Prof. Dr. Th. Härder
AG Datenbanken und Informationssysteme
Zi. 36/330, Tel.: 0631-205-4030
E-Mail: haerder@informatik.uni-kl.de
http://wwwdbis.informatik.uni-kl.de

Grundlagen Betrieblicher Informationssysteme

Sommersemester 2002

Universität Kaiserslautern
Fachbereich Informatik
Postfach 3049
67653 Kaiserslautern

Vorlesung:

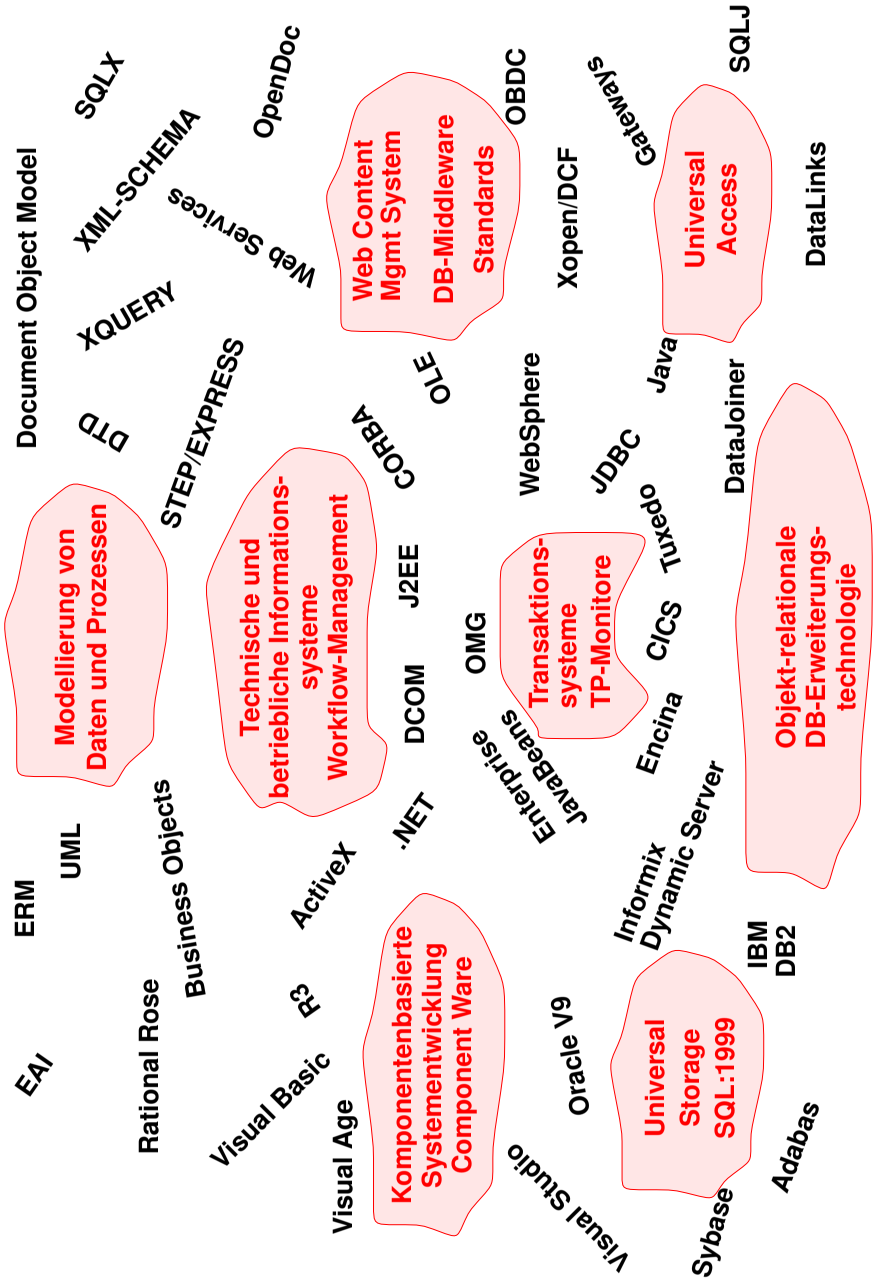
Ort: 46 - 220

Zeit: Di., 15.30 - 17.00 Uhr
und
Ort: 46 - 220

Zeit: Do., 10.00 - 11.30 Uhr

Übung:

n. V., s. Aushang



Betriebliche Informationssysteme[1]

Betriebliche Informationssysteme spiegeln die Geschäftsmodelle von Unternehmen wider und dienen dazu, deren Arbeitsabläufe zu organisieren und zu unterstützen. Darum sind sie wichtig.

Betriebliche Informationssysteme sind stark datenbankbasierte Anwendungen, oft mit sehr vielen Benutzern (Tausende und mehr). Es sind transaktionsverarbeitende Systeme, d.h., sie erbringen ihre Leistung in vielen, kleinen Schritten für die gleichzeitig zugreifenden Benutzer. Dabei müssen sie die Integrität der Daten gewährleisten sowie hohen Durchsatz und kurze Antwortzeiten schaffen. Betriebliche Informationssysteme laufen heutzutage typischerweise auf einem leistungsfähigen Server und präsentieren sich dem Benutzer mit einer grafischen Oberfläche als Client/Server-Systeme, zunehmend auch via Internet. Sie sind aber nicht nur Dialogsysteme, sondern benötigen meist auch einen Batch, der Massenverarbeitung effizient außerhalb des Dialogs abwickelt.

Ein Data Warehouse ist ein spezieller Aspekt betrieblicher Informationssysteme. Es akkumuliert und verdichtet die Daten aus den operativen Transaktionssystemen zu dem Zweck, die Geschäftsentwicklung zu analysieren und darauf aufbauend Entscheidungen zur Steuerung des Unternehmens zu treffen.

Die Software für betriebliche Informationssysteme ist meist sehr komplex und umfangreich (einige Hunderttausend bis Millionen Zeilen Programmcode), ihre Entwicklung verursacht erheblichen Aufwand und kostet viel Geld. Und sie lebt lange (zehn bis zwanzig Jahre und mehr), weshalb ihre Wartungskosten erheblich sind und die der Erstentwicklung oft übersteigen. Es lohnt sich folglich, Software von vornherein wartungsfreundlich und erweiterbar zu konstruieren.

Ohne betriebliche Informationssysteme könnten Unternehmen ihr Geschäft nicht betreiben. Das gilt in der industriellen Produktion ebenso wie im Dienstleistungssektor, ganz besonders jedoch für jene Unternehmen, die im Kern reine Informationsverarbeiter sind, z.B. Banken, Versicherungen, Touristikanbieter. Bei allen wirken die IT-Systeme tief auf die Unternehmensorganisation ein. Betriebliche Informationssysteme dienen dem Kerngeschäft von Unternehmen und damit Investitionsgüter von strategischer Bedeutung.

