

Anforderungen an Informationsobjekttypen als Basis von Architekturentscheidungen bei der Datenintegration

Reinhard Jung

Universität Bern, Institut für Wirtschaftsinformatik,
Engehaldenstrasse 8, 3012 Bern, Schweiz,
reinhard.jung@iwi.unibe.ch

Zusammenfassung. Für die Datenintegration sind inzwischen viele Architekturkonzepte und Technologien verfügbar, und es stellt sich die Frage, auf welcher Basis eine Architektur- und Technologieauswahl erfolgen sollte. Der Beitrag stellt aus fachlicher Perspektive (Sicht der Aufgabenträger) eine Reihe von nicht-funktionalen und funktionalen Anforderungen im Zusammenhang mit Informationsobjekttypen vor, aus denen sich potenziell technische Anforderungen an Datenintegrationsarchitekturen ableiten lassen.

1 Problemstellung

Der Wandel von Geschäftsmodellen und eine immer dynamischer werdende Produktpolitik zwingt die Unternehmen, in immer kürzeren Zeitabständen neue Softwarelösungen in Betrieb zu nehmen. In analoger Form gilt diese Feststellung auch für Behörden, da sich auch ihr Leistungsportfolio bei sinkenden Budgets ständig erweitert oder zumindest stark verändert.

Aus der geschilderten Situation ergibt sich die Herausforderung, neue Anwendungssysteme mit den bestehenden integrieren zu müssen, um die Effektivität des betrieblichen Informationssystems aufrecht zu erhalten oder herzustellen. Ein zentraler Schritt eines solchen Integrationsvorhabens ist die Integration der Daten aller beteiligten Systeme (Datenintegration).

Prinzipiell stehen eine ganze Reihe von Architekturkonzepten zur Verfügung, die für eine Integration grundsätzlich geeignet sind. Beispiele sind so unterschiedliche Konzepte wie Data-Warehouse-Systeme, Operational-Data-Stores und serviceorientierte Architekturen. Darüber hinaus ist auch im Bereich der Technologien inzwischen eine grosse Menge an Optionen verfügbar, beispielsweise Middleware-Produkte und -Standards, Integrations- oder Message-Broker und Web Services. Bevor eine konkrete Architektur gestaltet und eine Entscheidung für bestimmte Technologien und Software-Lösungen gefällt werden kann, muss allerdings eine genaue Spezifikation der Anforderungen erfolgen, denn viele Architekturkonzepte sind nicht universell einsetzbar. Ein Data-Warehouse-System beispielsweise ist – zumindest in der „klassischen“ Sichtweise dieses Architekturkonzepts (eine grundsätzlich andere Sichtweise wird vertreten bspw. in [15]) – aufgrund der Charakteristika von Datenextraktions- und Konsolidierungsprozessen nicht in der Lage, Echtzeit-Daten bereit-

zustellen oder etwa die Modifikation von Data-Warehouse-Daten in die operativen Anwendungssysteme zu propagieren. Der vorliegende Beitrag widmet sich der Frage, wie sich Anforderungen an Architekturen für die Datenintegration aus fachlicher Sicht und damit aus Sicht des Informationsbedarfs beschreiben lassen. Ferner wird diskutiert, inwieweit sich diese Merkmale potenziell auf die Auswahl einer technisch geeigneten Architektur für die Datenintegration auswirken.

2 Konzeptioneller Rahmen und Begriffe

Als Basis der weiteren Ausführungen wird hier zunächst ein konzeptioneller Rahmen beschrieben, der im Wesentlichen aus einer Basishypothese und einigen zentralen Begriffsdefinitionen besteht.

Für Zwecke des vorliegenden Beitrags ist zunächst festzulegen, worin das (Entwicklungs-)Ziel liegt: Es sei gegeben durch die Implementierung einer Datenintegrationsarchitektur. Der Zweck dieser Architektur liegt darin, den Aufgabenträgern innerhalb eines Untersuchungsbereichs einen effizienten Zugriff auf die Daten zu ermöglichen, die sie für ihre Aufgabenerfüllung benötigen. Dabei ist davon auszugehen (Basishypothese), dass ein effizienter Zugriff auf diese Daten mit Hilfe der vorhandenen Anwendungssysteme („legacy“) bzw. der vorhandenen Anwendungsarchitektur nicht möglich ist. Da eine vollständige Ersetzung der Anwendungsarchitektur aus Gründen des Investitionsschutzes nicht realistisch erscheint, wird statt dessen angestrebt, die Anwendungsarchitektur mit Blick auf den genannten Zweck zu ergänzen. Die Ergänzung besteht in so genannten Integrationskomponenten, die zusammen mit der Anwendungsarchitektur die Datenintegrationsarchitektur bilden (vgl. **Abb. 1**).

