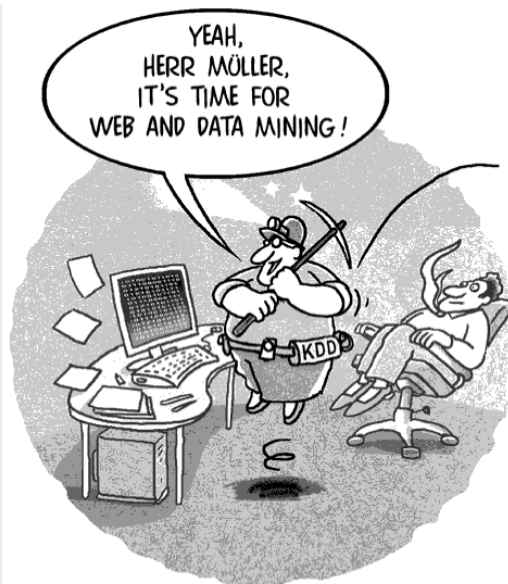


DW Data Warehousing – Funktionsweise, Einsatz- und Problemfelder

Andreas Meier, Universität Fribourg

DW Cartoon ...



Quelle: Knoll M., Meier A. (Hrsg): Web & Data Mining. Praxis der Wirtschaftsinformatik, HMD Nr. 268, August 2009, dpunkt Verlag, Heidelberg, Seite 2.

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 2

DW Inhaltsübersicht



- Unterschiede zwischen OLTP und OLAP
- Was versteht man unter einem Data Warehouse?
- Wie wird ein Data Warehouse modelliert?
- Welche Nutzungsfunktionen hat ein Data Warehouse?
- Anwendungsfall I: Kundenbeziehungsmanagement mit Web Analytics
- Anwendungsfall II: Medizinisches Communication Center
- Welches sind die wichtigsten Problemfelder?
- Welche Lösungsansätze bestehen?
- Ausblick

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 3

DW OLTP versus OLAP

	Online Transactional Processing	Online Analytical Processing
Daten	produktive Datenbanken	Datenbank für Entscheidungs- untersützung
Operationen	read, write, delete	read only
Zeitperiode	aktuelle Daten	historische, aktuelle und zukünftige Daten
Granularität	detailliert	unterschiedliche Aggregationsstufen
Antwortzeit	kurz (ms-s)	oft lang (s-min)
Verfügbarkeit	hohe	mittlere
Anzahl Benutzer	viele	wenige

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 4

DW Definition eines Data Warehouse

«A data warehouse is a **subject oriented, integrated, non-volatile,** and **time variant** collection of data in support of management's decisions.»

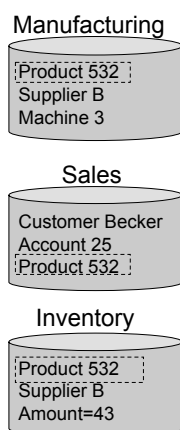
Quelle: Inmon W. H.: Building the Data Warehouse. Wiley 2005

