

Einstieg in das Thema

Integration, Interoperabilität, Standardisierung, Konversion

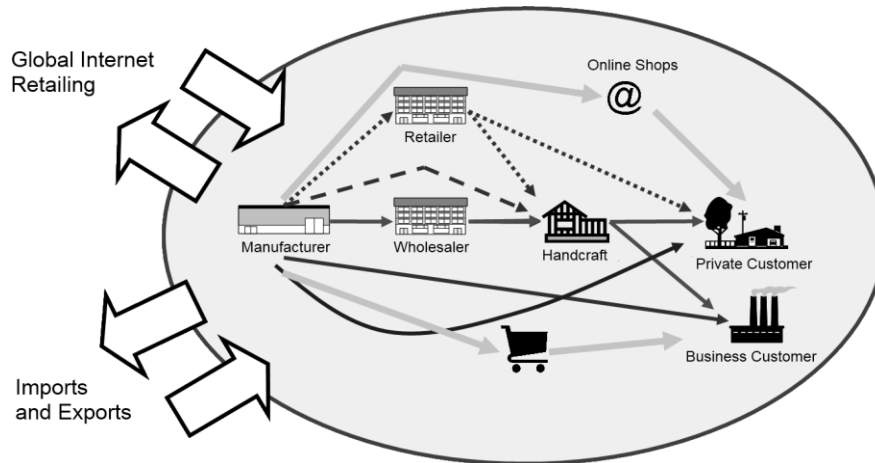
- Integration bezeichnet (Wieder)Herstellung einer Einheit, d.h. Teile, die eigentlich zusammengehören, aber voneinander getrennt sind, werden wieder zusammengefügt
- betriebswirtschaftliche Gründe
 - Arbeitsteilung, Spezialisierung, Differenzierung
 - Konzernbildung, dezentrale vs. zentrale Aufgabenerfüllung
 - Fusionen, Diversifizierung, Übernahmen, Outsourcing
 - Netzwerke (SCM), Kooperationen, Wertschöpfungsketten
- informationstechnische Gründe
 - Verbinden getrennter IS (entstehen z.T. zeitlich, Umstellen auf E-Business)
 - IS entstehen zeitlich versetzt (versch. Strukturen, Ressourcen, Alt-Systeme)
 - ungeeignete Lastverteilung, Redundanzen, mehr Ausfallsicherheit
 - Modularisierung, Objektorientierung, Flexibilität, Offenheit
 - sinkende Netzwerkkosten, bewährte Technik

Integrationsansätze beim St. Gallener Modell

- die IS-Architektur ist die Basis für die Integration der Applikationen und Daten, erhöht die Datenkonsistenz und die Wartbarkeit der Systeme
 - Analyse der Daten-, Funktions-, Organisations- und Kommunikationsmodelle
- Integrationsbereiche realisieren Synergiepotentiale durch die Informationsverarbeitung
 - Synergiepotentiale finden sich dort, wo hohe Interdependenzen bestehen zwischen Geschäftsfunktionen eines oder mehrerer Unternehmen
 - Synergiepotenziale primär in den Querschnittsfunktionen, wie Logistik, E-Business, ReWe, Kooperationen
 - in der Wertkette, aufgrund gemeinsamer Ressourcennutzung, rechtlicher Gründe, gemeinsamer Märkte, gemeinsamer Konkurrenz (gemeinsame Marktbeobachtung), informativischer Verflechtungen (Vertrieb arbeitet mit den gleichen Kundeninformationssystemen)
- Integrationsmechanismen
 - Zusammenfassung von Geschäftsfunktionen, Bildung eines Planungs- und Kontrollsystems, Bildung von Berichtswegen und -verteilern, rechtliche Strukturierung des Unternehmens (Töchter, Holding, ...), Bildung zentraler Funktionen (z.B. bei Matrix-Organisationen), Projekt mit Vertretern verschiedener Geschäftsbereiche, Bildung von Gremien und Ausschüssen
- potentielle Nutzen von Integration
 - (Transaktions-)Kostensenkung, schnellere Prozesse, Zusatznutzen für Kunden, erhöhte Flexibilität, stärkere Kundenbindung, erhöhte Differenzierung von der Konkurrenz, Aufbau von Markteintrittsbarrieren, bessere und aktuellere Informationen, geringeres Marktlieferantenrisiko durch bessere Prognosen, bessere Kooperationen in Supply-Chains
- potenzielle Probleme von Integration
 - einmaliger Aufwand, macht Prozessanalysen nötig, führt zu Abhängigkeiten und zur Offenlegung von Schnittstellen, erfordert gegenseitige Abstimmung und Vertrauen, verändert bestehende Strukturen und Machtbereiche
 - im E-Business nötig: Konzepte und Systeme, die überbetrieblich, offen und kompatibel sind; Informationsmanagement für Supply Chains und kooperative Systeme; intelligente Konzepte zur Standardisierung und Konvertierung von Informationen zwischen Unternehmen und Führungssystemen, Integration von Daten und Funktionen

zwischenbetriebliche und globale Interoperabilität (IOP) im E-Business

- Interoperabilität zwischen der Unternehmensvision (ICT, Strategien, Organisation, Prozesse) mit den strategischen (Ziele, Organisationstypen, Prozesse, Ressourcen) und operativen (Märkte, Aufträge, Geschäfte, Beziehungen) Bereichen
- Geschäftsbeziehungen entlang der Supply Chain



- morphologische Box
 - soziale Dimension: politische Einordnung, geographische Ausrichtung, Sektor, organisatorische Grenzen, Koordination (Markt, Hierarchie), Datenverantwortlichkeit (zentralisiert, delegiert, verteilt, unklar)
 - kommunikative Dimension: Semiotik (Syntax, Semantik, Pragmatik), Kommunikationspattern (B2B, A2N, ...), Anzahl der Partner, Wichtigkeit der Datenqualität
 - physikalische Dimension: Interoperationsmechanismus (horizontal, vertikal, Ring, Peer-To-Peer, Stern), Interoperationstypen

Grundlagen- Standards vs. Konversion

Begriff Standardisierung (Fokus E-Business)

