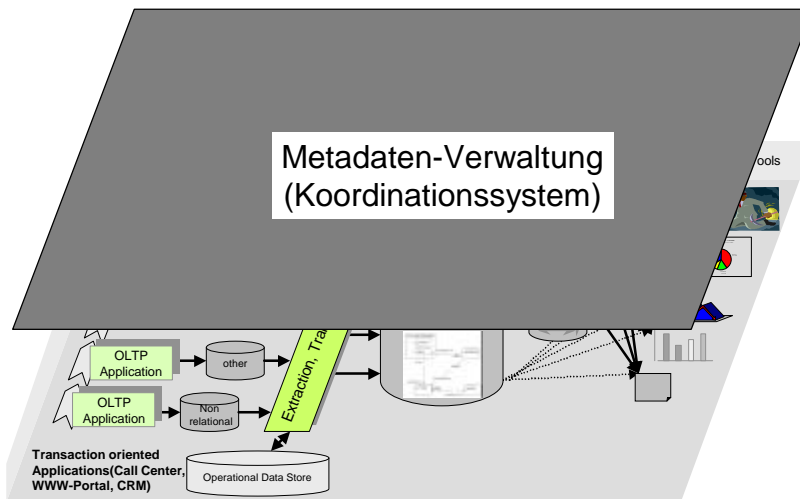
*Mängel bisheriger Forschungsansätze*

Zusammenfassung / Offene Fragen

## Data-Warehouse-Basissystem



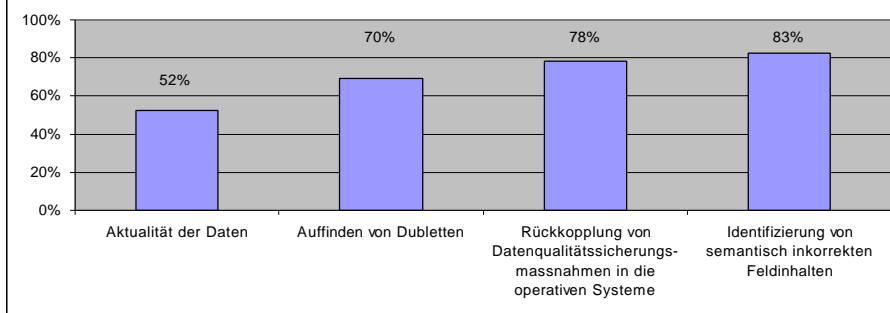
## Metadaten-Verwaltung



## DQ ein Hauptproblem in Data-Warehouse-Systemen

Folgende Aspekte werden als problematisch eingeschätzt:

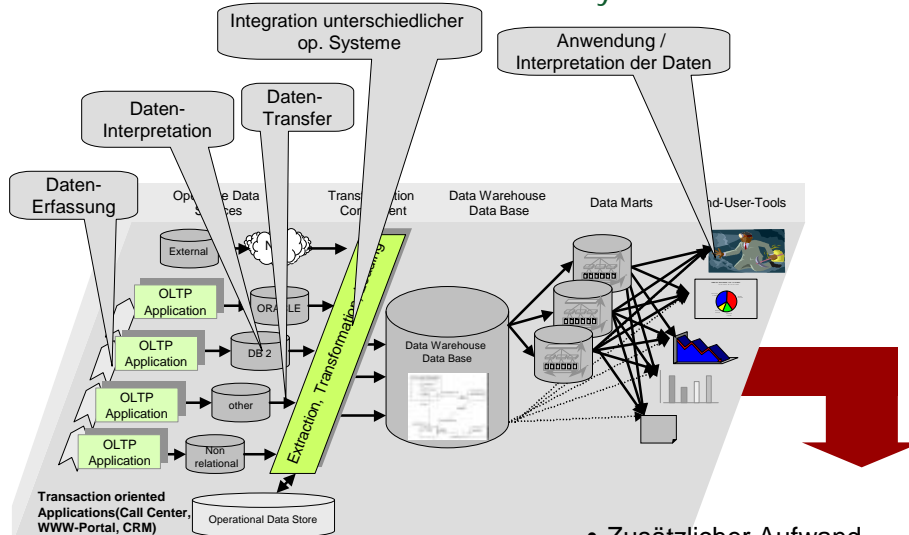
(% der erhaltenen Antworten)



