

Benlian, Alexander; Hess, Thomas

Working Paper

Identifikation und technische Bewertung von integrierten Datenverteilungsvarianten für eine effiziente Mehrfachnutzung multimedialer Medieninhalte

Arbeitsbericht, No. 3/2004

Provided in Cooperation with:

University of Munich, Munich School of Management, Institute for Information Systems and New Media

Suggested Citation: Benlian, Alexander; Hess, Thomas (2004) : Identifikation und technische Bewertung von integrierten Datenverteilungsvarianten für eine effiziente Mehrfachnutzung multimedialer Medieninhalte, Arbeitsbericht, No. 3/2004, Ludwig-Maximilians-Universität München, Institut für Wirtschaftsinformatik und Neue Medien (WIM), München

This Version is available at:

<https://hdl.handle.net/10419/60074>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WIM

Identifikation und technische Bewertung von integrierten Datenverteilungsvarianten für eine effiziente Mehrfachnutzung multimedialer Medieninhalte

Alexander Benlian / Thomas Hess

Arbeitsbericht Nr. 3/2004

Herausgeber: Prof. Dr. Thomas Hess

Institut für Wirtschaftsinformatik und Neue Medien
der Ludwig-Maximilians-Universität München
Ludwigstr. 28 VG, D-80539 München
Telefon: +49 89 2180-6390, Fax: +49 89 2180-13541
<http://www.wi.bwl.uni-muenchen.de>
E-Mail: wi-sekr@bwl.uni-muenchen.de

Inhaltverzeichnis

1 Einleitung	2
2 Die Integration multimedialer Daten als spezielles Integrationsproblem.....	2
2.1 Integration multimedialer Daten: Verortung innerhalb der klassischen Integrationsliteratur	2
2.1.1 Integrationsgegenstand	2
2.1.2 Integrationsrichtung	2
2.1.3 Integrationsreichweite	2
2.1.4 Automatisierungsgrad	2
2.1.5 Zusammenfassung: Einordnung der Content-Integration.....	2
2.2 Technische Anforderungen an die Mehrfachnutzung von Medieninhalten.....	2
2.2.1 Medieninhaltebezogene Anforderungen.....	2
2.2.2 Basisinfrastrukturelle Anforderungen.....	2
2.2.3 Zusammenfassung: Technische Anforderungen an eine Mehrfachnutzung von Medieninhalten.....	2
3 Integrierte Verteilungsvarianten von Content und deren empirische Ausprägungen	2
3.1 Datenverteilung: Eine Begriffsklärung	2
3.2 Integrierte Datenverteilungsvarianten.....	2
3.3 Empirisch beobachtbare Architekturausprägungen.....	2
4 Bewertung von Datenverteilungsvarianten aus technischer Sicht.....	2
4.1 Erfüllung medieninhaltebezogener Anforderungen	2
4.1.1 Zentrale vs. dezentrale Datenhaltung aus technischer Sicht	2
4.1.2 Zusammenfassung	2
4.2 Erfüllung basisinfrastruktureller Anforderungen.....	2
4.2.1 Zentrale vs. dezentrale Datenhaltung aus technischer Sicht	2
4.2.2 Zusammenfassung	2
5 Zusammenfassung und Ausblick	2
Literaturverzeichnis	II

Abbildungsverzeichnis

Abb. 1.1.1/1: Aufbau und Struktur des Arbeitsberichtes	2
Abb. 2.1.1/1: Disjunktive, adjunktive und konjunktive Verknüpfung von Medieninhalten	2
Abb. 2.1.1/2: Abgrenzung multimedialer von klassisch-monomedialen Integrationsobjekten	2
Tab. 2.1.2/1: Horizontale und vertikale Integration von Daten und Content	2
Abb. 2.2/1: Vier-Ebenen-Integrationsmodell nach Österle	2
Abb. 2.3.1/1: Content- und Metadatenproduktion und -bündelung im Content Management Workflow..2	
Abb. 2.2.3/1: Technische Anforderungskategorien an eine Mehrfachverwertung von Content.....2	
Abb. 3.1/1: Unterschiede zwischen integrierter und isolierter Datenhaltung	2
Abb. 3.1/2: Zentrale und dezentrale Datenhaltung	2
Abb. 3.2/1: Positionierung der Datenhaltungsszenarien	2
Abb. 3.2/2: Redundanzfreie, integrierte Datenverteilungsvarianten	2
Abb. 3.3/1: Zentralisierte Datenhaltung: CoreMedia's Systemarchitektur	2
Abb. 3.3/2: Hybrid, metadatenzentralisierte Datenhaltung: IBM DB2 Content Manager	2
Abb. 3.3/3: Dezentralisierte Datenhaltung: NXT 4-Plattform von NextPage.....2	

Tabellenverzeichnis

Tab. 2.1.1/1: Content-Management vs. Daten Mangement	2
Tab. 2.1.5/1: Einordnung der Content-Integration anhand klassischer Integrationsmerkmale.....2	
Tab. 3.2/1: Varianten der Datenverteilung	2
Tab. 4.2.2/1: Technische Bewertung unterschiedlicher Datenverteilungsvarianten	2