- Standardisierung als Ergebnis: logische und technische Vereinheitlichung
 - Standardisierung als Prozess der Entwicklung und Durchsetzung eines Standards (= Spezifikation)
 - Standardisierungsarten
 - marktliche Standardisierung: Unternehmen setzen defacto-Standards durch ihre Marktmacht
 - de jure-Standardisierung: übergeordnete, öffentlich-rechtliche Instanz definiert Standard (national: ANSI, DIN; europäisch: CEN, CENELEC, Ecma International; weltweit: ISO, IEC, ITU)
 - hybride Standardisierung: Konsortium/Komitee organisiert den Verhandlungsprozess zwischen unterschiedlichen Interessensgruppen
 - Anzahl Verbindungen ohne Standard: $\#Kunden * \#Lieferanten$
 - Anzahl Verbindungen mit Standard: $\#Kunden + \#Lieferanten$
 - Kommunikationsschnittstellen zwischen Lieferant und Kunde
 - Artikelkatalog, Anfrage, Rückfrage, Bestellung, Lieferung, Rechnung, Zahlung
 - historischer Verlauf der Standards
 - 80er: Prozessintegration/Insellösungen durch Bereichsstandards
 - frühe 90er: firmenweite Integration/Standard-SW durch Firmenstandards
 - Mitte 90er: konzernweite Integration/Standard-SW durch Konzernstandards
 - heute: integrierte B2B-GPs durch E-Business-Standards
- Kommunikation zw. Unternehmen ermöglichen, GPs automatisieren, gemeinsame Sprache

informationelle Sicht auf die Integration

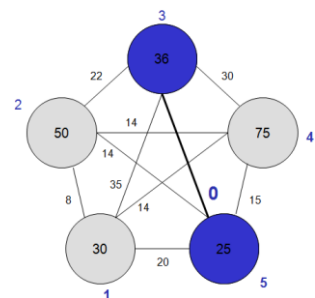
- nach Integrationsrichtung
 - zwischen Unternehmen (Supply-Chain-Sicht)
 - horizontal: zwischen Supply Chains (z.B. zwischen Großhändlern)
 - vertikal: innerhalb einer Supply Chain (Lieferant, Produzent, Großhandel, Kunde)
 - innerhalb eines Unternehmens
 - horizontal: zwischen Funktionsbereichen (Beschaffung, Produktion, Absatz, Transport)
 - vertikal: operative, administrative und strategische Bereiche integrieren
- nach Integrationsobjekten
 - Datenintegration
 - physische Datenintegration: nur eine Datenhaltungskomponente (Consolidation)
 - logische Datenintegration: ein gemeinsames Datenbankschema
 - Teil-Kopien an die Nachfrager (Propagation)
 - föderale Datenbeschaffung: virtuelle Zentralisierung (Federation)
 - morphologischer Kasten: Reichweite, Richtung, Lese-/Schreibrichtung, Schreibmodus, Datengranularität, Umfang der Datenhistorie, Erhebungsdauer, Metadatenverfügbarkeit
 - Funktionsintegration
 - Ziele: gemeinsam genutzte, umfassende Anwendungssysteme; Mehrfachverwendung gleicher Anwendungssysteme; Herstellung einer einheitlichen GUI
 - Lösungen: Integrationsplattformen (.NET, J2EE, CORBA)
 - Prozessintegration
 - Ziele: anwendungsübergreifende GPs unterstützen; Prozessbeschleunigung, -vereinfachung und -automatisierung
 - Lösungen: übergeordnete Steuerungsebene; Verwaltung von Workflow-Schemata, -Instanzen und Akteuren; Arbeitskorb-Metapher
- nach Integrationsreichweite
 - bereichsbezogen, innerbetrieblich/bereichsübergreifend, zwischenbetrieblich branchenintern/branchenübergreifend

ökonomische Sicht auf die Integration

- Standards verringern Integrationskosten
 - Netzeffekte: mit der Anzahl der Nutzer eines Standards steigt der einzel- und gesamtwirtschaftliche Nutzen
 - De facto-Standards
 - Unternehmen streben danach, mit diesen Monopolrenten zu erzielen
 - zusätzlich entstehen Wechselkosten für den Nutzer beim Wechsel → Lock in-Effekt
 - Standards als öffentliche Güter
 - stellen funktionierende Problemlösungen dar
 - nicht de facto-Standards erleichtern den Umstieg
 - erhöhen den Wettbewerb zwischen Lösungsanbietern
- die Ökonomie eines Standards bestimmt der Verbreitungsgrad, nicht nur der Standardisierungsprozess

systemischer Modellansatz zur Standardisierung

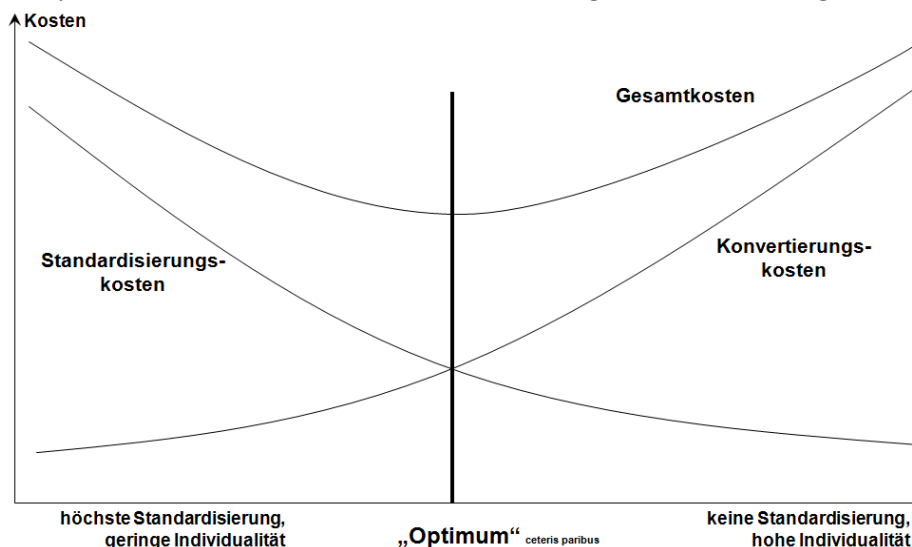
- Informationssystem als Menge von Systemelementen
- Beziehungen zwischen Systemelementen als Informationsaustausch



- Gesamtkosten = knotenbez. Kosten + kantenbez. Kosten
- Modellerweiterungen
 - zeitliche Dimension
 - einperiodig, einstufig: Vor- und Nachteile der Standardisierung sind zeitunabhängig
 - mehrperiodig, einstufig: die Vor- und Nachteile fallen zu verschiedenen Zeitpunkten an, die Entscheidung ist aber nur einmalig zu Beginn des Planungshorizontes zu treffen; die erwarteten Zahlungen/Nutzen sind zu diskontieren
 - mehrperiodig, mehrstufig: wie 2., die Entscheidungen sind jedoch für alle Perioden des Planungshorizontes zu treffen
 - sachliche Dimension
 - Entscheidung über alternative Standards möglich
- Lösung des Standardisierungsproblems: Methoden der Kombinatorik, ganzzahlige lineare Programmierung
- Problemfelder Modellbildung: Erfolg/Nutzen hängt ab von der Reihenfolge der Knoten, die den Standard übernehmen; Risiken, ob sich der Standard durchsetzt; Wechselkosten; Gesamt- vs. Knotenerfolg; Netzgesamtkosten als Zielfunktion (realistisch?)
- Modellkritik, -einschränkungen
 - Datenermittlung ex ante schwierig
 - ungerichtete Kanten unterstellen bilateralen Kommunikationsbedarf
 - hier entscheidet zentrale Instanz über alle Systemelemente
 - in der Realität Wettbewerb zwischen Standards, Nutzen abhängig von den Entscheidungen anderer Wirtschaftssubjekte und Informationsasymmetrien zwischen den Akteuren

