

Datenbanken und Datenmodellierung

Planet13 – uni von unten

Fredy Spring – 13. April 2015



CC BY-SA 4.0

Inhaltsübersicht

- **Grundbegriffe der Informationstheorie**
- **Informations- & Datenbanksysteme**
 - Motivation, Aufbau und Ziele
- **Datenbanken**
 - Modellierungsphasen
 - Konzeptionelles Modell
 - Entity-Relationship-Modell (ERM)
 - Logisches Modell
 - Relationale Datenbanken: Strukturelemente, Anomalien, Normalisierung
 - Physisches Modell
 - Integrität, Schlüssel, Datenkatalog

Daten – Information – Wissen

Daten

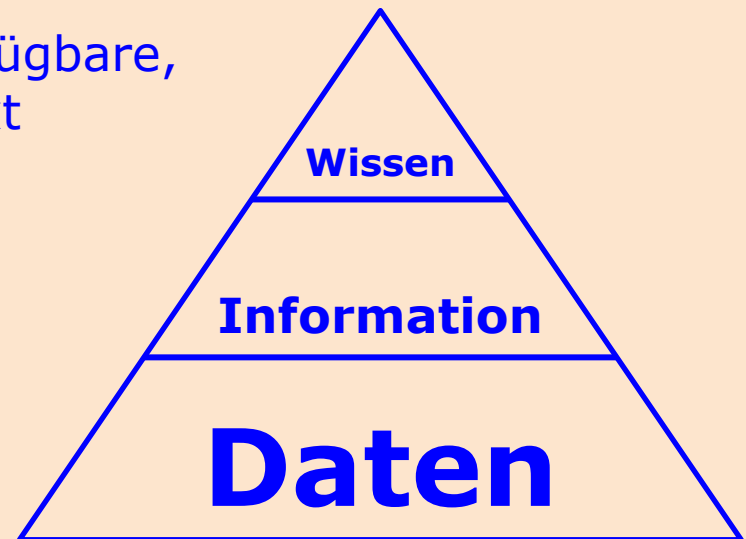
"Rohmaterial", "Fakten" – z.B. als Zeichen oder Funktionen – zur Darstellung von Information

Information

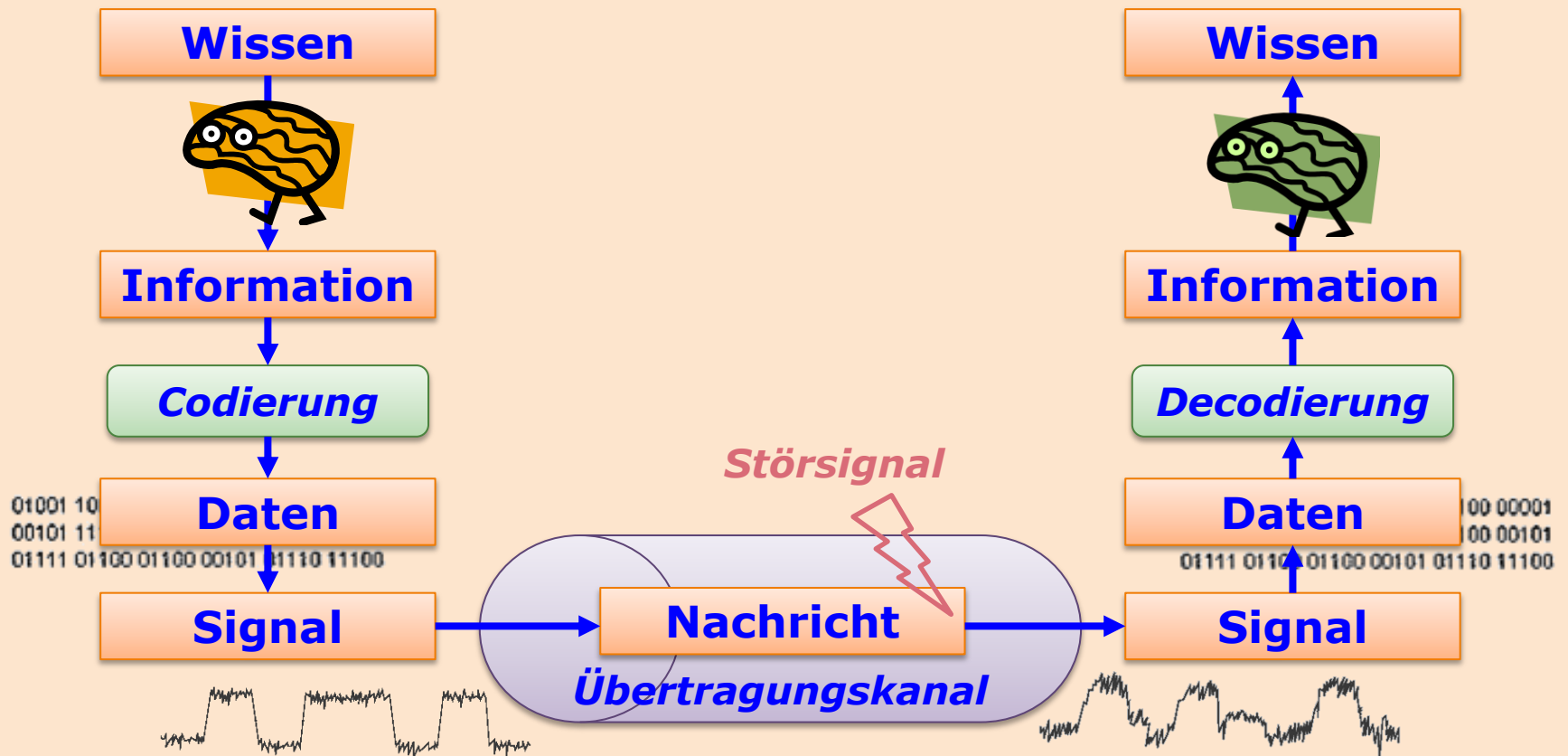
Informationen sind strukturierte und interpretierte Daten, die Antworten auf "einfache" Fragen geben.