1 Einleitung

Die zunehmende Entstehung von Medienverbundunternehmen mit integrierten Wertschöpfungsstrukturen (z.B. Time-Warner) ist insbesondere auf regulatorische, technische und ökonomische Kontextfaktoren zurückzuführen. Zum einen vergrößert das Zusammenwachsen der Branchen Telekommunikation, Informationstechnologie und Medien, die Deregulierung der Medien-, Kommunikations- und Informationsmärkte, der zunehmende Einsatz der Digitaltechnik und die Veränderung der Nutzerpräferenzen hin zu einer multioptionalen „Multitouchpoint“-Rezeption die Freiheitsgrade der Nutzung und Verwertung von Medieninhalten (vgl. Wirtz, 2001, S. 507ff.). Neben diesen marktorientierten Ursachen werden in der wissenschaftlichen Literatur zum anderen insbesondere unternehmensbezogene Ursachen angeführt. Wirtz unterscheidet hierbei marketingstrategische und unternehmensstrategische Ursachen (vgl. Wirtz, 2001, S. 507ff.). Marketingstrategische Ursachen der zunehmenden Integration sind vor allem auf Absatzsteigerungsmöglichkeiten und erhöhte Kundenbindungspotenziale zurückzuführen. Erste Untersuchungen über die Mehrfachverwertung von Medieninhalten in Medienunternehmen weisen auf eine verstärkte Ausdifferenzierung von Content-Verwertungsketten mittels neu hinzukommender, horizontal und/oder vertikal integrierter Verwertungsfenster (wie z.B. Online Content Syndication oder Cross-Media-Publishing) hin (vgl. Schulze/Hess, 2003). Unternehmensstrategische Gründe für die zunehmende Entstehung integrierter Medienunternehmen sind die Errichtung von Markteintrittsbarrieren durch größere Skaleneffekte sowie Kostendegressionseffekte durch Ausnutzung von Synergieeffekten (vgl. Schulze/Hess, 2003; Schumann, 2003; Martin/Eisenhardt, 2001).

Um eine Mehrfachverwertungsstrategie von Medieninhalten operativ in Form von Beschaffungs-, Erstellungs-, Bündelungs- und Distributionsprozessen effizient umsetzen zu können, benötigt ein Medienunternehmen nicht nur integrierte Wertschöpfungsstrukturen auf Geschäfts-, sondern auch integrierte Daten, Funktionen und Prozesse auf systemtechnischer Ebene. Eine Systemlandschaft mit integrierter Dateneinhaltung stellt dabei das *conditio sine qua non* einer effizienten Mehrfachverwertung von Inhalten dar. Obwohl sich die wissenschaftliche Literatur in der Wirtschaftsinformatik (WI) bisher sehr detailliert mit unterschiedlichen Daten-, Funktions- und Prozessintegrationsverfahren auseinandergesetzt hat, lag der bisherige Schwerpunkt der Betrachtung vorwiegend auf *atomar-monomedialen* Integrationsobjekten, die eher in analytisch-transaktionalen Entscheidungsunterstützungsszenarien Verwendung finden (vgl. Mertens, 2001, Mertens et al., 2004; Bunjes et al., 2002; Heine, 2000; Holten, 2003). Demgegenüber wurde die wissenschaftliche Analyse von multimedialen Integrationsobjekten innerhalb von gestalterisch-retrievalbasierter Anwendungsszenarien weitgehend vernachlässigt. Aufgrund der bei der Integration von

multimedialen Datenobjekten hinzukommenden technischen Anforderungen interessieren vor diesem Hintergrund insbesondere potenziell mögliche integrierte Datenverteilungsszenarien, die gerade für die effiziente Umsetzung einer Mehrfachnutzung von Medieninhalten unterschiedliche Vor- und Nachteile bereithalten.

Ziel des vorliegenden Arbeitsberichts ist die Beantwortung der Fragestellung, inwieweit unterschiedliche Datenverteilungsvarianten (technische) anwendungsspezifische und basisinfrastrukturelle Anforderungen für eine effiziente Umsetzung einer Mehrfachverwertungsstrategie erfüllen. Die nachfolgende Abbildung illustriert dabei den Aufbau des Arbeitsberichtes.