Andreas Meier, Departement für Informatik, Universität Fribourg

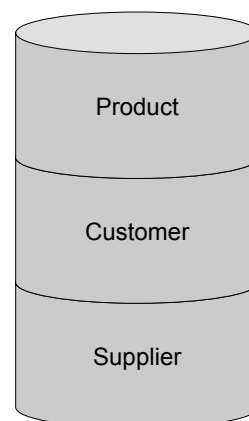
Folie 5

DW Subject Orientation

Operational Systems:
process orientation

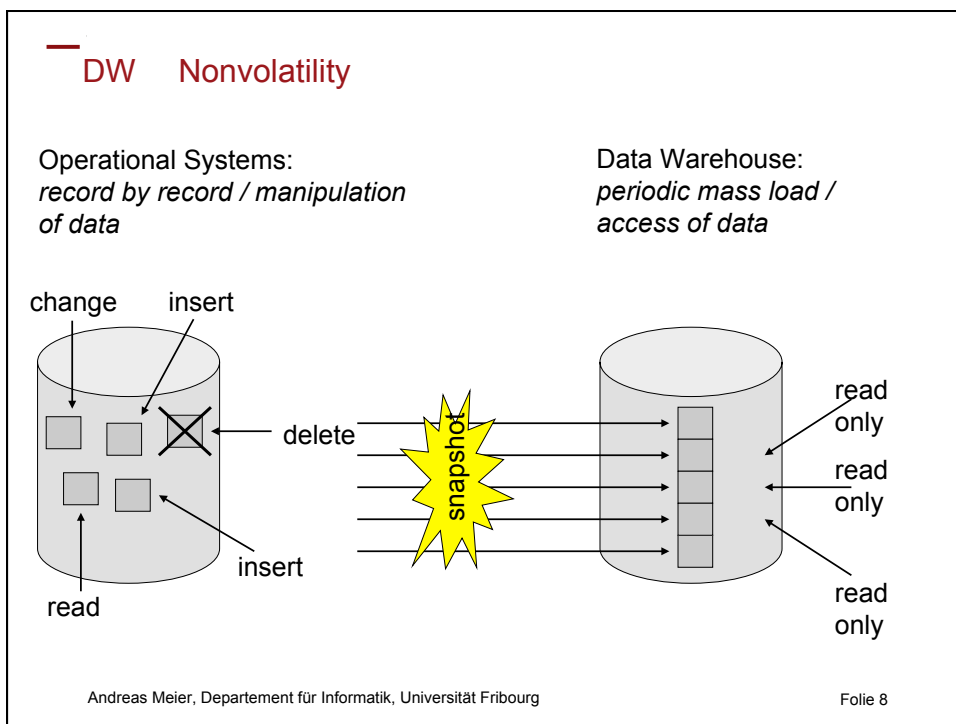
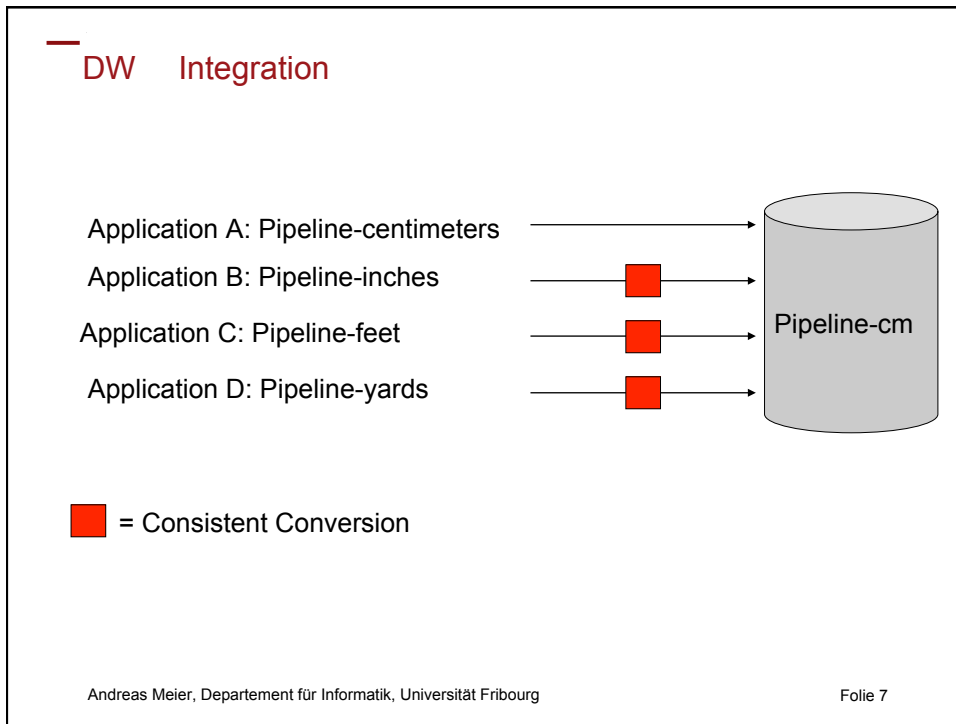


Data Warehouse:
subject orientation



Andreas Meier, Departement für Informatik, Universität Fribourg

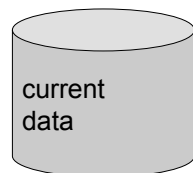
Folie 6



DW Time Variancy

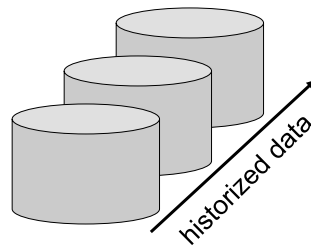
Operational Systems:

- time horizon - current or eventually 60 to 90 days
- update of records
- key structure may/may not contain an element of time



Data Warehouse:

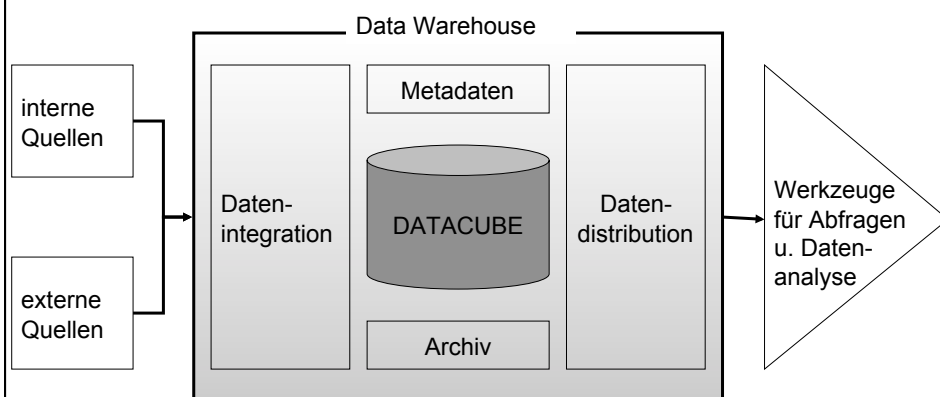
- time horizon - 5 to 10 years
- sophisticated snapshots of data
- key structure contains an *element of time* (data is associated with specific time periods)



Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 9

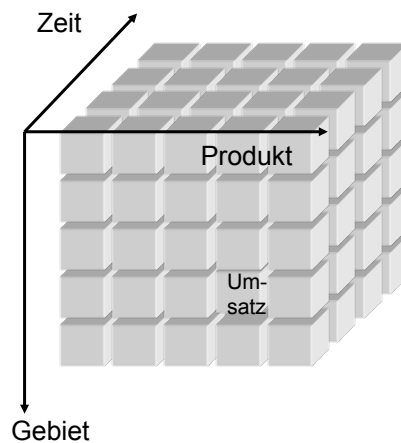
Architektur eines Data Warehouse



Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 10

Beispiel eines mehrdimensionalen Würfels



Fakten:

- Umsatz

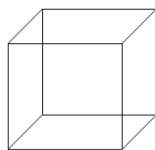
Dimensionen:

- Zeit
- Gebiet
- Produkt

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 11

DW Definition von Fakten



Fakten (Mass, Indikator, ...) sind Kennzahlen für die Analyse und fürs Reporting. Oft sind es numerische Werte.

Fakten können aggregiert und mit der Hilfe der Zeitdimension für Trendanalysen verwendet werden.

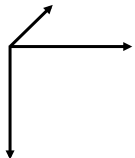
Beispiele:

- Umsatz
- Produktionskosten
- Verkaufspreis
- Kundenwert
- Marktanteil
- Mitarbeiterzufriedenheit
- ...

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 12

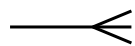

DW Definition von Dimensionen


Dimensionen sind Auswertungsmerkmale für unterschiedliche Aggregationsstufen.

Beispiele:

- Zeit
- Produkt
- Gebiet
- Kunde
- Lieferant
- ...

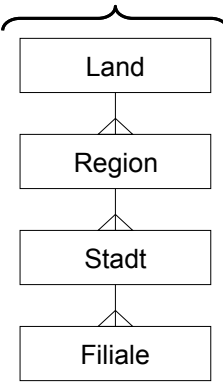
Aggregationsstufen:

 1:m
 1:1

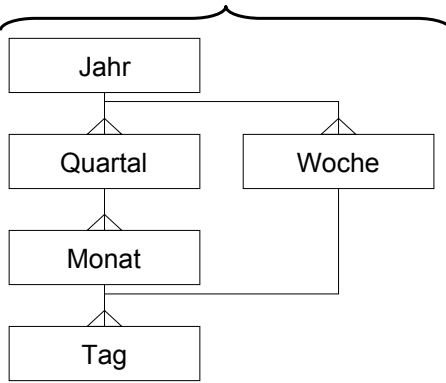
Andreas Meier, Departement für Informatik, Universität Fribourg Folie 13