Wissen

Wissen ist intelligent organisierte, verfügbare, anwendbare und mit Erfahrungskontext situierte Information.



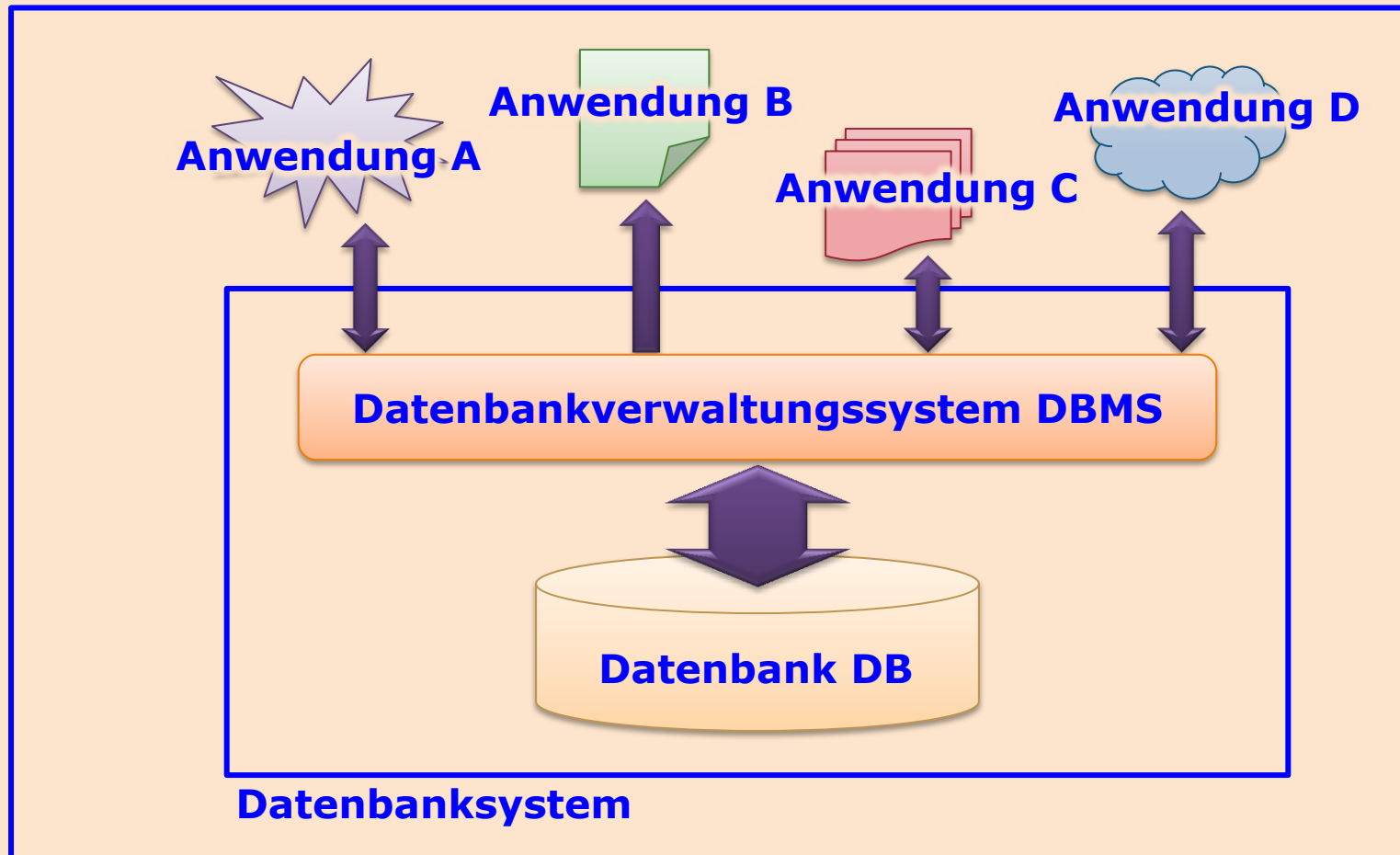
Übertragungsschema in der IT



Grundmotive für Datenbanken

- **Abbildung von Realobjekten und deren Beziehungen in computergerechter Form**
- **Daten strukturiert organisieren und für Informationsgewinnung bereitstellen**
 - **Ideal: eine Datenstruktur optimiert für alle denkbaren – und undenkbaren – Datenanalysen**
- **Daten dauerhaft ablegen und pflegen**
- **heute eine der wichtigsten IT-Anwendungen**
 - **Rechner orientierte Recherchiertechniken werden zu Kulturtechniken wie Lesen, Schreiben und Rechnen**

Informationssystem (1)



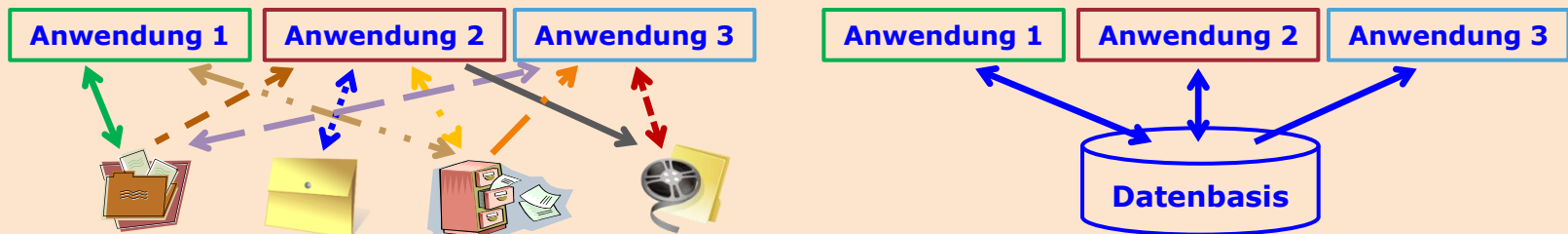
Informationssystem

Informationssystem (2)