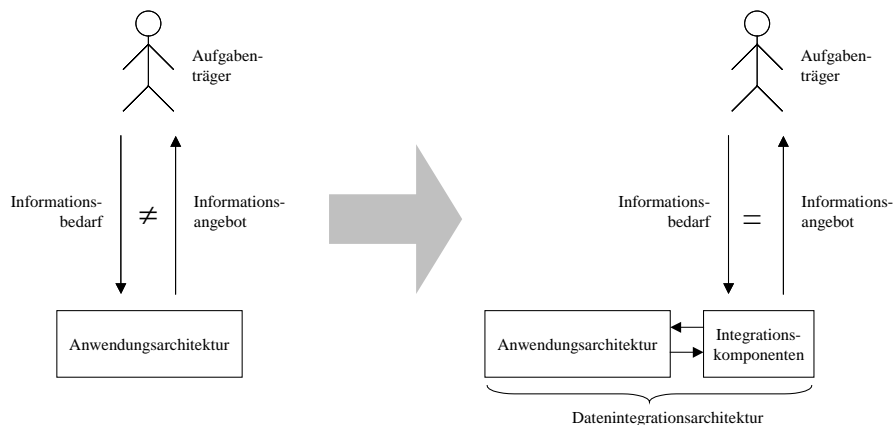


Abb. 1. Ausgangssituation beim Entwurf einer Datenintegrationsarchitektur

Zu konkretisieren ist der in der Abbildung verwendete Begriff „Information“. Die computergestützte Verarbeitung von Daten basiert auf Datenelementen (z.B. „Meier“, „400“) oder allgemein auf Datenelementtypen (z.B. Kundenname, Bestellmenge). Aus Sicht einer betrieblichen Transaktion sind die Datenelementtypen zu Datenobjekttypen zusammengefasst (z.B. Kunde, Bestellung), wobei Datenobjekttypen zuein-

ander in Beziehung stehen können; in Entity-Relationship-Modellen entspricht dieser Zusammenhang den Beziehungstypen, die Entitätstypen miteinander verbinden.

Die Sicht eines Aufgabenträgers orientiert sich nicht an der Menge der vorhandenen Datenelemente, sondern an dem Bedarf, der sich aus der Aufgabenausführung für ihn subjektiv ergibt. Im Vordergrund stehen also zweckorientierte Daten, die man entsprechend als Information bezeichnen kann. Auch die an Transaktionen orientierte Zusammenfassung der Daten entspricht nicht der Sichtweise eines Aufgabenträgers. Ein Beispiel ist die Analyse einer Verkaufsstatistik, in der miteinander in Beziehung stehende Ausprägungen von Informationselementtypen (z.B. Region, Verkaufsmenge, Verkaufspreis, Kundenname) enthalten sind. Ein verallgemeinertes Konzept, das als Ganzes den Gegenstand einer Aufgabe darstellt, wird im Folgenden als Informationsobjekttyp bezeichnet (z.B. eine Verkaufsstatistik). In **Abb. 2** ist der Zusammenhang zwischen den Konzepten der fachlichen Sicht (Informationsobjekttyp und Informationselementtyp) und der technischen Sicht (Datenobjekttyp und Datenelementtyp) dargestellt. Die Notation ist so zu verstehen, dass beispielsweise ein Informationsobjekttyp aus mindestens zwei Informationselementtypen besteht, während ein Informationselementtyp in mindestens einem Informationsobjekttypen enthalten ist, aber auch in mehreren enthalten sein kann.

