

Motivation und Grundbegriffe

Informationssystem

- dient der Erfassung, Aufbewahrung, Verarbeitung und Bereitstellung von Daten/Informationen
- Fokus für uns: rechnergestütztes IS, umfasst Software und Daten
- zielt auf die Abbildung von Eigenschaften bestimmter Domänen (Realitätsausschnitte)

Repräsentationsebenen

- Daten: symbolische Repräsentation, besitzt nur formale Semantik (z.B. math. Operationen)
- Informationen: Daten, die der Beschreibung realer oder gedachter Sachverhalte dienen; Bedeutung ergibt sich durch eine menschliche Interpretation in Bezug auf einen Kontext
- Wissen: Information mit höherer Abstraktion und Begründung

Grammatikalische Formen

- Syntax: Regeln, wie die Symbole einer Sprache zu Worten bzw. die Worte einer Sprache zu Sätzen (unabhängig von deren Bedeutung) angeordnet werden dürfen
- Semantik: Interpretation eines Satzes, muss nicht eindeutig sein
- formale Semantik: anwendbare Operationen

Systemintegration – vorläufige Begriffsabgrenzung

- zwei Informationssysteme werden dadurch integriert, dass sie auf gemeinsame Ressourcen zugreifen können (Daten, Funktionen, Ereignisse)
- Zugriff auf Ressourcen der zu integrierenden Systeme (Ressourcenverwaltung in der Verantwortung der einzelnen Systeme) oder auf gemeinsam genutzte Ressourcenverwaltungssysteme

Vorteile von Integration

- Vermeidung von Redundanzen
- höhere Integrität
- effizientere Prozesse
- bessere Ergonomie
- reduzierte Komplexität
 - dafür müssen stabile, echte Gemeinsamkeiten gefunden oder ggf. konstruiert werden (Qualität der Abstraktion zentral)

Integrationsbegriff: alltagsweltliche Bedeutung

- allgemein: Herstellung einer Einheit, Eingliederung in ein größeres Ganzes
- steht sowohl für einen Prozess wie für dessen Ergebnis

Aspekte der Integration im Kontext der Wirtschaftsinformatik

- Komponenten systemtechnischer Integration (Hardware, Daten, Funktionen etc.)
- Systemtechnische Integration aus Benutzersicht (Verfügbarkeit, Transparenz, Konsistenz etc.)
- Organisatorische Integration (Regeln, Normen, Werte; anschließend Abbildung auf das IS – soweit dies möglich ist)
- Integration vertikaler und horizontaler Sichten auf das Unternehmen
- Überbetriebliche Integration von Informationssystemen (Dokumentenaustausch, Prozesse etc.)
- Integration der Phasen des Life-Cycle

Begriffsbestimmung

- Begriffe im Umfeld
 - Interoperabilität (Austausch von Nachrichten, eher technisch)
 - Koordination (eher organisatorisch)
 - Kompatibilität (bspw. Kommunikationsfähigkeit)
 - Interdependenz (durch Integration entstehen Abhängigkeiten)
 - Überwindung von Heterogenität
- **Statische Integration**
 - zu integrierende Elemente: Komponenten, die Daten/Objekte aufnehmen und ausgeben können
 - Integration durch gemeinsames **semantisches Referenzsystem** (ein allen zu integrierenden Komponenten verfügbares System, in dem die zum Zweck der Kommunikation benötigten Begriffe/Konzepte für alle verbindlich festgelegt sind; Bsp: elementare Datentypen, Relationstypen, Klassen, spezielle Datenstrukturen)
- **Funktionale Integration**
 - Fähigkeit einzelner Komponenten, Funktionen/Prozeduren, die andere Komponenten anbieten, in sinnvoller Weise zu nutzen (Voraussetzungen: statische Integration und Interoperabilität, Erweiterung des sem. Referenzsystems um funktionale Aspekte)
- Objektorientierte Integration
 - statische + funktionale Integration, dazu Generalisierungs-/Spezialisierungsbeziehungen
- **Dynamische Integration** (oder zielgerichtete Kooperation)
 - Einbettung von Komponenten in Prozesse (Voraussetzungen sind funktionale Integration und Erweiterung des sem. Referenzsystems um dynamische Aspekte)

Integrationsniveau

- Integrationsniveau eines IS nicht exakt zu messen
- nimmt tendenziell mit der Semantik der Konzepte im sem. Referenzsystem zu
- die (extensionale) Semantik einer Darstellung hat umso mehr Gehalt, je geringer die Zahl möglicher Interpretationen ist
- eine gute Kommunikation ist durch Differenziertheit gekennzeichnet

Bewertung von Integration

- grundsätzlich positiv belegt: integrierte IS im Zweifel besser als nicht integrierte
- aber: in der Praxis sind warnende Stimmen zu verzeichnen
- eine Reduktion der Systemkomplexität impliziert meist zunächst einmal eine Erhöhung der Komplexität durch geeignete Abstraktionen → erzeugt Risiko
- es ist darauf zu achten, dass Komplexität und Risiko überschaubar bleiben

Etablierung semantischer Referenzsysteme

- Bilaterale Konventionen
 - Multilaterale Konventionen
 - „Industriestandards“
 - Standards. bzw. Normen
- Wahl hängt vom Anwendungsfall ab (Anzahl kommunizierender Systeme, mögliche zukünftige Entwicklung)
- XML kann ökonomisch sinnvoll sein, weil entsprechende Werkzeuge kostenlos sind

Die andere Seite der Medaille: Wiederverwendung

- Verwendung von Konzepten bzw. Artefakten über eine Anwendung hinaus (Software-Entwürfe, Software-Systeme, Entwurfswissen)
- zentrales Ziel der Wirtschaftsinformatik – attraktive Vision (bessere Qualität, geringer Kosten)

Voraussetzung für Wiederverwendung

- zentrale Voraussetzung: Gemeinsamkeiten mehrerer Systeme
 - übereinstimmende Anforderungen
 - gleiche/ähnliche Funktionen und Datenstrukturen
 - überzeugende Entkräftung von Einwänden
 - Erstellung wiederverwendbarer Software i.d.R. teurer
 - Wiederverwendung mitunter von Entwicklern abgelehnt
- Integration andere Perspektive, Systeme sind bereits vorhanden, erst dann Suche nach Gemeinsamkeiten

Wiederverwendbarkeit und Semantik

- Wiederverwendungskomfort: nimmt tendenziell mit der Semantik der angebotenen Artefakte zu (Voraussetzung: Artefakte entsprechen den Anforderungen des Verwenders)
 - Wiederverwendungsreichweite: nimmt tendenziell mit der Semantik der angebotenen Artefakte ab
- es muss daher ein geeignetes Abstraktionsniveau gefunden oder geschaffen werden (von den speziellen Ausprägungen muss für eine hohe Reichweite abstrahiert werden, gleichzeitig müssen Konzepte angeboten werden, die die jeweiligen Anpassungen komfortabel und sicher macht – analog zur Spezialisierung/Generalisierung bei der OOP)

Betriebliches Informationssystem

- Informationssystem, das auf die Unterstützung zweckgerichteten Handelns in einem Unternehmen gerichtet ist
- technische Sicht: Daten, die computergestützt verwaltet werden
- Nutzersicht: Verwaltung und Bereitstellung von Informationen
- aber: in anderer Form abgelegte Informationen damit nicht ausgeschlossen

Organisatorische Integration

- die effiziente Nutzung eines BIS erfordert seine wirksame Einbindung in die Unternehmensorganisation
- besitzen jeweils eigene Sprachen, gemeinsames Referenzmodell wäre wünschenswert

Fokus auf Gemeinsamkeiten

- die organisatorische Integration eines BIS erfordert Gemeinsamkeiten
 - Funktionen des Systems sollten mit Aufgaben in der Organisation korrespondieren
 - Gegenstände der Realwelt sollten mit Daten im System korrespondieren
- Integration wird gefördert durch gleiche/ähnliche Begrifflichkeit
 - erleichtert die Abbildung korrespondierender Sachverhalte
 - fördert das Verständnis des IS

Gegenseitige Anpassung

- Informationstechnologie schafft neue Möglichkeiten zur Automatisierung bzw. Unterstützung von Aufgaben in Organisationen

- Ausschöpfung der Potentiale erfordert allerdings häufig eine Reorganisation
 - führt dazu, dass Begriffe, die durch IS geprägt sind, Eingang in die Sprache der Organisationsmitglieder findet
- die gegenseitige Anpassung von IS und Organisationen gehört zu den Kernaufgaben der WInf

Anwendungssysteme

Betriebswirtschaftliche Anwendungssysteme

- Informationssysteme für bestimmte Anwendungsbereiche
- Vergangenheit: häufig Eigen- bzw. Auftragsentwicklung für individuell festgelegte Aufgabenbereiche im Unternehmen
- heute: Einsatz von Software-Produkten spezialisierter Anbieter
 - zunehmende Komplexität von Software
 - Wiederverwendungsvorteile, Skaleneffekte

Systeme für betriebliche Funktionen

- Fakturierung
 - Funktionen: Erstellung von Lieferscheinen, (Sammel-)Rechnungen, Mahnungen; Kontrolle des Zahlungseingangs; Kundenstammdatenverwaltung; Erstellung von Verkaufsstatistiken
- Lagerverwaltung
 - Funktionen: Abbildung der Lagerbestände und von Umlagerungen; Lagerplatzverwaltung; Fortschreibung/Abschreibung von Beständen; Unterstützung der Inventur
- Finanzbuchhaltung
 - Funktionen: Erfassung und Dokumentation aller Geschäftsvorfälle; Erstellung des Jahresabschlusses; Erstellung und Pflege von Kreditoren- und Debitorenkonten
- Vertrieb
 - Funktionen: Erstellung und Verwaltung von Angeboten; Verwaltung spezieller Kundendaten und von Vertriebsmitarbeitern; Gebietsverwaltung; Abbildung von Vertriebswegen
- Marketing
 - Funktionen: Unterstützung des Vertriebs; stärkere Berücksichtigung von Kundenprofilen und Eigenschaften von Märkten; Unterstützung der Marketing-Forschung
- Personal
 - Funktionen: Lohn- und Gehaltsabrechnung; Provisions- und Reisekostenabrechnung; Ermittlung der Personalkosten; Mietabrechnungen; Belegschaftsverkauf

Voraussetzungen der Datenintegration

- Syntax gemeinsam verwendeter Daten muss den zugreifenden Anwendungen bekannt sein
 - am besten auf der Ebene der verwendeten Implementierungssprache
 - aber: problematisch bei unterschiedlichen Implementierungssprachen
- Semantik sollte ebenfalls bekannt sein
 - am besten auf Anwendungsebene
 - Voraussetzung für maschinelle Weiterverarbeitung
 - aber: problematisch bei unterschiedlichen Implementierungssprachen
 - kann ggfs. durch eine Interpretation des Anwenders ersetzt werden

Warenwirtschaftssysteme (WWS)

- Fokus auf Datenintegration
 - Reduktion von Datenredundanz, Beschleunigung von Prozessen, höhere Aktualität und Integrität der Daten, Verringerung von Prozesskosten
- Entwicklungsmöglichkeiten: ein Standort → dezentrale Standorte → unternehmensübergreifend
- Bewertung
 - Begriff in der Praxis wesentlich von konkreten Systemen geprägt
 - Funktionsumfang wie auch Art und Umfang der Datenintegrität nicht einheitlich
 - i.d.R. von einem Anbieter erstellt, deshalb günstige Voraussetzung für Integration
 - möglicher Nachteil: Schaffung einer Insellösung ohne Integration mit anderen Teilen eines BIS

Produktionsplanungs- und Steuerungssysteme (PPS)

- schließen Funktionen der Materialwirtschaft ein
- Produktionsplanung dient der Vorbereitung, Produktionssteuerung der Steuerung und Kontrolle von Produktionsprozessen
- Bewertung
 - Begriff in der Praxis wesentlich von konkreten Systemen geprägt
 - Funktionsumfang wie auch Art und Umfang der Datenintegrität nicht einheitlich
 - physische Steuerung des Produktionsprozesses oft nicht enthalten
 - Integration verschiedener Systeme zur Unterstützung der Planung und Steuerung von Produktionsprozessen sehr sinnvoll (Beschleunigung des Gesamtprozesses, Reduktion von Fehlerwahrscheinlichkeiten)
 - ähnlich wie WWS: Systeme häufig von einem Anbieter, deshalb gute Integrationschancen
 - Integrationsbedarf mit anderen Anwendungen bleibt unberücksichtigt