- **Datennutzung (Anwendungsbereich)**
 - Methoden zur Datenabfrage, -pflege, -interpretation, -darstellung, -transformation etc.
 - in einer Client-Server-Architektur
 - professionelle, gelegentliche, spezielle Nutzer_innen
- **Datenbankverwaltungssystem (DBMS)**
 - Datenmodell definieren und verwalten
 - Datenmanipulation und Datenabfrage für Client
 - Zugriffsrechte und Mehrfachzugriff
 - Datensicherung (Transaktion, Backup, Recovery)
 - Effizienz (Performance, Skalierbarkeit, Verteilbarkeit)
- **Datenbank (DB)**
 - Datenbasis ("inhaltliche Rohdaten" ohne Systemdaten)
 - physisches Schema, Systemtabellen und -daten
 - physische Speicherung

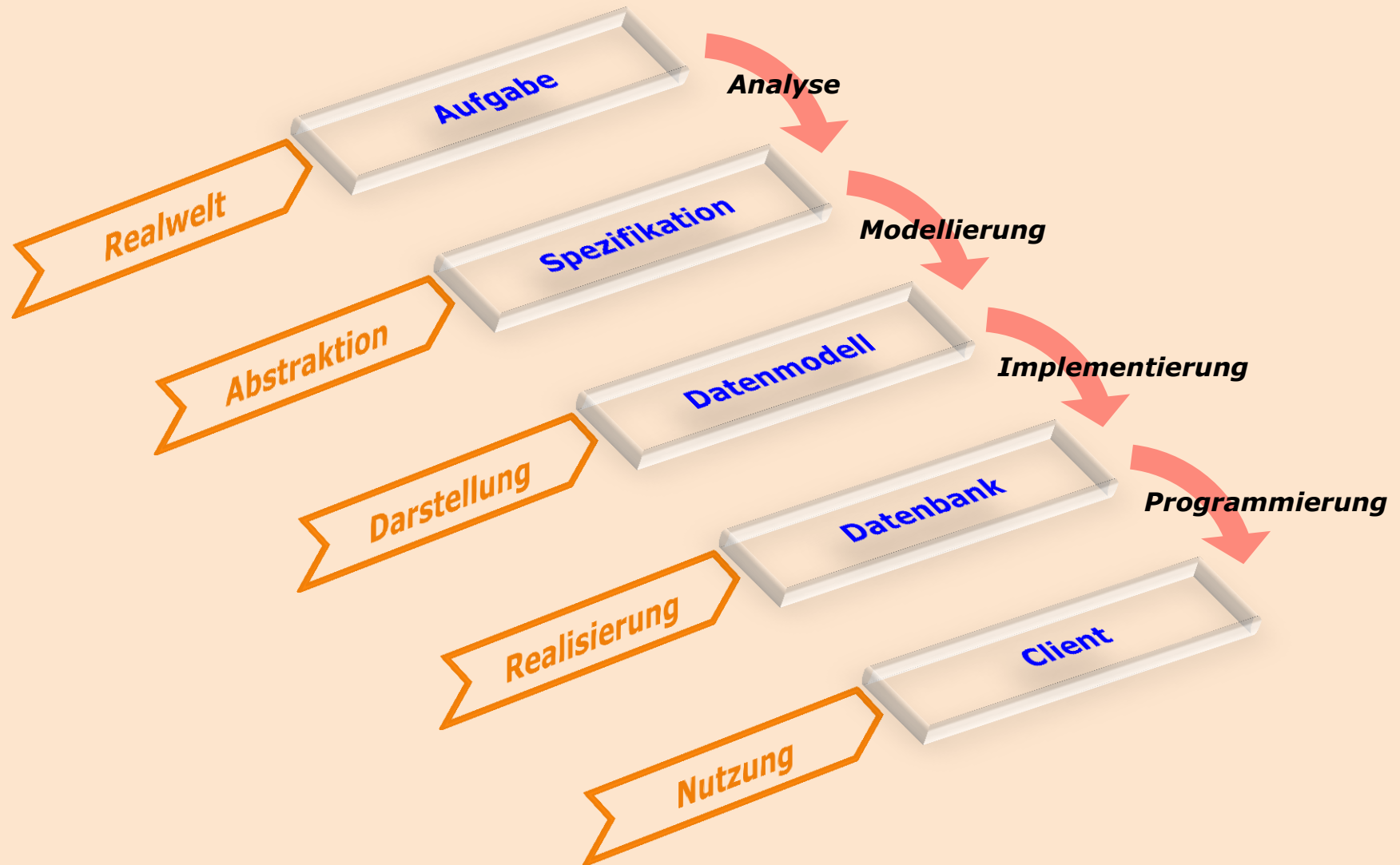
Hauptzweck/-ziele eines IS

- flexible Datenanalyse: Suche, Filterung, Berechnung, Ausgabe und Darstellung
- Werkzeug zur effizienten Verarbeitung riesiger Datenmengen bei kurzen Antwortzeiten
- verschiedenste Anwendungen auf Grundlage derselben Datenbasis und desselben Zugriffskonzepts



- gleichzeitiger Zugriff von verschiedenen Anwendungen (Mehrfachnutzung) mit ausdifferenzierter Zugriffskontrolle

Von der Aufgabe zur DB-Lösung



Was wirklich geschah (1)



**Was der
Anwender
wollte.**



**Was der
Anwender der
Entwicklerin
sagte.**



**Was die
Entwicklerin
verstand.**

Was wirklich geschah (2)



