Datenqualität in medizinisch-betriebswirtschaftlichen Informationssystemen

MedConf 2013

Endler Gregor, 16.10.2013

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Warum Datenqualität?

2002, USA:
600.000.000 $

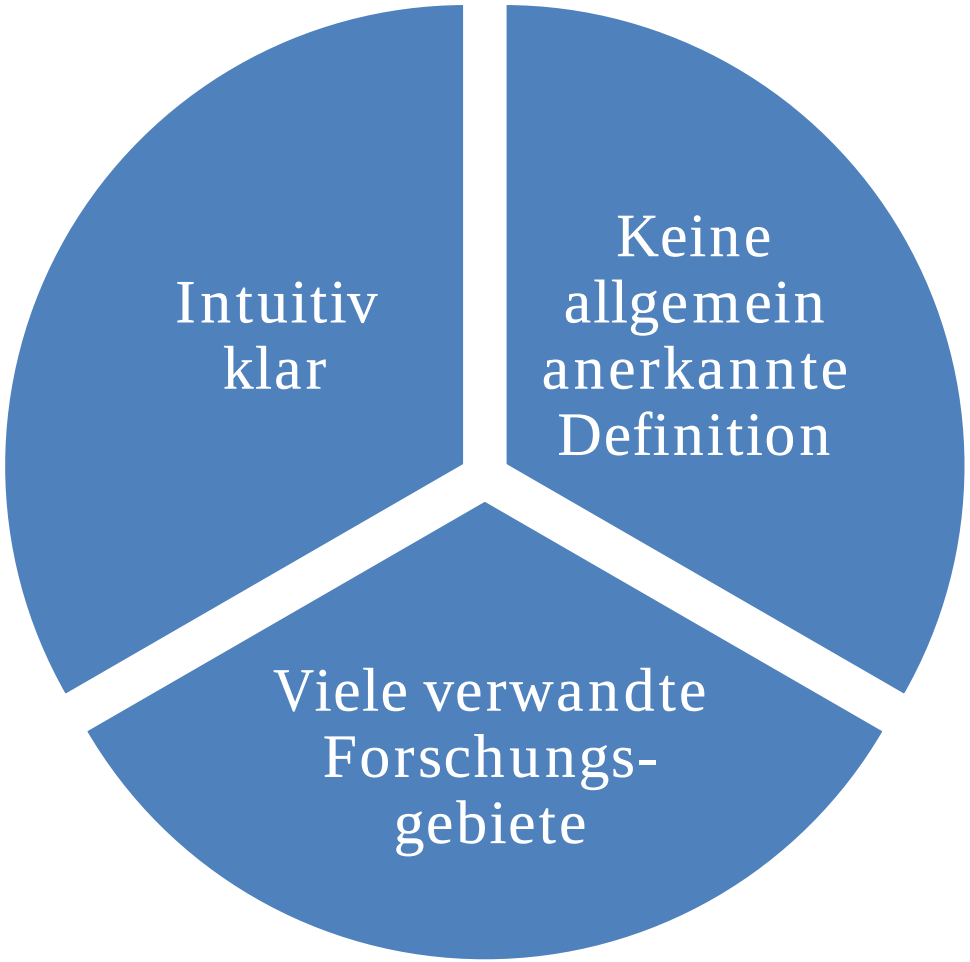
Y2k weltweit:
1.500.000.000 $

Kosten

44.000 – 98.000 Todesfälle durch Fehler

Fehlende Info: bis 81% der Fälle

Was ist Datenqualität?

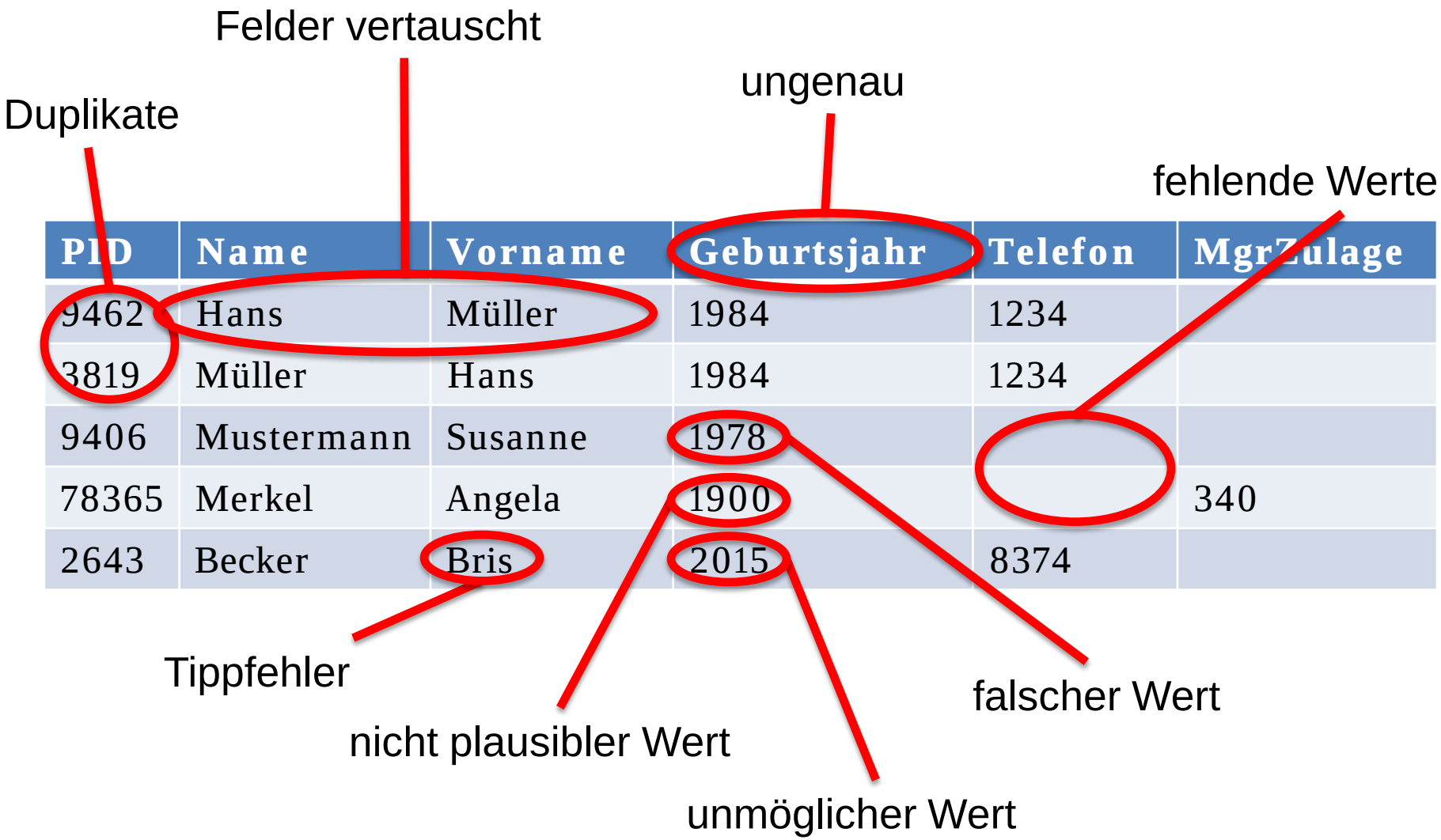


FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Generisch: „Fitness for Use“

1. Datenqualität ist subjektiv

Beispiel



PID	Name	Vorname	Geburtsjahr	Telefon	MgrZulage
9462	Hans	Müller	1984	1234	
3819	Müller	Hans	1984	1234	
9406	Mustermann	Susanne	1978		
78365	Merkel	Angela	1900		340
2643	Becker	Bris	2015	8374	

Was ist Datenqualität?