1.Diese Beschreibung der Aufgaben von betrieblichen Informationssystemen und von E-Business ist dem Buch “sd&m - Ein Software- und Beratungshaus stellt sich vor” (mit unwesentlichen Änderungen) entnommen (sd&m: software design & management, München, 2001).

E-Business

E-Business ist die Nutzung des Internets zu geschäftlichen Zwecken aller Art. Das fängt an mit der E-Mail und dem Herunterladen von Software per File-Transfer, geht weiter mit der Darstellung des Unternehmens, seiner Produkte und seiner Stellenangebote im World Wide Web sowie mit dem Intranet als elektronischem schwarzem Brett und Werkzeug des Wissensmanagements und schließt E-Commerce mit ein.

Bei E-Commerce fließt Geld, denn es geht um Handel, also den Abschluss und die Abwicklung von Kaufverträgen. Dabei werden Varianten unterschieden, je nachdem, wer mit wen handelt: ein Unternehmen mit seinen Endkunden (Business-to-Consumer, B2C), Unternehmen untereinander (Business-to-Business, B2B) oder Endkunden direkt miteinander über Börsen und Auktionen (Consumer-to-Consumer, C2C). Es lohnt sich zudem, nach dem „Aggregatzustand“ der gehandelten Produkte zu unterscheiden, d.h. zwischen dem Handel mit materiellen Produkten und jenem mit Informationen; denn die Geschäftsmodelle sind völlig verschieden. Bei materiellen Gütern kommt es entscheidend auf die Logistik an, also auf den Versand. Für unempfindliche Artikel (z.B. Bücher, Kleidung) genügt ein Paketdienst, für verderbliche und zeitkritische Ware (Blumen, die zur Geburtstagsfeier eintreffen sollen) muss ein spezieller Lieferservice eingerichtet werden. Für den reinen Informationshandel (z.B. mit Musik oder dem kostenpflichtigen Nachschlagen in einem E-Lexikon) ist das Internet der ideale Transportweg, nur zahlt dafür niemand gerne. Geschäftlich interessant dagegen ist Information, hinter der eine Dienstleistung steht (z.B. das Buchen eines Fluges, verbunden mit dem Kauf eines Tikkets, oder der Abschluß einer Versicherung).

E-Business braucht starke Softwaresysteme. Es sind komplexe Systeme, denn es genügt nicht, sich mit einer gut gestalteten Web-Oberfläche dem Benutzer zu präsentieren – werblich ansprechend, um ihn zu gewinnen, ergonomisch, um ihn nicht zu verlieren. Dahinter muss mehr stehen: eine flexible Anwendung, die sich schnell an geänderte Geschäftsprozesse anpassen lässt, und eine gehaltvolle Datenbank. Aber das ist noch nicht alles. Das Internet erfordert erhöhte Sicherheit und Bezahlung in neuen Formen. Tausende von Anwendern greifen gleichzeitig zu und erwarten rasche Reaktion. Weiterhin ist Anwendungsintegration von zentraler Bedeutung: Vorhandene (Legacy-) Systeme müssen über das Internet zugänglich gemacht und Anwendungen verschiedener Unternehmen darüber direkt verbunden werden. Kurzum, Software-Engineering für komplexe Systeme ist gefragt: @business @ngineering.

E-Business (2)

E-Business-Systeme sind nichts anderes als betriebliche Informationssysteme, allerdings mit besonderen Anforderungen: Erstens sind es eigentlich überbetriebliche Informationssysteme, denn sie verbinden über ein Unternehmen hinausgehend Mitarbeiter, Lieferanten und Kunden und werden vor allem von Menschen genutzt, die nicht Angestellte des betreibenden Unternehmens sind. Diese Anwender sind unbekannt, d.h., man kann sie beim Entwurf des Systems nicht befragen, man kann sie nicht schulen, ihnen keine Anweisungen erteilen, und man muss mit Ungeschicklichkeit oder gar böswilligen Attacken rechnen. Dennoch kommt es gerade auf sie an, denn es sind (potenzielle) Kunden (im B2C-Geschäft) oder Mitarbeiter von guten Geschäftspartnern (B2B). Die Ergonomie der Benutzeroberfläche ist also in bisher nicht bekannter Weise erfolgskritisch. Ebenso sind es die Organisation der Geschäftsprozesse, die nun über das eigene Unternehmen hinausreichen, sowie bestimmte Sicherheitsmaßnahmen (Firewall, Verschlüsselung, elektronisches Bezahlen).

Zweitens muss man beim Bauen und Integrieren von E-Business-Systemen eine noch nie da gewesene Komplexität der Technologie beherrschen. Man muss sich mit der Programmierung der Web-Oberfläche auskennen (HTML, XML, Java-Applets etc.), Netzprotokolle (z.B. HTTP) und Web-Server einzusetzen verstehen, Anwendungsprogramme in Java schreiben und unter der Transaktionskontrolle von Application-Servern zum Laufen bringen, Standard-Internet-Anwendungen (z.B. Intershop) sowie vorhandene (Legacy-) Systeme (CICS, Cobol, DB2) integrieren. Damit nicht genug: Systeme für E-Mail, Verschlüsselung, Zahlung und anderes müssen eingebunden werden. Schließlich ist das Ganze so zu integrieren, zu testen und zu tunen, dass es korrekt, zuverlässig und schnell läuft. Die zu verknüpfende Technologie reicht vom Mainframe über Client/Server-Systeme zum Internet.

Ziele

- **Vermittlung von Grundlagen- und Methodenwissen[1] sowie Erwerb von Fähigkeiten und Fertigkeiten im Bereich betrieblicher Informationssysteme:**
  - Nutzung von Informations- und Datenmodellen, insbesondere
    - Entity/Relationship-Modell und Erweiterungen
    - Relationenmodell und SQL
  - Modellierung von anwendungsbezogenen Realitätsausschnitten (Miniwelten, Diskursbereiche)
  - Entwerfen, Aufbauen und Warten von Datenbanken
  - Verstehen und Anwenden des Transaktionskonzeptes
  - Geschäftsprozessmodellierung und Abbildung auf Workflows

- **Voraussetzungen für Übernahme von Tätigkeiten:**
  - Entwicklung von betrieblichen Anwendungs- und Informationssystemen
  - Nutzung von Datenbanken unter Verwendung von (interaktiven) Datenbanksprachen
  - Planung und Realisierung von Geschäftsprozessen
  - Systemverantwortlicher für Datenbanksysteme, insbesondere Unternehmens-, Datenbank-, Anwendungs- und Datensicherungsadministrator

1. Grundlagenwissen ist hochgradig allgemeingültig und nicht von bestimmten Methoden abhängig. Die Halbwertzeit ist sehr hoch. Methodenwissen muß ständig an die aktuelle Entwicklung angepaßt werden. In der Informatik haben sich die entscheidenden Methoden alle 8-10 Jahre erheblich geändert. Werkzeugwissen ist methodenabhängig. Werkzeuge haben in der Informatik oft nur eine Lebensdauer von 2-3 Jahren.