**Was die
Entwicklerin
wollte.**



**Was die
Entwicklerin
baute.**

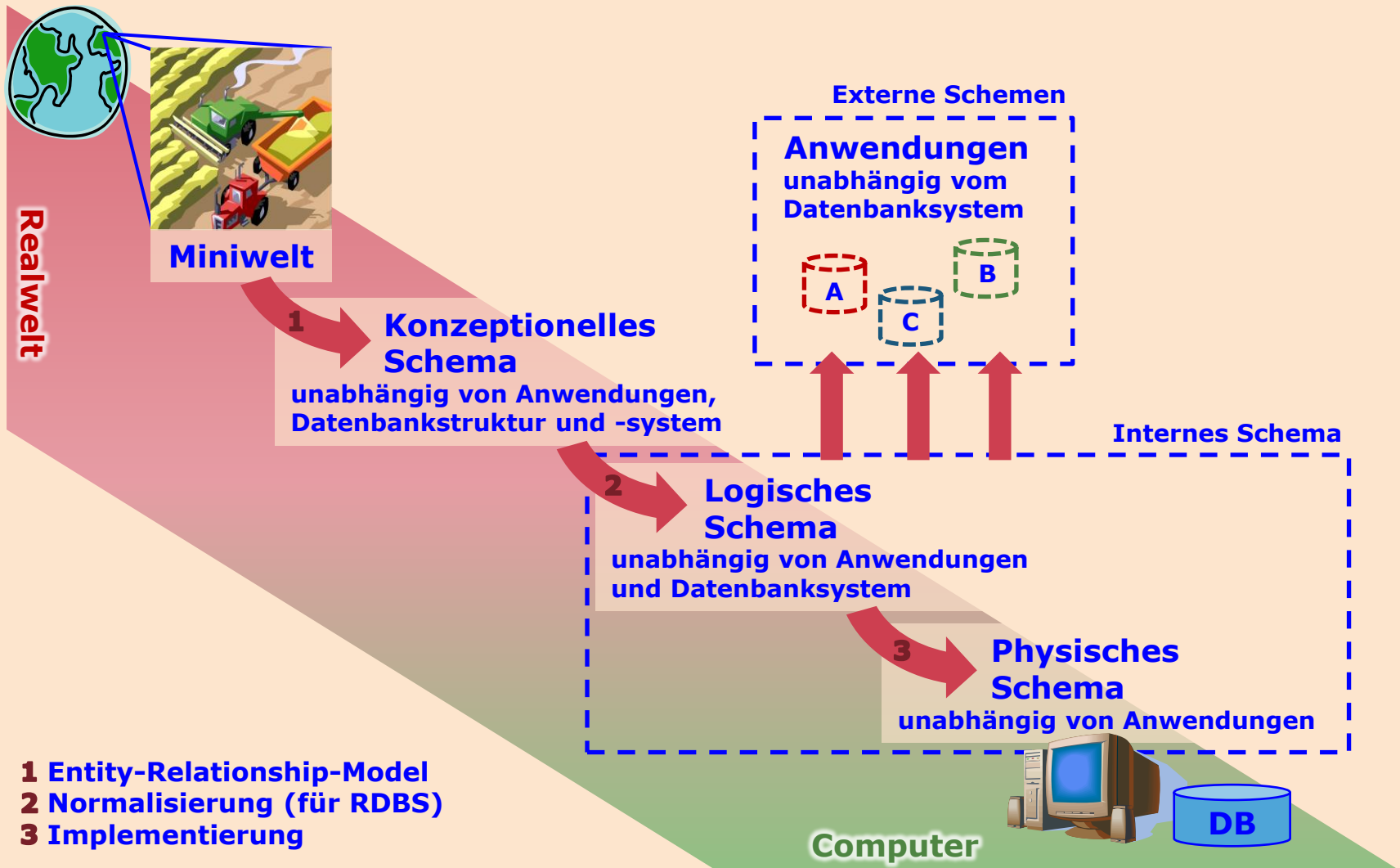


**Was der
Anwender
brauchte.**

Modell/Schema

- **Datenmodellierung ist der Vorgang zur wirklichkeitsgetreuen Abbildung eines Teilausschnitts der Realwelt (Miniwelt) in computergerechter Form**
- **Modell beschreibt die in der Datenbank enthaltenen Daten bezüglich ihrer Struktur und den Regeln der Datenhaltung**
 - **Beschreibung besteht aus Datenbankobjekten, Beziehungen und Integritätsregeln**
 - **Modell bleibt nach Erstellen relativ statisch**
 - **DBMS sorgt für die Einhaltung des Schemas**
- **mögliche Probleme der Modellierung sind Abbildungsverzerrungen und -fehler**

Modellierungsphasen & Schemen



Modellierungsphasen

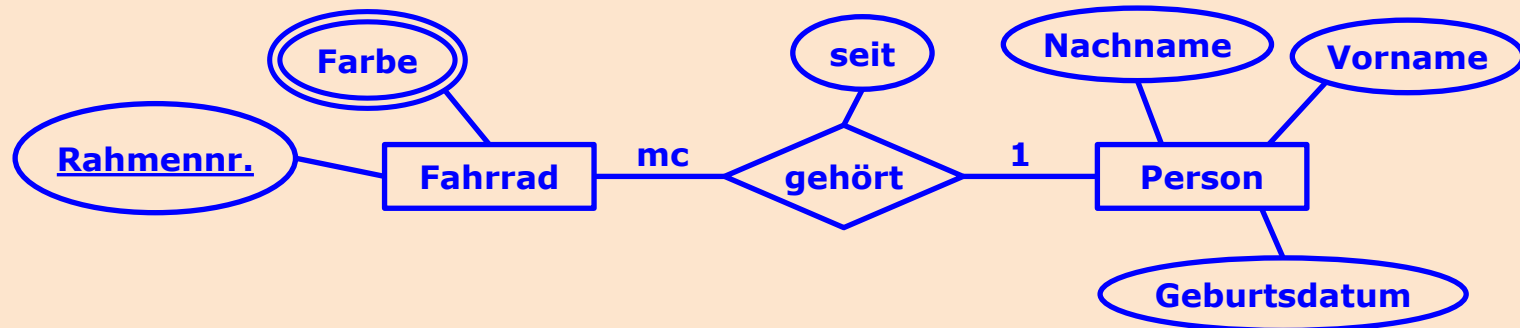
- **externe Phase/Anforderungsphase**
 - ermitteln des Informationsbedarfs der Nutzer_innen
 - Strukturierung der Information nach Output und Input
 - IS-Pflichtenheft
- **konzeptioneller Entwurf**
 - erstellen des Konzeptionellen/Semantischen Schemas
- **logischer Entwurf**
 - übersetzen des Konzeptionellen in das Logische Schema aufgrund der gewählten Datenbankstruktur
- **physischer Entwurf**
 - Datenschema mit DBMS implementieren
 - Optimierung der Datenhaltung (Indexe, Speicherstruktur)
 - Datenbanksprache

