

Prof. DI Dr. Erich Gams

# Datenbanken

Einführung & Motivation

informationssysteme htl-wels

# Übersicht ☞ Was lernen wir?



- › Warum INSY?
- › Ein leichter Einstieg



- › Überblick über den Jahresstoff
- › .. dann geht's los!

# Was ist INSY?

- › INSY – Kompetenzbereich „Datenbanksysteme“
- › Eigenschaften und Architekturen von Datenbanken/ Datenbanksystemen, Datenmodelle, Datenbank Anwendungen

Warum Datenbanken?

# Arten von Daten

## › Unternehmen

- Information über Kunden (Jö Karte), Lieferanten, Waren, Bestellungen, Buchhaltung, Marktstudien, Konstruktionsdaten (Auto, Motor, Kugellager,...) usw.

## › Verwaltung

- Meldedaten, Führerschein, Strafregister, Krankenversicherung, etc.

## › Wissenschaft

- Messdaten, fachspezifische Datensammlungen (z.B. Protein-Eigenschaften), Analyseergebnisse

## › Privat

- Telefonnummern, Email-Kontakte, Online-Zugänge, Familienfotos, MP3-Sammlung

## › Geoinformationsdaten (Straßen, Leitungen, Flüsse, Immissionen,...)

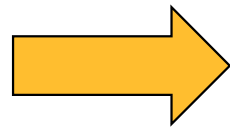
# Datenbanken

- › Daten sind wertvoll!
- › Daten sind ein Produktionsfaktor mit dem Firmen viel Geld verdienen.
  - **Facebook** hat 180k Server (Ende 2012)
  - **Google** hat 450k Server (2006), über 1 Million heute, 5,6 Milliarden Suchanfragen pro Tag

➡ Welche Produktionsfaktoren gibt es noch?

# Datenbanken

- › Daten müssen gespeichert und verwaltet werden



Datenbanken

- › In fast jeder Applikation steckt heutzutage eine Datenbank!
- › Ihr habt täglich mit einer Datenbank zu tun! Wo?

# Kriterien von Datenbanken

› Was ist eine Datenbank und welche Kriterien sollte eine Datenbank erfüllen?

› Erstes Kriterium:



Die Daten müssen **persistent** gespeichert werden.

# Kriterien von Datenbanken

## › Zur Erklärung des Wortes **persistent**:



1. Im Lateinischen bedeutet das Verb „persistere“ stehen bleiben, verharren.
2. Im Englischen bedeutet „persistent“ u.a. „andauernd, beständig“.
3. In der Informatik hat „persistent“ die spezielle Bedeutung, dass man eine Speicherung persistent nennt, wenn sie unabhängig von der Laufzeit von Programmen oder der Ein- und Abschaltung der Rechner ist.



# Beispiel: Excel Liste Adressbuch

- › Wie speichere ich die Adressen?
  - strikt alphabetisch **nach Nachnamen** oder
  - **in Gruppen** oder
  - **dienstliche und private Adressen**, oder
  - in **Bekannte aus dem Inland und aus dem Ausland**, oder
  - einige **Adressen unter dem Anfangsbuchstaben des Nachnamens**, andere wieder unter dem **Anfangsbuchstaben des Vornamens** usw.
- › Ich muss mich auf jeden Fall entscheiden!



# Beispiel: Excel Liste Adressbuch

- › **Was möchte ich nun mit den Adresseinträgen machen?**
- › Jetzt möchte ich mein Excel Adressbuch beispielsweise verschiedenste Sachen fragen können:
  - Zeige mir den Eintrag zu meinem Freund Charlie Chaplin.
  - Zeige mir alle Einträge, wo der Ort „Linz“ ist.
  - Zeige mir, ob es in meiner Datenbank einen Eintrag mit der Telefonnummer „069/123456“ gibt.
  - Zeige mir alle Einträge - sortiert nach Namen.
  - Zeige mir alle Einträge – sortiert nach Wohnorten, Straßen und Hausnummern
- › Und vieles andere mehr .....
- › Je komfortabler die **Suchfunktionalitäten** sind, über die die Datenbank verfügt, desto mehr ist sie dem Excel Adressbuch überlegen.

# Kriterien von Datenbanken

## › Zweites Kriterium:



Es muss die unterschiedlichsten **Suchzugriffe** auf den Datenbestand geben.