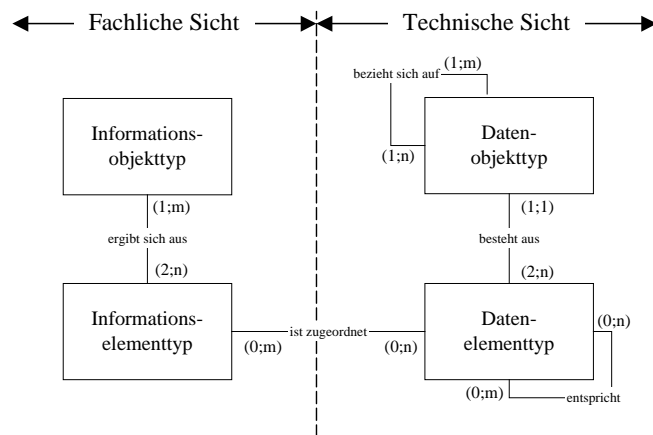


Abb. 2. Zusammenhang zwischen Informationsobjekt-/elementtypen und Datenobjekt-/elementtypen

Orientiert man sich an dem vorgestellten konzeptionellen Rahmen, so werden die Anforderungen eines Aufgabenträgers (fachliche Sicht) bezogen auf Informationsobjekte und -elemente oder allgemeiner auf der entsprechenden Typebene festgelegt. Aus technischer Sicht ist später zu prüfen, ob diese Anforderungen auf Basis der verfügbaren Datenelemente und -objekte erfüllbar sind bzw. welche Art von Datenintegrationsarchitektur zu realisieren ist.

3 Merkmale von Informationsobjekttypen und -elementtypen

Bereits eine Reihe von Autoren hat sich der Frage gewidmet, wie der Informationsbedarf eines Aufgabenträgers ermittelt werden kann, also wie die relevanten Informationsobjekttypen und -elementtypen zu erheben sind (vgl. z.B. [4], [10], [11], [13]). Als Ergänzung zu diesem inhaltlichen Aspekt von Informationsbedarfsanalysen wird hier die Fragestellung aufgeworfen, welche Merkmale der Informationsobjekt- und Informationselementtypen spezifiziert werden sollten. Als Anforderungen im Sinne dieses Beitrags werden (vom Aufgabenträger) gewünschte Ausprägungen dieser Merkmale aufgefasst.

In Anlehnung an das Requirements Engineering kann eine Klassifikation vorgenommen werden (siehe **Abb. 3**) einerseits in nicht-funktionale Merkmale, deren Ausprägungen die gewünschten Eigenschaften der Informationsobjekttypen und -elementtypen unabhängig von ihrer Verwendung (statische Sicht) repräsentieren. Eine zweite Klasse bilden andererseits die funktionalen Merkmale, deren Ausprägungen die gewünschten Eigenschaften der Informationsobjekttypen und -elementtypen bei der Verwendung (dynamische Sicht) abbilden.

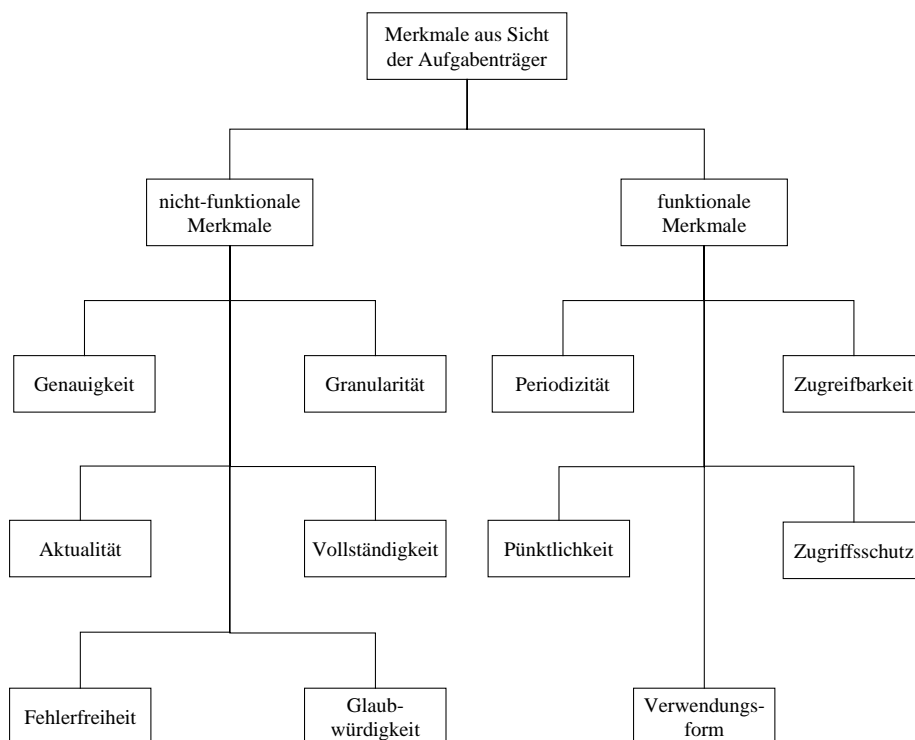


Abb. 3. Merkmale von Informationsobjekttypen und -elementtypen aus Sicht der Aufgabenträger