Konzeptionelles Schema/Modell

- **beschreibt die Miniwelt auf einer allgemeinen Ebene und auf hohem Abstraktionsniveau *unabhängig* von Datenbankstruktur, -system und Anwendungen mit folgenden Elementen:**
 - **Entitäten, Entitätsmenge: Realobjekte, Gedankenkonstrukte, Ereignisse**
 - **Beziehungen: logische Verknüpfungen zwischen den Entitäten**
 - **Attribute: Merkmale und Eigenschaften der Entitäten und Beziehungen**
 - **Primär-/Identifikationsschlüssel**
- **Darstellungswerkzeug für das Konzeptionelle Schema ist Entity-Relationship-Model (ERM)**
- **wird auch als Semantisches Schema bezeichnet**

Entity-Relationship-Modell (ERM)

- Darstellung des Konzeptionellen Schemas als Diagramm
- Darstellungselemente (Chen-Diagramm)
 - Entität (Rechteck)
 - Beziehung (Raute) mit Kardinalität
 - Attribut (Ellipse, mehrwertige mit doppeltem Rand)
 - Primär-/Identifikationsschlüssel (unterstrichen)



Umsetzungsschritte ERM

1. Entitäten bestimmen

2. Beziehungen definieren

- zwischen einer (rekursiv), zwei oder mehreren Entitäten
 - Anzahl der Verbindungen heisst Grad der Beziehung
- Kardinalität
 - einfach: 1 (genau ein)
 - konditionell: c (kein oder ein)
 - mehrfach: m (ein oder mehrere)
 - mehrfach-konditionell: mc (kein, ein oder mehrere)
- redundante Beziehungen vermeiden

3. Attribute für Entitäten und Beziehungen

- ein-, mehrwertig
- abgeleitet, berechnet

4. Primär-/Identifikationsschlüssel festlegen

Beispiel Adressliste

Personen-Nr.	Nachname	Vorname	Strasse und Hausnummer	PLZ und Wohnort	Kanton	Geburtsdatum
11-421-212	Pfister	Manuela	Stammweg 12	4230 Gerberach	LU	04.06.1988
12-524-077	Özgür	Belicai	Länggasse 5-9	7382 Hinterdorf	GR	12.03.1991
10-311-843	Havranek	Hansueli	Bahnhofstrasse 10A	1102 Wattendorf	SZ	05.11.1992
08-921-131	Schmitt	Armin	Hauptstr. 254	3200 Thunikon	BE	30.09.1985
10-634-574	De Vello	Sandra Anita	Rundweg 7	9799 Wattendorf	SG	14.01.1988
11-519-294	Schwab	Andrea	Jura-West-Quai 15/4	5785 Salitan	AG	22.11.1990
10-513-754	Meyer	Mariama	Meisenweg 11	5610 Mittelegg	AG	17.03.1991
09-168-712	Gigax	Rudolf	Höhenstrasse 5	7416 Gehrwil	GR	05.02.1989
12-154-577	Marx	Friedrich	Schildbürger Allee 50	7382 Vorderdorf	GR	31.09.1992
08-994-002	Engels	Karl	Fabrikstr. 47	4410 Finnwald	JU	22.01.1985