Computer Integrated Manufacturing (CIM)

- Systeme zur durchgehenden Unterstützung der Vorbereitung, Durchführung und betriebswirtschaftlichen Abbildung der Produktion in Industriebetrieben
- inspiriert durch die Vision einer vollautomatischen Produktionssteuerung
- Integration von CAD (zielt auf die Erstellung eines Konstruktionsplans eines Produktes), CAP (Arbeitsplanung, NC-Programmierung) und CAM (Fertigungs- und Transportsteuerung)
- dazu: Integration betriebswirtschaftlicher und technischer Aufgabenstellungen (Stückkostenrechnung, Erfassung von Ausschuss, Erfassung von Durchlaufzeiten und Lieferterminen)
- Bewertung
 - eröffnet beachtliche Rationalisierungspotentiale
 - Automatisierung von Teilfunktionen verspricht Kostensenkung bei gleichzeitig höherer Qualität
 - Integration von Teilfunktionen trägt zur Vermeidung kostenintensiver Friktionen bei
 - Integration technischer und betriebswirtschaftlicher Funktionen fördert die wirtschaftliche Planung und Durchführung von Produktionsprozessen
 - Automatisierungspotential durch die in der Produktion verwendeten Maschinen und Logistiksystem bestimmt
 - keine Integration mit weiteren betriebswirtschaftlichen Funktionen

Anwendungssysteme: Fazit

- historisch entstanden
 - Fokus zumeist auf Unterstützung ausgewählter Aufgaben
 - Integration mit anderen Anwendungen oft nicht berücksichtigt
- meistens zur Unterstützung von Funktionen im Unternehmen konzipiert
- deshalb: i.d.R. keine dedizierte Unterstützung von Geschäftsprozessen
- Integration von Anwendungssystemen durch unterschiedliche Sprachen und fehlende Dokumentation von Gemeinsamkeiten erschwert

Enterprise Resource Planning Systems (ERP)

- Begriff in der Praxis in vielfältiger, kaum einheitlicher Weise verwendet
- häufig auch betriebswirtschaftliche „Standardsoftware“ genannt
- entstanden durch die Integration von Teilsystemen zur Unterstützung betriebsw. Funktionen
- für eine Klasse von Unternehmen konzipiert
 - individuelle Anforderungen bleiben u.U. unberücksichtigt
 - i.d.R. Anpassungen erforderlich
- mögliche Vorteile
 - Kosten tendenziell besser kalkulierbar (nicht unbedingt günstiger)
 - breites Beratungs- und Schulungsangebot
 - ggf. Haftung des Anbieters
 - geringere Abhängigkeit von eigenen Entwicklern
 - Markt für Erweiterungskomponenten
 - mit der Software vertraute Anwender am Arbeitsmarkt verfügbar
- Herausforderungen
 - große Einsatzbandbreite vs. möglichst geringer Anpassungsaufwand (Konflikt zwischen Wiederverwendbarkeit und Semantik)
 - dabei zu berücksichtigen: große Varianz in den Anwendungsdomänen
 - Komplexität und Wartbarkeit

→ im Idealfall kennt man schon im Entwurf die Varianz der Anforderung und man bildet eine Abstraktion, die es bei der Anpassung ermöglicht, dass der einzelne Anwender z.B. durch Spezialisierung seine Anforderungen erfüllen kann, ohne dass das System im größeren Stil geändert werden muss
- Kosten und Risiken
 - Anpassung und Einführung mit erheblichen, schwer zu schätzenden Kosten verbunden
 - spezialisierte Dienstleister stehen zur Unterstützung bereit; deren Nutzen ist allerdings ex ante auch schwer zu bewerten
 - Ansätze in der Praxis zum Umgang mit dieser Herausforderung
 - Betrachtung von vergleichbaren Referenzprojekten
 - Berücksichtigung statistischer Auswertungen
 - Durchführung von Evaluationsprojekten, in denen vorab Aufwand und Nutzen eines ERP-Systems ermittelt werden soll
- Markt für ERP-Systeme
 - größere Zahl von Anbietern am Markt
 - gleichzeitig: Konzentrationstendenzen
 - Entwicklung und Markteinführung neuer Systeme mit erheblichen Investitionen verbunden (gilt besonders für große, multinationale Anwenderunternehmen)

- Bewertung
 - auch hier: keine einheitliche Funktionalität der am Markt angebotenen Systeme
 - tendenziell sehr gutes Integrationspotential, da alle Teilsysteme i.d.R. aus einer Hand
 - gleichzeitig – bei entsprechendem Entwurf – hohes Maß an Wiederverwendung (mit entsprechenden Wartungsvorteilen)
 - Teilsysteme häufig an betriebswirtschaftliche Funktionen orientiert; deshalb mitunter unzureichende Unterstützung von Geschäftsprozessen
 - für verbreitete Systeme am Markt umfangreiche Service-Angebote sowie geschulte Mitarbeiter verfügbar
 - Durchdringung des gesamten Unternehmens mit ERP-System fördert Abhängigkeit des Unternehmens
 - ambivalente Wirkung auf Wettbewerbsfähigkeit
 - einerseits Anschluss an gängigen Stand der Kunst
 - andererseits kaum Differenzierungsmöglichkeiten durch IT
 - tendenziell hohe Kosten für Anpassung und Einführung
 - Vergleich von Systemen wg. hoher Systemkomplexität und zahlreicher Unwägbarkeiten schwierig und risikobehaftet
 - Ermittlung von Kosten und zu erwartendem Nutzen ex ante überaus aufwändig

Entscheidungsunterstützungssysteme

Entscheidungsunterstützungssysteme

- sind häufig dedizierte Systeme, die unabhängig von operativen Systemen implementiert wurden
- Fokus liegt auf vertikaler Integration (Strategische Planung ↔ Controlling/Entscheidungsvorbereitung ↔ Abrechnungssysteme ↔ Operative Ebene)

Management Information Systems (MIS)

- Vorwiegend auf Führungskräfte der mittleren Ebene ausgerichtet, bzw. Stabsstellen
- frühe Euphorie (50er/60er Jahre)
 - durch erste betriebswirtschaftliche Software in den 50ern frühe Hoffnung auf wirksame Unterstützung des Management
 - Euphorie zusätzlich geschürt durch formale Entscheidungsmodelle aus der Management Science (Modelle rationalen Entscheidens, Operation Research ...)
 - zahlreiche Publikationen mit optimistischen Prognosen
- anschließend Ernüchterung
 - Systeme wurden von Führungskräften kaum genutzt (Restriktionen der damaligen IT: Stapelbetrieb; fehlende Daten; Verfügbarkeit von Informationen alleine reicht nicht aus; Modelle entsprachen nicht dem tatsächlichen Entscheidungsverhalten; durch erhebliche Abstraktionen kaum brauchbar)
 - z.T. polemische Kritik („MIES“, „MIST“)
- (bescheidenes) Come Back (80er Jahre)
 - MIS der „zweiten Generation“
 - wesentlich befördert durch grafische Benutzerschnittstellen von PCs und durch Netzwerke
 - diverse grafische Darstellungsmöglichkeiten
 - Möglichkeiten zu individuellen Abfragen

- Ziele
 - bessere Entscheidungen durch hohe Aktualität (schneller Zugriff auf operative Daten)
 - entscheidungsadäquate Auswahl und anschauliche Darstellung von Informationen
 - hohe Genauigkeit der Daten durch Vermeidung von Medienbrüchen
 - komplexe Aggregationen auf Knopfdruck, hohe Flexibilität von Auswertungen
 - kostengünstigere Entscheidungen durch
 - geringeren Zeitbedarf für Informationsbeschaffung und -auswertung
 - Entlastung (Einsparung) von Stabsmitarbeitern
 - geringeres Ausmaß an Kosten durch fehlerhafte Informationsaufbereitung
- Anforderungen
 - Erstellung und Auswertung von Kennzahlensystemen
 - diverse Aggregations- bzw. Disaggregationen (Drill-Down) für Kennzahlen
 - Ermittlung und Unterstützung der Diagnose von Soll-Ist-Abweichungen
 - Rückverfolgung von Ergebnissen und Abweichungen auf untergeordnete organisatorische Einheiten (Sensitivitätsanalysen, Exception Reporting)
 - diverse betriebswirtschaftliche Analysen (Investitionsrechnung, Break-Even-Analysen ...)
 - Planungsfunktionen (Prognosemethoden, What-If-Analysen, How-To-Achieve-Analyse, Planungsszenarien)
 - Präsentationsfunktionen
 - Traditionell: auf Papier ausgedruckte Berichte
 - graphische Präsentation am Bildschirm
 - anpassbare Benutzerschnittstelle
- Beurteilung
 - plausibler Ansatz zur Unterstützung von Management-Entscheidungen durch Aggregation/Auswertung operativer Daten und anschaulicher (grafischer) Aufbereitung
 - keine einheitliche Funktionalität
 - Integration mit Systemen der operativen Ebene zumeist unbefriedigend (unterschiedliche Datenstrukturen, keine Kommunikationsschnittstellen, keine gemeinsame Datenhaltung)
 - mittlerweile mehr und mehr verdrängt durch DWH-Systeme

Executive Information Systems (EIS)

- zielt auf eine umfassende Unterstützung für Führungskräfte der höchsten Ebene
- im Unterschied zu MIS Fokus eher auf hochaggregierten Zusammenhängen
- dazu: Integration von Bürokommunikation
- aber: vage abgegrenztes Konzept
- wesentliche Funktionsbereiche
 - Aufbereitung, Analyse und Darstellung von Unternehmensdaten (diverse Sichten)
 - persönliches Arbeitsumfeld, Aufgabenverwaltung, Terminkalender, Adressbuch ...
 - Textverarbeitung, E-Mail, WWW
 - externe Informationsdienste, Zugriff auf Fachdatenbanken
- Beurteilung
 - Grundidee plausibel: integrierter Rechnerarbeitsplatz für Unternehmensleitung
 - einheitliche Ergonomie für zugehörige Anwendungen
 - Bereitstellung entscheidungsrelevanter Informationen in geeigneter Aufbereitung
 - motiviert durch Rechnerphobie vieler Führungskräfte
 - Etikett auch durch Differenzierungsbemühungen des Marketing geprägt

- mittlerweile mehr und mehr verdrängt durch DWH- und Office-Systeme

Decision Support Systems (DSS)

- Entscheidungsunterstützung i.e.S.
- in den 70ern als Reaktion auf die Kritik am MIS entstanden
 - im Unterschied zu MIS auch für die Bearbeitung weniger strukturierter Probleme gedacht
 - durch Bemühen um Anpassung an Arbeitsstile von Führungskräften geprägt
- keine Bevormundung, sondern Unterstützung
- Funktionen: Planung, Prognose, Analyse, Simulation, Optimierung
- Abgrenzung problematisch
 - „wenig strukturierte Probleme“: abhängig vom Betrachter, keine Aussage über Grad der Strukturierung bzw. Formalisierung
 - „cognitive styles“: Forderung nach Anpassung daran durchaus sinnvoll, aber keine bewährten Theorien oder Modelle verfügbar → deshalb nur bescheidene Umsetzung der Forderung
 - Entscheidungsunterstützung: kein notwendiger Unterschied zu Informationsbereitstellung (wie bei MIS), gleiches gilt für Entscheidungsautomatisierung
- softwaretechnischer Hintergrund
 - keine einheitliche Architektur
 - mitunter sowohl deklarative, als auch prozedurale Teile
 - in den letzten Jahren verdrängt durch anderes Etikett: „Business Intelligence“
- Beurteilung
 - im Vergleich zu früheren MIS durch eine Reihe realitätsnäherer Annahmen gereift
 - Anpassung an kognitive Stile als sinnvolle Anforderung
 - nicht nur Optimierung, sondern auch Unterstützung diverser Analyseverfahren
 - erhebliche Varianz von Systemen mit diesem Etikett
 - keine einheitliche Menge abgedeckter Funktionen
 - keine eindeutige Abgrenzung gegenüber MIS und EIS
 - generelle Bewertung der Qualität von DSS deshalb nicht möglich
 - Etikett mittlerweile nur noch selten benutzt (mitunter im Zusammenhang mit XPS-Systemen)