Viele unterschiedliche Arten von „Qualität“

1. Datenqualität ist subjektiv

2. Datenqualität ist multidimensional

Korrektheit, Genauigkeit

Korrektheit

- Übereinstimmung Datenwert - Realwelt
- Syntaktisch
  vs.
  Semantisch

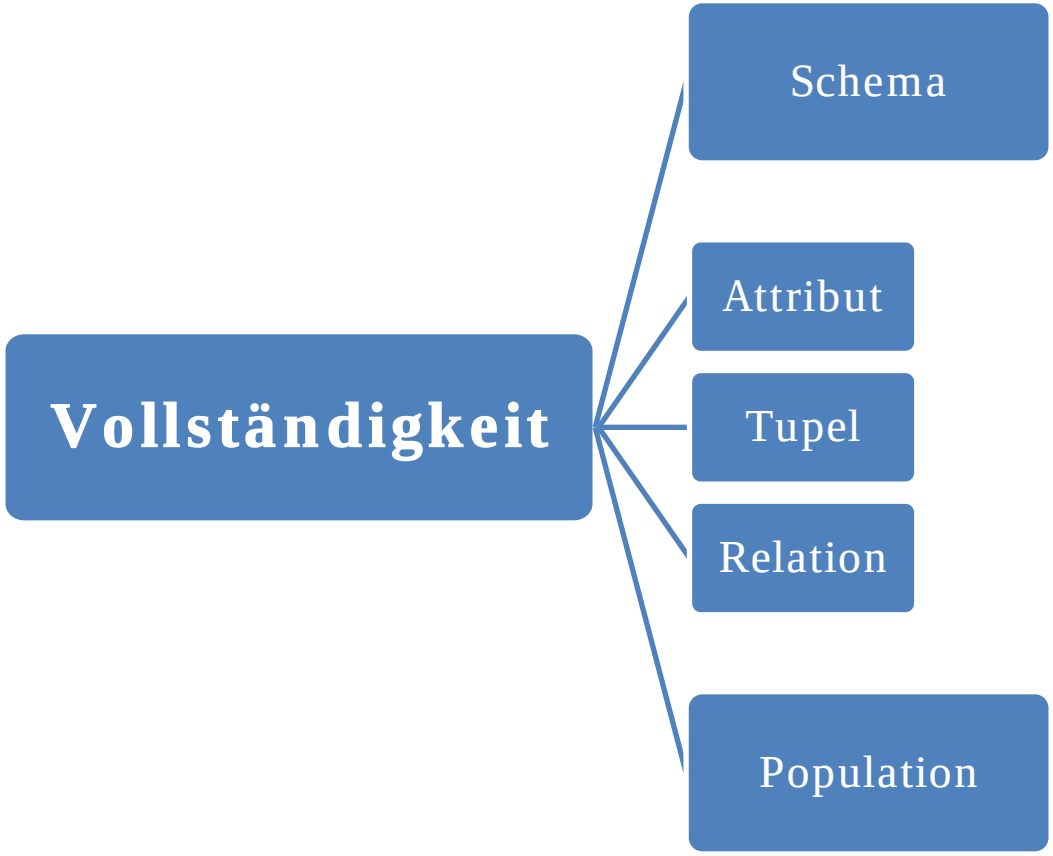
kontext*un*abhängig

Genauigkeit

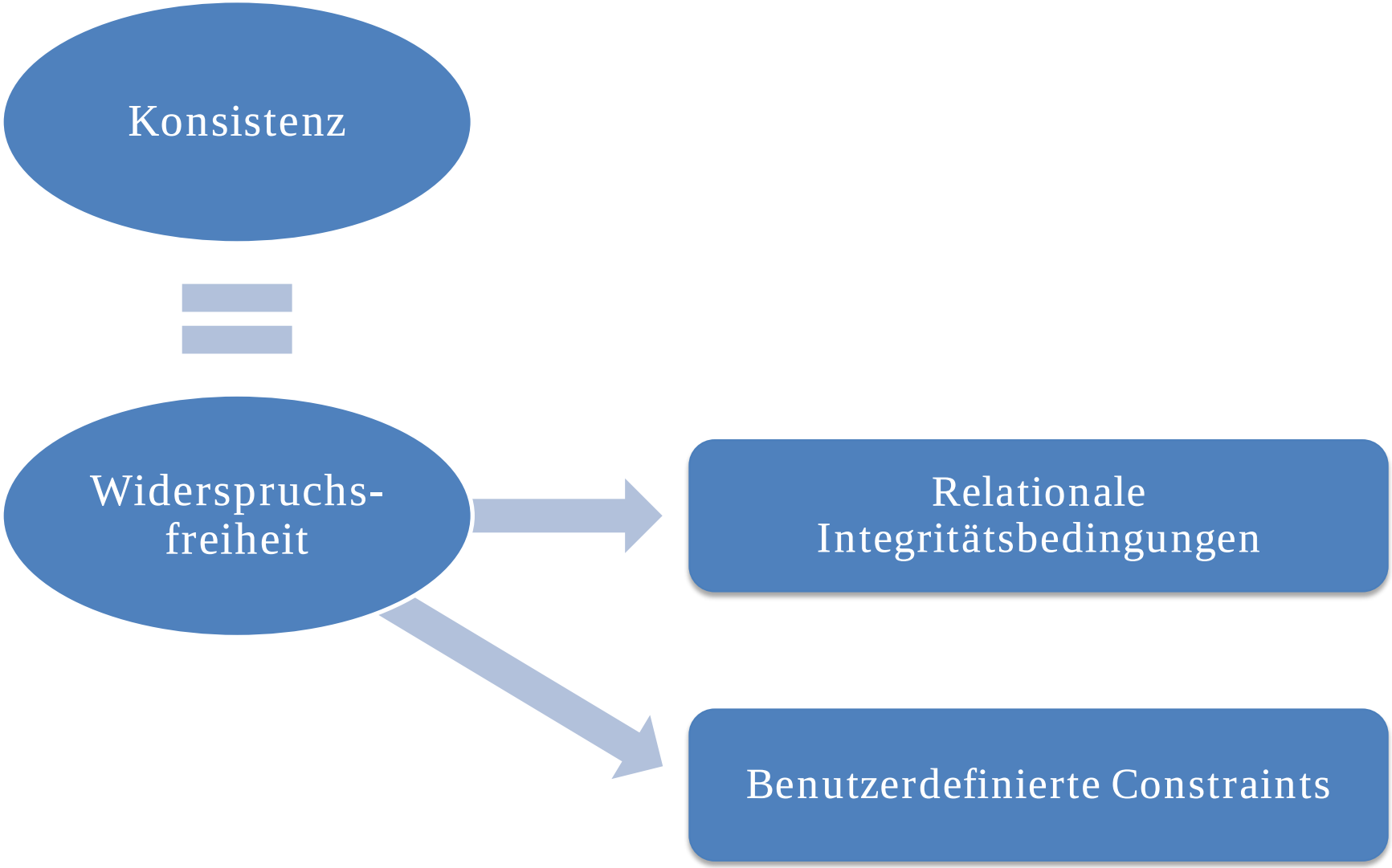
- Abstand Datenwert - Realwelt

kontextabhängig

Vollständigkeit



Konsistenz


Konsistenz
=
Widerspruchs-
freiheit
Relationale
Integritätsbedingungen
Benutzerdefinierte Constraints

Zeitbezogene Dimensionen

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Aktualität

Daten veraltet?