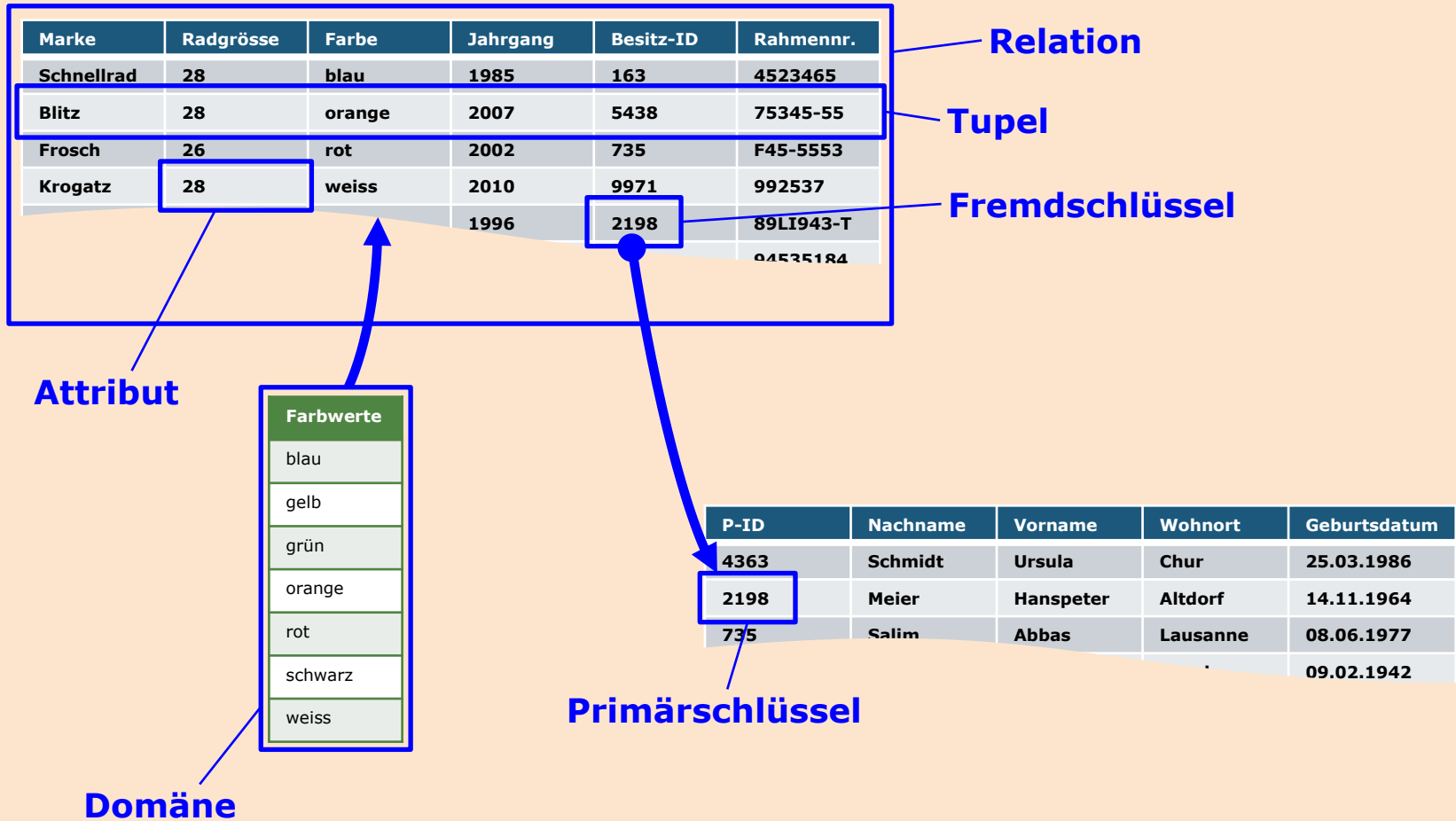
Logisches Schema / Modell

- **Das Logische Schema beschreibt die Daten aus dem ERM auf Grundlage einer Datenbankstruktur (Logisches Datenmodell)**
- **aktuelle und historische Datenbankstrukturen: Relationale Datenbanken (SQL-Datenbanken), Hierarchische DBs, Netzwerk-DBs, Objektorientierte DBs, Objektrelationale DBs, Einzeltabellen-DBs, XML-DBs, NoSQL-DBs**
- **Logisches Schema ist *abhängig* von der gewählten Datenbankstruktur**
- **Logisches Schema ist *unabhängig* von DBMS und Anwendungen**

Relationale Datenbanken (RDBS)

- **1969 entwickelt durch Edgar F. Codd (IBM), 1990 ergänzt durch die zwölf Codd'schen Regeln**
- **Strukturelemente eines RDBS**
 - **Relation (Tabelle mit den Sachdaten)**
 - **Tupel (Datensatz)**
 - **Attribut (Datenfeld)**
 - **Domäne (Wertebereich bzw. Datentyp eines Attributs)**
 - **Primärschlüssel**
 - **Fremdschlüssel (einfache Beziehungsabbildung)**
- **Structured Query Language (SQL) als weitgehend einheitliche Datenbanksprache**

Grundstrukturen eines RDBS



Vom ERM zum Relationalen Modell

- **Datenmodellierung durch Normalisierung**
- **Daten sollen redundanzfrei sein**
 - Redundanz besteht, wenn ein Teil der Daten ohne Informationsverlust weggelassen werden kann
 - keine mehrfache Datenpflege und keine Speicher-verschwendung
- **Probleme nicht normalisierter Daten**
 - Löschanomalie
 - Einfügeanomalie
 - Änderungsanomalie

Anomalien (Bsp. Adressliste)

- **Löschanomalie**

- Bsp.: Wird ein Person aus der Datenliste Adressen gelöscht, kann es passieren, dass eine Ortschaft ganz aus den Daten "verschwindet".

- **Einfügeanomalie**

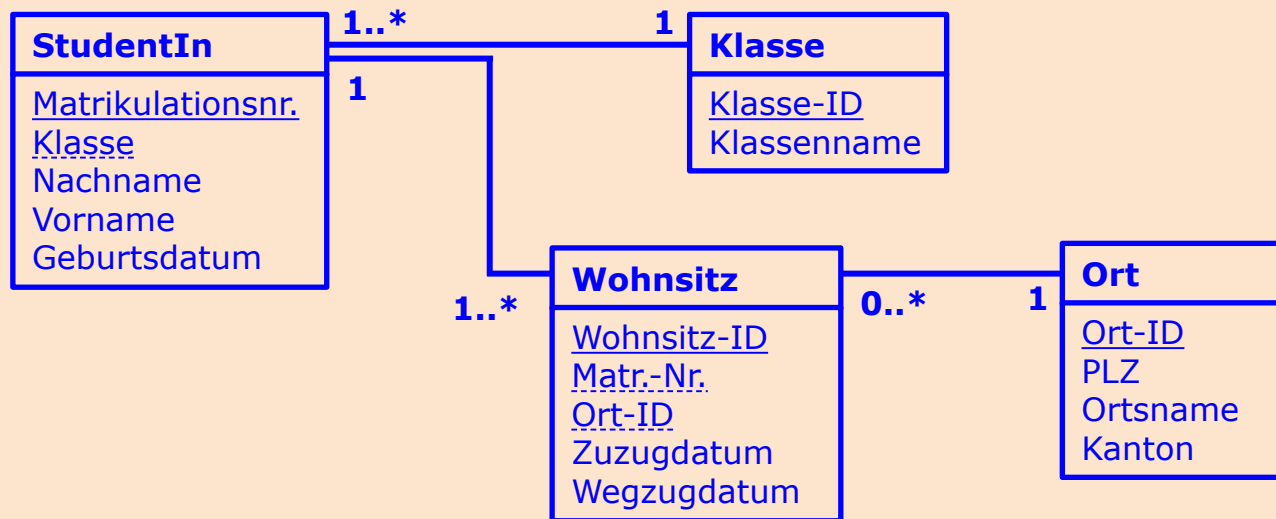
- Bsp.: Die Kantone können nicht lückenlos erfasst werden, da es Kantone gibt, zu denen keine Personendaten existieren.

- **Änderungsanomalie**

- Bsp.: Ändert eine Strasse ihren Namen, muss jeder Personen-Datensatz mit dieser Strassenanschrift einzeln geändert werden. Die Änderung *eines* Sachverhalts erfordert also *mehrere* Datenänderungen.