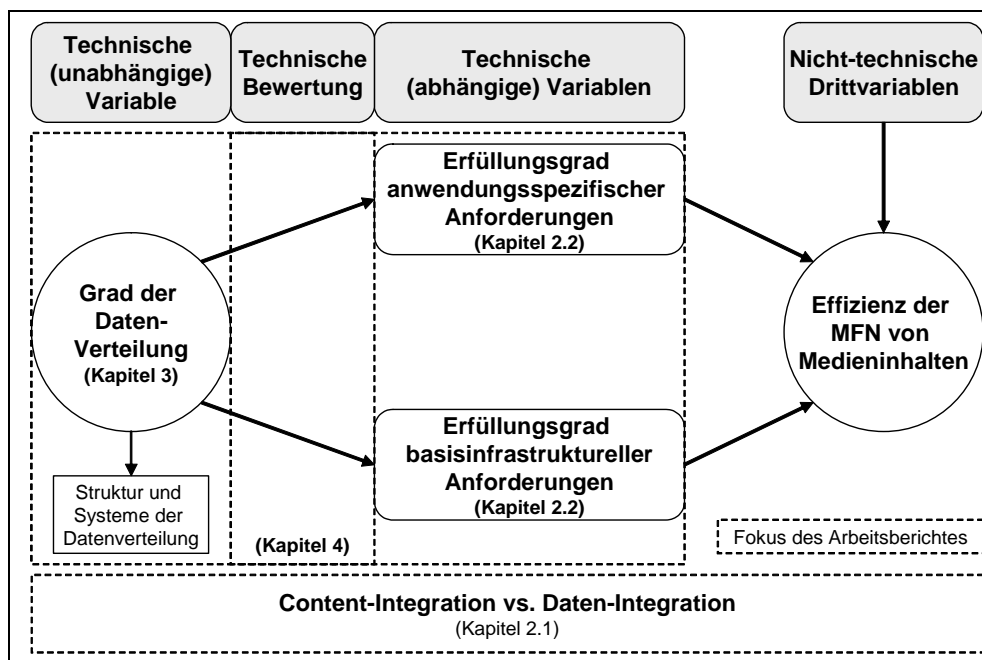


Abb. 1.1.1/1: Aufbau und Struktur des Arbeitsberichtes

In Kapitel 2.1 werden als Grundlage zur technischen Bewertung der zu identifizierenden integrierten Varianten der Verteilung multimedialer Daten zunächst das Konzept der Content-Integration vor dem Hintergrund der klassischen WI-Integrationsliteratur verortet. Darauf aufbauend werden technische Anforderungen an eine Mehrfachverwertung von Inhalten mittels einer Kombination aus empirisch gestützten konstruktivistischen Leistungen und einer Literaturanalyse systematisch abgeleitet (Kapitel 2.2). Kapitel 3 versucht anschließend, auf Basis einfacher kombinatorischer Überlegungen idealtypische integrierte Datenverteilungsszenarien zu identifizieren. Anhand von verschiedenen in der Praxis eingesetzten (Multimedia) Content Management Systemen, die gemäß ihrer Architektur spezielle Datenverteilungsszenarien nahe legen, soll die Existenz der identifizierten idealtypischen Datenverteilungsszenarien empirisch überprüft werden. In Kapitel 4 werden die idealtypischen Datenverteilungsszenarien schließlich auf Basis der in Kapitel 2.2 abgeleiteten, zur Umsetzung einer effizienten Mehrfachverwertungsstrategie zu erfüllenden, technischen Anforderungen

(qualitativ) bewertet. Die Darstellung der empirischen Architekturausprägungen sowie die Erklärungsbeiträge zur qualitativen Bewertung der unterschiedlichen Datenverteilungsszenarien stützen sich in der Argumentation einerseits auf einer Analyse der einschlägigen Literatur, andererseits auf drei im Zeitraum Dezember 2003 – März 2004 durchgeführte Leitfadeninterviews mit Content Management Systeme-Anbietern (IBM¹, Documentum² und CoreMedia³). Der Arbeitsbericht schließt mit einer Einordnung der Ergebnisse und einem Ausblick auf die künftige Forschungsarbeit.

¹ Zweistündiges Leitfaden-Interview am 06.12.2003 mit Dr. Wilhelm Gotthard (Manager DMD Content Management) und Dr. Sebastian Goeser (Senior Software Engineer) im Entwicklungszentrum IBM in Böblingen.

² Zweistündiges Leitfaden-Interview am 15.01.2004 mit Daniel Pelke (Business Development Manager, Central Europe) und Stefan Weiss-Weber (Technical Director, Central Europe) in München.

³ Schriftlicher Leitfaden-Fragebogen erhalten am 09.03.2004 von Björn Schäfer (technische Leitung).

2 Die Integration multimedialer Daten als spezielles Integrationsproblem

Die Medienindustrie grenzt sich gegenüber anderen Branchen (wie z.B. Banken und Versicherungen) dadurch ab, dass in ihrer Branchenwertschöpfungskette (digitale) Medieninhalte nicht nur als Produktionsfaktoren verarbeitet, sondern daraus auch digitale Medienprodukte produziert und gebündelt werden, bevor diese an Kunden verkauft und distribuiert werden (vgl. Schumann/Hess, 2002, S. 10). Die seit Ende des 20. Jahrhunderts fortschreitende Digitalisierung der Medieninhalte erweitert dabei im Gegensatz zur analogen Medieninhalteverarbeitung die Freiheitsgrade in der Verarbeitung und Manipulation von Medieninhalten bzw. Content⁴ erheblich. Aufgrund der eingangs bereits genannten marktorientierten Pull- und unternehmensbezogenen Push-Faktoren erscheint es auf der Unternehmensseite durch größere technische Spielräume effizienter zu sein, einmal erstellte Medieninhalte über Raum (z.B. in verschiedenen Medienformaten, in verschiedenen Medienkanälen (TV, Print), in verschiedenen Regionen bzw. Ländern) und Zeit (z.B. gleiches Format sequenziell in gleichen Medienkanälen (nur TV), u.U. jedoch auf anderen Medienträgern (z.B. TV-Sender)) mehrfach zu verwerten (vgl. Schulze/Hess, 2003).