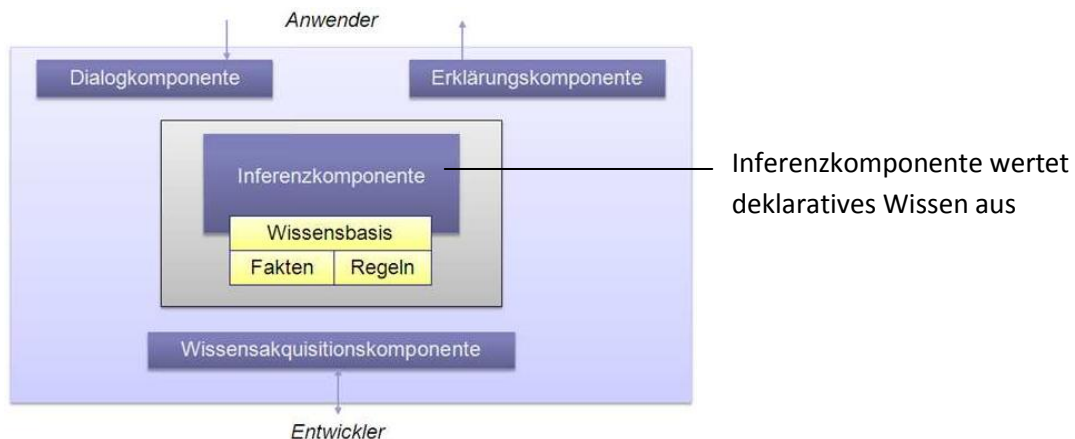
Zusammenfassende Beurteilung der Entscheidungsunterstützungssysteme

- Betonung vertikaler Integration (aber auch horizontale Integration als Grundlage wichtig)
- Integrationschancen hängen ab von
 - Verfügbarkeit von Schnittstellen in den operativen Systemen und deren Qualität
 - Heterogenität der verwendeten Implementierungssprache
- als eigenständige Systemklasse nur eingeschränkt zukunftsfähig, da Funktionalität einerseits von ERP-Systemen, andererseits von DWH-Systemen geboten wird

Wissensbasierte Systeme – Expertensysteme (XPS)

- Konzept in den 70er Jahren – große Euphorie in den 80er Jahren, aber auch kritische Stimmen
- traditionelle Künstliche-Intelligenz-Forschung: formale Rekonstruktion der Wirkungsweise menschlicher Intelligenz
- demgegenüber XPS: Automatisierung von Problemlösungskompetenz durch die Formalisierung explizit gemachten (Experten-)Wissens

- charakteristische Merkmale
 - eigenständige Architektur: Trennung formalisierten Wissens und der auf diesem Wissen operierenden Prozeduren



- anspruchsvolle Aufgaben: für Probleme gedacht, die üblicherweise von hoch qualifizierten Experten bearbeitet werden
- anwendungsbezogen: Lösung real existierender Probleme steht im Vordergrund
- anwenderfreundlich: komfortabler Dialog, vielfältige Erläuterungen
- Einsatzbereiche
 - Diagnose, Konfiguration, Planung, Prognose, Entwurf, Computer Aided Teaching
- Architektur
 - deklarative Wissensrepräsentation
 - Repräsentation in Form von Sätzen (Propositionen) einer formalen Sprache, die Zustände und Zusammenhänge (evtl. auch mit Kausalitätsanspruch) beschreiben
 - Auswertung erfordert Inferenzverfahren (wird i.d.R. von Entwicklungs- und Ausführungsumgebung bereitgestellt, Entwickler kann im besten Fall davon abstrahieren)
 - prozedural: traditionelles Programmiersprachenparadigma
 - aber: auch traditionelle Programmiersprachen enthalten deklarative Elemente (Variablen Deklaration)
 - wichtiges, aber nicht notwendiges Merkmal von XPS
 - Trennung von Wissensbasis und Inferenzkomponente
 - einfachere Pflege durch Monotonie der Wissensbasis (Hinzufügen neuer Propositionen hat keine Konsequenzen für den Wahrheitsgehalt bereits vorhandener Propositionen, also keine Gefahr durch Seiteneffekte)
 - allerdings : fatale Konsequenzen von Widersprüchen (daraus lässt sich jeder Satz ableiten)
 - ggfs. auch für Nicht-Programmierer verständlich
 - Abstraktion von Inferenzverfahren kann Performanzprobleme verursachen

Fakten	BMW_Aktie	kommt_aus_Bereich	Automobil
	BASF_Aktie	kommt_aus_Bereich	Chemie
	Opel_Aktie	kommt_aus_Bereich	Automobil
	Siemens_Aktie	kommt_aus_Bereich	Elektro
	Softlab_Aktie	kommt_aus_Bereich	Software
	Elektro	gehört_zu	Zukunftsindustrie
	Software	gehört_zu	Zukunftsindustrie
	BMW_Aktie	hat_letzte_Dividende	6
	BASF_Aktie	hat_letzte_Dividende	3
	Opel_Aktie	hat_letzte_Dividende	5
Regeln	Siemens_Aktie	hat_letzte_Dividende	8
	Softlab_Aktie	hat_letzte_Dividende	0
	Aktie ist_attraktiv		
	WENN	Aktie kommt_aus_Bereich	Branche
	UND	Branche gehört_zu	Zukunftsindustrie
	UND	Aktie hat_letzte_Dividende	Prozent
	UND	Prozent	> 4
	Aktie ist_attraktiv		
	WENN	Aktie kommt_aus_Bereich	Automobil
	UND	Benzinpreis	< 0.9

- logische Schlussfolgerungen
 - Deduktion: Schluss vom Allgemeinen auf das Besondere; bietet sich an, wenn man über ge-
haltvolle generelle Aussagen verfügt
 - Induktion: Schluss von Einzelfällen auf allgemeine Aussagen; bietet sich an, wenn keine gene-
rellen verlässlichen Aussagen zur Verfügung stehen
 - Abduktion: Schluss von der Konklusion auf die Prämisse (Wenn es geregnet hat, ist die Straße
nass. Die Straße ist nass. → Es hat geregnet.); bietet sich an, wenn die Konklusion nur wenige
Prämissen zulässt
- Inferenzverfahren = Suchverfahren
 - zielgetrieben (Rückwärtsverkettung)
 - Algorithmus zum Beweis einer vorgegebenen Aussage (Ziel) durch rekursives Beweisen
von Prämissen (Unterzielen), Beweis erfolgt durch Verifikation mittels Fakten und Re-
geln
 - Abbruch bei Verifikation oder nach vollständigem Durchsuchen der Wissensbasis
 - Einsatzbeispiele: Entscheidungsunterstützung, Diagnose (bei existierenden Annahmen
über mögliche Ursachen)
 - Beispiel in der VL: Soll Mitarbeiter befördert werden?
 - datengetrieben (Vorwärtsverkettung)
 - Algorithmus zur Suche/Generierung von Aussagen (Zielen), aus denen sich vorgegebene
Fakten und Regeln ableiten lassen
 - Start: Menge wahrer Sätze (z.B. Symptome)
 - Abbruch bei befriedigendem Ergebnis oder nach vollständigem Durchsuchen der WB
 - Einsatzbeispiele: Diagnose (wenn lediglich Symptome bekannt sind), Anlageberatung
- Voraussetzungen für Repräsentation von Wissen in XPS
 - für Problemlösung erforderliches Wissen kann in hinreichendem Umfang explizit gemacht
und ohne verzerrende Vereinfachungen formalisiert werden
 - Formalisierung in der Logik/Mathematik
 - Beschränkung auf die Form sprachlicher Ausdrücke
 - Abstraktion von Inhalt bzw. Bezug in Wirklichkeit
 - zielt auf systematische Beschreibung/Herstellung der syntaktisch und semantisch kor-
rekten Sätze einer Sprache durch einen Kalkül
 - Semantik von Aussagen entsteht allein durch die Zuordnung eines Wahrheitswerts (eine
Interpretation) innerhalb des formalen Systems
 - Formalisierung in der Wirtschaftsinformatik
 - Rekonstruktion der natürlichsprachlichen Beschreibung eines Sachverhalts mittels einer
formalen Sprache
 - Motivation: Nutzung der Vorteile einer formalen Sprache
 - formale Semantik immer auf die Form der formalsprachlichen Aussagen beschränkt
 - Darstellung in einer formalen Sprache aber niemals Selbstzweck
 - Formale Sprache
 - Menge vorgegebener Symbole, textuelle/grafische Notation, eindeutige Syntax, Seman-
tik (axiomatisch eingeführte wahre Aussagen, Definition von Äquivalenzen und zulässi-
gen Transformationen)
 - Beispiele: Aussagenlogik, Prädikatenlogik, Petrinetze

- Formalisierung und Wahrheit
 - Wahrheit von Propositionen (Sätzen) in einem formalen System immer auf das formale System beschränkt
 - Anwendung der wahren formalen Sätze auf Realität impliziert entsprechende Zuordnung und Interpretation
 - wahre formale Sätze können in einer auf die Realität bezogene Interpretation durchaus falsch sein
 - mögliche Vorteile einer Formalisierung: Präzisierung von Begriffen und Aussagen; wichtige Voraussetzung für Automatisierung; Nutzung „richtiger“, formaler Verfahren zur Analyse und Transformation formaler Aussagen (Beweise, Berechnungen)
 - Nachteil: Formalisierung benötigt immer eine Abstraktion, möglicherweise müssen wichtige Sachverhalte ausgeblendet werden, weil sie sich nicht formalisieren lassen
- Formalisierung – Vorgehen
 - Auswahl geeigneter formaler Sprache
 - Identifikation der wesentlichen Begriffe des zu formalisierenden Sachverhalts
 - widerspruchsfreie Beschreibung von Eigenschaften und Beziehungen der Begriffe mittels Aussagen der formalen Sprache
- Korrespondenz zu Realität
 - wichtig, da i.d.R. Analysen/Lösungen im formalen System auf die Realität übertragen werden sollen
 - weitgehend unproblematisch, wenn im formalen System beschriebene Objekte/Merkmale unmittelbar mit messbaren bzw. wahrnehmbaren Objekten/Merkmalen des betrachteten Realitätsausschnitts korrespondieren und diese im Hinblick auf die intendierten Problemlösungen vollständig sind (z.B. Kunde, Umsatz eines Kunden)
 - problematisch bei formalen Konzepten, deren Zuordnung zu Ausprägungen des betrachteten Realitätsbereichs mit Unsicherheit behaftet ist (z.B. Wettbewerbsfähigkeit)
→ Möglichkeiten: auf Urteilskraft des Benutzers vertrauen; Formalisierung weiter differenzieren; Formalisierung als nicht realisierbar aufgeben
- zentrale Herausforderung
 - Die Grenzen der Formalisierung in der Wirtschaftsinformatik sind dort zu sehen, wo eine semantisch äquivalente bzw. funktional/pragmatisch angemessene Rekonstruktion mehrdeutiger Sinngehalte durch eine endliche Menge messbarer/wahrnehmbarer Eigenschaften problematisch erscheint.
- Grenzen der Formalisierbarkeit von Wissen
 - implizites Wissen: was ist ein erfolgreiches Produkt; Bedeutung von bestimmten Statussymbolen für einen Mitarbeiter; Vertrauenswürdigkeit eines Kunden
 - Lösungsansatz: Induktion durch vergangene Ergebnisse, Expertenwissen
 - vages Wissen: Konfidenzmaß 90 %; wenn es günstig läuft dann ...
 - wichtig sind die Begründungen
 - nicht-monotones Wissen: Ausnahmen, wichtige Funktion für die Reduktion von Komplexität, Widerspruch wird i.d.R. im Kontext aufgelöst
 - Bsp: Alle Vögel können fliegen. Der Pinguin ist ein Vogel. Der Pinguin kann nicht fliegen (nicht monoton).
- Einsatzvoraussetzungen von XPS
 - Der betrachtete Realitätsausschnitt ist durch eine Fülle von Zusammenhängen gekennzeichnet, die sich in Form von Regeln anschaulich beschreiben lassen.