Ergebnisse einer CC DWS-Umfrage 1999

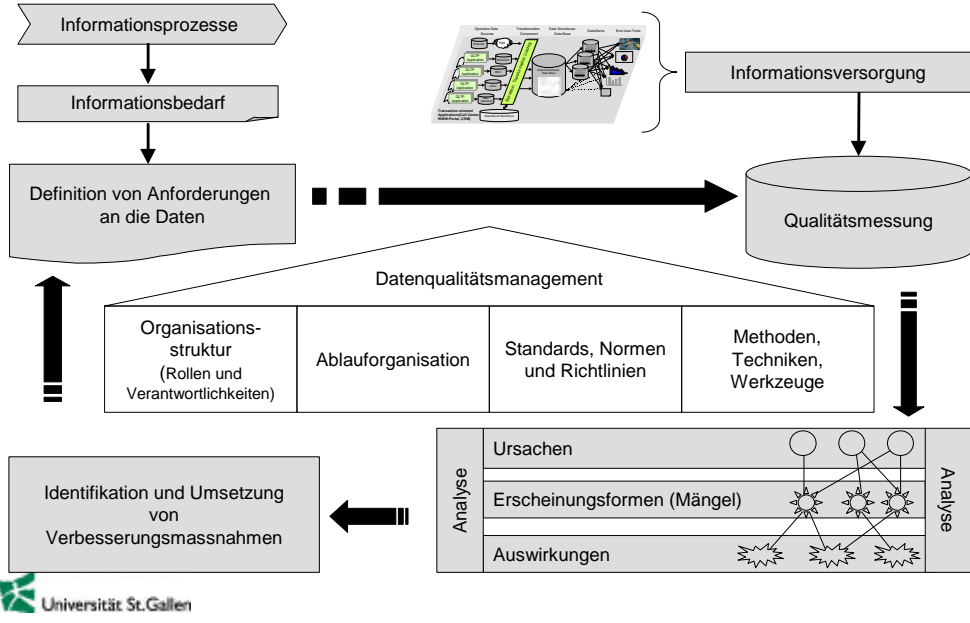
- Für nahezu alle Unternehmen stellt die Sicherstellung der Datenqualität beim Data Warehousing eine Herausforderung dar
- Semantik und die Rückkopplung in die operativen Systemen sind problematisch

## Problemfelder im DW-Basisystem



- Zusätzlicher Aufwand
- Reduzierte Akzeptanz / Nutzung
- Mangelnde Entscheidungen

## Datenqualitätsmanagement



## Datenqualität / Sichtweisen (entsprechend Garvin 1984)

**Anwenderbez. Ansatz:** Qualität wird durch den Produktnutzer festgelegt. Qualitative Produkte dienen dem Zweck der Benutzung durch den Kunden während des Gebrauchs.

Designqualität

Wie gut werden die Bedarfe durch eine Produktspezifikation erfüllt?

**Produktbasierte Ansatz:** Materielle Produkteigenschaften bestimmen die Qualität eines Produktes (inhärente Eigenschaft des Produktes)

Prozessqualität

Wie gut stimmt das aktuelle Produkt mit der vorgegebenen Spezifikation überein?