Standardisierung vs. Konvertierung

- Standardisierung: *alle* einigen sich auf *einen* Standard (idealtypisch)
- Konvertierung: Übersetzung erforderlich, trotz Standards bleiben Unterschiede
- potentielle Nachteile/Risiken von Standards
 - Verlust der Individualität
 - Lock-in-Effekt
 - mangelnde Möglichkeit oder Wahrnehmung der Einflussnahme auf Standards
 - Standardisierungsgeschwindigkeit, -zyklen, Anpassungsbedarf, -aufwand
- Schlussfolgerungen
 - Analyse des Trade-Offs zwischen Standardisierung und Konvertierung



- Kosten der Standardisierung
 - einzelbetriebliche Standardisierungskosten: Kosten für die Unterstützung des jeweiligen Standards, Nutzungsgebühren, Einführungskosten (Personal, HW, SW), lfd. Anpassungskosten, Standardwechselkosten
 - (außerbetriebliche) Kosten für die Standardentwicklung: Anforderungsanalyse und -definition, Spezifikation, Bereitstellung, Services, Wartung, Weiterentwicklung, Management des Standardisierungsprozesses, Verbreitung und Marketing
 - gesamtwirtschaftliche Kosten: Summe der einzelbetrieblichen Kosten plus Standardentwicklungskosten, u.a. abhängig von der Komplexität der S-Objekte und des S-Prozesses, der Anwendungsdomäne und der internationalen Verbreitung → letztendlich von der Nutzeranzahl

Konvertierung

- Konvertierungskosten
 - Analyse des Konvertierungsbedarfs, Erstellung des Konverters, Ausführung des Konvertierungsprozesses, Kosten aufgrund unzulänglicher Konvertierungsergebnisse, Anpassungserfordernisse
- Konvertererstellung
 - Analyse des fachlichen Konvertierungsbedarfs und der zu konvertierenden Informationsobjekte, Make-or-Buy, ggf. Entwicklung einer Software, Implementierung, Schulung, Optimierung der Nutzung
- Netzwerktopologien
 - vollständig vermaschte Struktur, Ringstruktur, Sternstruktur, Hybridstruktur

Kriterium	Peer-to-Peer	Ring	Stern
Anzahl Konverter	bis zu $n(n-1)$ Konverter	n Konverter	nur $2n$ Konverter, aber Entwicklung des Referenzinformationsobjektyps
Konvertierungsprozesse	immer nur ein Prozess, schnell, geringe Kosten	bis zu $n-1$ Konvertierungen abhängig von Entfernung zwischen 2 Knoten	immer zwei Konvertierungsschritte
Konvertierungsfehler	Fehler bleiben bilateral begrenzt	Informationsverluste/-fehler können sich addieren	Konvertierungsfehler unabhängig, entscheidend ist die Intermediärleistungsfähigkeit
Erweiterung um 1 Informationsobjektyp	$2n$ neue Konverter nötig, sehr aufwändig	2 neue Konverter nötig	2 neue Konverter nötig
Elimination von 1 Informationsobjektyps	problemlos	erfordert das Schließen der Lücke durch einen neuen Konverter	problemlos

Ansätze zur Verringerung der Standardisierungs-/Konvertierungskosten

- Ansätze zur Verringerung der Standardisierungskosten
 - schlankere Standardisierungsprozesse, leichtere Umsetzung/Anwendung von Standards, effektivere/betrieblich geeignetere Standards, Subventionierung, öffentliches Interesse an der globalen/nationalen Senkung der Integrationskosten

- Ansätze zur Verringerung der Konvertierungskosten
 - leistungsfähigere Technologien für die Konvertierung, einfachere Entwicklung, geringere Prozesszeiten/-kosten, „Meta-Standards“ als Orientierung
- potentielle Folgen: Verringerung der Gesamtkosten, Optimum verschiebt sich nach rechts, höhere Individualität möglich

Aussagen zur zwischenbetrieblichen Integration/Standardisierung

- Objekte: immer geschäftsbezogener Datenaustausch, fachliche Inhalte wichtig
- Kommunikation: M2M-Kommunikation, Automatisierung
- fachliche Standards
 - setzen i.d.R. andere, technische Standards voraus
 - schnelle, insbesondere marktliche Entwicklung ab 1998
 - seit 2001: Konsolidierung, aber auch Spezialisierung
- Unterschiede zur innerbetrieblichen Integration/Standardisierung
 - andere, eher marktliche Koordinationsmechanismen
 - größere Anzahl potenzieller Nutzer
 - geringer Integrations-/Standardisierungsgrad
 - jede Kommunikation erfordert gemeinsame Basiskonzepte und Terminologie/Fachsprachen
→ ZBIS ist komplexer, vielschichtiger, aufwändiger

Probleme von E-Business-Standards

- bilaterale Abstimmung ist schwierig, unzureichende Modelle zur Abstimmung der Inhalte, geringe Ausnutzung der Standards, Standards für komplexe Produkte und Dienstleistungen fehlen noch, EB-Standards gehen an der Realität in den Unternehmen vorbei, es gibt (noch) zu viele EB-Standards, die EB-Standards müssen miteinander stärker integriert werden