Die Herleitung der Merkmalsmenge kann hier nicht vollständig wiedergegeben werden (zu Details vgl. [5]). Sie basiert auf einer Recherche in der Literatur zur Daten-

qualität (vgl. [12], [14]) und in der betriebswirtschaftlichen Literatur (vgl. [1], [3], [6]). Bei den Merkmalen wird jeweils durch einen Buchstaben markiert, ob sich das jeweilige Merkmal auf Informationsobjekttypen („O“) und/oder -elementtypen („E“) bezieht.

3.1 Qualitative Anforderungen (nicht-funktionale Anforderungen)

Genauigkeit (O): Die Genauigkeit gibt an, wie detailliert Realweltobjekte durch Informationselemente beschrieben werden (vgl. [8] und auch [2]). Als Operationalisierung dieser Definition kann einerseits die Anzahl der für die Beschreibung verwendeten Informationselementtypen genutzt werden und andererseits auch das Skalenniveau dieser Typen; je höher das Skalenniveau desto grösser die Genauigkeit.

Aktualität (O/E): Die Aktualität drückt aus, inwieweit ein Informationselement die tatsächliche gegenwärtige Eigenschaft des zugrunde liegenden Realweltobjekts beschreibt.

Fehlerfreiheit (O/E): Ein Informationsobjekt ist bezogen auf einen bestimmten Zeitpunkt fehlerfrei, wenn die in ihm enthaltenen Informationselemente die Eigenschaften der zugehörigen Realweltobjekte in jenem Zeitpunkt widerspiegeln.

Glaubwürdigkeit (O/E): Unter der Glaubwürdigkeit lässt sich die Einschätzung subsumieren, ob bestimmte Informationsobjekte hinsichtlich ihrer übrigen Qualitätsmerkmale als geeignet eingestuft werden, ohne dass der betreffende Aufgabenträger für diese Merkmale die genauen Ausprägungen kennt.

Granularität (E): Die Granularität eines Informationselements drückt aus, ob und wie viele Einzelbeobachtungen in der Realwelt zusammengefasst wurden; in [2] wird für dieses Merkmal die Bezeichnung „Verdichtung“ verwendet. Von einer groben Granularität wäre beispielsweise im Fall eines Informationselementtyps „Jahresumsatz“ zu sprechen, der sich aus der Summierung von Einzelumsätzen ergibt.

Vollständigkeit (O): Die Vollständigkeit gibt an, inwieweit ein Informationsobjekt alle geforderten Informationselemente enthält. Im oben genannten Beispiel der Verkaufsstatistik wären fehlende Verkaufsmengen für bestimmte Regionen ein Indiz für Unvollständigkeit.

3.2 Funktionale Anforderungen

Periodizität (O): Im betrieblichen Berichtswesen werden drei Berichtsformen unterschieden, nämlich Standard-, Abweichungs- und Bedarfsberichte (vgl. [3]). Standardberichte sind regelmässig zu erstellen, und für Abweichungsberichte müssen in einem vorgegebenen Rhythmus Ist- mit Soll-Werten verglichen werden, sodass auch hier eine Regelmässigkeit zugrunde liegt. Bedarfsberichte werden hingegen in unregelmässigen Abständen von Aufgabenträgern angefordert. Auf die vorliegende Fragestellung übertragen, drückt die Periodizität aus, in welchem Rhythmus Informationsobjekte und -elemente von einem Aufgabenträger benötigt werden.

Pünktlichkeit (O): Die Pünktlichkeit setzt das Eintreffen eines Informationsobjekts beim Aufgabenträger (Zeitpunkt t_1) mit dem Zeitpunkt in Beziehung, zu dem das Eintreffen erwartet wurde (t_2). Liegt t_1 nach t_2 , dann ist von einer unpünktlichen Bereit-

stellung zu sprechen. Die Pünktlichkeit steht in enger Beziehung zur Antwortzeit bei der Anforderung eines Informationsobjekts, deren Länge die Pünktlichkeit massgeblich beeinflusst. Die Pünktlichkeit wurde als Merkmal vorgezogen, weil sie aus Sicht des Aufgabenträgers das wichtigere Merkmal ist.