- Der Realitätsausschnitt bzw. das Wissen über denselben ändert sich dynamisch. Eine monotone Wissensbasis verspricht hier Wartungsvorteile.
- Die Reihenfolge, in der die Regeln beim Programmablauf sinnvollerweise Anwendung finden sollten, streut in Abhängigkeit von den Anfragen und den aktuell gültigen Zusammenhängen erheblich.
- Integrationsbedarf
 - Entscheidungsunterstützungssysteme erfordern i.d.R. Daten, die in anderen Systemen verwaltet werden → empfiehlt die Integration mit solchen Systemen
 - Integration allerdings häufig erschwert durch unterschiedliche Sprachparadigmen und fehlende/unzureichende Schnittstellen
- Beurteilung der Expertensysteme
 - unter bestimmten Voraussetzungen gut geeignet, um Entscheidungen (teilweise) zu automatisieren (→ Einsatzvoraussetzungen)
 - deklarative Repräsentation relativ anschaulich und wartungsfreundlich (monotone Wissensbasis)
 - mystifizierende Terminologie allerdings wenig angebracht
 - wesentliches Merkmal: softwaretechnischer Ansatz
 - Kehrseite der Repräsentation (Abstraktion): Integration mit gängigen Implementierungssprachen aufwändig

Integration durch Datenbanktechnologien

Datenbanken – wichtige Begriffe

- DB-Schema: legt die (logische) Struktur der in einer DB zu verwaltenden Daten fest, definiert also die Semantik; wird i.d.R. auch in der DB abgebildet
- DB: eine Instanz eines DB-Schemas; enthält die zu verwaltenden persistenten Daten
 - Sichten: externe Sicht, logische Gesamtsicht, interne Sicht
- DBMS: dient der Verwaltung (Anlegen, Zugriff, Änderung) von DB

Ziele und Aufgaben des DBMS

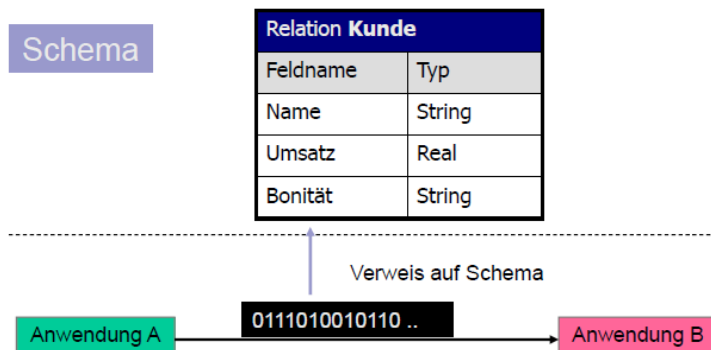
- Erzeugung und Verwaltung persistenter Daten
- physische Datenunabhängigkeit
 - die physische Organisation der Daten auf dem verwendeten Speichermedium bleibt dem Anwendungsprogramm verborgen
 - Vorteil: Veränderungen der physischen Datenorganisation wirken sich nicht auf die Anwendung aus
- logische Datenunabhängigkeit
 - die logische Sicht der Anwendungen auf die DB sollte von Änderungen des DB-Schemas (weitgehend) unberührt sein
 - Vorteil: nötige Änderungen des DB-Schemas können durchgeführt werden, ohne dass die zugreifende Anwendungen geändert werden müssen
- Unterstützung nebenläufiger Zugriffe auf gemeinsame Daten (Synchronisation, operationale Integrität)
- Überwachung von Integritätsbedingungen
- Zugriffskontrolle
- Backup/Recovery

Relationales Modell

- Modell für die Datenorganisation in DB
- ein relationales DB-Modell besteht aus einer Menge eindeutig benannter Relationstypen, die (mindestens) in der 1. Normalform sind
- eine Relation ist eine Menge von Tupeln, die aus einer festen Anzahl von Werten eines jeweils vorgegebenen Attributtyps bestehen
- Relationenalgebra definiert (Zugriffs-)Operationen über Relationen

Integration mittels RDBMS

- Austausch von Nachrichten
 - gemeinsames Schema legt die Bedeutung von Nachrichten fest, die zwischen Anwendungen ausgetauscht werden



- Voraussetzungen
 - es gibt ein gemeinsames Schema, auf das jederzeit von allen Anwendungen zugegriffen werden kann
 - Alternative: es existieren mehrere Kopien eines Schemas
 - die Typen der in den Anwendungen verwalteten Daten lassen sich auf die Datentypen im Schema ohne unzumutbare Verzerrungen abbilden
- gemeinsame Datenverwaltung
 - die zu integrierenden Anwendungen verwalten ihre (persistenten) Daten in einer gemeinsamen DB
 - der Zugriff erfolgt über entsprechende Funktionen des DBMS
 - das DBMS muss die Zugriffe synchronisieren
 - die Kommunikation zwischen Anwendungen kann sich auf das Versenden von Referenzen auf Daten in der gemeinsamen DB beschränken

Grenzen des relationalen Modells

- Objekte können nur als BLOB gespeichert werden → Verlust von Semantik, die nur mit Aufwand und Risiko rekonstruiert werden kann

Integration durch RDBMS - Beurteilung

- RDBMS stellen ausgereifte Technologie zur Verwaltung persistenter Daten im Mehrbenutzerbetrieb dar
- solide mathematische Fundierung
- SQL in hohem Maße standardisiert – fördert Investitionsschutz
- viele Programmiersprachen unterstützen Embedded SQL

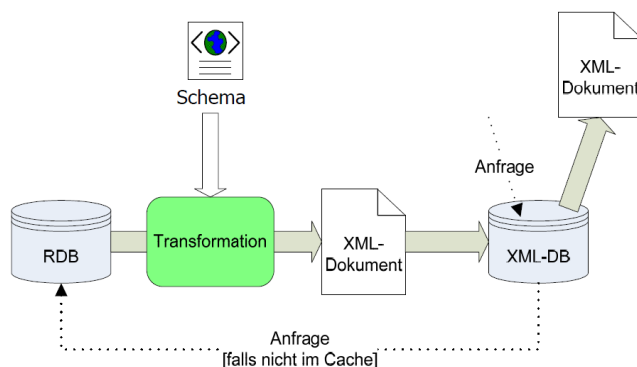
- aber: Abbildung von Programmiersprachen immer notwendig – mitunter semantische Verzerrungen
- SQL bietet nur wenige einfache Datentypen → eingeschränkte Integrität (Integrität sehr gut, aber mit sehr niedriger Semantik)

XML-Datenbanken

- dienen der Verwaltung persistenter XML-Dokumente
- Schemadefinition mittels XML-DTD oder XML Schema
- zunehmende Bedeutung durch wachsende Verbreitung von XML-Dokumenten
- unterschiedliche Ansätze zur Ablage und Verwaltung von XML-Dokumenten in DB

Ansätze zur Verwaltung von XML-Dokumenten in DB

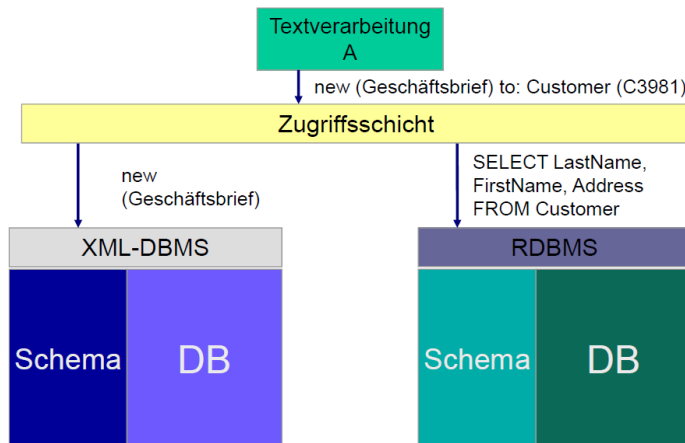
- spezieller Datentyp für die Speicherung kompletter XML-Dokumente in RDBMS
 - erlaubt Syntaxprüfung
- Generierung von XML-Dokumenten aus RDBMS
 - Semantik geht verloren ohne zusätzliche Schemata, da nur ein Datentyp vorhanden: PCDATA
 - bei Abbildung einer DTD auf RDBMS evtl. manuelle Eingriffe für korrekte Modellierung nötig
- Erweiterung von RDBMS („XML-fähige DBMS“)
 - erlauben Zugriff auf Teile des XML-Dokuments
- „native“ XML-DB
- XML-DBMS als Cache für RDBMS
 - XML-Dokumente werden aus RDB erzeugt
 - Problem: Generieren der XML-Dokumente aufwändig
 - deshalb u.U. unzureichende Performanz bei hoher Last
 - deswegen: temporäre Ablage der XML-Dokumente in XML-DB
 - schneller Zugriff auf ganze Dokumente
 - weniger: Zugriff auf einzelne Bestandteile von Dokumenten (keine Performanzvorteile)



XML-DB – Beurteilung

- XML-Dokumente zur Speicherung/Verwaltung von typisierten Daten schlecht geeignet („datenorientierte XML-Dokumente“)
 - Verlust an Semantik und Integrität, Verschlechterung gegenüber RDBMS
- XML-Dokumente ggfs. geeignet zur Speicherung von strukturierten Texten („dokumentorientierte XML-Dokumente“)
 - Semantikverlust hält sich in Grenzen
 - Vorteile gegenüber proprietären Dokumentformaten und einer Ablage mittels Dateiverwaltungssystem

- Integration durch XML-DB und RDBMS



- weitere Sinnvolle Einsatzszenarien
 - Archivierung von Geschäftsdokumenten
 - XML-DBMS als Cache

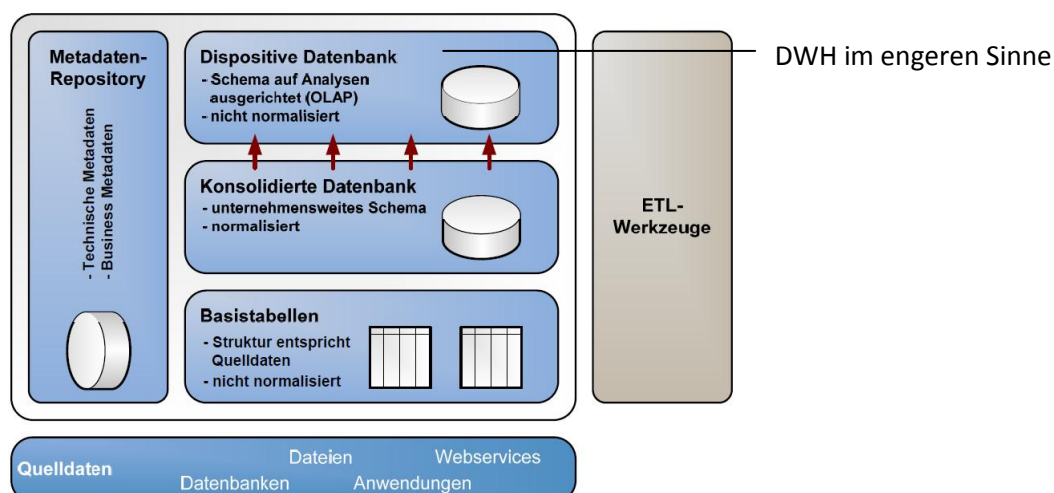
Data Warehouse

- keine einheitliche Begriffsdefinition
- kein monolithisches Software-Produkt; entsteht durch DBMS und ergänzenden Werkzeugen sowie durch die Spezifikation dedizierter Datenschemata und deren Instanzen
- wesentliche Merkmale
 - Unterstützung der Integration von Daten, die in heterogenen Quellen vorliegen
 - Auswertung bzw. Aufbereitung komplexer Datenmengen zur wirksamen Entscheidungsunterstützung → OLAP
 - entsteht durch einen (permanenten) Prozess der Datenextraktion und -aufbereitung