DW Beispiel zweier Dimensionen

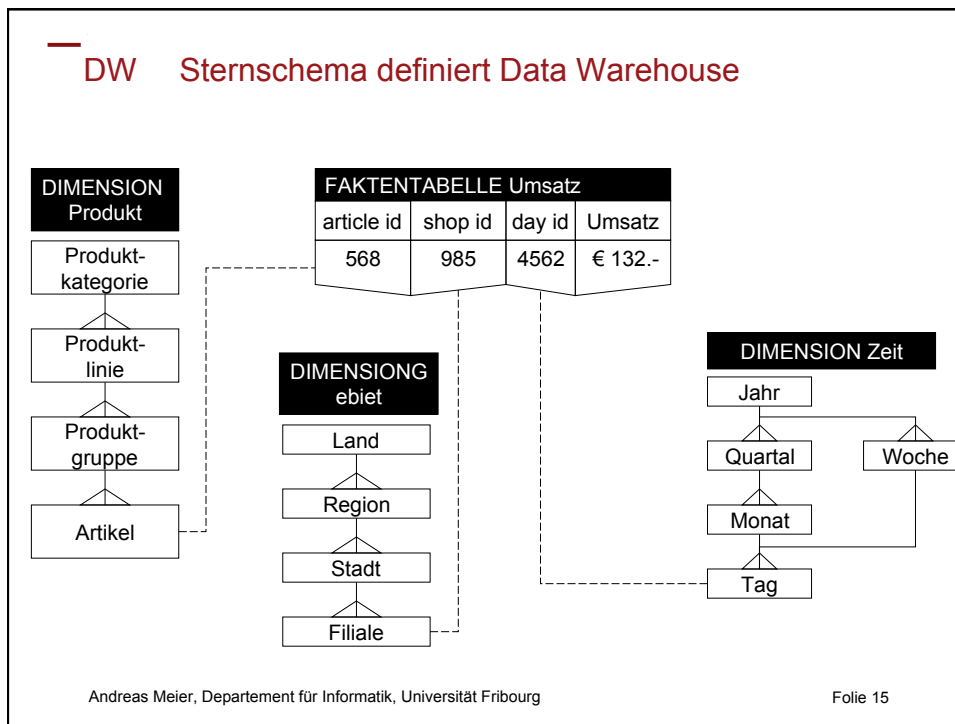
GEBIET



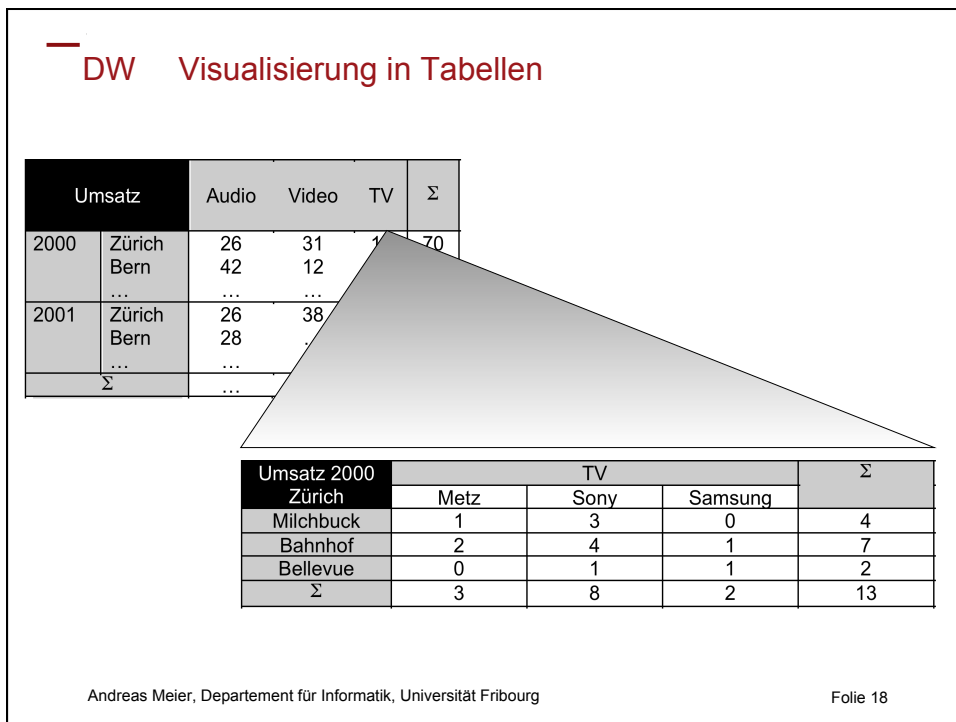
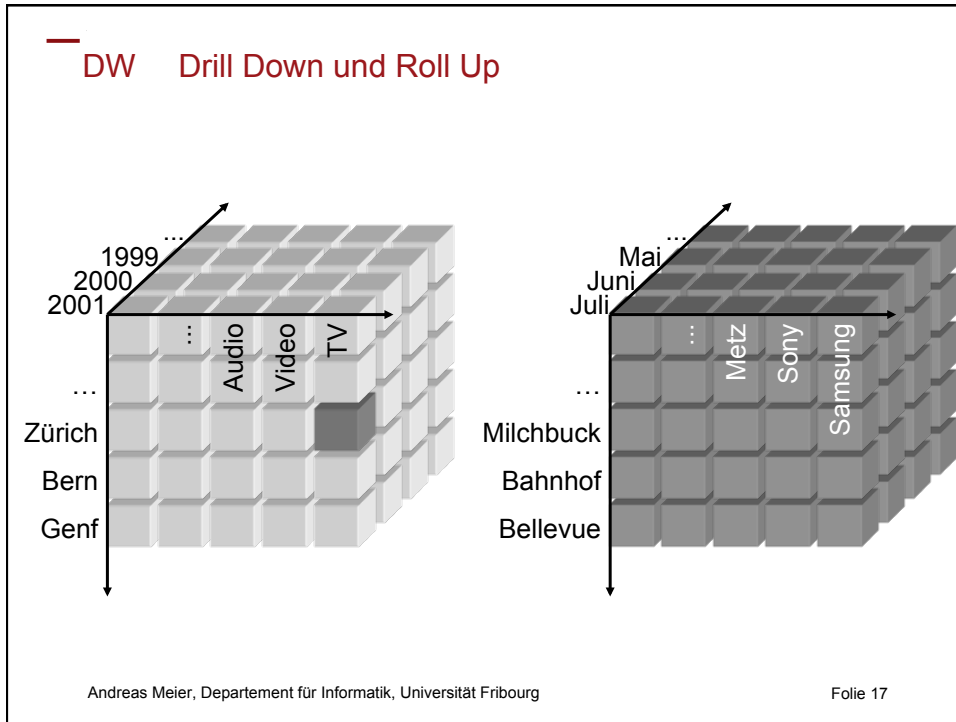
ZEIT