Verwendungsform (O/E): Die Verwendungsform gibt an, welche Operationen auf den Informationsobjekten vorgesehen sind. Eine Verkaufsstatistik wird typischerweise für rein informative Zwecke angefordert (Verwendungsform „lesend“). Es sind darüber hinaus durchaus Szenarien denkbar, in denen Informationsobjekte auch modifiziert werden müssen (Verwendungsform „lesend/schreibend“), beispielsweise Budgetierungsprozesse.

Zugreifbarkeit (O): Die Zugreifbarkeit drückt aus, ob der Aufgabenträger die (vorhandenen) Informationsobjekte auch tatsächlich erhalten kann. Bei diesem Merkmal besteht ein enger Zusammenhang zur Pünktlichkeit; eine fehlende oder eingeschränkte Zugreifbarkeit führt unweigerlich zur Unpünktlichkeit.

Zugriffsschutz (O): Integrierte Daten haben häufig sensitiven Charakter. Insofern ist auch festzulegen, inwieweit sie vor unberechtigtem Zugriff zu schützen sind. Dieses Merkmal ist nicht nur im Sinne von Anforderungen aus Sicht des Aufgabenträgers festzulegen, sondern gegebenenfalls auch aus Sicht der Organisation, denn unter Umständen können oder dürfen nicht jedem Aufgabenträger die gewünschten Informationsobjekte zugänglich gemacht werden.

4 Implikationen aus technischer Sicht

Geht man davon aus, dass ein Aufgabenträger seinen Informationsbedarf spezifiziert und jeweils angegeben hat, welche Ausprägungen die qualitativen Merkmale der Informationsobjekte und -elemente jeweils aufweisen sollten, dann stellt sich die Frage, welche Konsequenzen sich daraus aus technischer Sicht ergeben. Grundsätzlich ist dabei zu berücksichtigen, dass

- der Informationsbedarf durch Integration mehrerer Datenquellen zu decken ist, nämlich der verschiedenen Datenquellen innerhalb der vorhandenen Anwendungsarchitektur (Anwendungs- und Datenverwaltungssysteme);
- diese Datenquellen möglicherweise autonom sind und daher in ihrem Antwortverhalten nur begrenzt oder gar nicht beeinflusst werden können.
- zwischen den Datenquellen bezogen auf einzelne Datenelemente Redundanz bestehen kann (in **Abb. 2** dargestellt durch den Beziehungstyp „entspricht“).

Im Folgenden wird auf die Merkmale fokussiert, die potenziell einen Einfluss auf die Architekturauswahl haben.

Antwortzeit und Zugreifbarkeit: Antwortzeit und Zugreifbarkeit sind die massgeblichen Bestimmungsgrössen der Pünktlichkeit. Sobald eine der Datenquellen eines Informationsobjekts autonomes Verhalten aufweist, verlängert sich die Antwortzeit bis hin zur Nicht-Zugreifbarkeit. Eine Kompensation dieses Effekts ist (bei unterstellter Nicht-Modifizierbarkeit des Autonomieverhaltens) nur durch eine zusätzliche, redundante Speicherung der betreffenden Datenelemente möglich, wie sie beispielsweise in Data-Warehouse-Systemen vorgesehen ist.

Aktualität: Nicht redundante Datenelemente geben die erreichbare Aktualität der Informationselemente vor. Bei redundanten Datenelementen stellt sich hingegen zu-

nächst die Frage nach ihrer Konsistenz, d.h. ihrer Widerspruchsfreiheit. Bei gegebener Konsistenz kann ein beliebiges der Datenelemente genutzt werden. Bei inkonsistenten redundanten Datenelementen stellt sich hingegen die Frage, welches den aktuellsten Wert repräsentiert; diese Frage lässt sich allerdings nicht mit Sicherheit beantworten. Eine Möglichkeit ist, das Datenelement zu verwenden, das den spätesten Aktualisierungszeitpunkt aufweist.

Vollständigkeit: Die Menge der verfügbaren Datenelemente ist durch den Datenbestand innerhalb der Anwendungsarchitektur typischerweise begrenzt. In der Konsequenz sind einerseits „alte“ Datenelemente und andererseits Datenelemente aus organisationsexternen Quellen nicht in der Anwendungsarchitektur verfügbar. Wenn also entsprechende Informationselemente nachgefragt werden, muss die Datenintegrationsarchitektur einen eigenen Datenspeicher beinhalten, der „alte“ Datenelemente archiviert und Datenelemente aus externen Datenquellen verfügbar hält.