Standardisierung im E-Business

Ebenen Modell

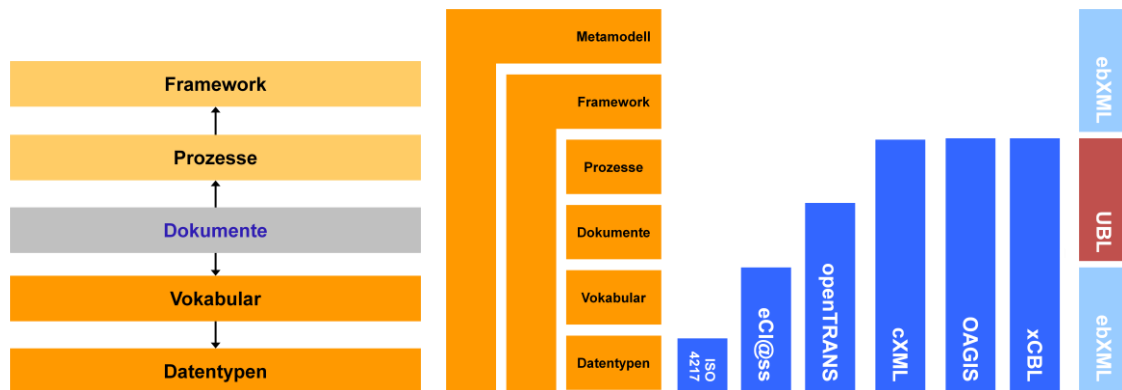
Standardisierungsobjekte

- fachlich-inhaltliche Standards: unterstützen Informations- und Leistungsaustausch (unser Fokus)
- technologische Standards: bieten technologische Infrastruktur (z.B. TCP/IP, XML Web Services)

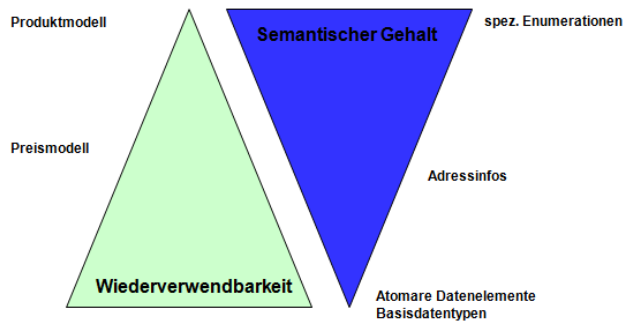
Standardisierungs-Ansatz: Schichten-Modell

- Vorgehen beim Schichten-Modell
 - Problem-Analyse – Lösungsweg → Lösungssequen
 - Zerlegung in Teillösungen → Zuordnung zu Ebenen
 - geordnete Verbindungen der Lösungen der Ebenen → Protokolle
 - EB-Standards oberhalb technischer Standards (8. Ebene im ISO/OSI-Modell)
- Ablauf-Organisation im Schichten-Modell
 - jede Schicht übernimmt definierte Aufgaben
 - untere Schichten stellen Dienste den höheren Schichten bereit
 - höhere Schichten nutzen Dienste der unteren Schichten
- Lösungen und Standards im Schichten-Modell
 - unterschiedliche Lösungen/Standards je Sicht
 - Austauschbarkeit von Lösungen/Standards der gleichen Sicht

BLI-Schichten-Modell zum Vergleich von E-Business-Standards



- Entwicklung aus 2 Richtungen: Aufbau der Dokumente (nach unten), Rolle der Dokumente in GPs (nach oben)
- ebenenweise Abstimmungsbedarf zwischen verschiedenen Firmen, falls unterschiedliche Datentypen, Vokabulare etc. eingesetzt werden
- eCl@ss ist ein Klassifikationsstandard, Vokabular wird zum Beschreiben der Eigenschaften benötigt, Rückgriff auf ISO-Datentypen
- Datentypen
 - Aufgaben
 - dienen der Codierung von Informationen *für Maschinen*
 - definieren zulässige Werte/Wertebereiche von Datenelementen
 - notwendig für jede Datenverarbeitung (Programmiersprachen, DBMS, Datenaustausch)
 - Unterschiede hinsichtlich Anzahl, Leistungsfähigkeit und Strukturierung
 - Basisdatentypen: Ganz-/Fließkommazahlen, Zeichenketten, Boolesche Werte, Zeit
 - Einschränkungen/Erweiterungen von Basisdatentypen
 - Aufzählungsdantentypen (Enumerationen)
 - Elemente standardisierter Aufzählungsdantentypen
 - Sprachen: Bezeichner, Texte, ... alle sprachabhängigen Datenelemente
 - Länder/Regionen: Adresse, Herkunftsland, Verfügbarkeitsgebiet, ...
 - Währungen: Preise, Versandkosten, Zollgebühren, Wertangaben, ...
 - Einheiten: Bestelleinheit, Preiseinheit, Verpackungseinheit, ... (Einheiten für Logistik anders als für den Verkauf, geliefert wird eine Palette, verkauft nur einzelne Kästen)
- Vokabular
 - Definition: Menge aller zulässigen Datenelemente (= Wortschatz)
 - atomare Datenelemente
 - kleinste Informationseinheit, nicht weiter zerlegbar
 - besitzt einen Datentyp (z.B. Preis mit Datentyp Währung)
 - komplexe Datenelemente
 - setzen sich aus mehreren atomaren oder komplexen Datenelementen zusammen (z.B. Preis als Betrag, Währung, Gültigkeit, Preistyp, Sonderpreis)
 - durch Hierarchisierung wird logischer Zusammenhang ausgedrückt
 - Definition/Standardisierung des Vokabulars
 - ist der fachliche Kern jedes EB-Standards
 - konsistent, redundanzfrei, zeitlich stabil, ...
 - Modellierungsaufgabe (Namenskonventionen, Wiederverwendung, Modularisierung, Generalisierung/Spezialisierung)



- EB-Vokabular vs. Referenzmodell
 - Referenzmodell: höhere Abstraktion als bei unternehmensspezifischen Modellen, praktische Anwendung durch Anpassung/Konfiguration/Erweiterung
 - EB-Vokabular: ist *nicht* Bezugsrahmen für die Entwicklung spezifischer Vokabulare, sondern ist im Gegenteil verbindliche Vorgabe (also Anpassung etc. vermeiden!)
- Vokabular für Lieferbedingungen: ShipmentMethodOfPaymentCoded, TransportDescription
- Dokumente
 - fassen logisch zusammenhängende Daten des Vokabulars zu Nachrichten zusammen
 - Dokumente sind Instanzen von Dokumenttypen
 - jeder Dokumenttyp ...
 - besitzt definierten Anwendungszweck (Rechnung als Aufforderung zum Saldo-Ausgleich)
 - bestimmt die Rolle der Kommunikationspartner (Rechnungsersteller/-empfänger)
 - bestimmt die Richtung der Kommunikation (unidirektional) (Auftrag an Auftragnehmer)
 - Transaktion: Übertragung eines Dokuments → Transaktionsstandards
 - Definition/Standards von Dokumenttypen
 - 1) Auswahl von Datenelementen der Vokabularschicht: Festlegung der Dokumenthierarchie, der Datenelementreihenfolgen, der Kardinalitäten der Items im Dokument
 - 2) Zweckgerichtete Einschränkung des Vokabulars: auf zulässige Wertebereiche einschränken, Datenelemente konkretisieren (z.B. PLZ des Lieferanten)
 - 3) Wiederverwendung durch strukturanaloge Dokumenttypen: gleiche Dokumentstrukturen für unterschiedliche Dokumenttypen (z.B. Kopf-/Detail-/Positionsdaten, Signatur)
 - Dokumenttypen: Invoice, ShippingSchedule, OrderConfirmation, PriceCheckRequest
- Prozesse
 - Definition: Folge von einer oder mehreren Transaktionen zwischen *zwei* Geschäftspartnern (*nicht* stets wechselseitig, explizite Prozessdefinition in Form eines Prozessmodells)