Relationales Modell in UML

- Relation als Rechteck mit Titel
- Attributliste (mit Domänenangabe)
- Primärschlüssel durchgezogen, Fremdschlüssel gestrichelt unterstrichen
- Beziehungen als Linie mit Kardinalität



Umsetzungsschritte ERM → RDM

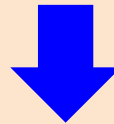
- 1. jede Entität mit ihren Attributen erzeugt eine eigene Relation**
 - meist künstlicher Schlüssel als Primärschlüssel anstelle eines "semantisch belasteten" Identifikationsschlüssel
- 2. Beziehung erzeugt eine eigene Relation, wenn**
 - a) sie zwischen mehr als zwei Entitäten besteht
 - b) es eine m:n-Beziehung ist
 - c) sie selbst Attribute besitzt
- 3. für jede Beziehung wird in der referenzierenden Relation ein Fremdschlüssel erstellt**
- 4. Entitäten/Relationen mit einer 1:1-Beziehung werden ggf. zu einer Entität/Relation zusammengefasst**

Normalisierung

- **1. Normalform (NF): alle Attribute besitzen atomare Wertebereiche**
- **2. NF: jedes Attribut ist voll funktional abhängig vom Primärschlüssel**
 - **B ist *funktional abhängig* von A, wenn zum Merkmal A nur ein Merkmal B existiert: "B ist durch A eindeutig bestimmt"**
 - **ist B *voll funktional abhängig* von A, ist B nicht nur von einem Teil von A abhängig**
- **3. NF: kein Attribut ist transitiv abhängig**
 - **C ist *transitiv abhängig* von A, wenn B von A und C von B funktional abhängig ist**
- **Boyce-Codd-NF, 4. und 5. NF → vgl. Literatur**

Beispiel 1. Normalform

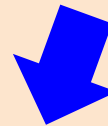
PID	Dozent_in	Fächer	Geburtsdatum	Alter
1	Müller Sandra	Mathematik, Physik, Lineare Algebra	23.02.1964	48
2	Meier Hans	Deutsch, Englisch	11.09.1975	37



PID	Dozent_in	Fach	Geburtsdatum	Alter
1	Müller Sandra	Mathematik	23.02.1964	48
1	Müller Sandra	Physik	23.02.1964	48
1	Müller Sandra	Lineare Algebra	23.02.1964	48
2	Meier Hans	Deutsch	11.09.1975	37
2	Meier Hans	Englisch	11.09.1975	37

Beispiel 2. Normalform

PID	Dozent_in	Fach	Geburtsdatum	Alter
1	Müller Sandra	Mathematik	23.02.1964	48
1	Müller Sandra	Physik	23.02.1964	48
1	Müller Sandra	Lineare Algebra	23.02.1964	48
2	Meier Hans	Deutsch	11.09.1975	37
2	Meier Hans	Englisch	11.09.1975	37

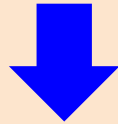


PID	Dozent_in	Geburtsdatum	Alter
1	Müller Sandra	23.02.1964	48
2	Meier Hans	11.09.1975	37

FID	PID	Fach
1	1	Mathematik
2	1	Physik
3	1	Lineare Algebra
4	2	Deutsch
5	2	Englisch

Beispiel 3. Normalform

PID	Dozent_in	Geburtsdatum	Alter
1	Müller Sandra	23.02.1964	48
2	Meier Hans	11.09.1975	37



**Alter wird bei Bedarf aus
Geburtsdatum berechnet**

PID	Dozent_in	Geburtsdatum
1	Müller Sandra	23.02.1964
2	Meier Hans	11.09.1975

FID	PID	Fach
1	1	Mathematik
2	1	Physik
3	1	Lineare Algebra
4	2	Deutsch
5	2	Englisch

Bsp. funktionale Abhängigkeit

AHV-Nr.	Name	Vorname	Geburtsdatum	Alter	PLZ	Wohnort	Kanton
756.4567.1234.99	Müller	Hans	13.12.1961	52	9513	Städtle	SG
756.5657.2647.91	Müller	Petra	05.05.1985	28	1225	Hauptwil	VD
616.6541.1254.87	Müller	Hans	12.02.1944	70	3451	Oberort	BE
756.1234.9814.00	Meier	Paul	08.09.1973	40	1234	Vorderstadt	VD
756.1265.1249.89	Meier	Yvonne	12.02.1944	70	6647	Hauptwil	LU

- **Geburtsdatum ist *funktional abhängig* von der Attributkombination <AHV-Nr., Name, Vorname>, nicht aber von <Name, Vorname>**
- **Geburtsdatum ist *voll funktional abhängig* von AHV-Nr., nicht aber von <AHV-Nr., Name>**
- **Alter ist *transitiv abhängig* von Geburtsdatum, nicht aber von <Name, Vorname>**
- **Kanton ist *transitiv abhängig* von <PLZ, Wohnort>, nicht aber von Wohnort**

Physisches Schema/Modell

- **physische Implementierung des logischen Modells auf einem konkreten DBMS**
- ***unabhängig* von den Anwendungen**
- **Elemente aus dem logischen Datenmodell werden physisch umgesetzt**
 - **Nebst den Elementen des logischen Modells sind dies Systemtabellen zur Modellbeschreibung (Metadaten), Integritätsbedingung, Indizes, Zugriffsrechte etc.**

für RDBS:

- **Relation → Tabelle**
- **Tupel → Datensatz (Tabellenzeile)**
- **Attribut → Datenfeld (Tabellenspalte)**
- **Domäne → Wertebereich (Wertetabelle), Datentyp**

Integrität (integer)

- **Integrität ist ein Begriff für die Qualität und Zuverlässigkeit von Daten**
 - **semantische Integrität: Miniwelt korrekt abgebildet**
 - **logische Integrität: logisch korrekt, widerspruchsfrei**
 - **Datenintegrität: inhaltlich korrekt und aktuell**
- **Datenbank ist konsistent, wenn sie in allen Aspekten integer, also widerspruchsfrei ist**
- **Integritätsanforderungen können auf allen Modellierungsebenen formuliert werden**
- **DBMS muss Integrität sicherstellen, z.B. mittels Transaktionen (Bündelung von atomaren Aktionen, die insgesamt den konsistenten Zustand der Datenbank erhält: "alles oder nichts")**