Verwendungsform: Sofern Informationsobjekte nur lesend verwendet werden sollen, ergeben sich kaum einschränkende Anforderungen an eine Datenintegrationsarchitektur. Ein anderes Bild bietet sich, wenn schreibende Zugriffe auf Informationsobjekten vorzusehen sind. Redundanzen zwischen Datenelementen aus unterschiedlichen Datenquellen bedingen bei der erforderlichen Schemaintegration nicht kompensierbare Einschränkungen hinsichtlich der zulässigen Operationen. Ein Beispiel sind bestimmte Formen von Attributwertkonflikten (vgl. [7], [9]), die schreibende Transaktionen unmöglich machen. Ferner ist zu berücksichtigen, dass schreibende Zugriffe bei Redundanz und autonomen Teilsystemen zusätzliche Schwierigkeiten bedingen.

5 Ausblick

Die in diesem Beitrag vorgestellten Merkmale von Informationsobjekttypen und -elementtypen stellen den Ausgangspunkt einer Bewertung von Datenintegrationsarchitekturen dar. Weitere Forschungsarbeit ist erforderlich, um zum einen die einschränkenden Eigenschaften einer Anwendungsarchitektur zu identifizieren; erste Ergebnisse mit Blick auf die Anwendungs- und Datenverwaltungssysteme sowie die Kommunikationsverbindungen (beispielsweise Autonomieverhalten) finden sich in [5]. Zum anderen ist zu untersuchen, welche ökonomischen Konsequenzen einzelne Architekturkonzepte mit sich bringen.

Literatur

1. Alpar, P., Grob, H.L., Weimann, P., Winter, R.: Anwendungsorientierte Wirtschaftsinformatik. 2. Auflage. Vieweg, Braunschweig Wiesbaden (2000).
2. Bleicher, K.: Das Konzept Integriertes Management. Campus, Frankfurt New York (1991).
3. Horváth, P.: Controlling. 6. Auflage. Vahlen, München (1996).
4. IBM Corporation: Business Systems Planning - Information Systems Planning Guide. 4. Auflage. IBM-Form GE20-0527-4, IBM Corp., Atlanta (1984).
5. Jung, R.: Ableitung und Bewertung von Integrationsarchitekturtypen auf Basis fachkonzeptueller Anforderungen, Habilitationsschrift. Universität St.Gallen (2004).
6. Küpper, H.-U.: Controlling. 3. Auflage. Schäffer Poeschel, Stuttgart (2001).
7. Larson, J.A., Navathe, S.B., Elmasri, R.: A Theory of Attribute Equivalence in Databases with Application to Schema Integration. IEEE Transactions on Software Engineering 15 (1989) 4, 449-463.
8. Lehner, W.: Datenbanktechnologie für Data-Warehouse-Systeme. dpunkt, Heidelberg (2003).
9. Lim, E.-P., Srivastava, J., Shekhar, S.: An Evidential Reasoning Approach to Attribute Value Conflict Resolution in Database Integration. IEEE Transactions on Knowledge and Data Engineering 8 (1996) 5, 707-723.
10. Rockart, J.F.: Chief Executives Define Their Own Data Needs. Harvard Business Review 57 (1979) 2, 81-92.
11. Strauch, B.: Entwicklung einer Methode für die Informationsbedarfsanalyse im Data Warehousing, Dissertationsschrift. Difo-Druck, Bamberg (2002).
12. Strong, D.M., Lee, Y.W., Wang, R.Y.: Data Quality in Context. Communications of the ACM 49 (1997) 5, 103-110.
13. Walpoth, G.: Computergestützte Informationsbedarfsanalyse – Strategische Planung und Durchführung von Informatikprojekten. Physica, Heidelberg (1992).
14. Wand, Y., Wang, R.Y.: Anchoring Data Quality Dimensions in Ontological Foundations. Communications of the ACM 39 (1996) 11, 86-95.
15. Zeh, T.: Data Warehousing als Organisationskonzept des Datenmanagements – Eine kritische Betrachtung der Data-Warehouse-Definition von Inmon. Informatik – Forschung und Entwicklung 18 (2003), 32-38.