- Framework
 - Rahmenwerk für die Geschäftsdokumentation
 - für den sicheren, verlässlichen und geordneten Austausch von Geschäftsdokumenten
 - Unabhängigkeit von den fachlichen Inhalten (Envelope)
 - Festlegung von
 - Kommunikationsprotokollen
 - Sicherheit: Authentifizierung und Verschlüsselung
 - Message Handling (Warteschlangenverwaltung, Ausnahmebehandlung, Bestätigungen)

- Metamodell
 - beschreibt die Objekte und Beziehungen der anderen Sichten
 - erlaubt die Konstruktion von EB-Standards (Einhaltung syntaktischer Regeln, Integrationsfähigkeit unterschiedlicher Standards auf Basis des gleichen Metamodells)
 - Elemente
 - Schichten, reflexive Bedingungen
 - Zuordnung von Objekten zu Schichten
 - Datentypen: Spezialisierungshierarchie
 - Vokabular: atomare vs. komplexe Datenelemente, Mehrfachverwendung, Zuweisung von Datentypen
 - Dokumente: auch Spezialisierung, Zuweisung von Datenelementen, ...
 - ...

Datenaustauschformate

- Definition
 - Art und Weise der Repräsentation von Daten
 - Anordnung der Datenelemente und -werte (Serialisierung)
 - Kodierung der Werte gemäß Datentypen und Zeichensätzen
 - Serialisierung ist Voraussetzung für Speicherung und Übertragung
- CSV-Formate
 - Merkmale
 - identisch strukturierte Datensätze in einer Liste
 - Datenwerte werden durch ein definiertes Trennzeichen separiert
 - Variante: kein Trennzeichen, Datenwerte in Spalten fester Breite anordnen
 - Dateiaufbau:
 - 1. Zeile: Nennung aller Datenelemente in der Reihenfolge ihres Auftretens
 - nachfolgende Zeilen: Inhalte der Datensätze
 - Serialisierung einer Tabelle
 - 1-n Beziehungen über Fremdschlüssel abbilden, jede Tabelle eine eigene CSV-Datei
 - Erweiterung der CSV-Struktur um IDs (PositionsNr1, PositionsNr2)
 - Datensatztypisierung: unterschiedlich zu interpretierende Datensätze (Typ1 = Auftrag, Typ2 = Auftragsposition, Typ3 = Artikel) mit generischen Datenelementen (z.B. Wert)
 - Bewertung
 - Vorteile: einfach zu erzeugen (insb. aus Dateiverwaltungssystemen), einfach zu verarbeiten, minimale Übertragungsgröße
 - Nachteile: komplexe Datenstrukturen nur schwer abzubilden, Formatspezifikation unzureichend (nur Reihenfolgen, keine Datentypen, keine formale Formatspezifikation)
- EDI-Formate
 - komplexe, hierarchische Strukturen
 - Abgrenzung unterschiedlich strukturierter Datensätze (Segmente) durch spezifische Codes
 - Abgrenzung der Datenelemente eines Datensatzes durch spezifische Codes
 - Bewertung
 - Vorteile: beliebig komplexe Datenstrukturen abbildbar, geringe Übertragungsgröße, erste minimale Auszeichnung von Datenwerten
 - Nachteile: Erzeugung und Verarbeitung erfordert spezielle Schnittstellen, keine formale weiterverarbeitbare Formatspezifikation

- XML-Formate
 - Meta-Sprache zur Definition von Sprachen
 - Auszeichnung der Datenwerte durch Tags (nicht vordefiniert, Definition eigener Tags)
 - Datei beginnt mit Prolog (Version, Zeichensatz, Dokumenttyp), dann folgt Wurzelement mit dem verwendeten Namensraum
 - Trennung von Inhalt, Struktur (Benennung von Tags und Formalisierung durch DTD) und Darstellung (CSS)
 - gleiche Semantik kann unterschiedlich modelliert werden (z.B. Attribute statt Elemente)
 - Anforderungen an XML-Dokumente
 - Wohlgeformtheit: Dokument enthält mindestens ein Element, korrekte Verschachtelungen und Schließung der Tags, alle Attribute mit Anführungszeichen
 - Gültigkeit: Dokument ist wohlgeformt und entspricht dem zugehörigen Schema
 - Reihenfolge der Elemente ist relevant (bei Attributen nicht)
 - Verarbeitung
 - über CSS als HTML-Seite oder über Importschnittstellen ins DBMS (für B2B relevant)
 - DTD-Sprachelemente
 - e^* (beliebig), e^+ (≥ 1), $e?$ (0 oder 1), $e_1 | e_2$ (Alternative), e_1, e_2 (Konkatenation)
 - #REQUIRED, #IMPLIED (optional), #FIXED „Wert“ (Konstante), „Wert“ (Default)
 - PCDATA/CDATA: Vorwärts-Deklaration von Element-Typ-Definitionen
 - Bewertung XML DTD
 - Syntax: kontextfreie Grammatik in abgewandelter Backus-Naur-Form, nicht XML-basiert und somit nicht mit XML-Werkzeugen zu verarbeiten
 - Datentypen: nur PCDATA als zu parsende Zeichenkette, keine echten Datentypen
 - XML Schema als deutlich mächtiger Alternative
 - Bewertung XML-Formate (strukturell, nicht inhaltlich)
 - Vorteile: beliebig komplexe Datenstrukturen abbildbar, formale weiterverarbeitbare Formatspezifikation, Unterstützung durch Datenverwaltungssysteme, Auszeichnung der Datenwerte
 - Nachteile: große Übertragungsgrößen, hierarchisches Datenmodell
- Zusammenfassung

	CSV	EDI	XML
Datenmodell	satzorientiert	satzorientiert	hierarchisch
Strukturkomplexität	gering	hoch	hoch
Dateien je Nachricht	z.T. mehrere notwendig	1	1
Selbstbeschreibung	nein	eingeschränkt	ja
Übertragungsgröße	minimal	gering	hoch
formale Spezifikation	nein	nein	ja
Werkzeugunterstützung	hoch	gering	mittel-hoch

