

Datenqualitätsmodelle

Stefan Willi Hart

3. Dezember 2009



INF

FAKULTÄT FÜR
INFORMATIK

1 Motivation

2 Modelle

- Systemseitiges DQ-Modell
 - Allgemein
 - Aufbau des DQ-Modells
- Empirisches DQ-Modell
 - Allgemein
 - DQ-Klassen und ihre Merkmale I
 - DQ-Klassen und ihre Merkmale II

3 Ausblick

4 Zusammenfassung

5 Literatur

Allgemein Teil I

Datenqualität

ist die Gesamtheit der Ausprägungen von Qualitätsmerkmalen eines Datenbestandes bezüglich dessen Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen.

[DMC04, Seite 4]

Modell

”Modell \langle ital. \rangle das, Muster, Vorbild; Nachbildung oder Entwurf von Gegenständen (vergrössert, verkleinert, in natürlicher Grösse). Modell können ausser wirklichen Gegenständen auch gedankliche Konstruktionen sein.”

[Brockh71,]



INF

FAKULTÄT FÜR
INFORMATIK

Allgemein Teil II

- Betrachtung über den gesamten Lebenszyklus
- müssen im Zusammenhang betrachtet werden

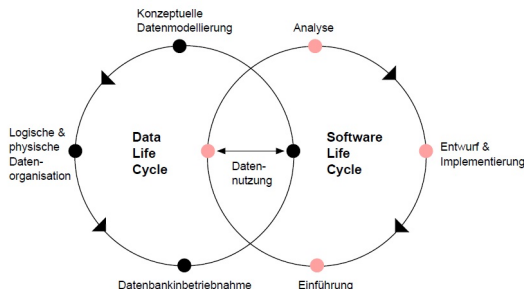


Abbildung: System-Bi-Cycle

[Hin01a, Kapitel 3 S.26]

Modelle

- **Systemseitiges Datenqualitätsmodell** von E. Borowski und H.J. Lenz
- **Empirisches Datenqualitätsmodell** von H. Hinrich

Systemseitiges DQ-Modell: Allgemein

Basis des Modells:

- Betrachtung von Zuständen des modellierten Abschnittes
- falsch zugeordnete Zustände im Informationssystem
- z.B. mehrdeutige/bedeutungslose Zustände oder unvollständige falsche Repräsentation

Systemseitiges DQ-Modell: Aufbau

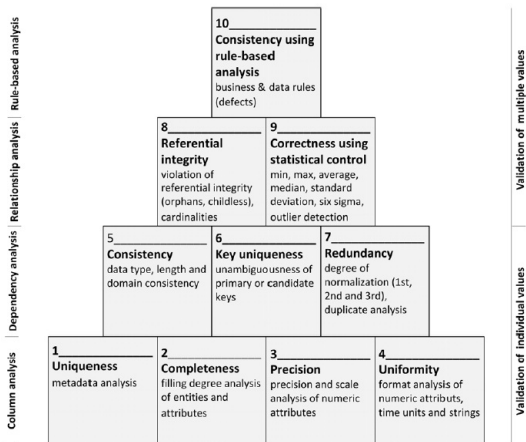


Abbildung: DQ-Dimensionen



Empirisches DQ-Modell: Allgemein

Basis des Modells:

- Beziehungen zwischen Merkmalen der Bedeutsamkeit
- von Merkmalen auf Basis von Kundenbefragungen
- Anzahl der DQ-Merkmale kann sehr groß sein

Empirisches DQ-Modell: Merkmale I

- Glaubwürdigkeit
 - Korrektheit
 - Konsistenz
 - Zuverlässigkeit
- Nützlichkeit
 - Vollständigkeit
 - Genauigkeit
 - Zeitnähe
 - Redundanzfreiheit
 - Relevanz

Empirisches DQ-Modell: Merkmale II

- Interpretierbarkeit
 - Einheitlichkeit
 - Eindeutigkeit
 - Verständlichkeit
- Schlüsselintegrität
 - Schlüsseleindeutigkeit
 - Referentielle Integrität



INF

FAKULTÄT FÜR
INFORMATIK

Ausblick

weitere Ansätze:

- intuitiver Ansatz:
 - Merkmale aufgrund intuitiven Verständnisses eines Experten
- Dienstleistungsorientierter Ansatz:
 - Quality of Service
 - Lieferung von Informationen an Informationsverbraucher
- Datenbanksystemorientierter Ansatz:
 - klassische 3-Ebenen Architektur
 - interne Schema (technische Merkmale)
 - logische Schema (Vollständigkeit und Konsistenz)
 - externe Schema (kontextabhängige Merkmale und Aktualität)

Zusammenfassung

- Modelle haben verschieden Ansichtsweisen
- Datenqualitätsmerkmale unterschiedlich Definiert

Ergebnis: Datenqualitätsmodelle sollen Datenqualitätsmerkmale definieren \Rightarrow nicht Allgemeingültig






Fragen?



INF

FAKULTÄT FÜR
INFORMATIK

Literatur

-  H. Hinrichs. Datenqualitätsmanagement in Data Warehouse-Umgebung, 2001
-  E. Borowski und H.J. Lenz, Design of a workflow system to improve data quality using Oracle warehouse builder 2008
-  www.wikipedia.org Stichwort: Modell Kategorie: Wissenschaftstheorie Stand 01.12.09
-  Dr. Ulrich Windheuser Vortrag auf den 5. Data-Mining-Cup Anwendertagen 2004
-  Brockhaus Enzyklopädie Band 12, 1971