ÜBERSICHT (vorl.)

0. Übersicht und Motivation

1. Einführung und Grundbegriffe

- Miniwelt – modellhafte Abbildung
- Information – was ist das?
- Aufgaben eines Informationssystems (IS)
- Daten in Informationssystemen
- Beispiele für Informationssysteme

2. E/A-Architektur und Zugriff

- E/A-Architektur von Informationssystemen
- Einsatz von Speicherhierarchien
- Datenstrukturen auf Externspeichern
- Informationssuche bei strukturierten und unstrukturierten Daten
- Probleme der Indexierung und Anfrageauswertung
- Suche und Dokumentenzugriff im Web

3. Informationsmodelle

- Vorgehensweise bei DB-Entwurf und -Modellierung
- Entity-Relationship-Modell (ERM)
- Erweiterungen
- Abstraktionskonzepte
  - Generalisierung
  - Aggregation
  - Assoziation

ÜBERSICHT (2)

4. Grundlagen des Relationenmodells

- Konzepte des Relationenmodells (RM)
- Relationenalgebra
  - Operationen
  - Algebraische Optimierung
- Abbildung ERM $\rightarrow$ RM

5. Die Standardsprache SQL (SQL2)

- Überblick
- Anfragesprache
- Datenmanipulation
- Datendefinition
- Abbildung von Beziehungen (Referentielle Integrität)
- Wartung von Beziehungen (Referentielle Aktionen)

6. Transaktionskonzept

- Überblick über ACID-Eigenschaften
- Gewährleisten von Atomärität
- Sicherung der Datenintegrität
- Korrektheit der Synchronisation im Mehrbenutzerbetrieb
- Aufgaben und Ziele der Fehlerbehandlung

7. Modellierung von Geschäftsprozessen

- Anforderungen
- Gschäftsprozeßmodellierung
- Modellierungsaspekte und -sprachen
- ARIS-Ansatz

ÜBERSICHT (3)

8. Petri-Netze und Workflows

- Elemente der Petri-Netz-Theorie
- Markierungssituationen und Ablaufstrukturen
- Systemklassen, Dynamische Eigenschaften
- Relationenmodell und Workflows

9. Workflow-Management

- Aspekte von Workflows
- Anforderungen an ein WfMS
- Funktionale Architektur eines WfMS
- Transaktionen und Workflows

10. Architekturmodelle

- Anwendungsentwicklung – allgemeine Probleme
- Sicht des Endbenutzers
- Entwurfsaufgaben
- Client/Server-Modell
- Zwei-und dreistufige C/S-Architektur
- TP-Monitor – Prinzipien und Einsatz
- Aufbau des DB-Servers (Drei-Schichten-Modell)

LITERATURLISTE

Baumgarten, B.: Petri-Netze – Grundlagen und Anwendungen, 2. Auflage, Spektrum Akademischer Verlag, 1996

Date, C.J.: An Introduction to Database Systems, Addison-Wesley Publ. Comp., Reading, Mass., 7th Edition, 2000

Kemper, A., Eickler, A.: Datenbanksysteme – Eine Einführung, 4. Auflage, Oldenbourg-Verlag, 2001

Korth, H.F., Silberschatz, A.: Database System Concepts, 3rd Edition, McGraw-Hill Book Comp., New York, 1997

Pernul, G., Unland, R.: Datenbanken im Unternehmen: Analyse, Modellbildung und Einsatz, Oldenbourg-Verlag, 2001

WEITERFÜHRENDE LITERATUR:

Appelrath, H.-J., Ritter, J.: R/3-Einführung – Methoden und Werkzeuge, Springer, 2000

*Orfali, R., Harkey, D., Edwards, J.: T*he Essential Client/Server Survival Guide, 3rd edition, Wiley Comp. Publ., New York, 1999

Scheer, A.-W.: ARIS – Vom Geschäftsprozeß zum Anwendungssystem, 3. Auflage, Springer,1998

Weikum, G., Vossen, G.: Transactional Information Systems, Morgan Kaufmann Publishers, San Francisco, CA, 2002

ZEITSCHRIFTEN:

TODS Transactions on Database Systems, ACM Publikation (vierteljährlich)

Information Systems Pergamon Press (6-mal jährlich)

The VLDB Journal (vierteljährlich)

Informatik - Forschung und Entwicklung (vierteljährlich)

TAGUNGSBÄNDE:

SIGMOD Tagungsband, jährliche Konferenz der ACM Special Interest Group on Management of Data

VLDB Tagungsband, jährliche Konferenz „Very Large Data Bases“

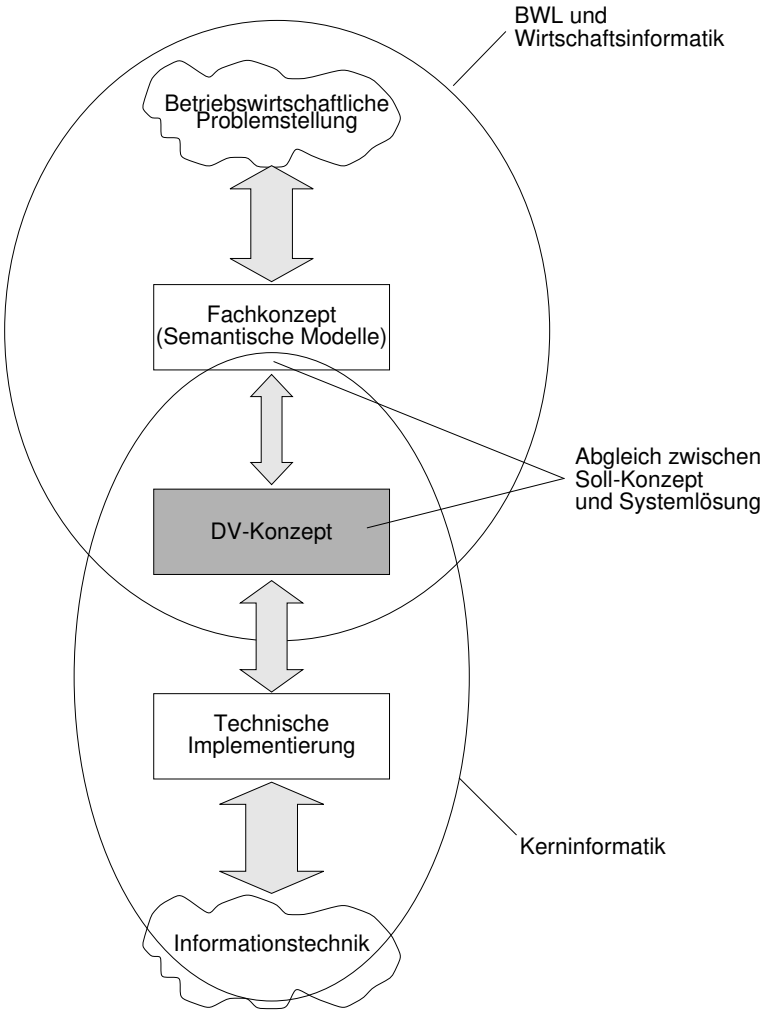
IEEE Tagungsband, jährliche Konferenz „Int. Conf. on Data Engineering“