Organisation

Organisation von Standards

- potenzielle Bereiche: Kodierung, Identifikation (ISBN), Klassifikation (eCI@ss), Datenaustausch (BMEcat), Prozesse (ebXML, UBL)

- Standardisierungsprozesse der einzelnen Organisationen sind auch standardisiert
- Vergleichskriterien für Standards (hier Organisation)
 - Aufbauorganisation: SDO/Konsortium/Einzelunternehmen, national/international, horizontal/vertikal (= branchenspezifisch)
 - Ablauforganisation (Standardisierungsprozess): Standardisierung, Formalisierung, Transparenz, Demokratie, Offenheit
 - Nutzung des Standards: Kosten, einmalig/laufend, inner-/überbetriebliche Nutzung
 - Dienstleistungen, Anwenderunterstützung: Support, Beispiele, Richtlinien, SW, Community
- Zertifizierung (Prüfung auf Einhaltung von Standards)
 - Geschäftsdokumente
 - Überprüfung gemäß Spezifikation
 - syntaktische Überprüfung auf Basis der Dokumenttypdefinitionen
 - semantische Überprüfung auf Basis inhaltlicher Anforderungen
 - SW-Produkte
 - Feststellen des Abdeckungsgrades
 - Überprüfung erzeugter Geschäftsdokumente und von Importfunktionen

Integration mittels Transaktionsstandards im E-Business

- Transaktion = Übertragung eines elektronischen Geschäftsdokuments
- Transaktionsdaten = Nicht-Stammdaten, z.B. Auftragsabwicklung
- Unterschiede zum Stammdatenaustausch (z.B. Produktdaten)
 - Inhalte sind zeitkritisch, Request-Response-Kommunikation, höhere Anzahl Dokumente je Periode, weniger komplexe Dokumenttypen, kleinere Dokumente

Vergleich der Transaktionsstandards

	cXML	xCBL	openTRANS	OAG
Entwickler	Einzelunternehmen	Einzelunternehmen	Konsortium	Konsortium
Herkunft/Ausrichtung	USA	USA	D	USA, International
Standardisierungsprozess: Transparenz	nein	nein	gering	ja
Standardisierungsprozess: Formalisierung	nein	nein	gering	ja
Mitgliedschaft: offen	nein	nein	(ja)	ja
Mitgliedschaft: Gebühren	-	-	(ja)	ja

Inhalte der Transaktionsstandards

- Überprüfung der inhaltlichen Abdeckung erfordert ...
 - detaillierte Untersuchung der Dokumenttypen und ihrer Datenelemente
 - Kenntnis der Anforderungen an diese Dokumenttypen
- Inhalte gemäß Schichtenmodell
 - Prozesse: Anzahl definierter Prozesse
 - Dokumenttypen: Anzahl unterschiedlicher Dokumenttypen
 - Vokabular: Datenmodelle für Kopf-, Transaktions- und Positionsdaten
 - Datentypen: Wiederverwendung standardisierter Datentypen

formale Spezifikation von Transaktionsstandards

- Welche XML-Schemasprachen? Wie werden die verfügbaren Modellierungskonzepte genutzt?
- Anforderungen an XML-Schemasprachen
 - Modularisierung der Spezifikation unterstützen, Datentypen/Elemente/Attribute definieren, Vererbung unterstützen, Datenkonsistenz sichern, Dokumentation unterstützen

Dokumentation von Transaktionsstandards

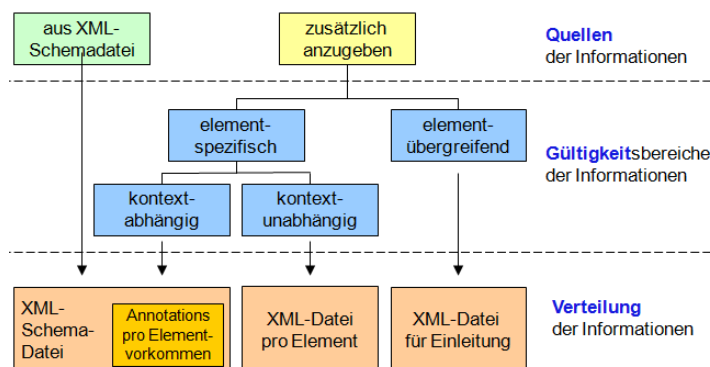
- XML-basierte EBS bestehen aus
 - Spezifikation: maschinenverarbeitbar, in XSD
 - Dokumentation: für menschliche Aufgabenträger, größtenteils Texte (idealerweise zielgruppenspezifisch, also multi-lingual)
- Aufgaben: Beschreibung der Semantik, Hinweise zur Architektur/Implementierung
- Instrumente, Kriterien
 - Dokumentationsformate, Prozessmodelle/grafische Schema-Repräsentationen (UML, ER, XML), strukturierte Datenelementverzeichnisse, Zielgruppendifferenzierung, Mehrsprachigkeit, integrierte Beispiele, integrierte formale Spezifikation, Fallstudien, etc.
- Dokumentationsformate
 - Kriterien: Navigation, Suche, Plattformunabhängigkeit, Druck-, Weiterverarbeitungsfähigkeit
 - Formate: PDF, DOC, HTML, kompilierte HTML-Hilfdatei, XML

Ansatz für eine bedarfsgerechte EB-Dokumentation (BLI-Entwurf)

- Qualitäts- und Effizienzsteigerung durch enge Verbindung von Dokumentation und Spezifikation
- Ergänzung der Spezifikation um ggf. fehlende Inhalte
- direkte Generierung der Dokumentation auf Basis der Spezifikation
- einheitlich angewandte technologische Basis (XML Schema, XSLT)
 - Datenbereitstellung durch XML Schema-Annotationen, d.h. einer rein formalen Repräsentation *aller* relevanten Informationen
 - Generator auf Basis von XSLT
- Analogien zu bewährten SE-Konzepten
 - Dokumentationsgenerierung vgl. Programmierdokumentation
 - Einsatz von Annotated Schema vgl. XML-Dateien in relationalen DBs

3 Schritte zu einer bedarfsgerechten Dokumentation

- 1) Ermittlung des Informationsbedarfs, d.h. was ist alles formal zu repräsentieren
- 2) Identifikation der Informationsquellen, d.h. welche Informationen sind bereits elektronisch verfügbar und welche müssen neu hinzugefügt werden
- 3) Festlegung der Informationsverteilung, d.h. wo werden die Informationen redundanzfrei gespeichert