Ich brauche verschiedenste **Auswahl-** bzw. **Suchfunktionen** für meine Datenbank, mit denen ich durch Beschreibung des gewünschten Ergebnisses die gewünschten Anzeigen erhalte.

# Kriterien von Datenbanken

- › Was kann ich noch alles mit mit einem Adressbuch bzw. den darin enthaltenen Daten machen?
  - Zunächst ist mein Adressbuch (Excel Sheet) völlig leer.
  - Sie können neue Einträge in dieses Adressbuch **einfügen**.
  - Sie können einzelne Einträge in diesem Adressbuch **verändern**.
  - Sie können Einträge herausstreichen, d.h. **löschen**.
  - Sie können das gesamte Adressbuch wegschmeißen/löschen.

# Kriterien von Datenbanken



Des Weiteren brauchen wir also sogenannte **Verwaltungsfunktionen** für unsere Daten.

- › Diese nennt man **CRUD** Funktionen
  - Create
  - Read
  - Update
  - Delete

# Beispiel Verwaltungsfunktionen

- › Die telefonische Vorwahl eines Ortes ändert sich beispielsweise von 01234 in 09876.
  - Jetzt möchtest du nicht alle Einträge einzeln durchgehen und ändern. Das ist fehleranfällig!
  - Bequemer ist: Liebe Datenbank ändere alle Einträge mit der Vorwahl 01234 in Einträge mit der Vorwahl 09876.

# Beispiel Verwaltungsfunktionen

- › Ich habe in meinem Adressbuch viele Einträge von Personen aus der Stadt Frankfurt am Main . Wenn ich für alle diese Personen immer wieder neu die Angabe „Frankfurt am Main“ machen würde, kämen unter anderem sicherlich die folgenden Varianten vor:
  - Frankfurt am Main
  - Frankfurt/Main
  - Frankfurt
  - Frankfurt a. M.
  - FFM
  - Frnkft.

# Kriterien von Datenbanken

- › Komplizierte (unmögliche)  
Realisierungsmöglichkeit der beschriebenen  
Such- und Verwaltungsfunktionen
- › Wie wollen Sie hier z.B. Ihrer Datenbank mitteilen:  
„Zeige mir alle Personen aus Frankfurt“???

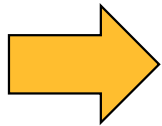


# Kriterien von Datenbanken

- › **Die Lösung dieses Problems:**
- › Datenbank so zu konzipieren, dass
  - › der Text „Frankfurt am Main“ **nur ein einziges Mal** gespeichert wird und
  - › bei allen Personen aus Frankfurt auf diesen einen Eintrag verwiesen wird.
- › Durch solch ein Vorgehen vermeidet man die **Redundanz (d.i. das mehrfache Vorkommen derselben Sache)** von Daten und damit mögliche Fehlerquellen.

# Um die Sache noch komplizierter zu machen ....

- › Eine Adresse ist doppelt eingetragen:
  - beispielsweise unter dem Vornamen einsortiert und auch unter dem Nachnamen eingetragen.
  - Unter „C“ steht: Charli D'Amelio, ..... , Handy: 0175/007
  - Unter „D“ steht: D'Amelio Charli, ..... , Handy: 0175/007



Handynummer wird geändert.

- › Dann kann es leicht passieren, dass man die Änderung nur bei einem Eintrag vornimmt.
- › Ihr Adressbuch enthält dann die Daten:
  - Unter „C“: Charli D'Amelio, ..... , Handy: 0175/0815
  - Unter „D“: D'Amelio Charli, ..... , Handy: 0175/007

# Kriterien von Datenbanken

- › Was Sie jetzt haben, nennt man **Dateninkonsistenz**
- › Auch sie kann durch die Vermeidung von **Datenredundanz** (das mehrfache Abspeichern ein und derselben Person) verhindert werden.