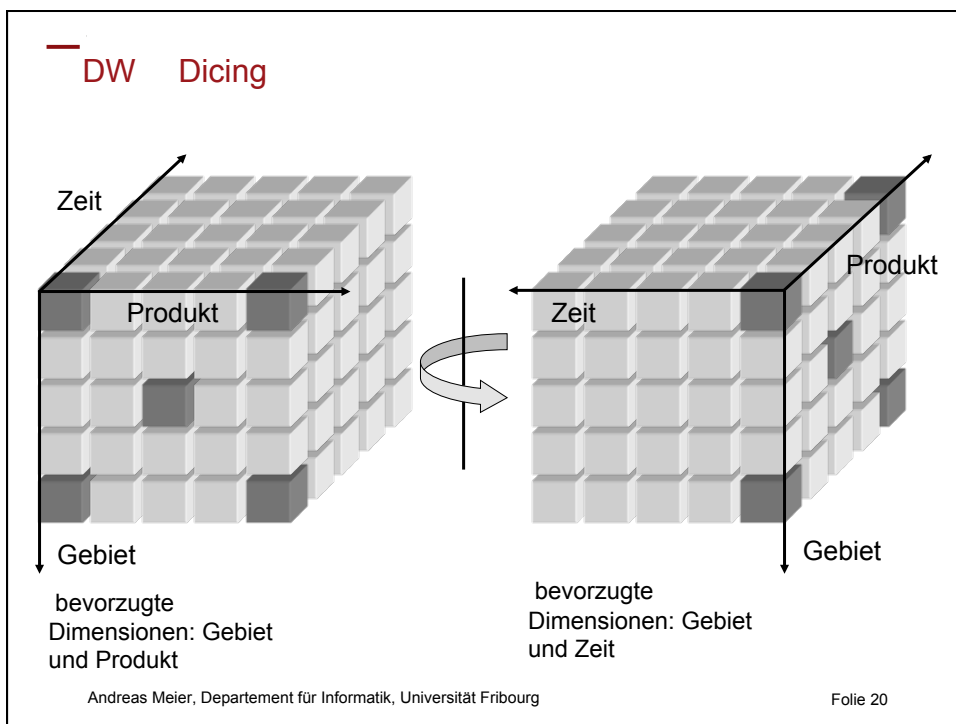
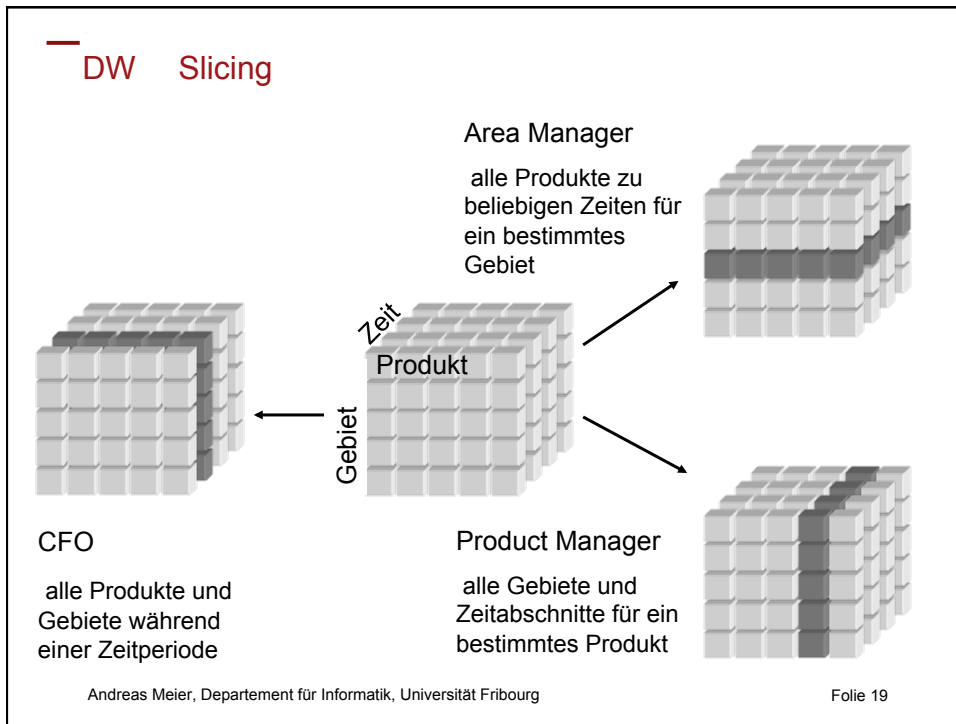