Ein Szenario der integrierten Mehrfachverwertung von Inhalten lässt sich rein technisch gesehen dadurch charakterisieren, dass nicht nur klassisch-monomediale, sondern multimediale Datenintegrationsverfahren angewendet werden müssen. Produktionsprozesse unterschiedlicher Medientypen (z.B. Videoschnitt, Bildbearbeitung und Textredaktion) müssen zusammenlaufen bzw. auf einander abgestimmt und technisch unterstützt werden. Dies hat zur Folge, dass sich nicht nur die Qualität der zu bearbeitenden Medientypen ändert und die Quantität der Medieninhalte steigert, sondern auch die Komplexität in der Verarbeitung und Verwaltung der multimedialen Datenobjekte erheblich zunimmt. Bevor in Kapitel 2.2 technische Anforderungen zur Umsetzung einer effizienten Mehrfachverwertung von Inhalten systematisch abgeleitet werden, sollen im nachfolgenden Abschnitt die Unterschiede der Integration multimedialer Daten gegenüber den in der klassischen Literatur zur Datenintegration diskutierten Konzepte skizziert werden.

⁴ Im Folgenden wird die Content-Definition von Anding und Hess herangezogen, die unter Content „eine aus menschlicher Intelligenz resultierende, durch maßgeblich redaktionelle Mittel angereicherte, individuell schützbar und zweckorientierte Abbildung impliziter Informationen“ verstehen (Anding/Hess 2003, S. 14).

2.1 Integration multimedialer Daten: Verortung innerhalb der klassischen Integrationsliteratur

Unter *Integration* wird in der klassischen WI-Literatur die „(Wieder-)Herstellung eines Ganzen“, d.h. die Zusammenführung oder Synthese von Einzelbestandteilen zu einem Gesamten bezeichnet (vgl. Mertens et al., 2004; Heilmann, 1989). Übertragen auf die betriebliche Informationsverarbeitung bedeutet Integration die Zusammenführung logisch und/oder physisch isolierter, meist heterogener Integrationsobjekte mit dem Ziel, eine effiziente Kommunikation zu ermöglichen. Voraussetzung dafür ist eine Einigung beteiligter Systeme über die Verwendung standardisierter Kommunikationsformen. Ein *Standard* bezeichnet in diesem Zusammenhang eine gemeinsame Sprache (z.B. Syntax, Grammatik, etc.), die es erlaubt, dass Botschaften in Form von Daten oder Funktionsaufrufen auf einer syntaktischen, semantischen und pragmatischen Ebene gegenseitig verstanden und interpretiert werden können (vgl. Voigtmann/Zeller, 2003).

Mertens charakterisiert die Integration der Informationsverarbeitung anhand der vier Dimensionen *Integrationsgegenstand*, *Integrationsrichtung*, *Integrationsreichweite* und *Automatisierungsgrad* (vgl. Mertens, 2001, S. 2). Anhand dieser vier Dimensionen soll die Integration multimedialer Daten innerhalb der klassischen Integrationsliteratur im Folgenden verortet werden.

2.1.1 Integrationsgegenstand

Als Integrationsgegenstände bzw. -objekte werden im Kontext der klassischen betrieblichen Informationsverarbeitung *Daten*, *Funktionen*, *Benutzeroberfläche*, *Prozesse* und *Programme* betrachtet (z.B. Mertens, 2001; Krcmar, 1991)⁵. Daten werden in diesem Zusammenhang als maschinell verarbeitbare formalisierte Repräsentationen (Gebilde aus Zeichen) von Fakten und Gedanken verstanden (vgl. Balzert, 2000).

In der Systematik der klassischen WI-Integrationsliteratur, in der Informationen bzw. Daten als reine Produktionsfaktoren zur Unterstützung von vertikalen Produktions- bzw. horizontalen Entscheidungsprozessen aufgefasst werden (vgl. Mertens, 2001, Mertens et al., 2004; Bunjes et al., 2002; Heine, 2000; Holten, 2003), wird die Integration multimedialer Inhalte (bzw. „Medienintegration“) als eine Variante der Programmintegration und damit als eine kundenseitige Zusammenführung von multimedialen Daten in einem Dokument bzw. Programm angesehen (vgl. Mertens, 2001, S. 4). Eine Unterscheidung von Daten nach statisch-diskreten, also zeitunabhängigen

⁵ Da im weiteren Verlauf des vorliegenden Arbeitsberichtes multimediale *Daten* als zentraler Integrationsgegenstand untersucht werden soll, werden im Folgenden alternative Integrationsobjekte einer weiteren Betrachtung entzogen.