Zeitnähe

Zeitgerechte Bereitstellung?

Wechselwirkungen

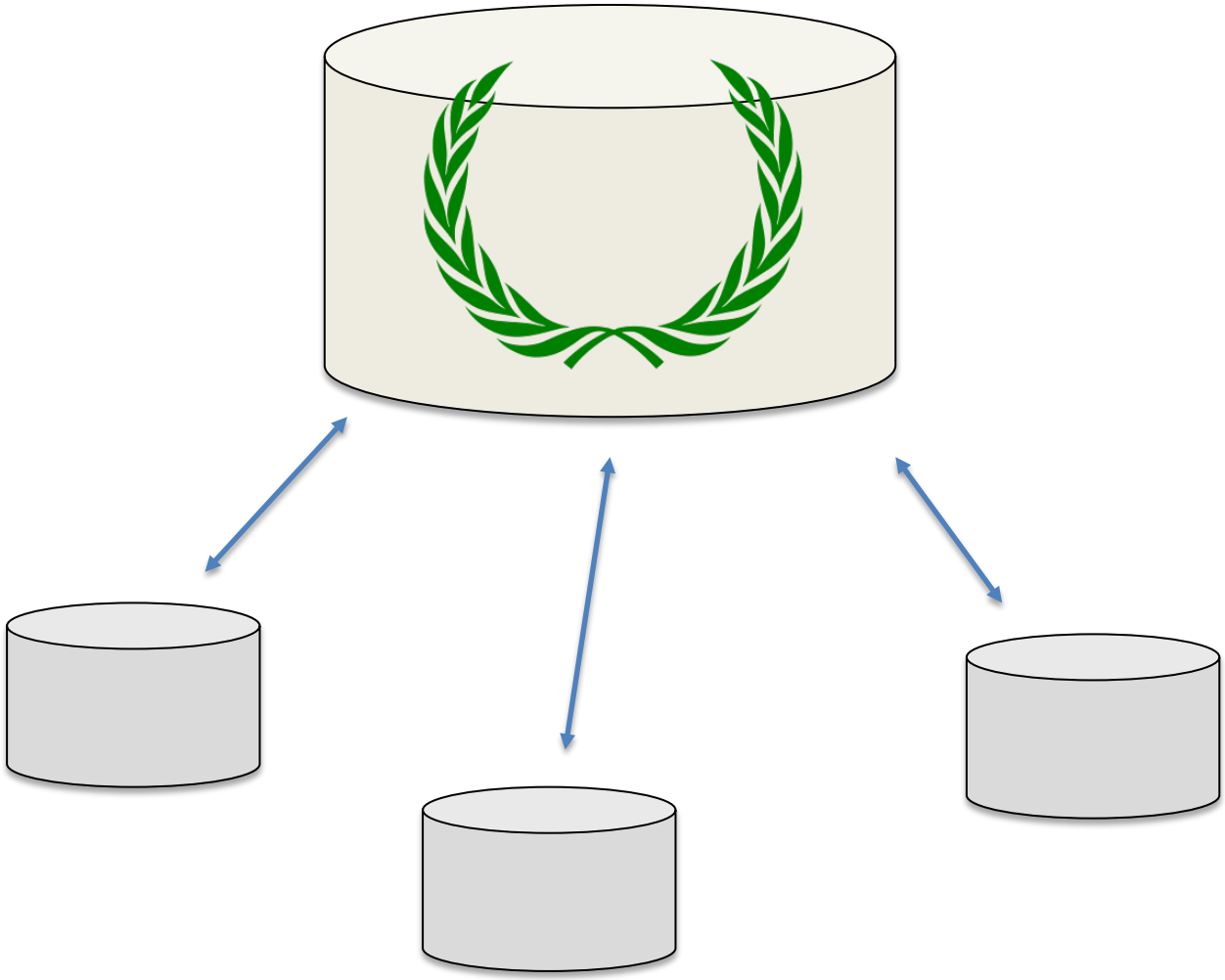
Wechselwirkungen zwischen Dimensionen

1. Datenqualität ist subjektiv
2. Datenqualität ist multidimensional
3. DQ-Dimensionen sind nicht unabhängig

Messen von Datenqualität

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Referenzdaten



Maße

Korrektheit

- Syntaktisch:
  Wertebereich, Rechtschreibung
- Semantisch: Realweltvergleich

Genauigkeit

- Realweltvergleich
- Spezialfall Distanzmessung an Realobjekt

Konsistenz

$$\frac{\#\text{ Tupel, die alle Integritätsbedingungen erfüllen}}{\#\text{ Tupel}}$$