OLTP vs. OLAP

Online Transaction Processing	Online Analytical Processing
<ul style="list-style-type: none"> - optimiert für kurze Transaktionen und bekannte Lastprofile - i.d.R. große Zahl paralleler Transaktionen - ACID-Eigenschaft und Performanz wichtig 	<ul style="list-style-type: none"> - Ziel: Entlastung der operativen Systeme, Beschleunigung der Ausführung komplexer Anfragen - Daten über die Zeit verfügbar machen - ACID unwichtig, nur Lesezugriff

Bestandteile eines DWS (nach Lehner)

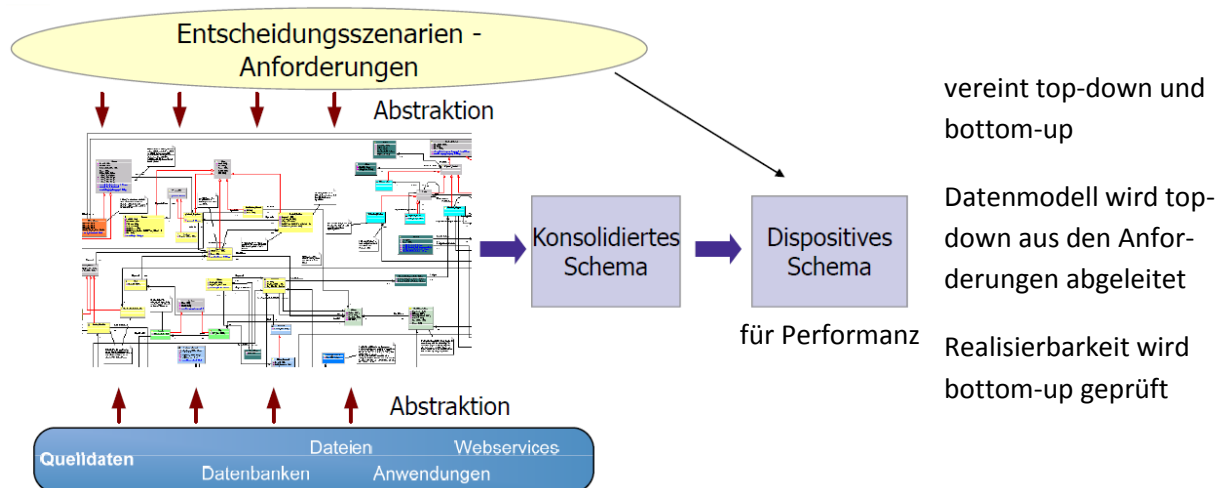


DWS: auch ein Prozess

- System zur Aufbereitung, Integration und Verwaltung von Daten
- gleichzeitig: Objekt und Objektivierung eines permanenten Prozesses
- aber: empfohlenes Vorgehen i.d.R. durch pragmatische Ausrichtung geprägt

Beispielhaftes Vorgehensmodell

1. Identifikation relevanter Quelldaten
 - Dateien, Datenbanken, Anwendungen, Webservices
2. Extraktion in Basistabellen
 - Struktur entspricht Quelldaten, Tabellen nicht normalisiert
 - Unterstützung durch Metadaten-Repository, welches Verzeichnisse von Basistabellen und Quelldaten beinhaltet
 - ETL-Werkzeuge: Erfassung von Quelldaten (wenn nötig) → Zugriff auf Metadaten-Repository → Erzeugen von Basistabellen aus Quelldaten
3. Reinigung der Daten
 - Divergenz zwischen Datenquellen (Semantik aufgrund unterschiedlicher Abstraktionen/Spezifikationen, Homonyme, Synonyme)
 - Aufgabe: Entdeckung und Beseitigung von Divergenzen (in den operativen Daten)
 - mangelhafte Datenqualität: Dubletten, fehlende Werte, inkonsistente Werte
 - Aufgabe: Entdeckung und Beseitigung von Qualitätsmängeln (in den operativen Daten)
 - Unterstützung durch Werkzeuge
 - Entdeckung von Strukturgleichheit bzw. -ähnlichkeit
 - dazu: Feststellung gleicher bzw. korrespondierender Datentypen
 - Entdeckung von Dubletten und möglicher Homonyme und Synonyme
 - inhaltliche („materielle“) Analyse von Daten
 - Entdeckung semantischer Inkonsistenzen durch Rückgriff auf generelle Spezifikationen (z.B. Ontologien) oder auf ergänzende, unternehmensspezifische Integritätsbedingungen
4. Konsolidierung der Daten
 - Normalisierung der bereinigten Basistabellen
 - ggfs. (Re-)Konstruktion eines neuen Schemas
 - Transformation der Daten aus Basistabellen in eine Instanz des konsolidierten Schemas
 - Dokumentation der Transformationsvorschriften im Metadaten-Repository
5. Erstellung dispositiver Datenbank
 - Anforderung an Entscheidungsunterstützung
 - Ermöglichung komplexer Abfragen und Auswertungen in kurzer Zeit – OLAP
 - auch: Berücksichtigung zeitlicher Entwicklungen
 - dazu: möglichst keine Belastung des operativen, transaktionsorientierten Systems
 - Maßnahmen
 - Erstellung von Schemata, die Analysen effizient unterstützen
 - dazu: bewusste Einführung von Redundanzen, „materialisierte“ Daten, z.B. aggregierte/berechnete Daten (dadurch natürlich Informationsverlust)
 - setzt Wissen um typische Auswertungen voraus
 - Konsequenz: Beschränkung auf lesenden Zugriff

Methodisches, modellbasiertes Vorgehen**Metadaten-Management**

- aktuelle Situation: Insel-Lösungen mit werkzeugspezifischen (lokalen) Repositories
- proprietäre, spezifische Metadatenmodelle
- Inkompatibilitäten machen den Metadaten-Austausch schwierig
- Ziel: zentrales unternehmensweites Metadaten-Repository zur Speicherung von technischen und Business-Metadaten für die gesamte Infrastruktur
- Standard: Common Warehouse Metamodel (CWM) der OMG

Benutzersicht

- Differenzierung von Fakten und Dimensionen
- möglichst anschauliche Zugriffsmetaphern mit anschaulichen Analyseoperationen
 - Datenwürfel („Data Cube“)
 - Fakten durch Zahlen innerhalb eines Würfels abgebildet
 - Dimensionen sind qualifizierende, außerhalb des Würfels abgebildete Informationen
 - Zugriff visualisiert durch Auswahl von Elementen des Würfels (z.B. Slice → Reduktion auf zwei Dimensionen; Dice → Reduktion auf eine Dimension)
 - Pivot-Tabelle
 - Analysedimensionen werden auf (zweidimensionale) Tabelle abgebildet
 - erlaubt Berücksichtigung beliebig vieler Dimensionen
 - unterstützt komfortablen interaktiven Zugriff
 - Auswertungsoperationen
 - Slice und Dice durch Ausblenden von Tabellenspalten/-zeilen (Projektion)
 - Roll-Up: Aggregation der Daten über eine weitere Dimension
 - Drill-Down: Disaggregation aggregierter Daten zur Unterstützung einer differenzierteren Betrachtung (Ursachenanalyse)
- technische Realisierung
 - Zugriffsschicht häufig durch Tabellenkalkulationsprogramme realisiert
 - setzt geeignete Schnittstellen voraus
 - bietet Benutzern i.d.R. vertraute Umgebung
 - bietet ergänzende diverse grafische Darstellungsmöglichkeiten

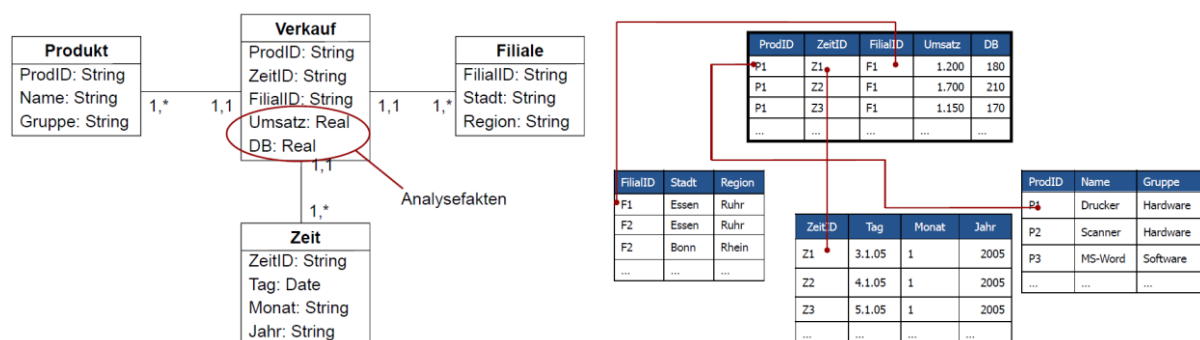
Datenbanktechnologie

- RDBMS nur eingeschränkt geeignet
 - relationales Modell unterstützt keine Differenzierung von Dimensionen und Fakten
 - keine Optimierung der Datenrepräsentation für spezielle Analyseoperationen
- Erweiterung von RDBMS (ROLAP = Relationales OLAP)
 - Umsetzung von OLAP-Operationen in SQL
 - Nutzung spezieller Schemata (Star-/Snowflake-Schemata)
- dedizierte DBMS
 - MOLAP = multi-dimensionales OLAP
 - basiert auf speziellen multidimensionalen Datenmodellen
 - optimiert für OLAP
 - OLE DB für OLAP als aufkommender Schnittstellenstandard

MOLAP – Bewertung

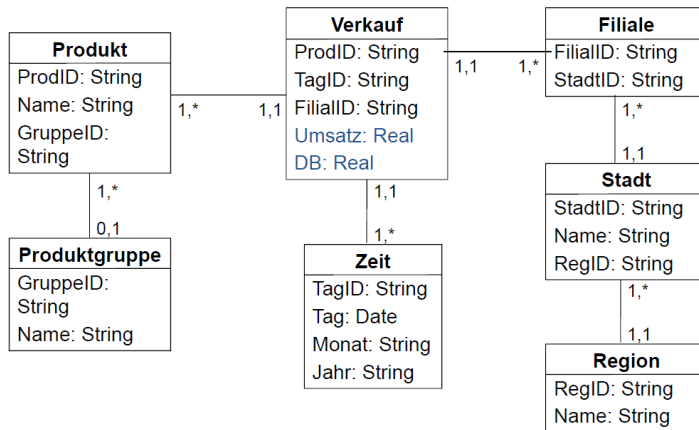
- schnell aufgrund optimierter Datenhaltung
- Speichersystem und OLAP-Engine stark miteinander verwoben, Datenzugriff proprietär
- inkrementelle DB-Aktualisierungen nur schwer möglich, da das multidimensionale Array i.d.R. komplett neu generiert werden muss, wenn sich die Struktur ändert
- Problem in der Dünnbesetztheit (Sparsity)
 - in typischen Datenwürfeln sind weniger als 5 % der Zellen besetzt
 - leere Zellen benötigen jedoch ebenfalls Speicherplatz
 - deshalb Probleme bei sehr großen Datenbeständen
 - Kompression als Lösungsansatz
- es existiert keine einheitliche multidimensionale Algebra wie in der relationalen Welt
- multidimensionale Modelle entwickelten sich aus den Implementierungen in kommerziellen OLAP-Produkten
- wissenschaftliche Arbeiten versuchen nachträglich eine Formalisierung und saubere Definition
- Hybrid OLAP (HOLAP) als Kompromiss, relationale Ablage der Detaildaten, multidimensionale Ablage der Aggregate

ROLAP – Star Schema



- Verbesserung der Performanz durch weitere Redundanz – nur sinnvoll bei großen Datenbeständen und gleichgearteten Berichten