Medientypen (wie z.B. Texte, Bilder, Zahlen) und dynamisch-kontinuierlichen bzw. zeitabhängigen Medientypen wird nicht vorgenommen.

Daten bzw. Informationen werden insbesondere in klassischen Nutzungsszenarien verwendet, in denen menschliche Aufgabenträger operative bzw. strategische Informationen in unterschiedlichen Aggregationsniveaus zur Entscheidungs- (z.B. im Querschnittsbereich des Rechnungswesens) bzw. zur Produktions- (z.B. Lagerhaltungsinformationen) und Distributionsunterstützung (z.B. zur Bewerbung des Produktes) benötigen. Zur Unterstützung dieser Prozesse werden dispositive Management-Informationssysteme (bzw. Planungs- und -kontrollsysteme) bzw. operative Systeme eingesetzt (vgl. Mertens et al., 2004, S. 86), die z.B. auf integrierten Datenaggregations- und -aufbereitungssystemen (z.B. Data Warehouses) beruhen. Die zu integrierenden Daten lassen sich als atomare, fein-granulare und normalisierbare Datenwerte beschreiben, die sich mittels statistischer, mathematischer und logischer Operationen aggregieren lassen können (z.B. Berechnung einer aggregierten Umsatzsumme). Typische Operationen auf „klassischen“ Daten, welche durch triviale Datentypen (z.B. einfache Zeichenketten, Wahrheitswerte oder Integer-Zahlenräume) in relationalen Datenbanken repräsentiert werden, sind Abfragen sowie Manipulationen zur Analyse, Aufbereitung und Präsentation der Daten. Zur Optimierung der Datenorganisation und somit auch der Operationen auf den Datenbeständen werden die Daten in sog. relationalen Tabellen gehalten, welche reale Entitäten einer abzubildenden Anwendungsdomäne darstellen sollen (vgl. Balzert, 2000, S. 224ff.; historische Grundlagen siehe Chen, 1976). Merkmale bzw. Attribute der Entitäten, die man in diesem Sinne auch als triviale Metadaten interpretieren kann, werden als sog. Felder innerhalb einer relationalen Tabelle organisiert. Konkrete Wertausprägungen der Attribute werden als Tupel gespeichert. Indem Entitäten bzw. relationale Tabellen über Schlüsselattribute, welche Entitäten in einer speziellen Anwendungsdomäne eindeutig identifizieren sollen, verbunden werden, werden sog. übergreifende semantische Datenmodelle unterschiedlicher Anwendungsbereiche entwickelt (vgl. Stahlknecht, 1995, S. 190).

Im Gegensatz zu den klassischen Integrationsdomänen (z.B. in Banken, Industrie, etc.) werden Medieninhalte im Bereich der Medienindustrie sowohl als Produktionsfaktoren als auch als Endprodukte angesehen, welche für das Unternehmen einen *strategischen Wert* besitzen. Dieser strategische Wert stellt sich dabei als direktes bzw. indirektes Wertschöpfungspotenzial durch die Veräußerung⁶ der Medieninhalte bzw. als deren Verwendung als Werbeträger dar. Der Fluß der Medieninhalte richtet

⁶ Bei einem Verkauf der Medieninhalte an Dritte sind grundsätzlich Urheber- und Verwertungsrechte zu beachten (vgl. vgl. Breyer-Mayländer/Werner 2003, S. 44; Fetscherin 2002, S. 166;).

sich dabei strikt horizontal entlang der Wertschöpfungskette Produktion, Bündelung und Distribution aus (vgl. Schumann/Hess, 2002).

Zentrale Nutzungsszenarien bzw. Funktionen innerhalb des Medieninhalteflusses stellen die Erstellung, Bearbeitung, Bündelung, Suche und Archivierung der Medieninhalte dar, welche mit kreativ-gestalterischen Autorenwerkzeugen bzw. Archivierungs- und Verwaltungssystemen unterstützt werden.

(Multimediale) Medieninhalte als Integrationsobjekt sind bezüglich ihrer Struktur im Gegensatz zu den klassisch-monomedialen Daten hoch komplex und speicherintensiv (vgl. Somani et al., 2002). Da in der klassisch-monomedialen Datenintegration Daten in einem feingranularen Zustand gespeichert werden, ist der semantische Gehalt der Daten zur Interpretation im Information Retrieval zumeist klar und eindeutig. Unstrukturierte Medientypen (wie z.B. Texte, Bilder, Video- oder Audio-Informationen) hingegen sind in ihrer Struktur grobgranularer und somit auch bezüglich ihrem semantischen Gehalt nicht eindeutig interpretierbar. Aus diesem Grund müssen zur Optimierung der semantischen Interpretierbarkeit und letztlich auch der Auffindbarkeit der Medieninhalte zusätzliche textbasierte, inhaltsbeschreibende (interpretierende) Metadaten⁷ bereitgestellt werden (vgl. Schmitt, 2002, S. 32). Aufgrund der semantischen Reichhaltigkeit und Vernetztheit der Metadaten multimedialer Medieninhalte erweisen sich traditionelle Normalisierungsversuche oftmals als unzureichend bzw. unmöglich, unstrukturierte Medieninhalte geeignet zu strukturieren bzw. zu modellieren. Ontologien, welche abstrakte Netzwerke von fachdomänenrelevanten Begriffen darstellen und vielfältige Beziehungstypen (z.B. Vererbungs-, Aggregations-, Assoziationsbeziehungen) zwischen Entitäten zulassen, erweisen sich für eine Abbildung semantischer Metadaten häufig als effizienter als einfache relationale Tabellen (vgl. Fluit et al., 2002; Benlian et al., 2004).