GI Tagungsbände der Tagungen der Gesellschaft für Informatik, Tagungen innerhalb des Fachausschusses 2.5 Informationssysteme

und viele weitere Konferenzreihen

Entwicklung von Informationssystemen

- **Sicht auf den Themenbereich** (nach Scheer)
  - Was ist die Gesamtaufgabe?
  - BWL-Anforderungen an die Weiterentwicklung der Informationstechnik
  - IT eröffnet neue Möglichkeiten für BWL-Konzepte und -Anwendungen



Entwicklung von Informationssystemen (2)

- **Betriebliche Anwendungs- oder Informationssysteme**
  - Eigenentwicklungen von SW (zu teuer, nur noch < 20%)
  - Einsatz integrierter Standard-SW (hat sich durchgesetzt)
  - neuerdings: Ergänzung von speziellen Teilsystemen durch Componentware (es werden SW-Komponenten für bestimmte AW-Ausschnitte zu Anwendungssystemen „montiert“)

  ➥ Vielfalt (Wirrwarr) von Methoden mit starken Überlappungen

- **Wie sieht ein SW-Entwicklungsprozeß aus, der**
  - alle drei Formen unterstützt
  - verschiedene Sichtweisen zuläßt: z. B. daten-, ereignis-, funktionsorientiert
  - eine Übersichtlichkeit und Trennschärfe bei den Methoden bietet?

  ➥ Gesucht ist also eine Methodologie (Lehre von den Methoden) für die Entwicklungsmethoden

- **ARIS – Architektur integrierter Informationssysteme**
  - erlaubt die einzelnen Bausteine eines IS hinsichtlich ihrer Art, funktionalen Eigenschaften und ihres Zusammenwirkens zu beschreiben
  - will eine Methodologie als Rahmenkonzept (Framework) bieten, die typische Fragen beantworten helfen soll:
    - Gibt es einen besten Weg bei der Entwicklung von IS?
    - Wo beginnt der Entwicklungsprozeß und wo endet er?
    - Wieviel Stufen der Entwicklung sind erforderlich?

  ➥ Zielsetzung solcher Fragen ist Einordnung und Bewertung von Methoden

Entwicklung von Informationssystemen (3)

- **ARIS verfolgt einen ganzheitlichen Beschreibungsansatz für Informationssysteme**
  - **Datenmodellierung:** war Ausgangspunkt und ist aus Sicht von DBIS immer noch dominierender Aspekt
  - **Weitere Bereiche:** Funktions-, Steuerungs- und Organisationssicht
  - **Beschreibungsebenen** wollen alle Phasen des „Life Cycle"-Prozesses eines Informationssystems umfassen
    - Analyse (weitgehend informell)
    - Fachkonzept (formale Beschreibungssprachen)
    - DV-Konzept (logische Konkretisierung)
    - Implementierung (abstrakte Realisierung)
    - Konkrete Realisierung (Systementwicklung, -evolution, -wartung)

- **Vorgehensweise von ARIS**
  - Unterstützung der Beschreibung, Realisierung und Integration von Geschäftsprozessen
    (engl. *business process*, synonym: Unternehmensprozeß)
  - Entwicklung eines Modells für Geschäftsprozesse
  - Zerlegung dieses Modells in einzelne Sichten
    (wegen seiner hohen Komplexität)
  - Konzept der unterschiedlichen Beschreibungsebenen
    - von der betriebswirtschaftlichen Problemstellung
      bis zur rechnergestützten Realisierung
    - umfaßt Aufgaben der BWL und Wirtschaftsinformatik
      bis zur Kerninformatik

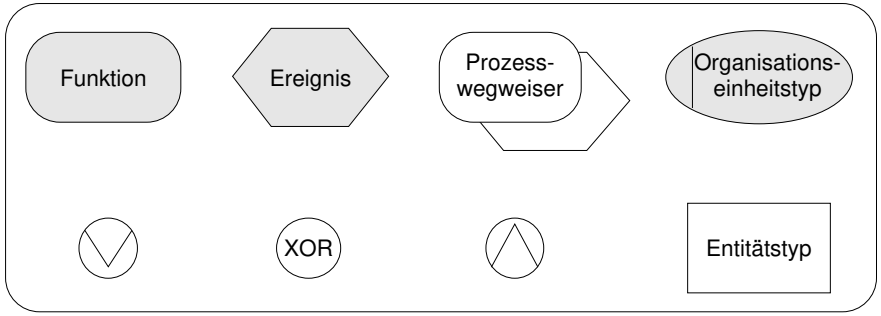
  ➥ ARIS will Informationssysteme zur Unterstützung von Geschäftsprozessen **ganzheitlich** beschreiben