# Ein paar Begriffsdefinitionen

- › **In·kon·sis·tenz**, die; -, -en 1. Uneinheitlichkeit, Zusammenhanglosigkeit, Widerspruch,  $\leftrightarrow$  Konsistenz(1) 2. keine Mehrzahl  $\rightarrow$  inkonsistente (2) Beschaffenheit,  $\leftrightarrow$  Konsistenz (2)
- › **Kon·sis'tenz**, die; -, keine Mehrzahl 1. Einheitlichkeit, Zusammenhang, Widerspruchsfreiheit,  $\leftrightarrow$  Inkonsistenz(1)
- › **Re·dun'danz**, die; -, -en 1. Überfluss, Überzähligkeit 2. überflüssige Mehrfachkennzeichnung einer Information 3. Weitschweifigkeit 4. TECHNIK Mehrfachauslegung von Systemen zur Sicherung

aus Langenscheidts Fremdwörterbuch

# Warum Datenbanken?



Wir fordern also von unserer Datenbank:  
Sie soll so konstruiert sein, dass man die  
Daten **redundanzfrei** speichern kann.

# Kriterien von Datenbanken

- › **Ein „öffentliches“ Beispiel: Online-Händler:**
- › Wenn Sie z.B. bei einem Online-Händler ein Buch suchen und bestellen wollen:
- › Sie greifen auf eine Datenbank der Bücher (und deren Eigenschaften) des Online-Händlers zu.
- › Gleichzeitig mit Ihnen greifen möglicherweise noch viele andere potentielle Kunden ebenfalls auf diese Datenbank und auch auf exakt dieselben Informationen zu.

# Warum Datenbanken?



Ein Datenbanksystem muss für  
einen **Mehr-Benutzer-Betrieb** ausgelegt sein.

# Überblick Jahresstoff Fach INSY

1. Grundbegriffe, Fachbegriffe, „Fremdwörter“: DB, DBMS, DBS, IS
2. Datenabstraktion, Datenunabhängigkeit, Datenbankarchitektur, Datenbankentwurf
3. Datenbankmodellierung
  - Entity-Relationship-Modell und Erweiterungen (Chen und EERM)
    - Beziehungen mit den verschiedenen Kardinalitäten
    - Weak Entities
    - Ein- und mehrstufige Rekursion
    - Generalisierung und Spezialisierung
  - Modellierung mit UML



# Überblick Jahresstoff Fach INSY

## 1. Das relationale Modell

- Das relationale Datenmodell nach Codd
- Relationale Datenbanken: Synonyme und Bedeutungen
- Übliche skalare Datentypen von Attributen
- Wichtige Eigenschaften von Relationen
- Überführungsstrategie: Abbildung von ER-Modelle auf das relationale Datenmodell

## 2. Anomalien und die Normalformen 1,2,3 und BCNF

## 3. Die relationale Datenbankabfragesprache SQL

# Wie kriege ich eine positive Note ?



- › SMÜPs, PLFs (Hauptgewicht)
- › Referate (z.B.: Vorstellung eines ER-Tools)
- › Mitarbeit (Mitschrift, Wortmeldungen, Aufmerksamkeit! können die Note verbessern oder verschlechtern)
- › (Hausübungen)
- › (WH)

# Daten, Information, Wissen



## › Aufgabe:

- Finde **mindestens** 2 Definitionen je Begriff
- Erkläre jeden Begriff mit **eigenen** Worten
- Finde zu jedem Begriff ein selbsterklärendes **Beispiel**

## › Zitieren!!!!

# Definition: Daten

## › Daten

- bestehen aus einem oder mehreren Zeichen, welche im Zusammenhang gesehen einen sinnvollen Inhalt ergeben.
- Auf dieser Ebene der Begriffshierarchie kann jedoch noch keine Aussage über den Verwendungszweck der Daten getroffen werden.
- Menge von Fakten, die formatiert und maschinell verarbeitbar festgehalten und explizit bekannt sind, z.B. elektronisch in der vorgegebenen Struktur einer Datenbank oder als unstrukturierter Volltext