Statistische Operationen auf multimedialen Daten lassen sich im Bereich der Medieninhaltsanalyse, welche versucht, semantische Metadaten aus den gegebenen Inhalten automatisiert zu extrahieren, durchführen (z.B. Häufigkeitsverteilungen von Begriffen oder Farbwerten). Eine mathematische oder logische Aggregation der Medieninhalte ist dagegen in den meisten Fällen nicht sinnvoll. Da es sich bei multimedialen Datentypen um verschiedenartige diskrete und kontinuierliche Medientypen mit spezifischen Charakteristika handelt, werden an Content Repositories für multimediale Daten größere Anforderungen gestellt als an Datenbanken, die ausschließlich triviale Datentypen bzw. nur einen Medientypen verwalten müssen. Vereinfachend lässt sich eine adjunktive bzw. konjunktive Verknüpfung der maximal vier Medientypen im Ge-

⁷ Metadaten, nach der ISO-Spezifikation definiert als „[...] information and documentation which makes data sets understandable and sharable for users“ (vgl. Kretzschmar/Dreyer 2004), dienen der automatisierten Klassifizierung und Identifizierung der häufig unstrukturiert vorliegenden Content-Inhalte sowie für einen schnellen Zugriff im Information Retrieval (Bechtold 2002, S. 35ff.).

gensatz zur disjunktiven Verknüpfung monomedialer (Einfach-) Verwertung mit Abbildung 2.2.1/1 darstellen.

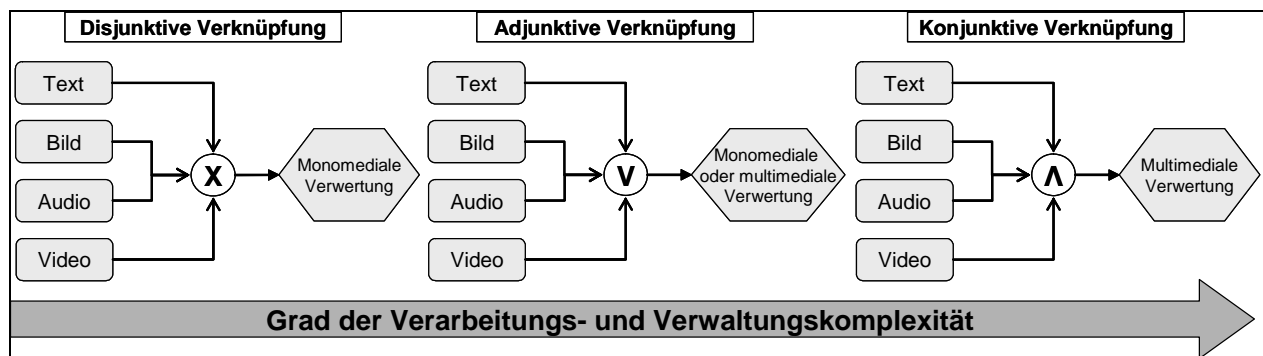


Abb. 2.1.1/1: Disjunktive, adjunktive und konjunktive Verknüpfung von Medieninhalten

Hinzukommt, dass im Vergleich zu Anwendungsszenarien der Einfachverwertung von Medieninhalten nicht nur die Eigenheiten und Spezifika *eines* Medientyps zu berücksichtigen sind, sondern ebenso diejenigen der jeweiligen anderen Medientypen. Die Komplexitätssteigerung⁸ der Verwaltung und Verarbeitung multimedialer Daten drückt sich dabei in der Vielfalt an erforderlichen medienspezifischen Technologien zur Realisierung gewünschter Funktionalitäten aus (siehe Abbildung 2.1.1/2). So existieren z.B. zur inhaltsbasierten Generierung von Metadaten aus den bestehenden Medieninhalten unterschiedliche Technologien (Spracherkennungs-, Video-Sequenz-Analyse, Text Mining, Bild-Analyse-Verfahren), die ein multimediales Gesamtsystem beherrschen muss.