**Prozessbez. Ansatz:** Qualität bedeutet die Einhaltung von Spezifikationen und die Abwesenheit von Fehlern

Qualitätsplanung

Qualitätslenkung



Informationsbedarfe

Daten und Funktionen des DWH-Systems

Datenversorgung



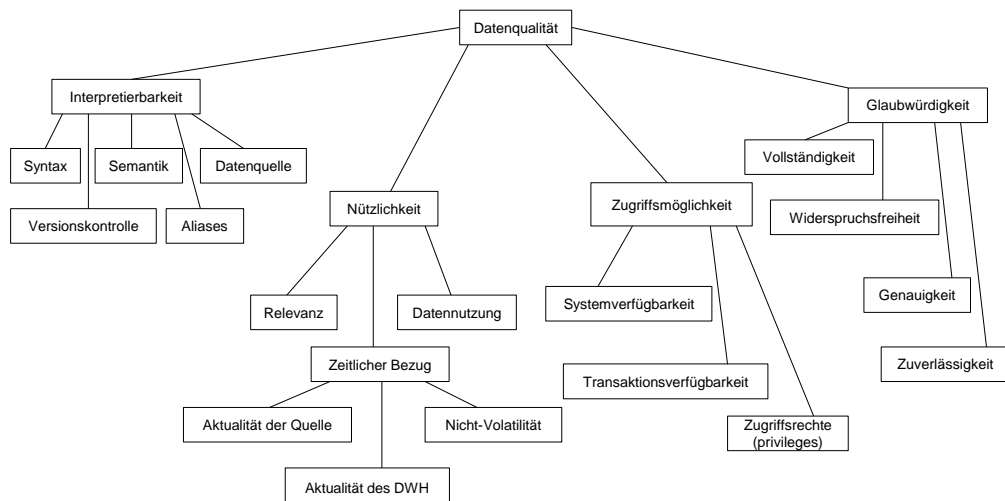
- Wertbezogene Ansatz: Stellt Qualität i. S. v. Nutzen in Bezug zu den Kosten
- Der *transzendente Ansatz* kennzeichnet Qualität als angeborene Einzigartigkeit oder Superlative (eher abstrakt philosophisches Verständnis)

## Ansätze in der Literatur (Beispiele)

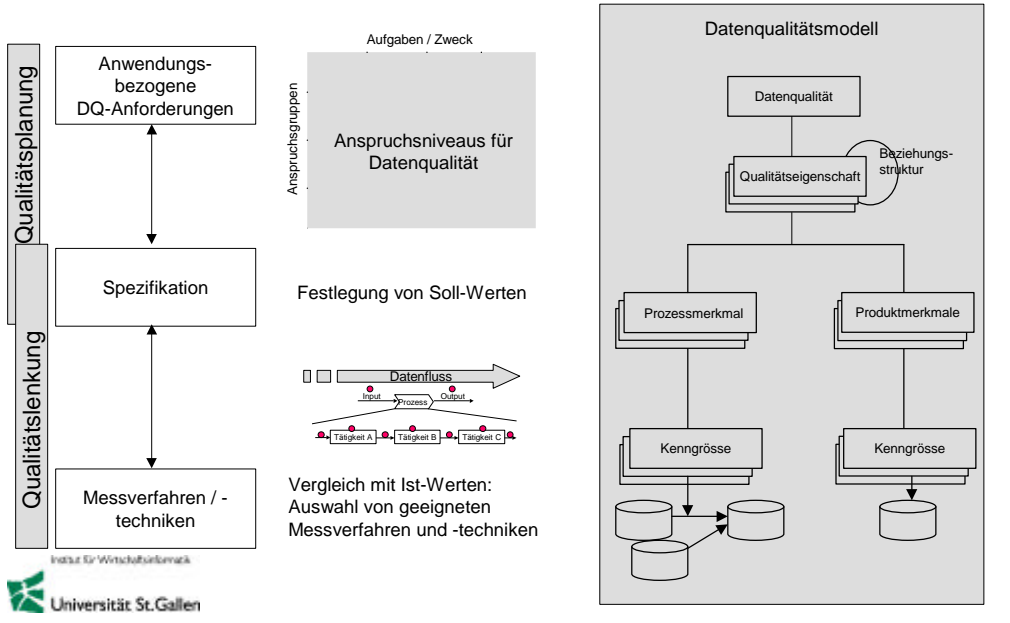
Autor and Jahr der Publikation	Anwendungskontext
Augustin / Reminger 1990	Management Information Systems
Morris et al. 1996	Management
Redmann 1996	Datenbanken / Informationssysteme
Miller 1996	Informationssysteme
Wang / Strong 1996	Datenbanken / Informationssysteme
Davenport 1997	Informationsmanagement
Ballou et al. 1998	Data-Warehouse-Systeme
Kahn / Strong 1998	Informationssysteme
Rittberger 1999	Informationsanbieter
English 1999	Data-Warehouse-Systeme
Huang et al. 1999	Knowledge Management
Jarke et al. 2000	Data-Warehouse-Systeme