ARIS als Rahmenkonzept

- **Geschäftsprozeß**

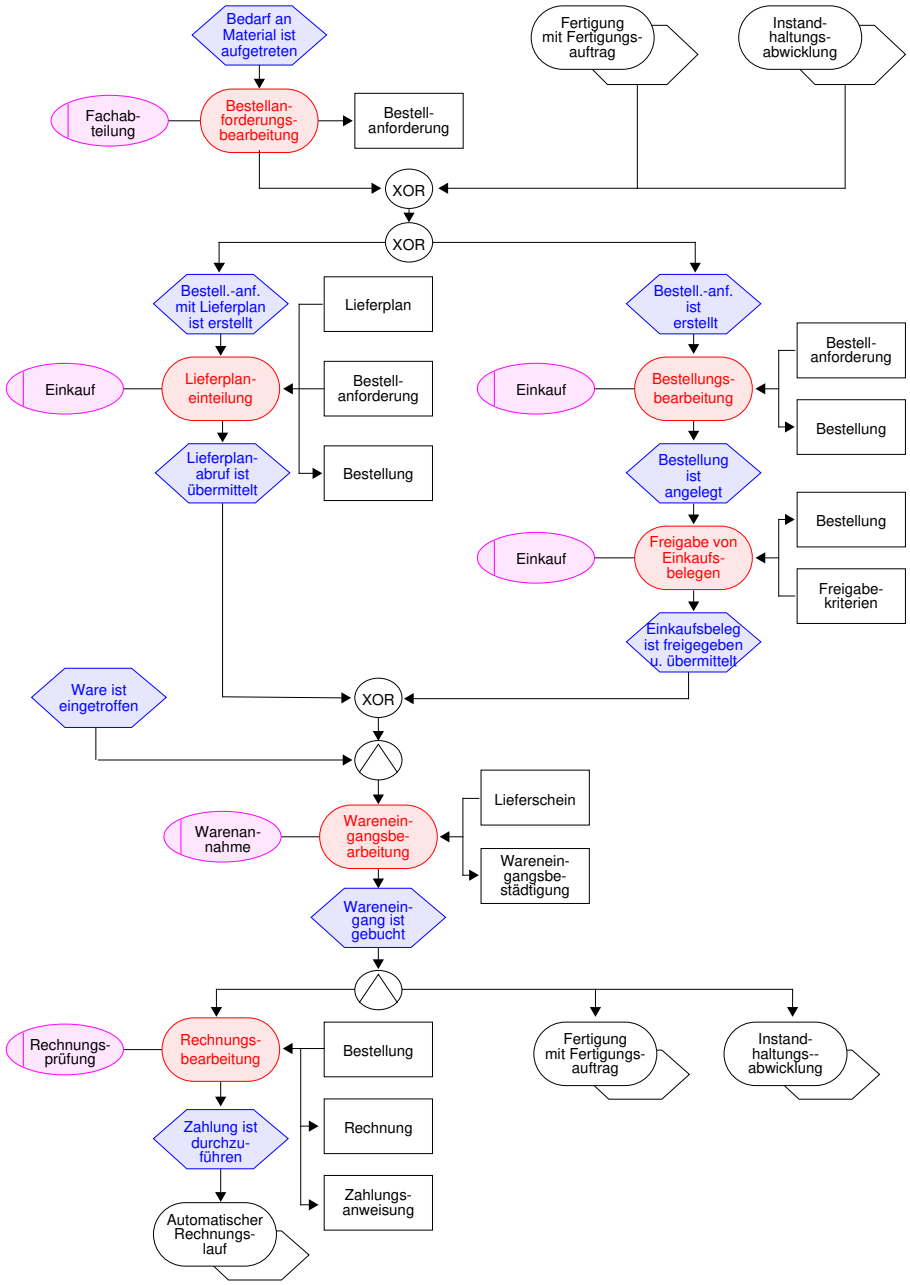
  ist eine Abfolge von Aktionen bzw. Interaktionen, die von Objekten bzw. zwischen mehreren Objekten durchgeführt werden und einem betrieblichen Ziel dienen
  - Objekte (Akteure), die Aktionen durchführen: Mitarbeiter, Kunden, Lieferanten, aber auch technische Systeme (Anwendungssysteme)
  - Durchführung von Aktionen: Erzeugung, Transformation und Verbrauch von Objekten, die Materialien, Produkte, Informationen und allgemeine Dienstleistungen sein können

- **EPK-Modell zur Beschreibung von Geschäftsprozessen**
  - Ereignisgesteuerte Prozeßketten (EPK)
  - Beschreibungselemente



- **EPK-Beispiel**
  beschreibt typische Geschäftsabläufe des Beschaffungswesens

EPK-Beispiel



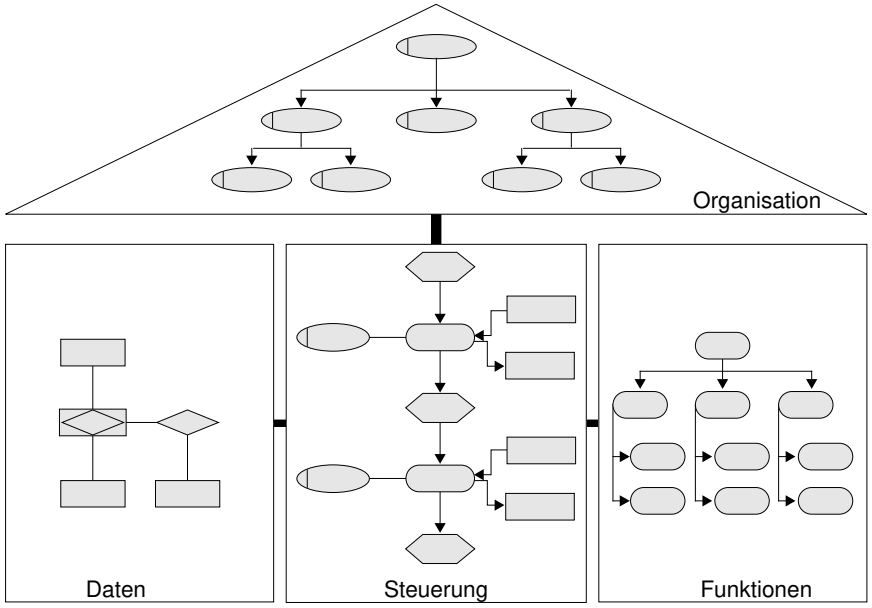
EPK-Beispiel: Erklärungen

- **Funktionen**
  - bezeichnen fachliche Aufgaben oder Tätigkeiten, die in (Inter-) Aktionen von einem oder mehreren Objekten durchgeführt werden
  - erzeugen Ereignisse (Zustände oder Situationen), auf die wiederum in anderen Funktionen reagiert werden kann
- **Ereignisse**
  - spezifizieren betriebswirtschaftlich relevante Zustände von Informationsobjekten, die vor der Durchführung von Funktionen erfüllt sein müssen
  - ein oder mehrere Ereignisse (Startereignisse) lösen eine Funktion aus, die ein oder mehrere Ereignisse als Ergebnis haben kann

  ➥ Alternierende Folgen von Ereignissen und Funktionen, die durch gerichtete Kanten (Kontrollfluß) zu einem Prozeß verbunden werden
- **Verknüpfungsoperatoren (Konnektoren)**
  - Ereignisverknüpfungen erlauben den Kontrollfluß aufzuspalten oder zusammenzuführen
  - Funktionsverknüpfungen modellieren das Ergebnis von mehreren Funktionen als ein Ergebnis. Außerdem kann ein Ergebnis mit mehreren nachfolgenden Funktionen konjunktiv verknüpft werden
- **Verschiedene Abstraktionsebenen**
  - eine Funktion kann EPKs verbergen
  - Prozeßwegweiser beschreiben Abhängigkeiten zwischen EPKs
- **Organisationseinheiten**
  - Spezielle Knotentypen erlauben die Modellierung der an einer Interaktion beteiligten Objekte (Abteilung, Gruppe, Stelle)
  - Rollenkonzept erlaubt Abstraktion
- **Entitätstypen**
  - beschreiben die verbrauchten, erzeugten und transformierten Objekte
  - werden in der Datensicht verwaltet

ARIS als Rahmenkonzept (2)