Schlüssel

- ***Schlüssel***: ein Attribut/eine Attributkombination, das/die eine Entität identifiziert
- ***Superschlüssel***: ein Schlüssel, der die Entität eindeutig identifiziert
- ***Identifikationsschlüssel (Schlüsselkandidat)***: minimale Attributkombination, die die Entität eindeutig identifiziert
- ***Primärschlüssel***: 1 gewählter Schlüsselkandidat
 - übrige Identifikationsschlüssel sind *Alternativschlüssel*
 - als Attributkombination auch *Verbundschlüssel* genannt
 - meist als *künstlicher Schlüssel* umgesetzt
- weitere Schlüssel-Begriffe: *Suchschlüssel, Sekundärschlüssel, Fremdschlüssel, Sortierschlüssel, partieller Schlüssel*

Integritätsbedingungen in RDBS

- **Eindeutigkeitsbedingung: jedes Tupel in einer Relation ist unverwechselbar**
- **Domänenbedingung und Gültigkeitsregeln (Constraints): Attribute dürfen nur Werte aus dem gültigen Wertebereich enthalten und weitere Bedingungen aus dem Anwendungskontext**
- **referentielle Integrität: jeder Fremdschlüsselwert referenziert einen gültigen Primärschlüssel**

Datenkatalog, Data-Dictionary

- **beschreibt das Datenbankschema umfassend**
- **bildet alle Tabellen, Attribute und Beziehungen mit erklärenden Beschreibungen ab, benennt alle Integrationsbedingung sowie Prüfinformationen, Schlüsselbeziehungen, Zugriffsregeln, Wertebereiche, Indizes etc.**
- **dient der DB-Administration als Referenz und Nachschlagewerk bei der Erstellung, Wartung und Nutzung der Datenbank**

weitere Begriffe aus der DB-Welt

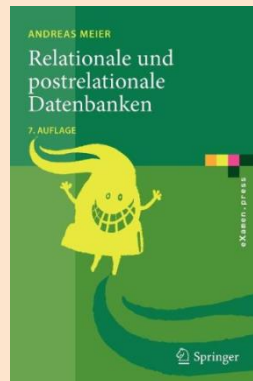
- **Metadaten: Daten über Daten**
- **Persistenz: Fähigkeit der permanenten Datenspeicherung und -bereitstellung**
- **Mirroring: Datenspiegelung zur Lastverteilung**
- **Replikation: Datenspiegelung mit Synchronisationsmechanismen**
- **Recovery: Datenwiederherstellung**
- **Skalierbarkeit: charakterisiert das Leistungsverhalten bei zunehmender Last**

... und noch mehr Begriffe

- **Benchmarking: empirischer Leistungstest für DBS**
 - Antwortzeit (Response Time)
 - Transaktionen pro Sekunde (tps)
- **Data Mining**
 - Datenanalyse ohne vorhergehende bzw. "vage" Fragestellung und Datenstrukturierung
 - Informationen ergeben sich (hoffentlich) durch Anwenden von systematischen (aber nicht gezielten) Strukturierungsmethoden (statistisch, heuristisch, visuell, Aggregation, Pivoting)
 - gewonnene Resultate werden vorerst als Hypothesen betrachtet

Literatur – Einstieg und Vertiefung

- E** Eirund Helmut, Kohl Ullrich: *Datenbanken – leicht gemacht. Ein Arbeitsbuch für Nicht-Informatiker* (2010)
- E&V** Elmasri Ramez et al.: *Grundlagen von Datenbanksystemen* (2009)
- E** Meier Andreas: *Relationale und postrelationale Datenbanken* (2010)
- E&V** Saake Gunter et al.: *Datenbanken – Konzepte und Sprachen* (2013)
- E&V** Vossen Gottfried: *Datenmodelle, Datenbanksprachen und Datenbankmanagementsysteme* (2008)
- E** Zehnder Carl August: *Informationssysteme und Datenbanken* (2005)



weitere verwendete Literatur

Arbeitskreis Datenbanken im Unterricht, *Datenbank*. Augsburg 2000

http://www2.ludwig-geissler-schule.de/materialien/Fachbereiche-und-Materialien/Datentechnik/Datenbanken/_mysql/datenbanktheorie_bayern.pdf

Blümel Bernd et al., *Datenbanken und Tabellen aus dem ERM, SQL*. 2012

www.fb6.info/winf/bernd_bluemel/scriptewi1/datenbank.pdf

GITTA, *Database Management and Systems Module*. 2011

www.gitta.info/IntroToDBS/de/html/index.html, www.gitta.info/DBSysConcept/de/html/index.html,
www.gitta.info/LogicModelin/de/html/index.html

Hartmut Ernst, *Grundkurs Informatik. Kap. 14.2 Datenbanken*. 2003

<http://tiefe-wun.de/~adarma/Studium/Gr%20InfV/Lehrbuch/inf14.pdf> (zuletzt online erreichbar am 17.09.2012)

Holzinger Andreas, *Basiswissen IT/Informatik, Band 2: Informatik*. Würzburg 2003

Kemper Alfons, Eickler André, *Datenbanksysteme: Eine Einführung*. 7. Auflage, München 2009

Kersken Sasha, *IT-Handbuch für Fachinformatiker. Kap. 12 Datenbanken*. Bonn 2011

http://openbook.galileocomputing.de/it_handbuch/kap_12_datenbanken_001.html

Nebiker Stephan, *Datenbankverwaltungssysteme*. Vorlesungsmanuskript, Muttenz 2002

Schuldt Heiko, *Datenbanken. Vorlesungsmanuskript*, Basel 2013

<http://informatik.unibas.ch/index.php?id=128>

Steiner René, *Grundkurs Relationale Datenbanken: Einführung in die Praxis der Datenbankentwicklung für Ausbildung, Studium und IT-Beruf*. 7. Auflage, Wiesbaden 2009

Letzter Aufruf 04.02.2015