Zeitnähe

- Fehlende Daten für Arbeitsschritt?
- Prozessmonitoring

Messen von Vollständigkeit

Vollständigkeit

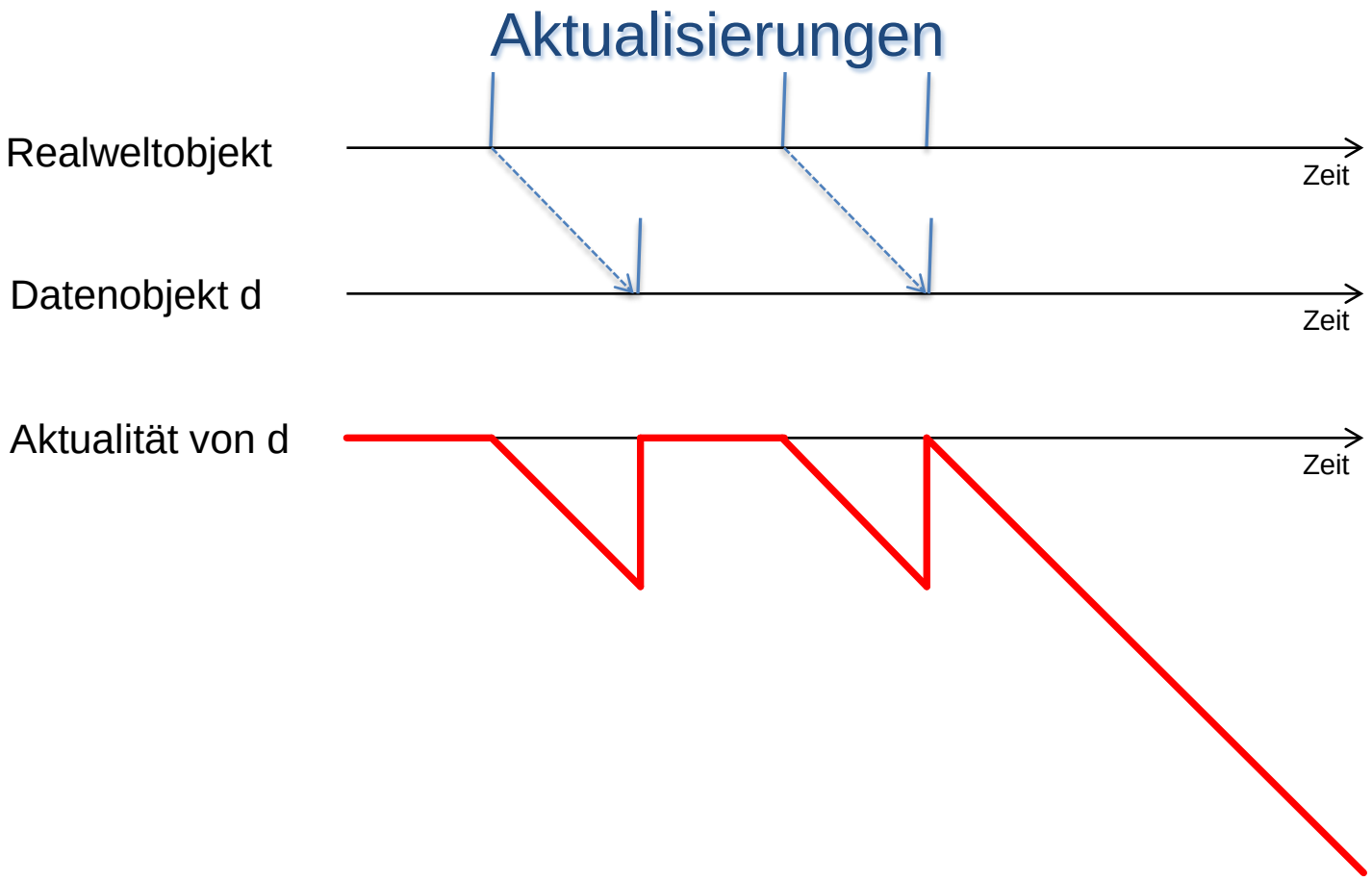
- Attribut, Tupel, Relation: Anteil *NULL*
- Schema: Schema- & Bedarfsanalyse

- Populationsvollständigkeit: Expertenwissen, Realweltvergleich

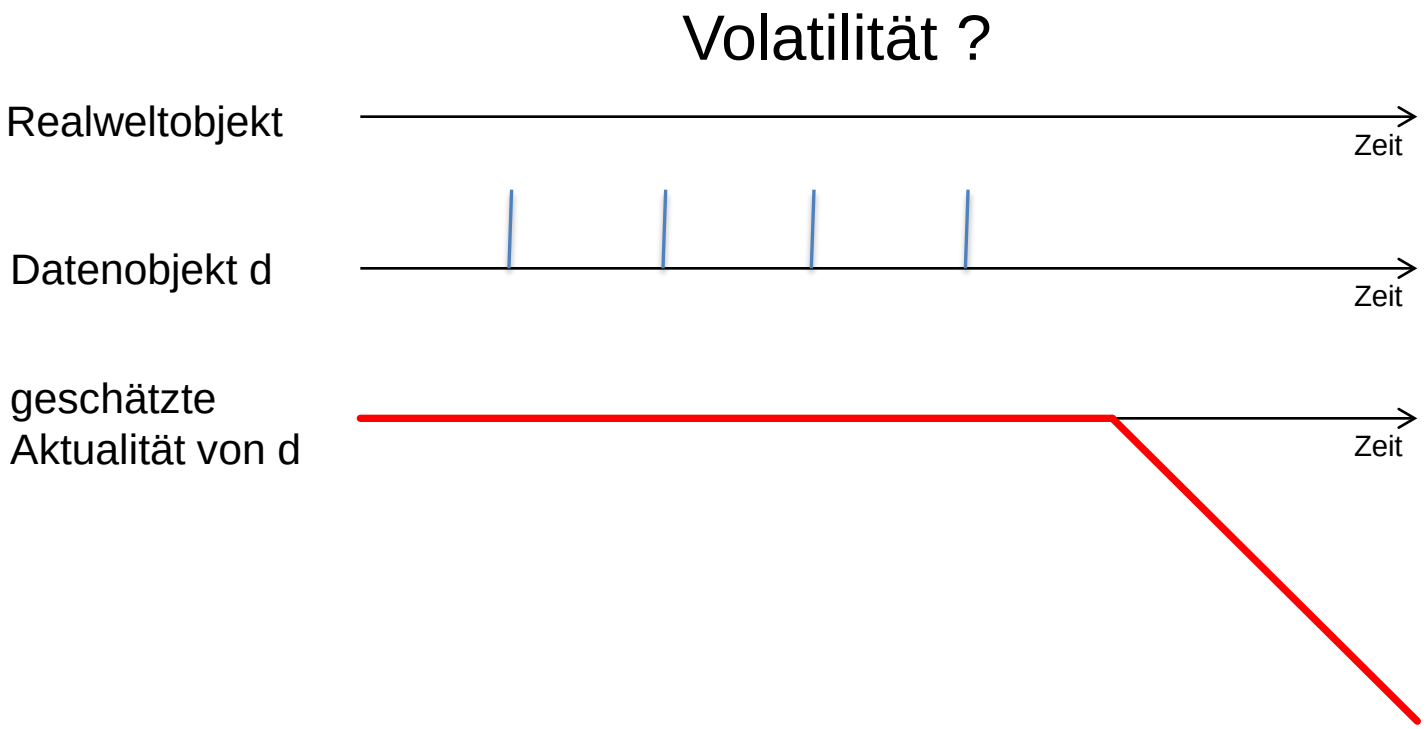
NULL

- Wert existiert nicht
- Wert existiert, ist aber nicht bekannt
- Nicht bekannt, ob Wert existiert