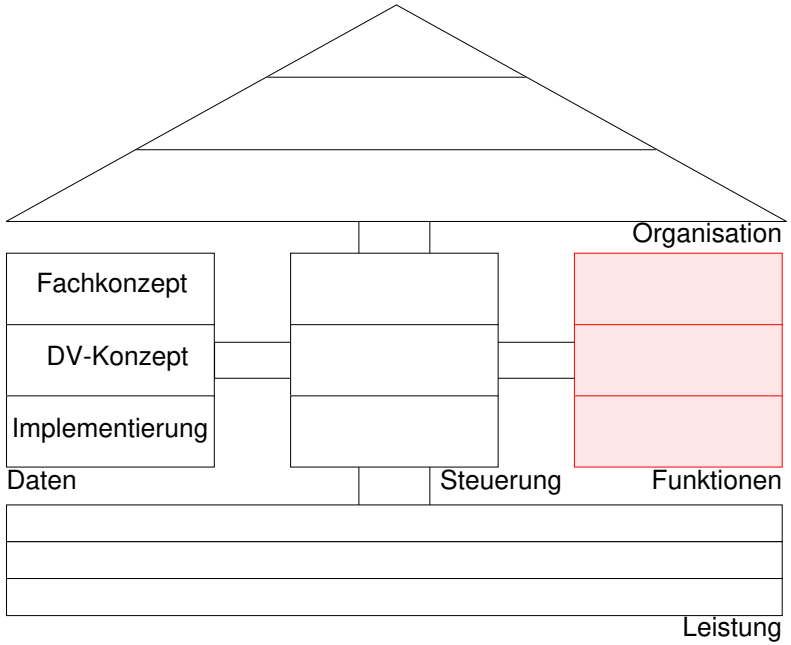
- **Idee von ARIS: Zerlegung in Sichten**
  - **Organisationssicht:** beschreibt die Mitarbeiter (menschliche Akteure) und ihre Organisationsstrukturen
  - **Datensicht:** beschreibt die Daten und ihre Abhängigkeiten untereinander
  - **Funktionssicht (oder Aufgabensicht):** stellt die Aufgaben in ihrer hierarchischen Abhängigkeitsstruktur (Aufgabe kann aus mehreren Teilaufgaben bestehen) dar
  - **Steuerungssicht:** integriert die in den anderen Sichten definierten Objekte in den Geschäftsprozeß
- **Veranschaulichung der Sichtenzerlegung von ARIS**



- **Erweiterung um Leistungssicht**
  - In neueren ARIS-Versionen werden noch die betrieblichen Leistungen (Produkte, Sachleistungen, Dienstleistungen) beschrieben

ARIS als Rahmenkonzept (3)

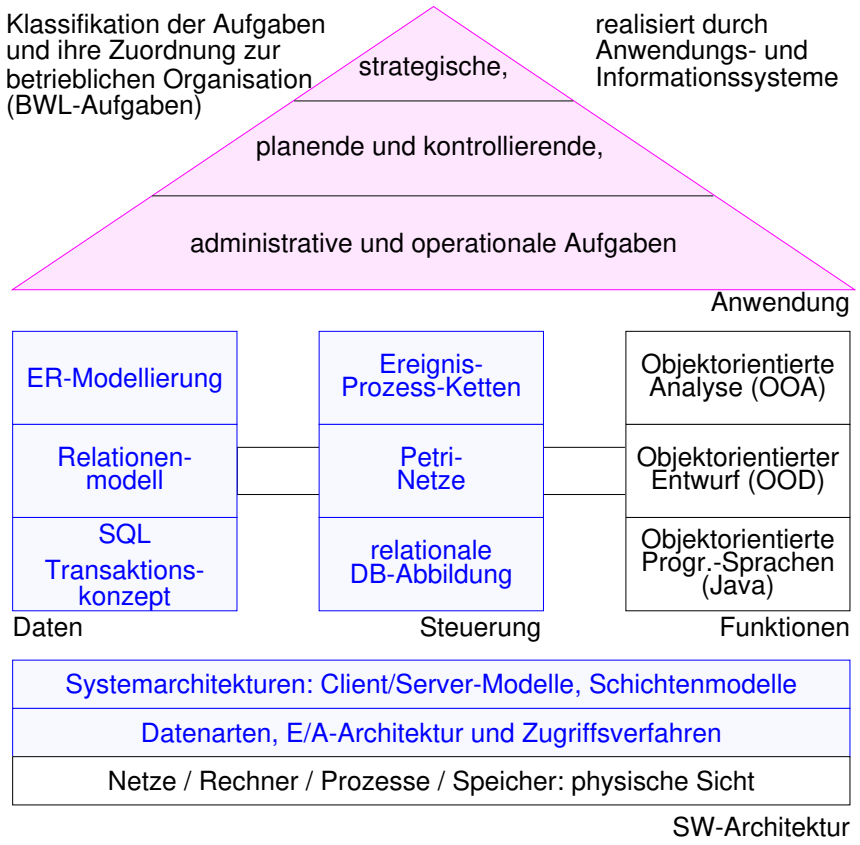
- **ARIS: Abstakte Beschreibung der Vorgehensweise**
  - Zerlegung in verschiedene Sichten
  - Verfeinerung durch unterschiedliche Beschreibungsebenen



- **Einteilung in Sichten ist abhängig von der Zielsetzung**
  - ARIS beschreibt mehr die BWL-Sichtweise, während wir mehr die Informatik-Sichtweise der Entwicklung von Informationssystemen betonen wollen
  - Einheiten der Organisations- und Leistungssicht werden deshalb als Ausschnitte der Datensicht verstanden
  - Die Leistungssicht wird ersetzt durch eine Sicht der SW-Architekturen zur Bereitstellung von Rechnerplattformen, Prozessen und Infrastrukturen zur Realisierung von Informationssystemen
  - Anstelle der Organisationssicht führen wir (um das Haus mit Dach beizubehalten) eine Anwendungssicht ein, welche die Zuordnung der Aufgaben in der Unternehmenshierarchie charakterisiert