ROLAP – Snowflake Schema



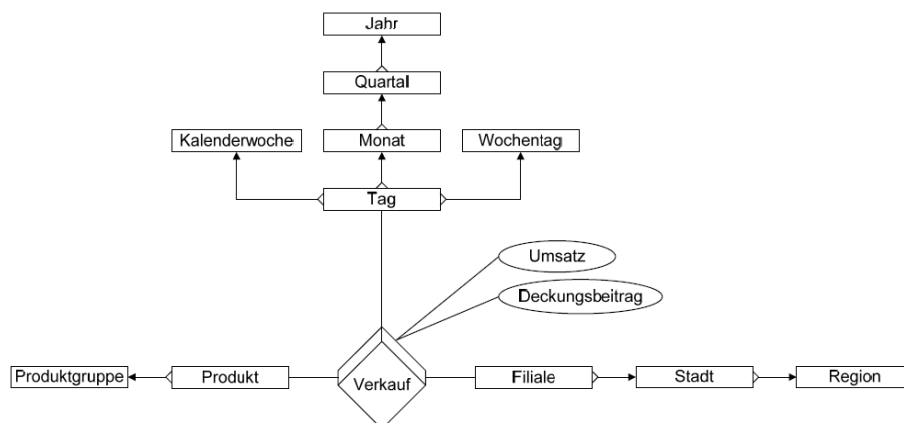
- Faktentabelle weiterhin aggregiert, nur Dimensionen normalisiert

Star/Snowflake-Vergleich

- für beide gilt
 - zentrale Faktentabelle zur Speicherung der quantifizierenden Daten (Kennzahlen)
 - zusammengesetzter Primärschlüssel aus den Dimensionen
 - wenige Spalten, viele Tupel
- Star-Schema
 - denormalisierte Dimensionstabellen
 - relativ wenig Tupel
 - Verwendung für Selektion und Aggregation
- Snowflake-Schema
 - Dimensionstabellen sind normalisiert (d.h. die DB ist in 3NF)
 - leichtere Aktualisierungsmöglichkeit, schlechtere Performanz

Entwurf konzeptueller Datenmodelle

- ER-Modell für multidimensionalen Entwurf schlecht geeignet, da keine Unterscheidung zwischen qualifizierenden Entities (Dimensionen) und quantifizierenden Entities (Kennzahlen/Fakten)
- es gibt verschiedene Vorschläge für spezielle Sprachen zur Darstellung multidimensionaler Datenmodelle (M/ER, ADAPT, Erweiterungen von UML) → aber z.Z. kein Standard in Sicht



Data Mart

- betont den Umstand, dass mitunter eine unternehmensweite konsolidierte bzw. dispositive DB zu aufwändig ist
- zielt auf dispositive DB für eingeschränkten Bereich eines Unternehmens (z.B. eine Abteilung) → separation of concerns
- mitunter als weitere Schicht über dispositiver DB eingeführt
- Begrifflichkeit nicht eindeutig bzw. nicht überzeugend

Abhängige Data Marts

- Extraktion eines Teils des DWH zur Nutzung in einer Abteilung oder für eine bestimmte Benutzergruppe/Anwendung
- zur Erinnerung: das multidimensionale Datenmodell ergibt sich aus den Analyseanforderungen der Anwender; eine DWH sollte für verschiedene Anwendungen geeignet sein
- Verwendung von (virtuellen oder materialisierten) Data Marts zur Erzeugung einer multidimensionalen Sicht auf das unternehmensweite Datenmodell; vorberechnete Aggregate oder MDDB-Technologie nur in OLAP Data Marts

Unabhängige Data Marts

- Versuch der Verwendung eines Abteilungs-Data-Marts als „Proof of Concept“ bzw. Prototyp für ein DWH-Projekt
- Problem: der Aufwand für den Ladevorgang (Data Cleaning, Datenextraktion), die Hauptproblematik in einem DWH-Projekt, werden nicht reduziert
- unabhängige Data Marts führen tendenziell zu Integritäts- und Skalierbarkeitsproblemen, wenn sie später zusammengeführt werden sollen
- Vorteil: schnell, billig, dezentral zu entwickeln, überschaubar zu verwalten

DWH – Beurteilung

- Pragmatischer Ansatz zur Integration heterogener Datenquellen
- unausgereifte und uneinheitliche Terminologie erschwert Verständnis
- Unterstützung schneller Analysen durch redundante Datenhaltung (geringe Aktualität der Daten)
- Qualität des DWH hängt wesentlich von der Aufbereitung und Rekonstruktion vorhandener Datenquellen ab
- auf Lesezugriffe beschränkt; deshalb kein Beitrag zur Integration operativer (transaktionsorientierter) Systeme
- Aufwand zur Realisierung hängt ab von
 - Heterogenität der Datenquellen
 - Qualität der Schnittstellen zu den Datenquellen
 - Varianz der Analysetypen
- (unkontrollierte) Änderungen der Syntax und/oder Semantik der Datenquellen gefährden Funktionsfähigkeit bzw. verursachen erheblichen Pflegeaufwand
- Beitrag zur Integration betrieblicher Informationssysteme ambivalent
 - keine Integration auf operativer Ebene: Anwendungen werden nicht kommunikationsfähig
 - virtuelle Integration durch konsolidiertes Schema

Data Mining

- zielt auf die Extraktion nützlicher Informationen aus großen, u.U. heterogenen Datenmengen

- dazu: automatisierte Ermittlung genereller Zusammenhänge aus einer Vielzahl von Einzelfällen durch Induktion
- realisiert durch diverse Verfahren der induktiven Statistik oder mittels neuronaler Netze

Wichtige Formen des Data Mining

- Assoziationsregeln
 - Bsp: Wenn ein Kunde Lachs kauft, dann kauft er auch Weißwein (Korrelation: 45 %)
- Muster in Zeiträumen
 - Bsp: Die Entwicklung der Aktienkurse der Unternehmen A, B und C verläuft nach dem gleichen Muster
- Clustering
 - Bsp: Bildung von Clustern prospektiver Konsumenten, die sich hinsichtlich der Anforderungen an einen PKW wesentlich unterscheiden

Induktiver Fehlschluss

- eine größere Zahl von Einzelfällen kann eine generelle Aussage bestätigen bzw. stützen, aber nicht beweisen
- ergo: der Schluss von einer endlichen Menge von Einzelfällen auf generelle Aussagen ist prinzipiell fehlbar
- anders: „vollständige Induktion“ in der Mathematik, weil vollständige Iteration durch unendliche Mengen

Data Mining und DWH – Beitrag zur Entscheidungsunterstützung

- regelmäßige ‚blinde‘ Suche nach interessanten Mustern
- gezielte Suche nach Mustern
 - Bsp: Welche Kriterien beeinflussen die Kaufentscheidung für Festplattenrekorder?
- Überprüfen von Hypothesen
 - Bsp: Das Kaufinteresse für Festplattenrekorder ist signifikant vom Geschlecht und dem Einkommen der Kunden abhängig

Data Mining und DWH – Unterstützung der (Re-)Konstruktion heterogener Datenquellen

- auf Schema-Ebene
 - Suche nach gleichen oder ähnlichen Datenstrukturen
 - ergibt u.U. Hinweise für Gestaltung eines konsolidierten Schemas
- auf Instanzen-Ebene
 - Suche nach Mustern in Daten gleichen Typs
 - ergibt u.U. Hinweise für eine sinnvolle Differenzierung von Datenstrukturen
 - Bsp: viele Organisationsnamen beinhalten „Universität“ → Hinweis auf Bildung eines entsprechenden Entitätstyps

Data Mining – Beurteilung

- u.U. wirksamer Beitrag zur Analyse
 - Entdeckung bisher nicht bekannter, entscheidungsrelevanter Zusammenhänge zwischen Daten
 - Überprüfung entscheidungsrelevanter Hypothesen
- Unterstützung bei der Entwicklung gemeinsamer Schemata
- Unterstützung bei der Verbesserung der Datenqualität

- damit: Unterstützung der Integration heterogener Datenquellen; sinnvolle Ergänzung von DWH
- Erfolgchancen abhängig von
 - differenzierte Beschreibung von Datenstrukturen
 - Verwendung möglichst einheitlicher Bezeichner zur Beschreibung von Datenstrukturen
 - Repräsentativität der eingehenden Daten
 - Qualifikation der Nutzer → Evaluierung ob die Muster überhaupt Sinn machen
- Unterstützung durch
 - Qualifizierung der Nutzer → theoretischer Hintergrund zur Dateninterpretation
 - Verzeichnisse (ggfs. Ontologien) von Bezeichnern für Datentypen/Klassen und Objekten/Eigenschaften in den betrachteten Domänen

Integration durch Datenaustauschformate

Integration durch Nachrichtenaustausch

- technische Kommunikationsinfrastruktur muss vorhanden sein
- Systeme benötigen ein gemeinsames semantisches Referenzsystem
- Werte vs. Referenzen
 - Referenzen: Anwendungen haben eine gemeinsame Datenbasis, ökonomischer Vorteil
 - Werte: selten gemeinsame (unternehmensübergreifende) Datenbestände vorhanden
- wird eine Standardstruktur als Integrationsmedium verwendet, benötigt man bei n Anwendungen nur n (bzw. 2 n) Abbildungen auf diese Struktur → dafür muss die Struktur aber so flexibel sein, dass sie alle zu verschickenden Nachrichten auch abgemessen abbilden kann

Dokument

- etymologische Herkunft
 - „documentum“ → Beweis, „wodurch man etwas lehren kann“
 - „docere“ → lehren, unterrichten, nachweisen
- traditionelle Bedeutung
 - persistente (schriftliche) Darstellung
 - dient dem Nachweis bzw. der verbindlichen Darstellung eines Sachverhalts
 - Original und Kopie sind zu unterscheiden

Elektronisches Dokument

- Aufweichung des traditionellen Dokumentbegriffs
- Sammelbegriff für digitale Objekte (zumeist als Datei abgelegt) mit folgenden Eigenschaften
 - Semantik der enthaltenden Daten wird von der jeweiligen Bearbeitungssoftware gar nicht oder nur eingeschränkt berücksichtigt (z.B. Textverarbeitung und Vertrag)
 - Inhalte sind für die Interpretation durch einen menschlichen Betrachter bestimmt
 - deshalb: Präsentation ist wichtiger Bestandteil des Dokumentbegriffs
 - Original und Kopie i.d.R. nicht zu unterscheiden

Unterschied zu Daten

- Differenzierung nicht eindeutig (elektronische Dokumente sind Daten!)
- Dokumente zumeist mit einer Präsentation verbunden
- wichtig: ein Dokument ist ein sequenzialisierbarer Datencontainer, der als Nachricht fungieren kann
- i.d.R. Ablage von Werten, nicht von Referenzen

- die enthaltenen Daten sind i.d.R. nicht in differenzierter Weise typisiert
- die Struktur eines solchen Container ist i.d.R. variabel

Typen von Sprachen im Kontext von Dokumenten

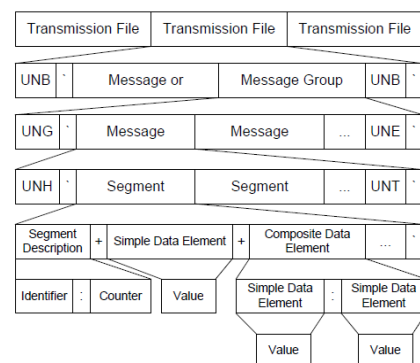
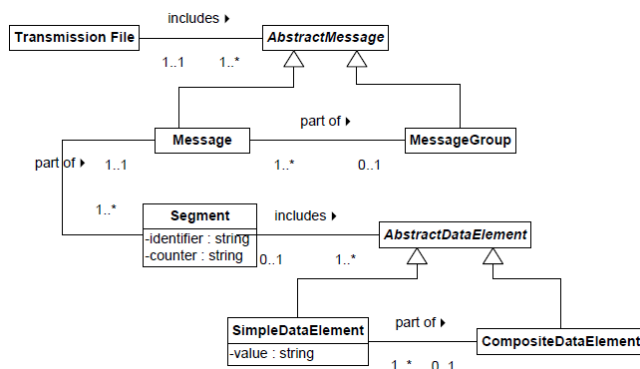
- Dokumentsprachen
 - werden benutzt, um Informationen strukturiert zu beschreiben
 - einzelne Dokumente werden damit beschrieben
 - Bsp: HTML, RTF, EDIFACT, TeX/LaTeX etc.
- Metasprachen
 - dienen der Definition von (Dokument-)Sprachen
 - Bsp: SGML, XML, EBNF (Extended Backus-Naur-Form)