## › Langenscheidt: **elektronisch gespeicherte, verarbeitbare Werte.**

# Daten

|          |          |          |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|
|          | 5.506,78 | 5.805,28 | 6.103,87 | 6.402,42 | 6.701,00 | 6.999,60 | 7.298,10 | 7.596,68 |
|          | 5.075,78 | 5.307,81 | 5.539,74 | 5.771,75 | 6.003,75 | 6.235,77 | 6.467,83 | 6.699,77 |
|          | 4.512,41 | 4.735,45 | 4.958,50 | 5.181,52 | 5.404,55 | 5.627,60 | 5.850,63 | 6.073,66 |
|          | 3.999,78 | 4.204,63 | 4.409,56 | 4.614,37 | 4.819,22 | 5.024,11 | 5.228,95 | 5.433,84 |
|          | 3.729,42 | 3.916,12 | 4.102,86 | 4.289,62 | 4.476,40 | 4.663,14 | 4.849,91 | 5.036,66 |
| 3.554,76 | 3.729,42 | 3.904,01 | 4.078,65 | 4.253,31 | 4.427,95 | 4.602,61 | 4.777,21 | 4.951,85 |
| 3.222,33 | 3.382,16 | 3.541,94 | 3.701,72 | 3.861,53 | 4.021,32 | 4.181,11 | 4.340,92 | 4.500,74 |
| 2.946,31 | 3.073,12 | 3.199,84 | 3.326,62 | 3.453,32 | 3.580,11 | 3.706,86 | 3.833,64 | 3.960,41 |
| 2.605,22 | 2.705,64 | 2.806,04 | 2.914,54 | 3.025,94 | 3.137,41 | 3.248,87 | 3.360,31 | 3.471,78 |
| 2.605,22 | 2.705,64 | 2.806,04 | 2.914,54 | 3.025,94 | 3.137,41 | 3.248,87 | 3.360,31 | 3.471,78 |
| 2.462,67 | 2.553,17 | 2.643,79 | 2.738,83 | 2.833,89 | 2.932,95 | 3.038,38 | 3.143,92 | 3.249,35 |
| 2.332,10 | 2.402,06 | 2.471,96 | 2.541,93 | 2.611,81 | 2.683,84 | 2.757,28 | 2.830,71 | 2.905,45 |
| 2.332,10 | 2.402,06 | 2.471,96 | 2.541,93 | 2.611,81 | 2.683,84 | 2.757,28 | 2.830,71 | 2.905,45 |
| 2.160,52 | 2.217,31 | 2.274,13 | 2.330,91 | 2.387,74 | 2.444,52 | 2.501,31 | 2.558,16 | 2.614,93 |
| 1.998,69 | 2.050,60 | 2.102,60 | 2.154,52 | 2.206,48 | 2.258,42 | 2.310,42 | 2.362,35 | 2.414,31 |
| 1.933,28 | 1.984,96 | 2.036,61 | 2.088,27 | 2.139,90 | 2.191,54 | 2.243,16 | 2.294,83 | 2.346,33 |
| 1.860,82 | 1.907,98 | 1.955,09 | 2.002,21 | 2.049,34 | 2.096,50 | 2.143,63 | 2.190,75 | 2.230,61 |
| 1.727,89 | 1.775,02 | 1.822,20 | 1.869,30 | 1.916,45 | 1.963,56 | 2.010,70 | 2.057,86 | 2.104,95 |

# Daten

- › Daten haben nur eine **syntaktische** Dimension.
- › In Unternehmen mit gewachsenen IT-Infrastrukturen sind Daten in großen Mengen auf Datenträgern vorhanden.
- › Dadurch kommt es zu einer Daten-Überflutung der Akteure.
- › Konzepte wie Data Warehouse oder das Informationsmanagement sind der Versuch, aus der Datenflut die Informationen, die für die Organisation wichtig sind, zu extrahieren und sinnvoll aufzubereiten.
- › Erst wenn aus Daten Information und Wissen wird, erhalten sie ihren Wert.

# Information


- › **Daten sind die Bausteine von Informationen.**
- › Beispielsweise kann ein Datenfeld "Name" je nach der Sichtweise des Betrachters als Vorname, Nachname, Firmenname, etc. betrachtet werden. Dies hängt einerseits von dem Betrachter selbst- andererseits von dem jeweiligen Kontext ab, in den die Information verpackt ist, wenn der Betrachter mit ihr konfrontiert wird.
- › Grundsätzlich kann man **Information als mit Kontext angereicherte Daten** betrachten,,
- › Information ist eine spezielle Kenntnis über Entitäten wie Fakten, Ereignisse, Dinge, Prozesse, Personen, Ideen oder Konzepte.
- › Informationen haben eine **semantische Dimension.**