Rahmenkonzept für die Vorlesung

- **Zuordnung der Themen der Vorlesung**

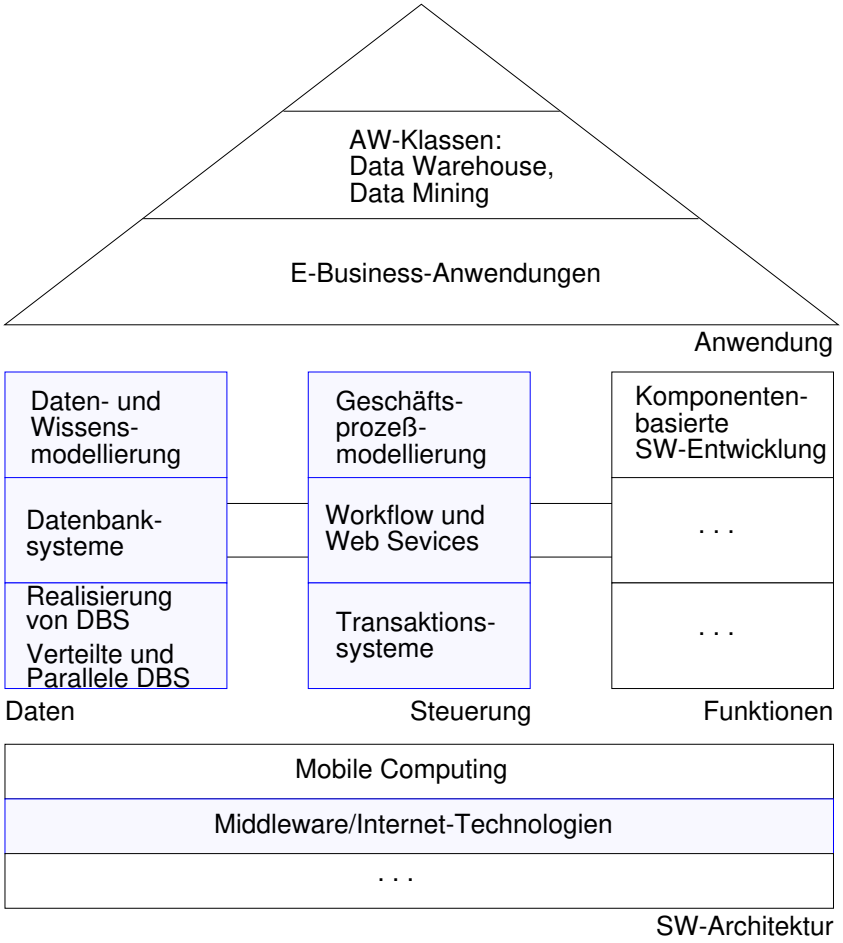


- **GBIS-Rahmen ist Wegweiser** durch die Vorlesung
  - Zerlegung in Sichten
  - Verfeinerung durch hierarchische Beschreibungsebenen
  - Ergänzung durch Anwendungssicht als „Dach“

➥ **Zuordnung** enthält unsere Interpretation der Sichten und ihrer Aufgaben, Zusammenspiel von 5 Aspekten: W^5-Modell

Rahmenkonzept für die Vorlesung (2)

- **Vertiefende Vorlesungen (Beispiele)**
  - Unterscheidung nach Beschreibungsebenen ist hier nicht sinnvoll
  - Zuordnung zu den einzelnen Sichten steht im Vordergrund



Zum Nachdenken – Wie viele Informationen gibt es auf der Welt?[1]

- **Was ist überhaupt Information?**
  - z. T. schwierige Abgrenzung, individuelle Bewertung
  - Was wird aufgezeichnet, was lohnt sich aufzuheben?
  - Redundante Information (Exemplare eines Buchs, Kopien eines Films) wird hier ausgeschlossen
  - Nicht alle Daten, die weltweit durch Sensoren, Meß- und Aufnahmegeräten usw. (Experimente, Überwachung, ...) erzeugt werden, führen zu relevanten Informationen (in unserem Sinne)

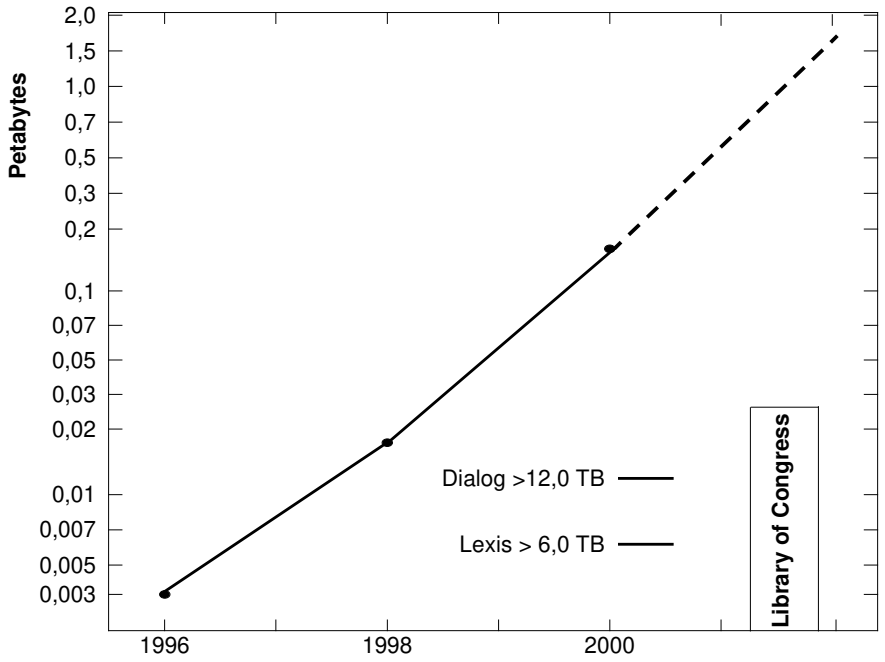
- **Annahme**
  - Um alle relevanten Informationen aufzuheben, genügt die Speicherung von wenigen Tausend PBytes[2]
  - Die Produktion von Magnetplatten und -bändern erreicht diese Speicherkapazität um das Jahr 2000

- **Behauptung:**

  In wenigen Jahren
  - sind wir in der Lage, „alles“ aufzuheben, d. h., keine Information muß weggeworfen werden
  - werden die „typischen“ Informationen nur noch von Rechnern aufbewahrt, gesucht und aufbereitet; der Mensch sieht weder die Daten, noch kennt er den Aufbewahrungsort und die genauen Ableitungsverfahren.

1. http://www.lesk.com/mlesk/ksg97/ksg.html
2. 1 Gigabyte (GByte) = 1.000 Megabytes = 10^9 Bytes
   1 Terabyte (TByte) = 1.000 Gigabytes
   1 Petabyte (PByte) = 1.000 Terabytes
   1 Exabyte (EByte) = 1.000 Petabytes
   1 Zettabyte (ZByte) = 1.000 Exabytes
   1 Yotabyte (YByte) = 1.000 Zettabytes

Web-Größe



nur Text-Dateien (Ascii-Daten)

Vergleichsmaß ist Library of Congress (LC): 20 Mio Bücher * 1 MByte ~ 20 TBytes
Dialog: http://www.dialog.com/
Lexis: http://www.lexis-nexis.com/

- Google indexiert 2 Mrd. Web-Dokumente (in 2002)
- Deep Web (Schätzung): 7.5 PBytes

- **Wachstum des Web**
  - Schätzung 1998 Faktor 10 in 2 Jahren
    Wie lange kann es so weiter gehen?
  - Anzahl der Benutzer: Faktor 10?
    2000: ~ 200 Mio. → 2 Mrd.?

➥ 20 PBytes ASCII-Daten im Web?