Electronic Data Interchange (EDI)

- Ursprung: 60er Jahre, amerikanische Transport- und Einzelhandelsindustrie
- wachsender Dokumentenaustausch: Aufträge, Rechnungen, Lieferscheine, Schecks
- Ziel: Kosteneinsparungen
- über die Jahre verschiedene, branchenspezifische Standards daraus entstanden (z.B. EDIFACT)

EDIFACT

- mittlerweile als ISO-Standard verabschiedet (ISO 9735)
- vorgegebene Struktur des EDIFACT-Standards, um ein Geschäftsdokument auf einen Datenstrom abzubilden
 - Bezeichnung der benötigten Datenelemente durch Rückgriff auf das United Nations Trade Data Element Dictionary (UNTDDED)
 - eine formale Syntax zur Bildung von Segmenten aus Datenelementen
 - Nachrichtentypen (message types), die den formalen Aufbau bestimmter Dokumente (z.B. einer Rechnung) aus vorgegebenen Segmenten standardisieren (z.Z. ca. 200)



- Erweiterbarkeit
 - mit Hilfe des (erweiterbaren) UNTDDED sowie unter Beachtung der Syntaxregeln können bei Bedarf neue Nachrichtentypen entworfen und ggfs. nach aufwändiger Prüfung standardisiert werden
 - Nachrichtentypen sollten branchenübergreifend und international sein, Segmente sollten wiederverwendbar gestaltet werden
- Message Function Codes dienen der inhaltlichen Verknüpfung von Nachrichten (z.B. Cancellation, Addition, Response)

- Interchange Agreement
 - vertragliche Vereinbarung zwischen Unternehmen, die die vom Standard nicht festgelegten Konventionen des Datenaustauschs regelt
 - das EDIFACT-Board stellt für den Entwurf einer solchen Vereinbarung eine Checkliste bereit
- Vorteile
 - pragmatischer Ansatz, der eine schnelle Verbreitung fördert
 - Einsatz in betrieblichen IS durch viele (integrierte) Konverter relativ unproblematisch
- Nachteile
 - aus der Sicht der Informatik eher bedenklich: Enumeration statt Abstraktion, Erweiterung statt Spezialisierung
 - Layout der Geschäftsdokumente wird von EDIFACT nicht berücksichtigt – ist aber für Unternehmen u.U. wichtig
 - verliert durch zunehmende Verbreitung XML-basierter Standards an Bedeutung

Dokumente in Unternehmen – Einsatzzwecke

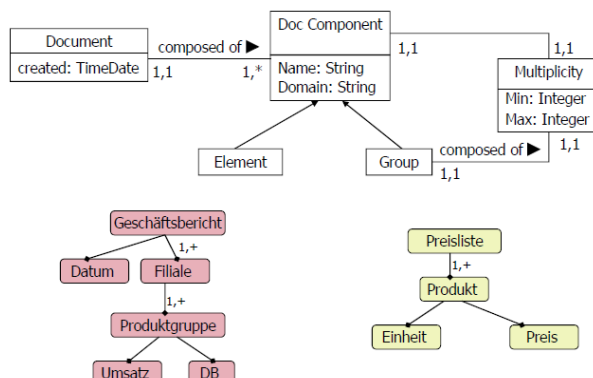
- Austausch strukturierter Daten
 - Geschäftsdokumente zwischen Unternehmen
 - Entwurfsdokumente (CAD oder CASE)
- Realisierung einheitlicher Strukturen für bestimmte Dokumentationsaufgaben
 - Geschäftsberichte, Projektberichte
- Realisierung einheitlicher Präsentationsformen
 - nach Maßgabe der corporate identity
 - im Hinblick auf bestimmte Ausgabemedien

Unterscheidung von Typ und Instanz

- Spezifikation von Dokumenttypen wichtige Voraussetzung für intendierte Strukturierung/Vereinheitlichung
- jede Instanz weist die im Typ definierte Struktur auf
- Instanzen unterscheiden sich durch ihren Zustand
- wichtige Voraussetzung für die Wahrung der Integrität von Dokumenten

Identifikation von Gemeinsamkeiten

- große Varianz möglicher Dokumenttypen und Typdefinitionen
- in Zukunft weitere Dokumenttypen zu erwarten
- durch starre Generalisierungs-/Spezialisierungsbeziehungen nicht abbildbar
- Strukturen von Dokumenttypen teilweise sehr unterschiedlich, aber möglicherweise lassen sich alle Typen auf eine generische (Meta-)Struktur zurückführen (z.B. Baumstruktur)



Vereinheitlichung – Gemeinsame Dokumentbeschreibungssprache

- Sprachmittel entsprechen korrespondierendem einheitlichen Metamodell
- erlauben die Definition beliebiger Dokumenttypen in einer Sprache
- fördert die Wiederverwendung von Werkzeugen
 - Werkzeuge können für Dokumentbeschreibungssprache konzipiert sein
 - z.B. Parser, die Typkonformität prüfen oder Werkzeuge zur Unterstützung der Transformation in andere Repräsentationen
- reduzierter Einarbeitungsaufwand für neue Dokumenttypen

Beispiel für eine Metasprache: SGML

- entwickelt in den 60er Jahren, seit 1986 ISO-Standard
- Sprache zur Beschreibung von Dokumenttypen
 - Dokumenttypen als sprachliche Struktur zur Beschreibung einer Klasse von Dokumenten
 - SGML deshalb auch Metasprache genannt (Sprache zur Beschreibung von Sprachen)
- Fokussierung auf Inhalt
 - Markups für logische Dokumentstrukturen (DTD), Layout nicht berücksichtigt
- umfangreiche, komplexe Spezifikation; Nutzung deshalb sehr aufwändig

XML

- abgespeckte Fassung von SGML
- sehr populär, u.a. durch HTML
- XML erlaubt die Definition von Dokumenttypen mittels einer Document Type Definition (DTD)
- für viele, ganz unterschiedliche Dokumenttypen eingesetzt
- mitunter als besonders fortschrittliche Art der Datenverarbeitung dargestellt
- Dreiteilung der Dokumente
 - Struktur: DTD (Typ)
 - Inhalt: Daten + Markup (Inhalt)
 - Darstellung: Stylesheet (Repräsentation)

Sprachebenen

- XML ist eine Metasprache
 - Metasprache dient der Beschreibung von Dokumenttypen
 - aber: Unterscheidung zwischen Objektsprache und Metasprache ist relativ, eine Sprache kann sowohl als Objektsprache als auch als Metasprache dienen
- XML dient der Beschreibung von Dokumenttypen mittels der Konzepte Element, Attribut und Element-Inhalt (Funktion als Metasprache)
- Instanzen dieser Dokumenttypen (Objektebene) sind durch Verwendung von Elementen, Attributen und Element-Inhalten aufgebaut: XML-Dokumente (Funktion als Sprachklasse)

Semantisches Niveau

- für die Inhalte der Datenelemente eines XML-Dokumenttyps nach DTD stehen keine differenzierten Datentypen zur Verfügung
 - lediglich ein allgemeiner Zeichenkettenstring (#PCDATA)
 - die Beschreibung der Inhalte bleibt deshalb auf einem relativ niedrigen semantischen Niveau

Extensible Stylesheet Language (XSL)

- eine Familie von Sprachen zur Transformation bzw. Präsentation von XML-Dokumenten
- besteht aus drei Teilen
 - XSL Transformations (XSLT): Definition von Transformationen von Dokumenttypen
 - XML Path Language (XPath): Adressierung von Teilen eines XML-Dokuments
 - XSL Formatting Objects (XSL-FO): Definition der Präsentation eines XML-Dokuments

XML-Schema

- beinhaltet einfache und komplexe Datentypen sowie Constraints (z.B. Kardinalitäten)
- bessere Voraussetzung für Integrität und Wiederverwendung definierter Elemente
- Namensräume für die Definition von Klassen
- Bewertung
 - deutlich bessere Möglichkeit, die Semantik von Dokumenten auszudrücken
 - deshalb besser geeignet für Dokumente mit hohem Integritätsbedarf
 - Handhabung allerdings deutlich aufwändiger
 - Investitionsschutz: weitere Nutzung von DTDs ohne großen Aufwand möglich
 - Transformation in flache XML-Schema-Dokumente
 - Nutzung der Vorteile von XML-Schema dann aber nur eingeschränkt möglich

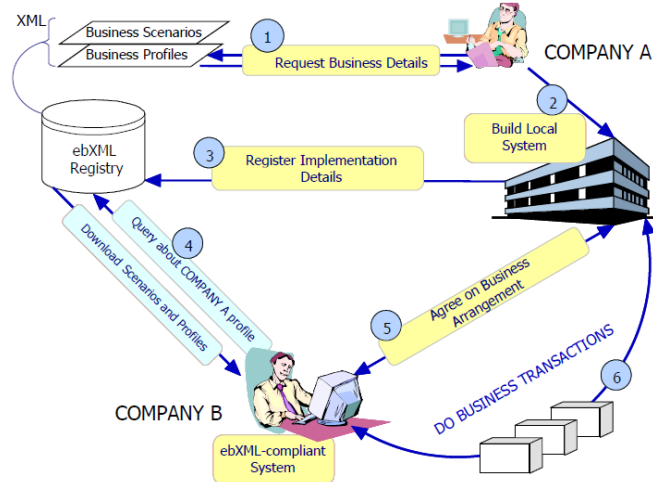
XML – Beurteilung

- unterstützt die Trennung von Inhalt und Layout
- verschiedene Ansichten leicht realisierbar
- aber: Layout nicht völlig unabhängig vom Inhalt zu ändern (oder umgekehrt) → da das Layout auf den Inhalt referenziert
- Metasprache – deshalb große Einsatzbandbreite
- allerdings keine Datentypen bei DTDs, deshalb eingeschränkte Möglichkeiten, Semantik auszudrücken und Integrität sicher zu stellen
- geringe Semantik der Dokumente fördert allerdings Wiederverwendung
- wirtschaftlich reizvoll durch Standardisierung
 - viele kostengünstige Werkzeuge, umfangreiches Know-How verfügbar
- aber: Standardisierung gilt nur für Metasprache, Dokumenttypen nicht standardisiert
- mitunter unseriöse Überzeichnung der Bedeutung von XML empfiehlt kritische Distanz

ebXML

- gefördert von OASIS und der UN/CEFACT
- Gegenstand und Ziel
 - unternehmensübergreifende Geschäftsprozesse, die durch den Austausch elektronischer Dokumente realisiert werden
 - Unterstützung der Anbahnung, Gestaltung und Durchführung solcher Geschäftsprozesse
 - Bereitstellung geeigneter Infrastruktur und Standards

Modell der Anwendung von ebXML



Szenarios

- dienen dazu, typische Einsatzszenarien von ebXML zu beschreiben
- sollen die Realisierung von ebXML-Infrastrukturen unterstützen
- Bsp: 2 Handelsunternehmen schließen einen Vertrag und starten den Austausch

Unterstützung unternehmensübergreifender Geschäftsprozesse

- Bereitstellung standardisierter Konzepte in zentralem Verzeichnis („registry“)
 - Geschäftsprozessstypen
 - Nachrichtentypen/Dokumententypen
- Ablage individueller Profile in Registry
- Registry unterstützt Suchfunktionen
- strukturierte Vereinbarungen unter Handelspartnern auf der Basis individueller Profile
- Durchführung geschäftlicher Transaktionen durch den Austausch von XML-Dokumenten auf der Basis von Vereinbarungen