Bewertung Dokumentationskonzept BMEcat

- Aufwand für Qualitätssicherung wurde deutlich reduziert durch Vermeidung struktureller Inkonsistenzen und Effizienzsteigerung durch Einsatz von Reports
- Entwickler des Standards konnten sich auf die fachliche Spezifikation und die inhaltliche Beschreibung konzentrieren
- Vorteile der verteilten Datenhaltung: Wiederverwendung der Elementbeschreibungen, zeitgleich verteiltes Arbeiten, Aufteilung in logische Module
- zeitnahes Feedback durch schnelle automatisierte Generierung von Zwischenversionen
- Umsetzung des Konzeptes konnte sowohl die Qualität der entwickelten EB-Standards als auch die Qualität des zugrunde liegenden Entwicklungsprozesses erhöhen
- es fehlt noch eine benutzerfreundliche GUI und ein Templatekonzept

Probleme bei der Bewertung von Standards

- wie wird die Verbreitung gemessen, welche Unternehmen werden betrachtet, welche Dokumenttypen sind relevant

abschließende Bewertung

	cXML	xCBL	openTRANS	OAG
Organisation	-	-	+	-
Inhalte	o	o	+	+
formale Spezifikation	+	-	+	o
Dokumentation	+	-	o	o
Markstellung (Deutschland)	-	o	-	-

Integration im E-Business

Schema Mapping

Datentransformation

- EB-Standards verringern die Unterschiede und erleichtern die Kommunikation
- aber Konkurrenz der EB-Standards, Standardnutzung verändert nicht per se die innerbetriebl. IS
- daher Konvertierung zwischen Standards bzw. Standards und innerbetrieblicher IS
- das Basiskonzept der Konvertierung sind sogenannte Datentransformationen
- Aufgabe
 - Daten von einem Datenformat in ein anderes überführen
 - nicht den Informationsgehalt, sondern die Darstellung/Kodierung ändern
- Anwendungen
 - Schema-Integration: mehrere relationale Schemata in ein gemeinsames zusammenführen
 - Migration von Datenbeständen aus Legacy-Systemen
 - Aufbau eines DWH
- Entwicklung von Datentransformationen → Mapping
 - View-Konzept als Ersatz für Schema-Integration
 - Schema-Matching: korrespondierende Schemata-Elemente in Beziehung bringen
 - Ergebnis: Mappings, Transformationsanweisungen

Anwendungsbereiche

- unterschiedliche Schemata
 - unterschiedliche Anzahl und Struktur der Datenelemente
 - unterschiedliche Benennung der Datenelemente
 - unterschiedliche Datentypen korrespondierender Datenelemente
- unterschiedliche Serialisierungsformen
 - CSV-, EDI-, XML-basierte Datenformate
 - Problem: Serialisierung bei verschiedenen Datenmodellen
- unterschiedliche Datenmodelle
 - relationales vs. objektorientiertes vs. hierarchisches Datenmodell
 - Transformationsprobleme: Vererbung, Rekursion, reflexive Beziehungen, Primärschlüssel ...
- Konvertierungsfälle
 - Datenimport: EB-Standard → Schema des importierenden IS
 - Datenexport: Schema des exportierenden IS → EB-Standard
 - direkte Konvertierung: EB-Standard 1 → EB-Standard 2
- Mapping erfolgt auf Schema-/Typ-Ebene (z.B. ERM ↔ XML Modell) und kann auf der Instanzen-Ebene veranschaulicht und umgesetzt werden

Typisierung (nach Anzahl der Datenelemente)

		Zielformat	
		ein Datenelement	Mehrere Datenelemente
Quellformat	ein Datenelement	1:1-Mapping	1:N-Mapping
	Mehrere Datenelemente	N:1-Mapping	N:M-Mapping

- 1:1-Mapping
 - Anforderungen des XML-Datenmodells
 - Unterscheidung von Tags und XML-Attributen
 - Berücksichtigung der Dokumenthierarchie (Hierarchiestufe, Container)
 - Parameter des 1:1-Mappings
 - Umfang des Datenelements im Quell- und Zielformat (Tag oder Tag + Attribut)
 - Benennung, Kodierung und Hierarchiestufe der Elemente (identisch oder verschieden)
- 1:N-Mapping
 - Operation: Verteilung
 - höhere Detaillierung im Zielformat, Formulierung einer Zerlegungsvorschrift
 - Operation: Hinzufügung
 - Belegung von Datenelementen im Zielformat mit Werten, die nicht explizit im Quellformat vorhanden sind (z.B. ergänzende Währungsangaben, Sprachen, Preistypen)
- N:1-Mapping
 - Operation: Konkatenation
 - Werte werden verknüpft, ggf. Änderung der Reihenfolge, Kodierung
 - Operation: Aggregation
 - (Neu-)Berechnung des numerischen Zielwerts, z.B. Nettoendpreise
 - Operation: Selektion
 - nur eine Teilmenge der Quelldatenelemente wird herangezogen

- N:M-Mapping
 - komplexe Operation
 - kann sich aus Verteilung, Hinzufügen, Konkatenation, Selektion und Aggregation zusammensetzen
 - eine Mapping-Seite nimmt Bezug auf unterschiedliche Objekttypen, d.h. nicht nur Preise-zu-Preise, Merkmale-zu-Merkmale etc.
 - zwischen den Objekttypen bestehen inhaltliche Beziehungen
 - XML-Datenformate modellieren diese Beziehungen unterschiedlich, insbes. durch Dokumenthierarchie und Container-Bildung

Schemaunterschiede

- Datentypen
 - Problem: unterschiedliche Datentypen im Quell- und Zielsystem (Wertveränderungen erforderlich), insbesondere bei standardisierten vs. nicht-standardisierten Enumerationen
 - Lösungen: Zeichenkettenoperationen, Berechnungsvorschriften, Mapping von Werten zu Enumerationsdatentypen (z.B. „Deutschland“, „D“ und „BRD“ zu „DE“)
- partitionierte Datenformate
 - Mapping erstreckt sich über mehrere Quell- und/oder Zielformate
 - Zusammenhang erkennen und berücksichtigen (Datenelemente auf Primär- und Fremdschlüssel prüfen)
 - z.B. Verarbeitung mehrerer CSV-Dateien zu einer XML-Datei
- instanzabhängiges Mapping
 - Zuordnungen können nicht mehr allgemeingültig beschrieben werden, da sie von den Werten der Quelldatenelemente (Instanzen) abhängen
 - erst bei der Ausführung der Datentransformation kann über die richtige Zuordnung entschieden werden
 - Fall 1: Operationen werden instanzabhängig angewendet (z.B. wenn „Tax“ > 0, dann „steuerpflichtig“ = true)
 - Fall 2: Mapping-Kardinalität ist instanzabhängig (z.B. <organame type="???">??</organame> kann auf „suppliername“ oder „buyername“ gematcht werden, Inhalt bestimmt schließlich das Zielelement)