Andere Informationsarten

- **LC hat neben gedrucktem Text weitere Informationsquellen, die weit mehr Speicherplatz benötigen:**
  - 13 Mio. Photos ergeben, selbst komprimiert auf 1 MBytes JPG, noch 13 TBytes
  - 4 Mio. Landkarten lassen sich auf ~ 200 TBytes einscannen
  - 0,5 Mio. Filme mit jeweils 1 GByte (komprimiert, meist nicht in Farbe und voller Länge) benötigen 500 TBytes
  - 3,5 Mio. Tonaufzeichnungen, jeweils auf einer Audio-CD, erreichen fast 2,000 TBytes.

  → Gesamter Umfang der LC-Informationen: ~ 3 PBytes

- **Vergleich Web – LC**
  - LC enthält vorwiegend publiziertes Material
  - Im Web stehen bereits mehr englischsprachige Texte, die in den letzten 18 Monaten geschrieben wurden, als in der LC
  - Viele URLs enthalten Informationen, die man auch in einer großen Bibliothek finden kann (~ 28 %)
  - Inhalt dieser Dokumente (Lexis-Nexis und Dialog) ist jedoch besser aufbereitet (Hypertext) und einfach zu durchsuchen

- **Erinnerung**
  - Vor Gutenberg (~ 1450) gab es in ganz Europa nur ~ 30,000 Bücher, fast ausschließlich Bibeln und Bibelkommentare[1]
  - Um das Jahr 1500 gab es über 9 Mio Bücher über alle möglichen Themen. Flugschriften und sonstige Druckwerke beeinflußten Politik, Religion, Wissenschaft und Literatur

  → Der *Information Highway* wird unsere Kultur ebenso dramatisch verändern wie Gutenbergs Druckerpresse die Welt des Mittelalters

1. James Burke: „In dieser Welt war jegliche Erfahrung persönlich: Die Horizonte waren eng, die Gemeinschaft nach innen gekehrt. Die Außenwelt kannte man nur vom Hörensagen."

Menschliches Gedächtnis

- **Wieviel Information kann das menschliche Gehirn aufnehmen?**
  - 200 MBytes an Information

    T.K. Laudauer: „How much do people remember? Some estimates of the quantity of learned information in long-term memory", Cognitive Science 10 (4) pp. 477-493 (Oct-Dec 1986).
    - Studie zur Aufnahme- und Vergessensrate sowie zur Nutzung bei normalen Tätigkeiten
    - Bei 10^{12}-10^{14} Neuronen kann man davon ausgehen, daß das Gehirn 10^{3}-10^{5} Neuronen für jedes Bit des Gedächtnisses enthält
    - Der größere Teil des Gehirns wird für die Wahrnehmung, Bewegungskontrolle usw. genutzt

  → Even if only 1 % of the brain is devoted to memory, it looks like your head accepts considerable storage inefficiency in order to be able to make effective use of the information (Landauer).

- **Mit der Landauer'schen „Schätzung" und den ~ 6 Mrd. Menschen auf Erden erhält man als gesamtes Speichervermögen aller menschlichen Gedächtnisse 1.200 PBytes.**

  → Rechner können heute alles digital speichern, was alle Menschen wissen oder an was sie sich erinnern können

- **Diese Schätzungen sind ungenau. Jedoch wäre es auch bei einem anderen „Aufnahmemodell" kein Problem, die Informationsmenge eines Menschen auf eine Laptop-Platte zu speichern.**
  - Menschliche Aufnahmerate beim Langzeitgedächtnis: 1 Byte/sec nach Landauer
  - Gesamtmenge in einem menschlichen Leben: 2 GBytes (2 Mrd. Sek.)

Menschliches Gedächtnis (2)

- **Selbst wenn der Mensch sein Leben lang kein Wort vergessen würde, ist die Speicherung der Informationsmenge heute schon machbar (auf einer Platte)**

  Bilanz (durchschnittl. Zahlen für US-Bürger/Jahr)

	Stunden	Wort/Minute	Wörter/Jahr	MBytes
TV	1578	120	11 Mio.	50
Film	12	120	-	-
Lesen	354	300	6,4 Mio.	32

  → ~ 6 GBytes von ASCII-Daten (Text) in 75 Jahren

  Die automatische Erfassung dieser Informationen (Spracherkennung, und OCR oder elektronische Bücher/Zeitungen und ASCII-Skripte von TV-Sendungen) ist in naher Zukunft mit tragbaren Geräten möglich. Was bedeutet das beispielsweise für Schüler/Studenten?
  Wie weit gehen diese Möglichkeiten? (Hörsaal mit Laptop-/Notebooks)

  → Das gilt nicht für Ton, Bild und Bewegtbild

- **Aber in 20 Jahren?**

  Speicherbedarf für eine Person (nach Jim Gray)

Datentyp	**Datenrate (Bytes/ Sek.)**	**benötigter Speicher pro Stunde und Tag**	**benötiger Speicher für eine Lebenszeit**
gelesener Text	50	200 KB; 2-10 MB	60-300 GBytes
gesproch.Text @120 wpm	12	43 K; 0,5 MB	15 GBytes
Sprache (komprimiert)	1.000	3,6 MB; 40 MB	1,2 TBytes
Bewegtbild (komprimiert)	500.000	2 GB; 20 GB	1 PByte

  → Aufzeichnung der gesamten Lebensgeschichte wird möglich!

Schlußfolgerungen

- **Es wird genug Platten- und Bandspeicher geben, um alles zu speichern, was alle Menschen schreiben, sagen, tun oder photographieren.**
  - Für das Schreiben gilt dies bereits heute
  - In einigen Jahren trifft das auch für die restlichen Informationen zu
  - Wie lange wird es noch dauern, bis alle VITA-Dokumente eines Menschen als Lebensgeschichte aufgezeichnet werden können?

- **Rechner speichern und verwalten Informationen besser und effektiver als Menschen**
  - Nach dem Jahr 2000 werden viele Platten und Kommunikationsverbindungen direkt Informationen aus Rechner-zu-Rechner- und nicht mehr (nur) aus Mensch-zu-Mensch-Kommunikation speichern
  - Wie lange wird es noch dauern, bis der Mensch die meiste gespeicherte Information gar nicht mehr zu sehen bekommt?
  - Wir müssen lernen, wie alles automatisch ausgewertet werden kann und was bei unserer knappen Zeit unserer besonderen Aufmerksamkeit bedarf.

- **Künftige Entwicklung**
  - Heute konzentriert man sich bei den „Digitalen Bibliotheken“ auf die Eingabe: auf das Scanning, Komprimieren und OCR von Informationen.
  - Morgen wird anstelle der Eingabe die „relevante Auswahl“ die wesentliche Rolle spielen: Selektion, Suche und Qualitätsbewertung von Informationen

  ➥ Wir können eine reale „World Encyclopedia“ mit einem echten „planetary memory for all mankind“ aufbauen, wie H.G. Wells bereits 1938 in seinem Buch „World Brain“ geschrieben hat!