Geschäftsprozessstyp

- beinhaltet Beschreibungen eines unternehmensübergreifenden Prozesses, vor allem der darin enthaltenen Transaktionen sowie der beteiligten Rollen
- sollte als XML-Dokument erfasst sein
- graphische Darstellung möglich (z.B. use case diagram, activity diagram)
- Transaktion definiert als Austausch von Geschäftsdokumenten nach einem bestimmten Protokoll
- Bsp: Durchführung einer Bestellung

```
<ProcessSpecification name="Simple" version="1.1" uuid="[1234-5678-901234]">
  <!-- Business Documents -->
  <BusinessDocument name="Catalog Request"/>
  <BusinessDocument name="Catalog"/>
  <BusinessDocument name="Purchase Order"/>
  <BusinessDocument name="Invoice"/>
  <BusinessDocument name="Payment"/>
  <BusinessDocument name="Inventory Report Request"/>
  <BusinessDocument name="Inventory Report"/>
  <Package name="Ordering">
    ...
  </Package>
</ProcessSpecification>
```

benötigte Nachrichtentypen

Darstellung einer Transaktion mit Fokus auf beteiligte Rollen

```
<BinaryCollaboration name="Fulfillment Payment">
  <InitiatingRole name="payee"/>
  <RespondingRole name="payor"/>
  <BusinessTransactionActivity name="Process Payment"
    businessTransaction="Process Payment"
    fromAuthorizedRole="payee"
    toAuthorizedRole="payor"/>
</BinaryCollaboration>
```

Nachrichtentyp

- in Form von XML-DTDs abgelegt
- dienen dazu, die Syntax des auszutauschenden Geschäftsdokumente zu definieren
- intendierte Semantik wird natürlichsprachlich beschrieben

- Profile
 - jede teilnehmende Organisation legt ihr Profil, das sog. Collaboration Protocol Profile (CPP) in der Registry ab
 - ein CPP beinhaltet die unterstützten Geschäftsprozesse, die Fähigkeit die enthaltenen Transaktionen durchzuführen sowie die benötigten bzw. unterstützten Schnittstellen
- Collaboration Protocol Agreement (CPA)
 - Vereinbarungen zwischen Partnern
 - dienen als Grundlage für die Durchführung von Transaktionen zwischen zwei oder mehr Partnern
 - legen unter Verweis auf die Profile der beteiligten Partner die jeweils involvierten Geschäftsprozesse und die darin abzudeckenden Transaktionen fest
 - legen die genaue, verbindliche Semantik der dazu ausgewählten Nachrichtentypen fest
- Beurteilung
 - hohe Flexibilität durch die Verwendung von Registries
 - durch Betonung von Geschäftsprozessen Einsatz für Anwender leichter nachzuvollziehen
 - durch graphische Modelle von Geschäftsprozessen (bisher nur rudimentär verfügbar) noch höhere Anschaulichkeit
 - relativ breite Unterstützung
 - aber: Spezifikation umfangreich und schwer verständlich

Austauschformate und Integration

- unterstützen wirksam die Kommunikation zwischen Anwendungen durch den Austausch von Nachrichten
- kostengünstiger Einsatz setzt Standardisierung und hinreichend offene Anwendungen voraus
- i.d.R. keine Referenzsemantik, deshalb Redundanz
- bei Verwendung von XML: Integrität der ausgetauschten Nachrichten eingeschränkt

Integration durch Workflow Management Systeme

Der Ausgangspunkt: Geschäftsprozesse

- zunehmende Bedeutung von Geschäftsprozessen für eine wettbewerbsorientierte Unternehmensorganisation
- informationstechnische Unterstützung von Geschäftsprozessen deshalb wünschenswert
- traditionelle betriebswirtschaftliche Systeme allerdings i.d.R. auf Unternehmensfunktion ausgerichtet
- deshalb Bedarf an einer Technologie, die existierende Anwendungen prozessorientiert integriert
- Definition Geschäftsprozess: Ein Geschäftsprozess ist eine wiederkehrende Abfolge von Aktivitäten, die mehr oder weniger rigiden Regelungsmustern genügt. Er ist zielgerichtet und steht in einem direkten Zusammenhang mit der marktgerichteten Leistungserstellung eines Unternehmens. Solche Prozesse, die unmittelbar und in bedeutsamem Umfang zur Erzeugung von Erträgen führen und gleichzeitig wettbewerbsrelevante Kompetenzen eines Unternehmens eines Unternehmens fordern, bezeichnet man als Kernprozesse. Die Ausführung von Geschäftsprozessen erfordert den Einsatz knapper Ressourcen.
- Definition Geschäftsprozessstyp: Ein Geschäftsprozessstyp beschreibt eine Klasse gleichartiger Geschäftsprozesse.

- Definition Geschäftsprozessmodell: Ein Geschäftsprozessmodell ist eine zweckgerichtete Abstraktion eines Geschäftsprozesstyps, die häufig – aber nicht notwendig – mit einer grafischen Darstellung einhergeht.

Workflows

- Definition Workflow: Ein Workflow ist eine Abstraktion eines Geschäftsprozesses, die vor allem auf den Fluss digitalisierter Dokumente bzw. Objekte gerichtet ist. Menschliche Aktivitäten bzw. Entscheidungen im Rahmen eines Geschäftsprozesses werden dabei weitgehend ausgeklammert bzw. auf Interaktionen mit Anwendungssystemen reduziert.
- Definition Workflow-Modell: Ein Workflow-Modell stellt einen Workflow-Typ (als Klasse gleichartiger Workflows) dar. Dabei sollte die Darstellung den Anforderungen der letztlich beabsichtigten Automatisierung genügen und kann deshalb Aspekte enthalten, auf die in einem korrespondierenden Geschäftsprozessmodell verzichtet wird.

Workflow Management: Ziele aus Anwendersicht

- Steuerung von Workflows
 - Überwachung der Ausführungsregeln
 - Überwachung ordnungsgemäßer Terminierung
- Förderung der Integrität von Workflows
 - Vermeidung redundanter Aktivitäten
 - Vermeidung unzulässiger Zustände und Zustandsänderungen
- Förderung der Transparenz von Workflows
 - Darstellung der aktiven Workflows
 - Darstellung des jeweils aktiven Zustands
 - Darstellung der Workflow-Historie
- Randbedingung: Beibehaltung existierender, im Prozess benutzter Anwendungen
- Workflow Management System (WFMS): Ein System zur Beschreibung, Ausführung und Kontrolle eines Workflows unter Verwendung vorhandener Software und ggfs. eigener Anwendungsteile.

Voraussetzungen einer automatisierten Steuerung

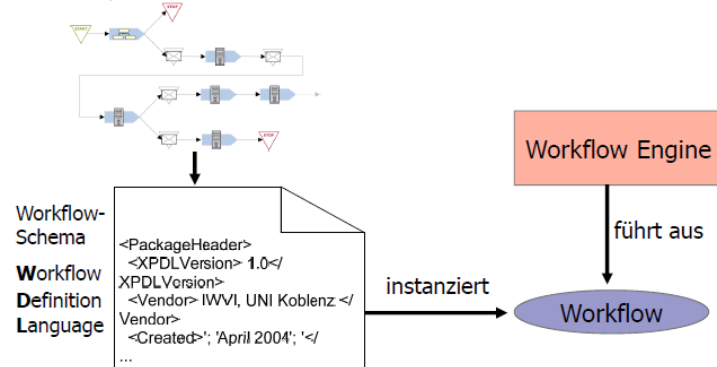
- formalisierte Ausführungsregeln (Kontrollfluss)
 - Reihenfolge der Aktivitäten
 - Integritätsbedingungen
- Erkennen relevanter Ereignisse in den Anwendungen
 - Zustandsveränderungen
 - Bearbeitungsschritte abgeschlossen
 - Verstöße gegen Integritätsbedingungen
- Ausführung von Funktionen der Anwendungen
- Kenntnis der verwendeten Dokument- bzw. Datenstrukturen
- Dienste zur Autorisierung und Benachrichtigung von Benutzern

Workflow Schema

- Spezifikation eines Workflow-Typs mittels einer Workflow-Spezifikationssprache
- Instanzen des Schemas sind Repräsentationen konkreter Workflows
- die Workflow-Engine des WFMS überwacht Ausführung nach Maßgabe des jeweiligen Schemas

Vom Prozessmodell zum Workflow-Schema

Geschäftsprozessmodell



Analogie zur Datenbanktechnologie:
aus Datenmodell wird SQL gemacht,
daraus können Instanzen erstellt
werden

- Modelle gut zur Analyse und Kommunikation, der Rest läuft weitgehend transparent automatisch ab

Qualität der Integration

- die Steuerungsmöglichkeiten und damit die Potentiale für dynamische Integration steigen
 - je detaillierter die verwendeten Dokumente und Daten im Schema spezifiziert sind
 - je mehr relevante Ereignisse vom WFMS erfasst werden können
 - je mehr Funktionen der Anwendungen vom WFMS ausgeführt werden können
 - gemeinsamer Namensraum (2 Ansätze: gleiche Datenbank mit eindeutigen Primärschlüsseln oder über Middleware-Systeme mit einem zentralen Verzeichnis)

Erkennen von Ereignissen

- idealtypisch: generiert durch Anwendung → von WFMS feststellbar
 - Voraussetzung für Steuerung/Integritätskontrolle auch innerhalb der Anwendung
 - aber: i.d.R. nicht verfügbar
- Alternative: indirekte Feststellung von Ereignissen
 - Anlegen, Löschen, Ändern persistenter Daten
 - Zugriff über Dateiverwaltungssystem oder DBMS
 - differenzierte Behandlung von Ereignissen setzt Spezifikation von Datenstrukturen im Schema voraus (und entsprechende Zugriffsmöglichkeiten)

Ausführen von Funktionen

- idealtypisch: Zugriff auf alle Funktionen (und Daten) der beteiligten Anwendungen (z.B. API)
 - erlaubt Steuerung/Integritätskontrolle auch innerhalb der Anwendung
 - aber: differenzierte Schnittstellen i.d.R. nicht verfügbar
- Alternative: Zugriff auf Funktionen der Infrastruktur
 - Starten und ggfs. Terminieren von Anwendungen über Betriebssystem
 - Erzeugen von Dokumenten bzw. Daten über Dateiverwaltungssystem oder DBMS
- Problem: Schnittstellen/APIs können sich im Zeitverlauf ändern

Benachrichtigungsdienst

- zur Durchführung eines Workflows ist häufig eine Benachrichtigung der Benutzer erforderlich, z.B. Information, Aufforderung
- verschiedene Realisierungen über WFMS-Client, E-Mail oder Anwendungen (API)

Standards

- Standardisierung im Kontext von WFMS von großer Bedeutung
 - Investitionsschutz, Wiederverwendbarkeit (vergleichbar mit DBMS)
- einige Initiativen: XPD, WS-BPEL, JBoss jBPM

WFMS – Beurteilung

- Unterstützung von Geschäftsprozessen stellt hohes Nutzenpotential in Aussicht
- Fokus auf prozessorientierte Integration – orthogonal zur datenorientierten (bzw. objektorientierten) Integration (Datenintegration als Voraussetzung für Prozessintegration)
- Architektur geeignet, Wartung zu unterstützen (Anpassung der Modelle)
- gegenwärtig zusätzlicher Auftrieb durch Web Services
- Basis für prozessorientiertes Controlling
 - Unterstützung von Prozesskostenrechnung
 - automatisierte Erfassung von Kennzahlen (Qualität, Laufzeiten, ...)
- Standardisierung von Workflow-Spezifikationssprachen
 - fördert Investitionsschutz
 - beschränkt sich allerdings auf einen Teil der Gesamtfunktionalität
- aber: gegenwärtige WFMS erlauben nur eine eingeschränkte Integration
 - keine Berücksichtigung von Integritätsbedingungen innerhalb von Webanwendungen
 - keine Berücksichtigung von Ereignissen, die innerhalb von Anwendungen generiert werden
 - hängt von der Offenheit der beteiligten Anwendungen/Komponenten ab

Herausforderung

- Gestaltung eines brauchbaren Kompromisses zwischen kontrollierten Prozessen und individueller Freiheit bei der Beschreibung von Abläufen
- komfortable Konfiguration durch Anwender
- Funktionsfähigkeit in heterogenen Hardware-/Betriebssystem-Umgebungen
- Integration existierender Software
- Spezifikation und Verwaltung langer Transaktionen