Andreas Meier, Departement für Informatik, Universität Fribourg Folie 14



- ### DW Die vier wichtigsten Operationen
- Drill Down
 - Roll Up
 - Slicing und
 - Dicing
- Andreas Meier, Departement für Informatik, Universität Fribourg Folie 16





DW Beispiel einer Dicing-Operation

Turnover July		TV			Σ
Zurich	Metz	Sony	Samsung		
Milchbuck	5	3	1		9
Bahnhof	2	4	1		7
Bellevue	0	1	4		5
Σ	7	8	6		21

Turnover June		TV			Σ
Zurich	Metz	Sony	Samsung		
Milchbuck	1	3	0		4
Bahnhof	2	4	1		7
Bellevue	0	1	4		5
Σ	3	8	5		16

Turnover May		TV			Σ
Zurich	Metz	Sony	Samsung		
Milchbuck	0	6	3		9
Bahnhof	2	4	1		7
Bellevue	0	1	4		5
Σ	2	11	8		21

Turnover "Samsung" Zurich		Month			Σ
	May	June	July		
Milchbuck	3	0	1		4
Bahnhof	1	1	1		3
Bellevue	0	0	0		0
Σ	4	1	2		7

Turnover "Sony" Zurich		Month			Σ
	May	June	July		
Milchbuck	6	3	3		12
Bahnhof	4	1	1		6
Bellevue	0	0	0		0
Σ	10	4	4		18

Turnover "Metz" Zurich		Month			Σ
	May	June	July		
Milchbuck	0	1	5		6
Bahnhof	2	1	2		5
Bellevue	0	0	0		0
Σ	2	2	7		11

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 21

DW Nutzenpotenziale

- Ein Data Warehouse ist für die Steuerung des Geschäfts unabdingbar.
- Ein Data Warehouse lässt sich für unterschiedliche Dimensionen und Aggregationsstufen auswerten.
- Ein Reporting ohne DW ist aus Sicht des Controllings kaum realistisch, da eine gemeinsame Datenbasis konsistent gehalten werden muss.
- Auswertungen können bereichsübergreifend oder subjektorientiert vorgenommen werden.
- Flexible Auswertungsfunktionen für den mehrdimensionalen Würfel erlauben den Führungskräften detaillierte Analysen.
- Vergleiche, Analysen und Prognosen mit Zeitbezug sind machbar.
- Operative Systeme werden entlastet.

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 22

DW Inhaltsübersicht



- 😊 Unterschiede zwischen OLTP und OLAP
- 😊 Was versteht man unter einem Data Warehouse?
- 😊 Wie wird ein Data Warehouse modelliert?
- 😊 Welche Nutzungsfunktionen hat ein Data Warehouse?
 - Anwendungsfall I: Kundenbeziehungsmanagement mit Web Analytics
 - Anwendungsfall II: Medizinisches Communication Center
 - Welches sind die wichtigsten Problemfelder?
 - Welche Lösungsansätze bestehen?
 - Ausblick

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 23

DW Definition von Web Analytics

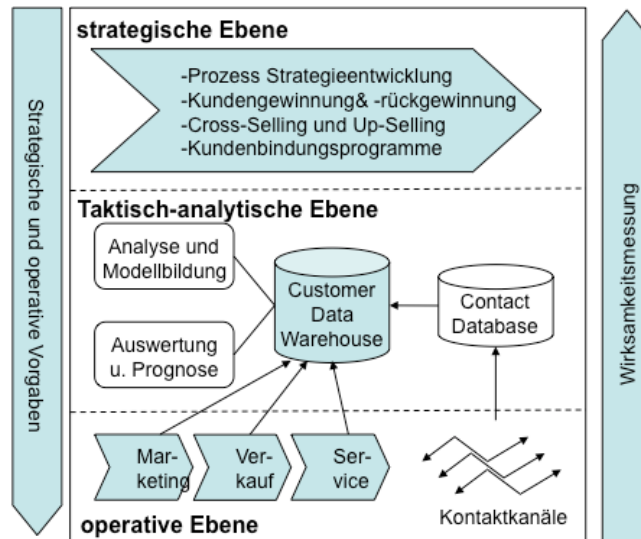
*“Web Analytics is the **measurement, collection, analysis and reporting** of Internet data for the purposes of understanding and optimizing Web usage.”*