Messen von Aktualität



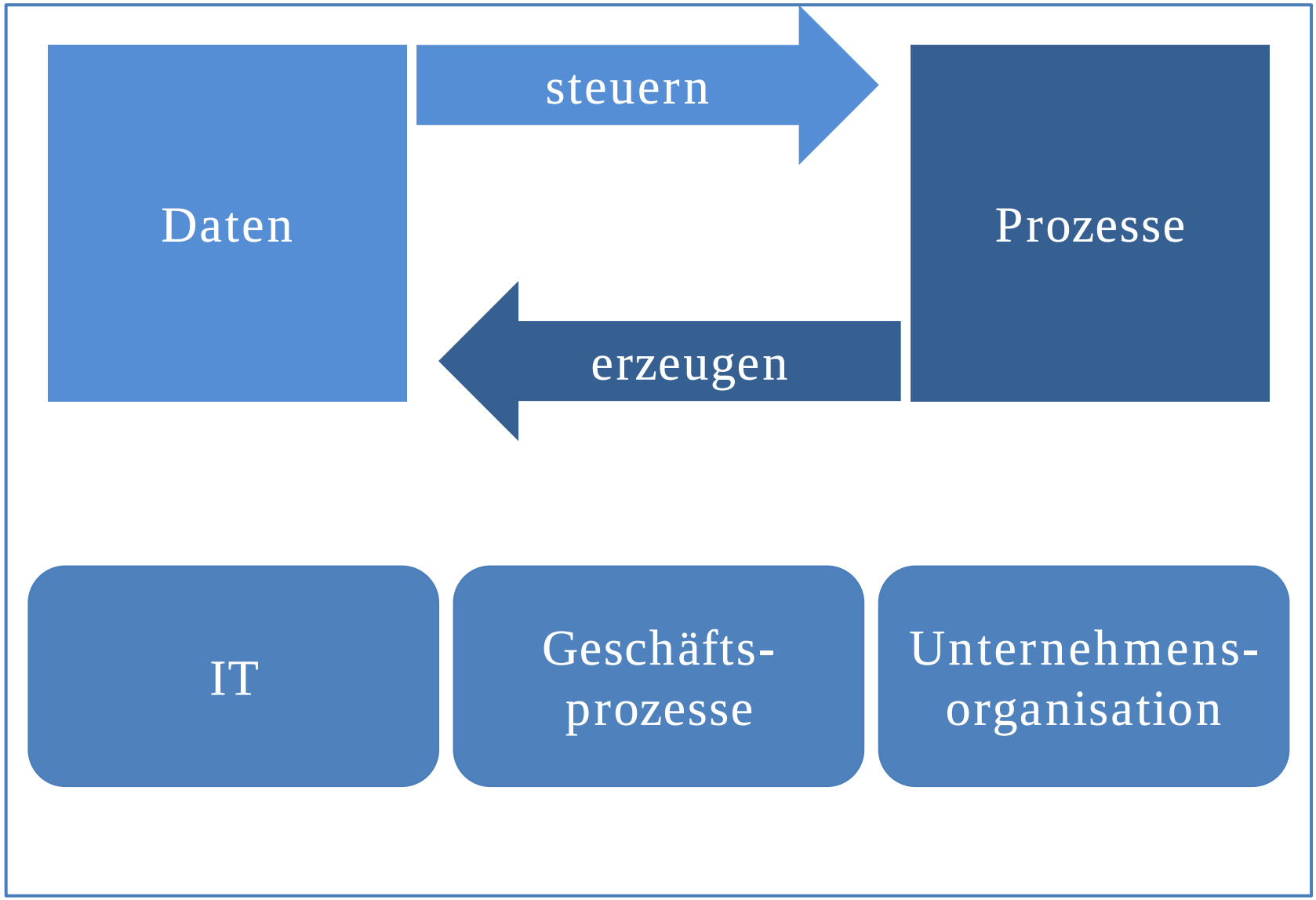
Messen von Aktualität ctd.



Verbesserung von Datenqualität

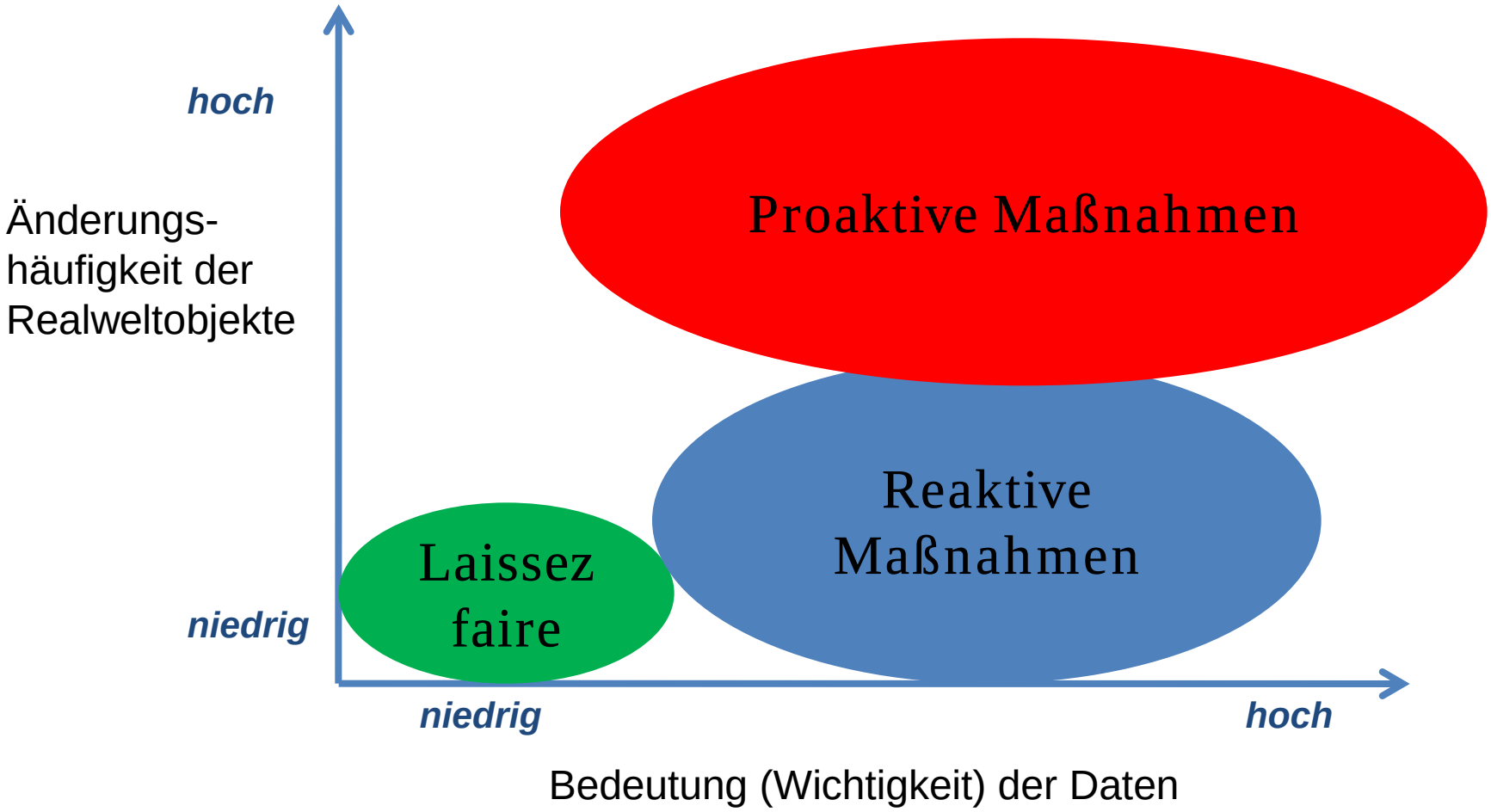
FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