Berücksichtigung unterschiedlicher Informationsgehalte

- bisher
 - Quell- und Zielschemata unterscheiden sich *strukturell*
 - der durch sie abgebildete Informationsgehalt ist nahezu identisch
- aber
 - Informationsgehalte von EB-Standards sind selten deckungsgleich
 - unterschiedliche Schwerpunkt/Branchenanforderungen
 - unterschiedliche Mindestanforderungen an Dokumente (Pflichtfelder)

- mögliche Folgen für Konvertierung
 - Informationen gehen verloren
→ Informations-Overhead/-Verluste
 - Informationen fehlen
→ Informationsdefizite

		Datenelemente im Zielformat		
		Pflicht	Optional	nicht vorhanden
Datenelemente im Quellformat	Pflicht	Mapping	Mapping	Informations-(verluste) overhead
	Optional	Informations-defizit	Mapping	Informations-(verluste) overhead
	nicht vorhanden	Informations-defizit	-	-

- Ursachen für Informations-Verluste
 - fehlende Datenelemente im Zielformat: es existiert kein korrespondierendes Datenelement im Zielformat
 - geringer Datenelement-Kardinalität im Zielformat: während das Datenelement im Zielformat nur einmal auftreten kann, hat es im Quellformat die Kardinalität N
 - unzureichender Wertebereich im Zielformat: Wertebereich im Quellformat > als Zielformat
- Informationsverluste zeigen, dass das Zielformat inhaltlich nicht den Einsatzanforderungen entspricht
- Informationsverluste sind daher bei Auswahlentscheidungen von EB-Standards auf jeden Fall zu untersuchen
- Informationsdefizite
 - Problem: die im Quellformat vorhandenen Informationen sind nicht ausreichend, um inhaltlich genügende oder wenigstens gültige XML-Zieldokumente zu erstellen, Konvertierung muss jedoch alle Pflichtdatenelemente des Zielformats mit Werten belegen
 - Lösungsansatz: die fehlenden Informationen manuell oder automatisiert hinzufügen
 - Einfügung von Dummy-Werten
 - dienen nur der Herstellung der Dokumentgültigkeit, Werte ohne Informationsgehalt
 - Werte dürfen nicht zur Prozesssteuerung verwendet oder Benutzern präsentiert werden
 - Einfügung von Vorgabewerten
 - sind für einen Zeitraum für eine Organisation gleich bleibende Werte
 - werden im exportierenden System verwaltet (z.B. für Currency, Language, Territory)
 - Einfügung individueller Werte
 - während der Dokumentkonvertierung werden benutzerseitig fehlende Werte ergänzt
 - aber nicht geeignet für produktbezogene Stamm- oder zeitkritische/viele Transaktionsdaten

Entwicklung von Datentransformationen

- bei Anwendung auf EB-Standards
 - Analyse der formalen Spezifikation & der Dokumentation, Hinzuziehung von Domänenwissen
 - auch: Dokumentinstanzen
- Entwicklungsreihenfolge
 - mindestens alle Pflicht-Ziel-/Quelldatenelemente
 - auch: inhaltliche Kriterien
- Dokumentation der Datentransformation
 - Aufgaben: formale Spezifikation als Grundlage für SE, Fortschrittskontrolle, Versionsanpassung bei EB-Standards
 - Mapping-Tabellen: Gegenüberstellung der Dokumenttypstrukturen, Typisierung und Beschreibung jeder Datentransformation
- automatisierte Entwicklung von Datentransformationen
 - Input: formale Spezifikation (Datenelemente, Datenstruktur), (Dokument-)Instanzen
 - Instrumente: Termvergleiche, Taxonomien, Synonyme/Dictionary für Datenelemente, manuelles Pre- und Post-Matching, lernende Verfahren
 - aber: vergleichsweise geringe Konkretheit

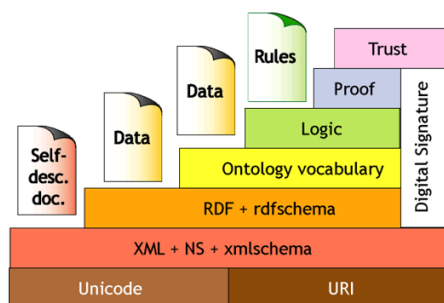
Ontologien – semantische Integration

Ontologien in der Informatik

- Definitionen
 - explizite Spezifikation einer Konzeptualisierung
 - gemeinsames Verständnis einer betrachteten Domäne
- Ontologien sind Artefakte
 - sie bestehen aus einem Vokabular, um Sachverhalte zu beschreiben
 - sie enthalten eine Menge von expliziten Annahmen über die beabsichtigte Bedeutung des Vokabulars
- Ontologien beschreiben Domänenwissen möglichst explizit, formal, maschinenverarbeitbar
- verwandte Begriffe
 - Taxonomien: Einteilung von Elementen in ein Klassifikationssystem anhand ihrer Beziehungen zueinander
 - Thesauri: zusätzliche Beziehungstypen „synonym“ und „ähnlich“
 - TopicMaps: Beziehungstypen, Verknüpfung mit Dokumenten, für Navigation/Visualisierung

XML ist für die Wissensrepräsentation unzureichend

- Erweiterbarkeit: verlangt nach vorheriger Vereinbarung des gemeinsamen Vokabulars
- keine explizite, formale, maschinenverarbeitbare Semantik: Semantik allein durch Dokumentstruktur-/hierarchie, Semantik und Struktur sind ineinander verwoben
→ einsetzbar prinzipiell nur für enge Kooperationsszenarien
- keine is-a-Beziehungen
- keine Vererbungshierarchien in XSD
 - keine Mehrfachvererbung
 - unzureichende Top-down-Attributvererbung (type-Konstrukt kann Oberklasse *entweder* erweitern *oder* einschränken)
 - keine Bottom-up-Instanzvererbung (z.B. Instanzen von „Beamter“ sind nicht zugleich Instanzen von Person“)
- XML ist nur ein Baustein in der semantischen Pyramide



Resource Description Framework (RDF)

- beschreibt Web-Ressourcen durch Metadaten
- Aussagen als Triplet: Subject → Predicate → Object
- XML-Syntax, Graph
- RDF Schema (RDFS) definiert das Vokabular für RDF
 - Vokabular als typisierte Hierarchie (Class, subclassOf, type, Property, domain, range)
- Vorteile: Trennung von Syntax und Semantik (unabhängig von Dokumentstruktur, sehr einfach zu verarbeiten), standardisiertes Vokabular für Metadaten
- Nachteile: kein Inferenzmodell, keine Regeln