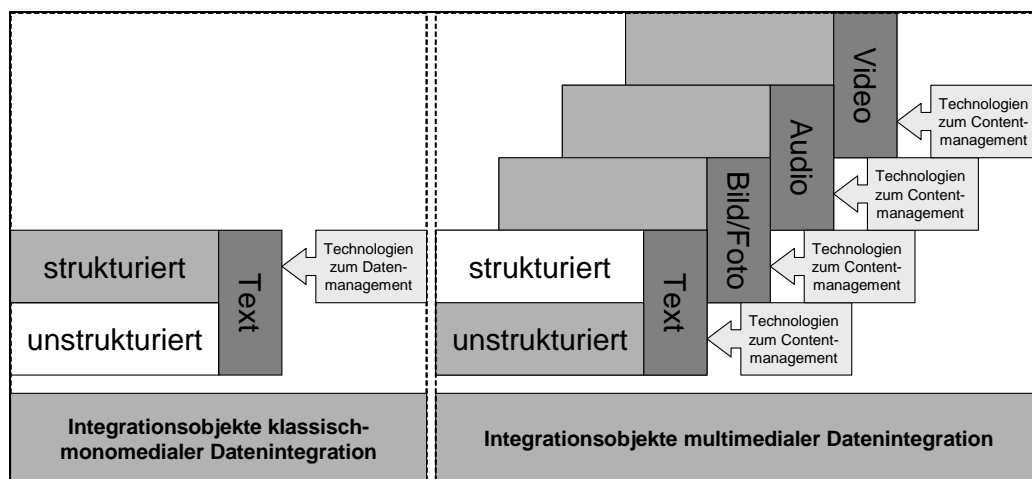


Abb. 2.1.1/2: Abgrenzung multimedialer von klassisch-monomedialen Integrationsobjekten

⁸ Die Komplexitätssteigerung lässt sich insgesamt auf eine gestiegene Technologiekomplexität (Beherrschung unterschiedlicher Anwendungssysteme, Formate und Medientypen für Endgeräte und Medienkanäle; Koordinationstechnologien zur Organisation und zum Timing in Bündelungstechnologien; höhere Hardwareanforderungen wie z.B. Multiprozessorenteknik) sowie auf eine gestiegene Produktkomplexität (Modularisierung von Inhalten zur Mehrfachkombination bereits existierender Module und Erhöhung der Zugänglichkeit und Übertragungsvolumina von Inhalten) zurückführen (vgl. Wildemann 1998, S. 48).

Multimediale Daten können in Form abstrakter Datentypen zum einen als Byte-Strom in (zumeist objektrelationalen) Datenbanken, zum anderen in Dateisystemen abgelegt werden. Bezüglich der Verwaltung in Datenbanken existieren die Möglichkeiten, multimediale Daten in medienspezifisch optimierten Datenbanken (z.B. Bilder- oder Videosequenzdatenbanken) oder sogar in allumfassenden Multimedia-Datenbanken abzulegen.

Tabelle 2.1.1/1 fasst die wesentlichen Unterschiede zwischen Content Management und Datenmanagement überblicksartig zusammen (in Anlehnung an Somani et al., 2002).

	Content Management	Daten Management
Einsatz der Information	Information als Produktionsfaktor <i>und</i> Produkt	Information als Produktionsfaktor
Nutzungsszenario	Produktion, Bündelung, Distribution (Archivierung, Retrieval)	Produktions- und Entscheidungsunterstützung
Auswertungsziel, Ziel des Datenzugriffs	Bearbeitung der Medieninhalte (des Content)	Manipulation, Informations- und Entscheidungsunterstützung
Anwendungssysteme	Kreativ-gestalterische Autorenwerkzeuge, Archivierungs- und Verwaltungssysteme	Dispositive und operative Anwendungssysteme
Datenmodell und Daten Repository	Objektrelationale Datenbanksysteme, Dateisysteme, medienspezifische oder Multimedia Datenbanken	Relationale (eher selten: objektorientierte oder objektrelationale) Datenbanksysteme
Organisation der semantischen Metadaten	Ontologien, Taxonomien	Datenschema, -modelle
Daten- bzw. Medientyp	Komplexe Datentypen: diskret-unstrukturiert und kontinuierlich-unstrukturiert (Rich Media)	Triviale Datentypen (v. a. Zahlen-, String- und Wahrheitswerte): diskret-strukturiert
Datengranularität	Gesamte Dokumente bzw. Medienmodule	Individuelle, atomare Datenwerte

Tab. 2.1.1/1: Content-Management vs. Daten Mangement

2.1.2 Integrationsrichtung

Hinsichtlich der Integrationsrichtung innerhalb der Aufbauorganisation eines Unternehmens wird in der klassischen WI-Integrationsliteratur zwischen der horizontalen und vertikalen Integration unterschieden. Horizontale Integration bezieht sich dabei auf die Verbindung der Teilsysteme in der betrieblichen Wertschöpfungskette, vertikale Integration auf die Datenversorgung höhere Unternehmensebenen zur Entscheidungsunterstützung (vgl. Mertens, 2001, S. 4). Innerhalb klassischer Branchen (wie z.B. Handel, Banken oder Industrie) werden Informationen horizontal integriert,

um die Funktionsabteilungen einerseits mit funktionsübergreifenden Informationen zur Verrichtung der wertschöpfenden Funktionen zu versorgen, andererseits um Informationen über das produzierte Hauptprodukt (z.B. Eigenschaften wie Mengen- oder Preisangaben) zu sammeln bzw. zu kommunizieren. Vertikale Integration in Form der Aggregation entscheidungsrelevanter Informationen findet in klassischen Integrationsdomänen insbesondere zur Entscheidungsunterstützung des mittleren und höheren Managements statt.