Quelle: Web Analytics Association,
siehe <http://www.webanalyticsassociation.org>

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 24

DW **Kreislauf für das Kundenbeziehungsmanagement**



Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 25

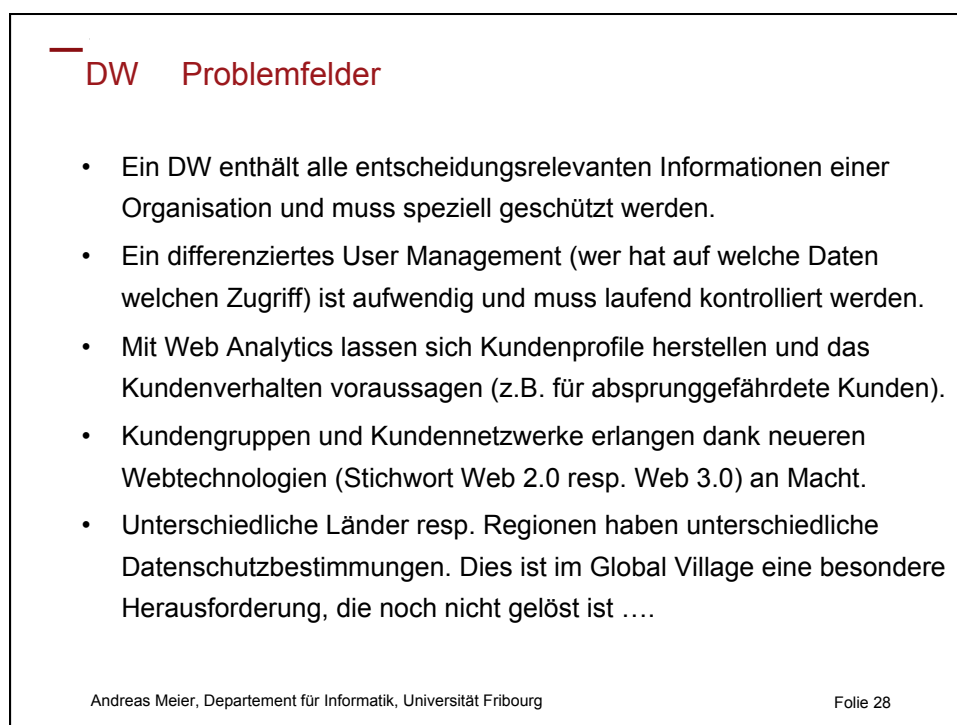
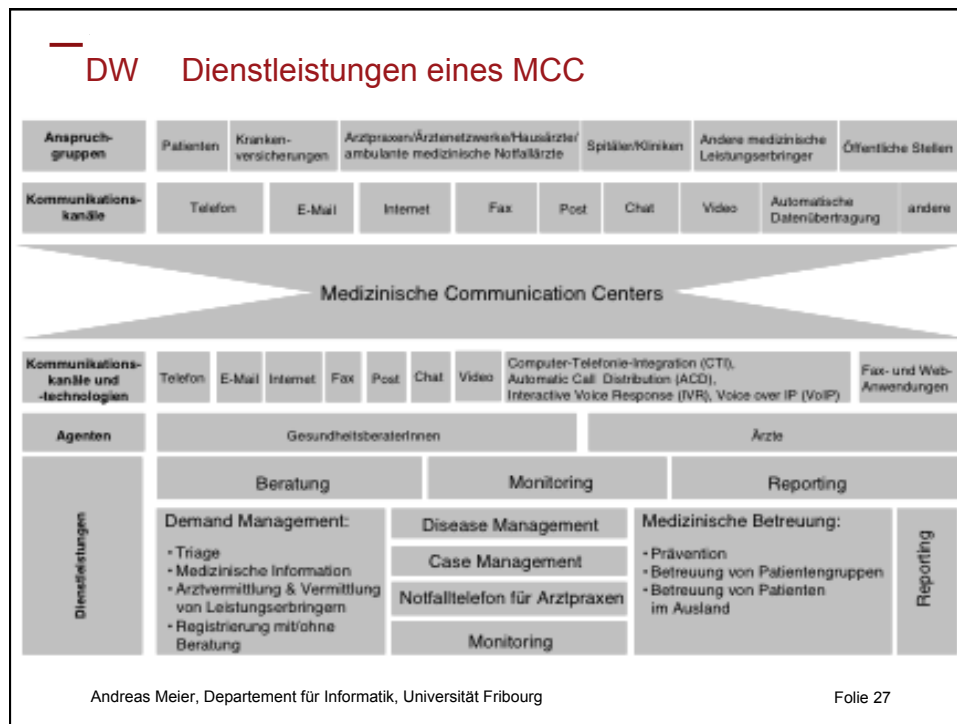
DW **Definition medizinisches Communication Centers**