DQ und Prozesskontrolle

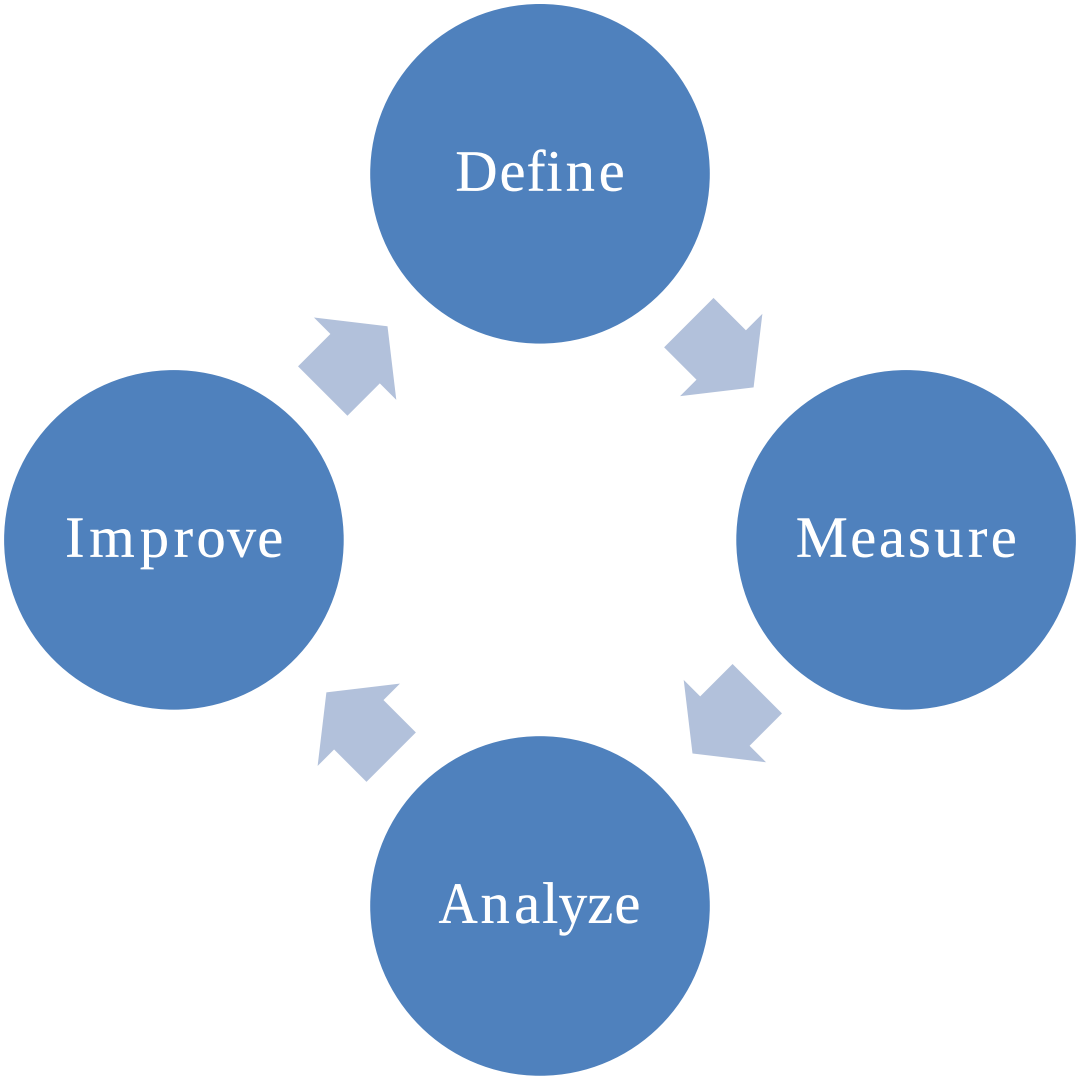


Maßnahmenportfolio

Maßnahmenportfolio zur Verbesserung der DQ
[Redman 1996]



TDQM Zyklus



TDQM: Wang et al: "Data Quality", Kluwer, 2000

Transparenz

… der DQ

… der Datenproduktion

A1, QS1	A2, QS2	A3, QS3
w1, qw1	w2, qw2	…
…	…	…

QS2	q1	q2	…
qw2	qind1	qind2	…
…	…	…	…

Transparenz

… der DQ

… der Datenproduktion

Provenance

Why | Where | How

Maßnahmen

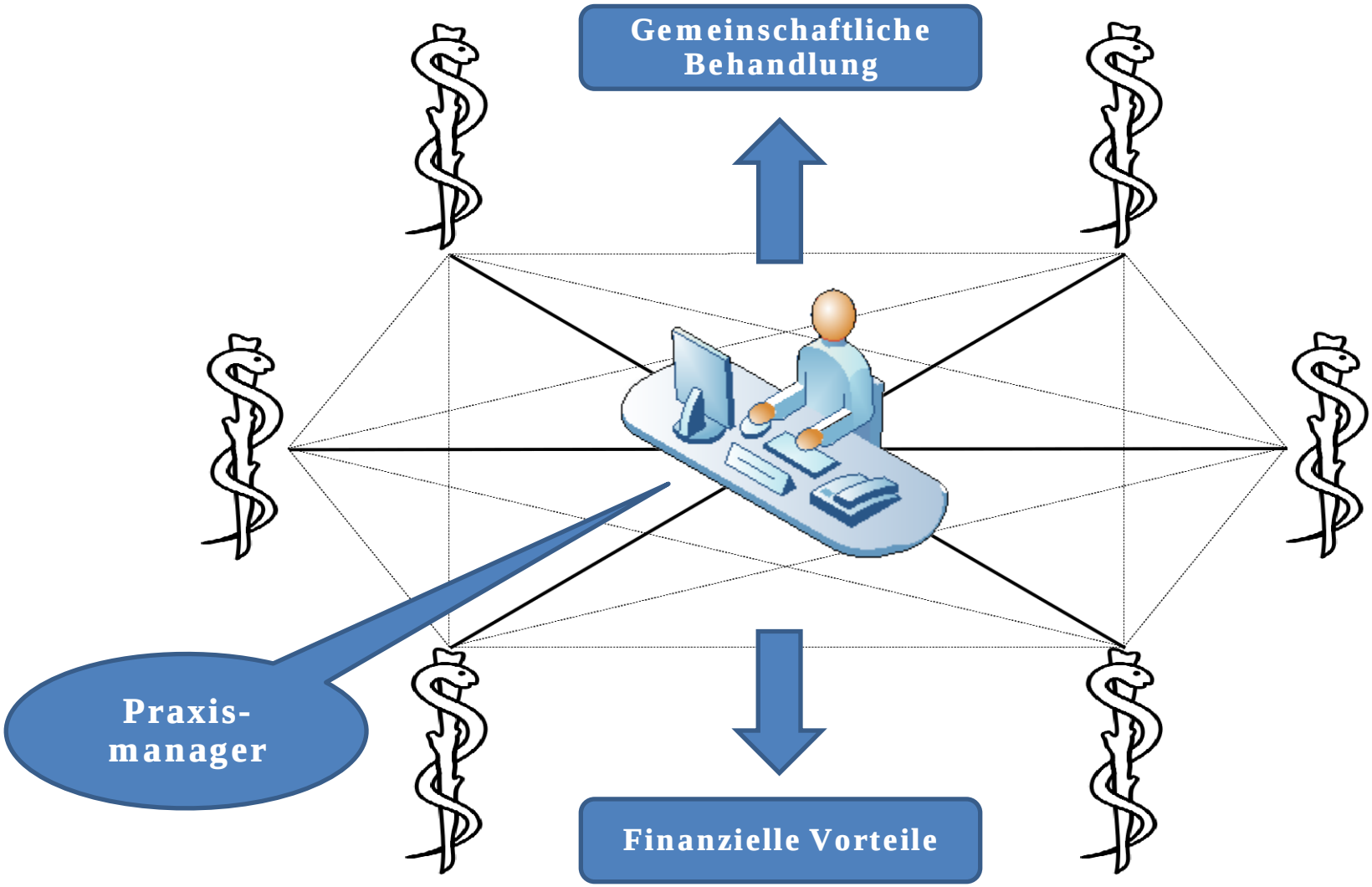
Profiling	Manuelle Korrektur	(semi-)automatische Korrektur
Attributanalyse	...	Ausreißer
Abhängigkeiten		Identity Matching
Fremdschlüssel-beziehungen		Record Linkage
Redundanzen		