- ♦ Häufig für einen spezifischen Anwendungskontext
- ♦ Unterschiedliche Definitionsansätze (IQ / DQ) und -sichtweisen (Meist eine anwendungsbez. Sichtweise)
- ♦ Allgemein: Genauigkeit, Konsistenz, Vollständigkeit, Relevanz, Objektivität, Zugriff, Zeitlicher Bezug (Aber häufig keine empirische Überprüfung)
- ♦ Meist werden keine Beziehungen zwischen den Kriterien berücksichtigt
- ♦ Meist keine Messung / Operationalisierung der Kriterien (Qualitätsindikatoren)

## Datenqualitätskriterien für DWH-Systeme



## Modell zur Qualitätsplanung und -lenkung



## Mängel bisheriger Forschungsansätze (1)

### Erfassung (bzw. Modellierung) von Datenqualitätsanforderungen (Qualitätsplanung)

- Bisher wenig Beachtung
  - Erweiterung des ER-Modellierung (Wang et al.)
  - Basierend auf dem Goal-Question-Metric-Ansatz (Jarke et al.)
  - Quality Function Deployment (Kundenanforderungen~Qualitätskriterien) (Redman)
- Nur grobe Vorgehensweise
- Keine Aussage über die generelle Zielgrößen (?) oder deren Ermittlung (Aufgaben- und Anspruchsgruppenspezifisch ?)
- Kein Bezug zwischen den Qualitätsvorgaben und deren Messung

Wie könne Datenqualitätsanforderungen erfasst und beschrieben werden?

## Mängel bisheriger Forschungsansätze (2)

### Produktbezogene Messverfahren (Qualitätslenkung)

- > Beruhen auf der Zuordnung von Werten zu bestimmten Qualitätskriterien
- > Häufig basierend auf Fragebogen (subjektiv) oder bezogen auf datenbankspez Aspekte (Integritätsbedingungen)
- > Wenig Kombination verschiedener Messverfahren
- > Messverfahren sind lediglich beschrieben und wenig konkret ausgestaltet

### Prozessbezogene Messverfahren (Qualitätslenkung)

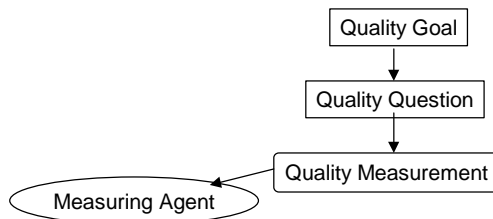
- > Bislang wenig berücksichtigt.
- > Kenntnis der Eingangsdatenqualität und der transformationsspezifischen Eigenschaften problematisch
- > Ermittlung der DQ in Abhängigkeit der vorgenommenen Datentransformationen

Wie kann Datenqualität in Data-Warehouse-Systemen gemessen werden?

## Mängel bisheriger Forschungsansätze (3)

### Integration in die Metadaten-Verwaltung

- > Repräsentation in einem Metamodell für DWH-Systeme (Jarke et al.)



- > Nicht primär am Datenfluss und den Datentransformationsprozessen orientiert.
- > Die Messung findet anhand von Entitäten und Relationen statt. (Produktkriterien)
- > Sowohl im Common Warehouse Metamodel (CWM) als auch in Produkten nicht berücksichtigt.

Wie kann die Qualitätsplanung und -lenkung in die Metadaten-Verwaltung integriert werden?

## Zusammenfassung / Offene Fragen



- > Datenqualität eines der wichtigsten Probleme beim Data Warehousing
- > Unterschiedliche Qualitätssichtweisen
- > Unterscheidung zwischen Design- und Prozessqualität
- > Kombination verschiedener Messtechniken (Produktmerkmale / Prozessmerkmale)

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

- > Welche Anforderungen haben Endbenutzer an die Daten (Datenqualität?, anwendergruppen- und aufgabenbezogene Anspruchsniveaus)?
- > Wie können die Qualitätsanforderungen in Qualitätsvorgaben transformiert werden?
- > Wie können die Qualitätsvorgaben gemessen werden (Messpunkte, Messzeitpunkte sowie Messtechniken)?
- > Wie kann die Qualitätsplanung und Qualitätslenkung innerhalb eines DWH-Systems berücksichtigt werden (Integration in die Metadaten-Verwaltung)?