# Information

|        | <i>CHIP</i> | <i>Computer<br/>bild</i> | <i>c't</i> | <i>PC Welt</i> | <i>Summe</i> |
|--------|-------------|--------------------------|------------|----------------|--------------|
| 1998-I | 278.727     | 923.432                  | 331.098    | 343.875        | 1.877.132    |
| 1999-I | 321.282     | 1.224.107                | 370.429    | 391.292        | 2.307.110    |
| 2000-I | 345.239     | 1.155.124                | 366.792    | 394.002        | 2.261.157    |
| 2001-I | 404.791     | 1.031.921                | 390.037    | 462.080        | 2.288.829    |
| 2002-I | 431.111     | 990.680                  | 394.626    | 511.315        | 2.327.732    |
| 2003-I | 511.297     | 920.915                  | 398.960    | 512.048        | 2.343.220    |
| 2004-I | 412.237     | 949.888                  | 394.218    | 487.273        | 2.243.616    |
| 2005-I | 402.328     | 841.374                  | 388.708    | 466.473        | 2.098.883    |
| 2006-I | 404.098     | 771.350                  | 358.834    | 429.906        | 1.964.188    |
| 2007-I | 406.412     | 732.223                  | 350.720    | 422.118        | 1.911.473    |
| 2008-I | 402.382     | 758.257                  | 355.813    | 397.203        | 1.913.655    |
| 2009-I | 408.149     | 731.261                  | 352.796    | 392.423        | 1.884.629    |



# Definition: Wissen

- › Informationen können als Produktionsfaktor (Ressource) angesehen werden und damit entsprechend beplant, beschafft und ihr Einsatz wirtschaftlich gesteuert werden.  Informationsmanagement
- › **Wissen (höchste Stufe der Begriffshierarchie):**
  - Wissen verbindet alle Eigenschaften der unteren Ebenen und erweitert diese.
  - "Wissen kann beschrieben werden als in einen bestimmten Kontext gestellte Information, die für denjenigen, der über diese Information verfügt, von Wert ist und ihn dazu befähigt, etwas zu tun, wozu er ohne dieses Wissen nicht in der Lage gewesen wäre."
  - Es ist seinem Wesen nach **mit menschlichem Handeln verbunden**.
  - Von **Wissen kann nur in Verbindung mit dem Betrachter** gesprochen werden.

# Jede Menge Abkürzungen

## DB,DBMS,DBS,IS

- › **Datenbank (DB) :** **Datenmodell zusammen mit einer Sammlung von Daten,**
  - die einen Ausschnitt der realen Welt beschreiben.
- › Die Aspekte der realen Welt, die dargestellt werden, nennt man auch Miniwelt oder Universe of Discourse.
- › ist eine logisch zusammenhängende Sammlung von Daten.
- › wird für bestimmte Zwecke entworfen entwickelt und mit Daten befüllt.
- › **Strukturierte Datensammlung.**

# Jede Menge Abkürzungen

## DB,DBMS,DBS,IS

- › Datenbankmanagementsystem (DBMS)
  - Software zum Anlegen und Verwalten von Datenbanken.
- › Datenbanksystem (DBS)
  - Ein Datenbank-Management-System zusammen mit einer oder mehrerer konkreten Datenbanken.
  - Sozusagen die Kombination aus Datenbank, DBMS und Hardware

# Informationssysteme (IS)

## › Bank-IS

- Daten: Kunden, Konten, Devisen, Wertschriften, Kredite, Hypotheken, ...
- Anwendungen: Buchhaltung, Zahlungsverkehr (> 50000 pro Tag), Anlageberatung, Kreditprüfung, Risikomanagement, ...

## › IS einer Fluggesellschaft

- Daten: Flüge (>1000/Tag), Passagiere, Reservationen & Buchungen (> 5 Mio./Jahr), Gepäckstücke, Flugzeuge (> 1000), Flugpersonal, ...
- Anwendungen: Reservationen & Buchungen (von > 10000 Terminals, bis zu 20/Sek.), Check-In, Gepäckzustellung, Einsatzplanung, Flugplanoptimierung, ...

## › Bibliotheks-IS

- Daten: Dokumente: Bücher, Zeitschriftenbände, Forschungsberichte, Videos (z.T. > 20 Mio. Dokumente), Titel, Autoren, Kategorien, Deskriptoren, Abstracts, ...
- Anwendungen: Recherchen nach relevanter Literatur (z.T. 100/Sekunde)