Datenqualität im Projekt MEDITALK

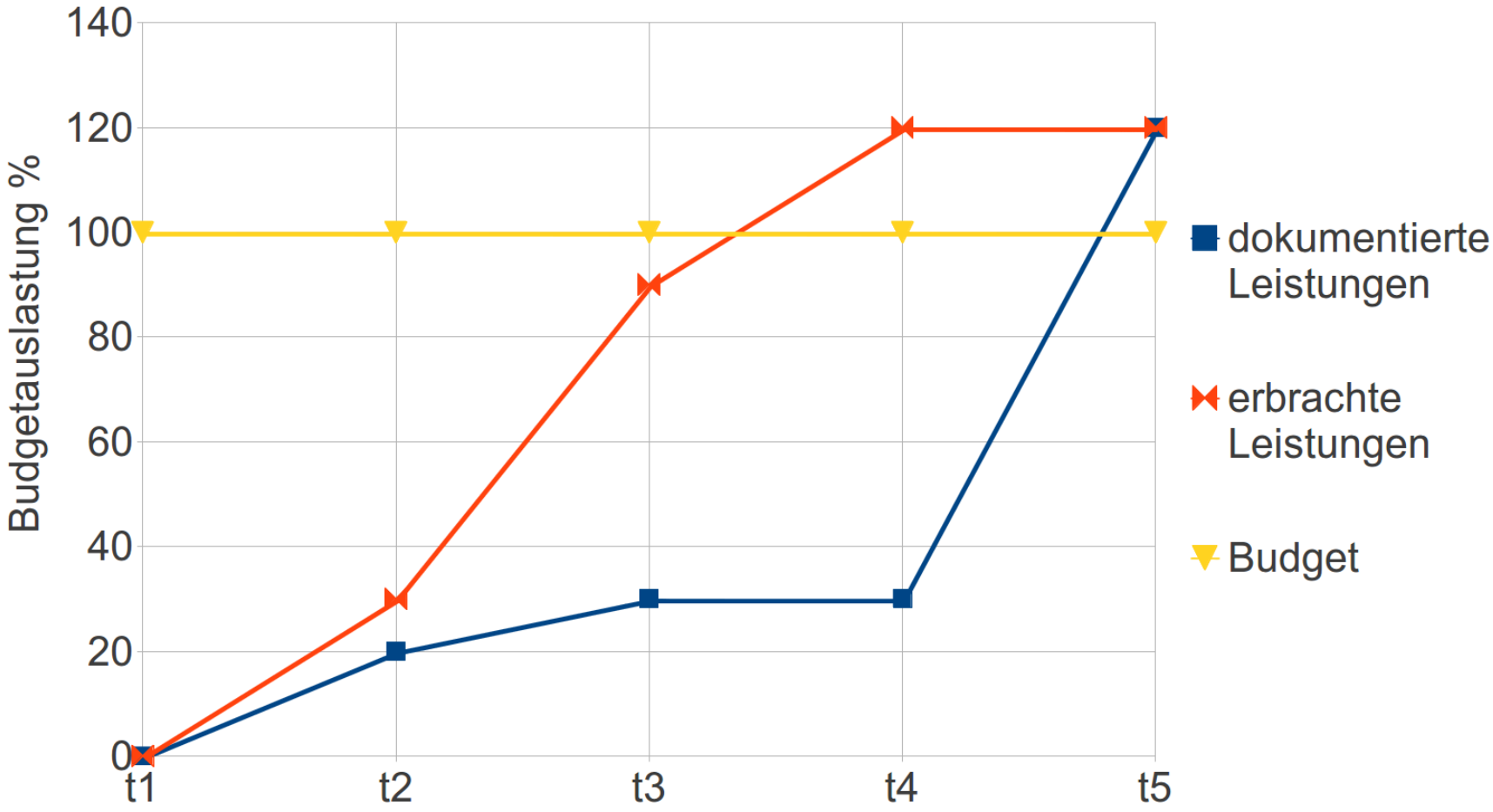
FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