*“Ein medizinisches Communication Center bietet der Bevölkerung **Demand-, Disease- und Case-Management** über unterschiedliche Kommunikationskanäle an.”*

Quelle: Ionas A.: Modellierung, Entwicklung und Nutzung eines Data Warehouse für medizinische Communication Centers. Dissertation, Universität Fribourg, 2008 (als eThesis verfügbar)

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 26



DW Lösungsansätze

1. Geschäftsrelevante Daten im Land des Domizils speichern.
2. Identifikation, Authentifikation und Autorisierung aller Anspruchsgruppen regeln.
3. Datenmanagement mit Datenarchitektur, Datenadministration und Datentechnik organisatorisch verankern.
4. Datenschutzbeauftragte im Unternehmen bei der Revision resp. beim Controlling ansiedeln (Vieraugenprinzip).
5. Deklarationspflicht für personenbezogene Daten (Umfang, Dauer) auch auf Websites wahrnehmen.
6. Technische Sicherheits- und Backup-Systeme nutzen und regelmässig auf Tauglichkeit testen.
7. Kryptographische Verfahren für Speicherung und Übermittlung von Daten anwenden.

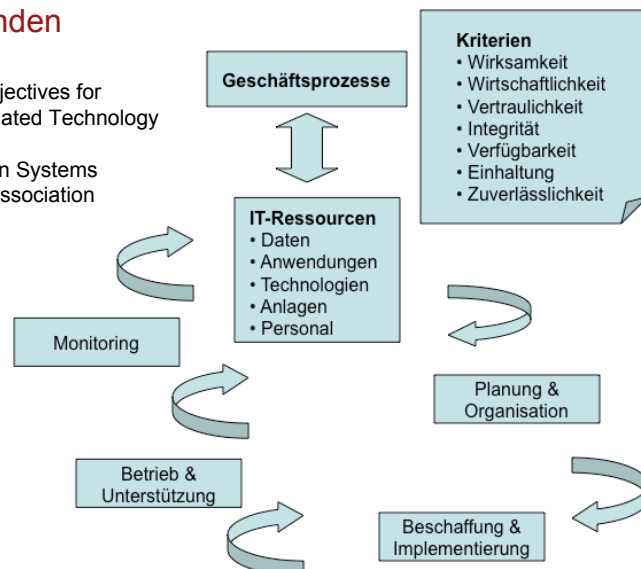
Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 29

DW Framework CobiT für Ordnungsmässigkeit verwenden

CobiT = Control Objectives for Information and Related Technology

ISACA = Information Systems Audit and Control Association



Andreas Meier, Departement für Informatik, Universität Fribourg

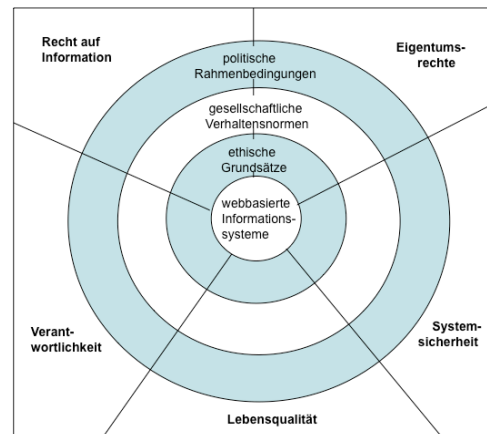
Folie 30

DW Ausblick

Die Informations- und Wissensgesellschaft ist auf die Nutzung von Internet-technologien angewiesen. Auch hier gilt im Sinne von Immanuel Kant:

**“Handle nur nach derjenigen
Maxime, durch die du zugleich
möchtest, dass sie ein
allgemeines Gesetz werde”.**

Die fünf Dimensionen ethischen Handelns nach Laudon/Laudon



Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 31

DW Literaturangaben

- Bauer A., Günzel H.: Data-Warehouse-Systeme – Architektur, Entwicklung, Anwendung. dpunkt 2008
- Gabriel R., Gluchowski P., Pastwa A.: Datawarehouse und Data Mining. W3L-Verlag 2009
- Inmon W. H.: Building the Data Warehouse. Wiley 2005
- Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P.: Fundamentals of Data Warehouses. Springer 2000
- Kimball R., Ross M.: The Data Warehouse Toolkit. Wiley 2002
- Meier A.: Relationale und postrelationale Datenbanken. Springer 2010

Weitere Angaben unter

- *Forschungsgruppe Information Systems* <http://diuf.unifr.ch/is>
- *Forschungszentrum Fuzzy Marketing Methods* <http://www.FMsquare.org>

Andreas Meier, Departement für Informatik, Universität Fribourg

Folie 32