## › Weitere:

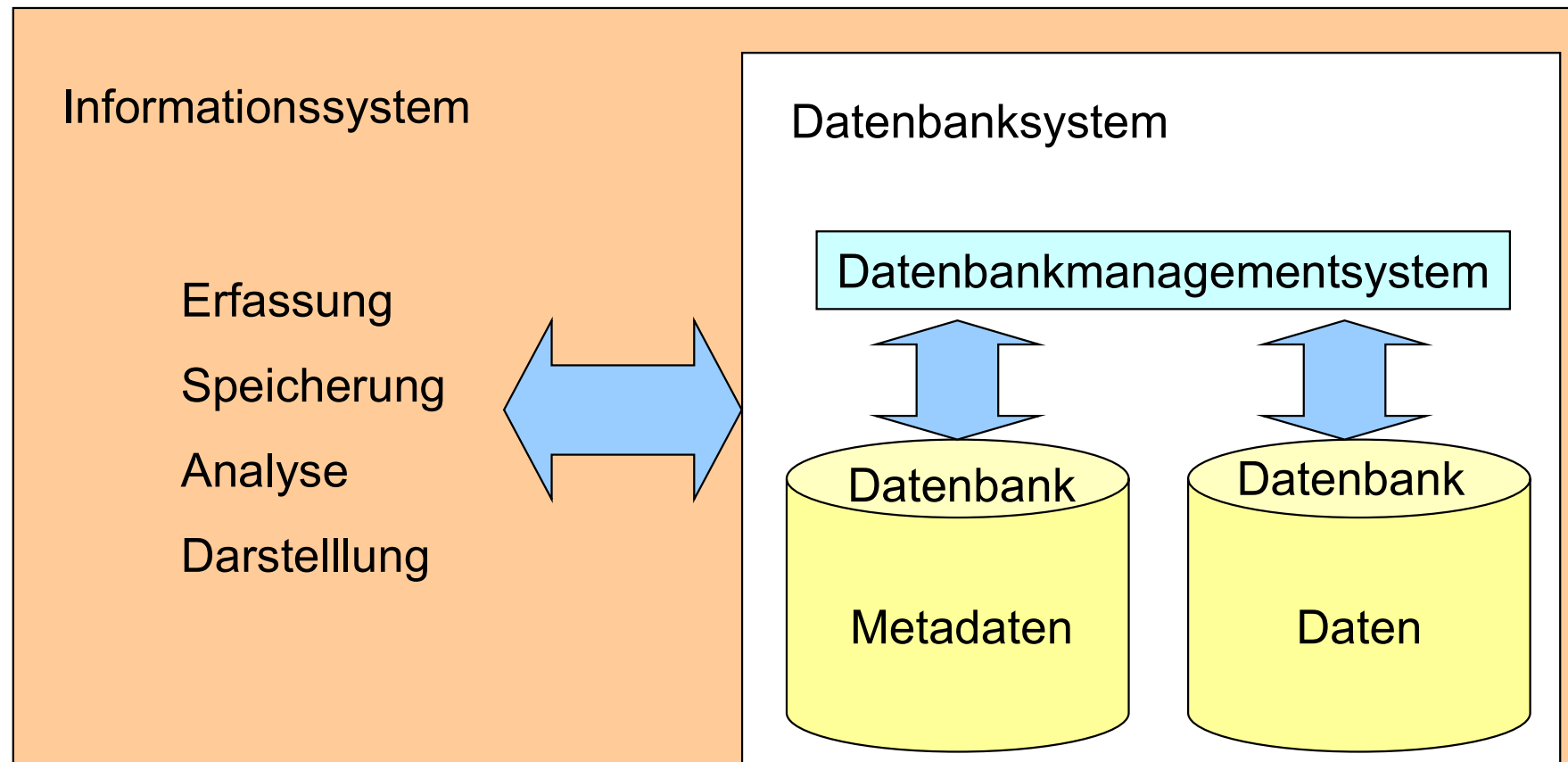
- Enterprise Resource Planning ERP (Personal, Buchhaltung, Lager, Produktion, Prozesskontrolle)
- Decision Support (statistische Auswertungen)
- E-Business (Kataloge, Aufträge)
- Websuche (Volltext, Struktur von Webpages)

# Informationssysteme

- › Ein **Informationssystem** erweitert die Datenbank um eine Reihe von Werkzeugen (engl. software tools) zur Abfrage, Darstellung, Transformation und Analyse von Daten.

[http://www.gitta.info/IntroToDBS/de/html/DefinOfTerms\\_dataAndInfo.html](http://www.gitta.info/IntroToDBS/de/html/DefinOfTerms_dataAndInfo.html)

# Überblick







Auf los geht's los :-)