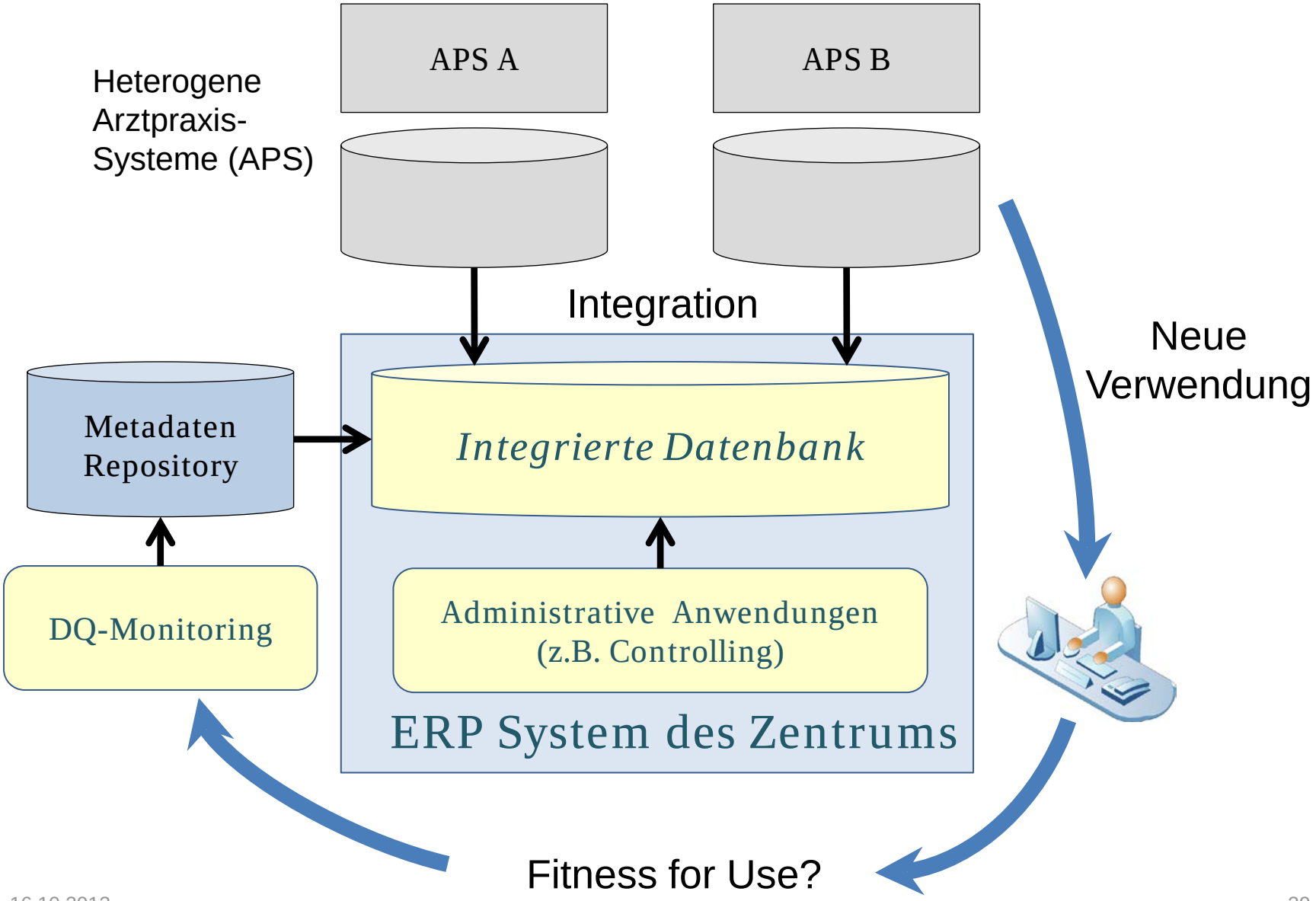
Kontext


Gemeinschaftliche Behandlung
Praxis-manager
Finanzielle Vorteile

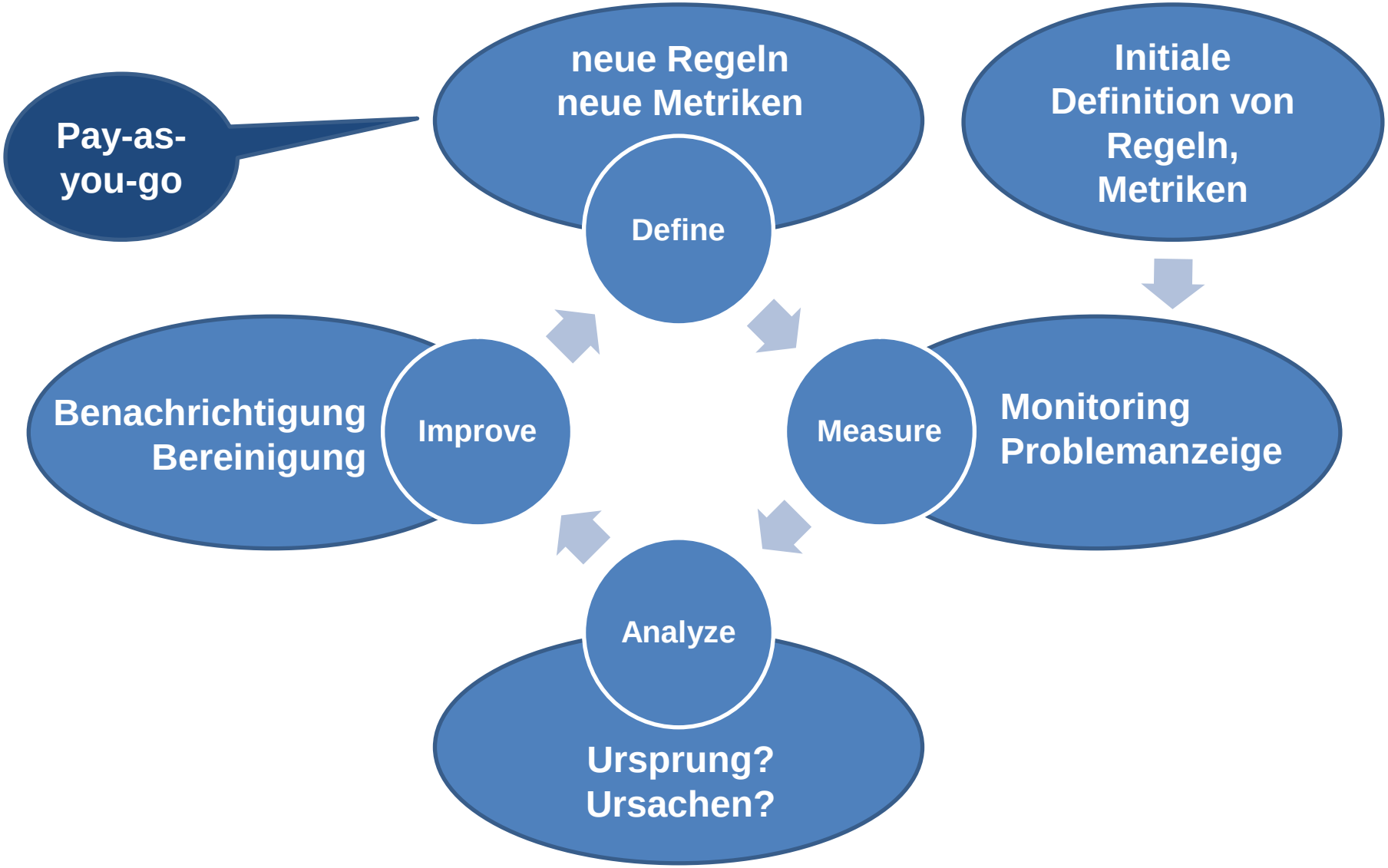
Wo drückt der Schuh?



MVZ und Praxisnetz



Pay-as-you-go DQ-Monitoring



TDQM: Wang et al: "Data Quality", Kluwer, 2000

Zusammenfassung

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

- Datenqualität ist subjektiv
- Datenqualität ist multidimensional
- DQ-Dimensionen sind nicht unabhängig
- Messen entlang der Dimensionen
- Oft kontinuierliches DQ-Management nötig

Kontakt

Gregor Endler
Lehrstuhl für Informatik 6 (Datenmanagement)
FAU Erlangen-Nürnberg

gregor.endler@fau.de

www6.cs.fau.de/people/greg/

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

TECHNISCHE FAKULTÄT

Quellen

Batini, C. and Scannapieco, M.:
Data Quality. Concepts, Methodologies and Techniques, Springer, 2006

Eckerson, W.:
Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data, The Data Warehouse Institute, Repost Series, 2002

Fisher, C.W. and Kingma, B.R.:
Criticality of Data Quality as Exemplified in Two Disasters, Information Management, 2002

English, L.P.:
Improving Data Warehouse and Business Information Quality, Wiley & Sons, 1999

Institute of Medicine:
IOM Report 1999
IOM Report 2001

Lenz, R.Y.:
Vorlesungsmaterial *Evolutionäre Informationssysteme*, 2012

Miller, D.W., et al.:
Missing prenatal records at a birth center: A communication problem quantified, AMIA Annu. Symp. Proc., 2005

Redman, T.C.:
Data Quality for the Information Age, Artech House, 1996

Wang, R. et.al.:
Data Quality, Kluwer, 2000