Zwar werden in Medienunternehmen Informationen über den Umsatz, Kosten, etc. ebenso zur Entscheidungsunterstützung vertikal über Unternehmensebenen hinweg integriert. Der Schwerpunkt liegt jedoch insbesondere auf der horizontalen Integration, in der Medieninhalte sowohl als Produktionsfaktoren verarbeitet als auch als Hauptprodukte veräußert werden. Entlang des Wertschöpfungsprozesses müssen die einzelnen Schritte der Beschaffung, Produktion, Bündelung und Distribution von Medieninhalten dabei aufeinander abgestimmt werden. Nachfolgende Tabelle fasst die jeweiligen schwerpunktmäßigen Einsatzszenarien (siehe angegraute Felder) von immateriellen Informationen bzw. Daten in den jeweiligen Branchen überblicksartig zusammen.

Integrationsrichtung Datenart und Branchenfokus	Horizontal	vertikal
Daten in klassischen Anwendungsszenarien (Banken, Industrie, Handel)	Produktionsfaktor bzw. Produktinformationen	Entscheidungsunterstützung
Daten und Content in Medienunternehmen	Produktionsfaktor und Hauptprodukt, das am Markt verkauft wird	Entscheidungsunterstützung

Tab. 2.1.2/1: Horizontale und vertikale Integration von Daten und Content

2.1.3 Integrationsreichweite

Hinsichtlich der Integrationsreichweite werden innerhalb der klassischen WI-Integrationsliteratur sowohl innerbetriebliche (Bereichs-, einzelbetriebliche und konzernweite Integration, z.B. Data Warehouses oder EAI) als auch zwischenbetriebliche Integrationsvarianten (z.B. mittels EDI) sehr ausführlich diskutiert. Innerhalb von Medienunternehmen liegt der Schwerpunkt dagegen auf der innerbetrieblichen Integration, die insbesondere den Abbau künstlicher (technischer) Barrieren bei der Unterstützung von Produktionsprozessen zum Ziel hat. Anwendungsbereiche für den zwischenbetrieblichen (bzw. endkundenseitigen) Medieninhalteaustausch existieren insbesondere in der physischen oder digitalen Distribution von Medieninhalten (wie z.B. im Geschäftsmodell des (Online) Content Syndication (vgl. Anding, 2004)).

2.1.4 Automatisierungsgrad

Das Potenzial, Vorgänge oder Prozesse automatisieren zu können, richtet sich grundsätzlich an der Art der Aufgabe aus, welche von Informations- und Kommunikationstechnologien unterstützt werden soll. Gerade in klassischen Integrationsdomänen (wie z.B. Banken, Handel, Automotive, etc.) findet man in Produktions- und Distributionsprozessen einen hohen Anteil an hoch-strukturierten und routinisierten Aufgaben vor, die zum Teil vollautomatisiert abgewickelt werden können. Planungs- und Entscheidungsprozesse lassen sich dagegen nur begrenzt durch IT-Systeme unterstützen. Im Bereich der Medienindustrie spielen Kreativität und Differenzierung aufgrund des hohen Wettbewerbsgrades eine große Rolle. Aufgaben während des Produktions- und Distributionsprozesse können deshalb nur teilweise (z.B. Recherche-, Layout- und Druckprozesse) von Anwendungssystemen unterstützt werden.

2.1.5 Zusammenfassung: Einordnung der Content-Integration

Resümierend lässt sich festhalten, dass sich die Integration multimedialer Medieninhalte durch die Spezifika der Medienbranche und der multimedialen Medientypen in zentralen Punkten von der Integration monomedialer, trivialer Daten unterscheidet. In folgender Tabelle wird die Einordnung der multimedialen Datenintegration anhand der klassischen Integrationsmerkmale noch einmal überblicksartig zusammengefasst.

Klassifikationsmerkmal	Schwerpunktmäßige Merkmalsausprägung			
	Daten	Funktionen	Prozesse	Programme
Integrationsrichtung	Horizontale Integration	Vertikale Integration	Diagonale Integration	
Integrationsreichweite	Innerbetrieblich (Unternehmensbereich bzw. Einzelunternehmen)	Innerbetrieblich (Konzern)	Zwischenbetrieblich	
Automatisierungsgrad	Nicht-	Teil-	Voll-Automation	

Tab. 2.1.5/1: Einordnung der Content-Integration anhand klassischer Integrationsmerkmale

2.2 Technische Anforderungen an die Mehrfachnutzung von Medieninhalten

Technische Anforderungen an ein Informations- bzw. Kommunikationssystem zur Unterstützung einer Geschäftsstrategie bzw. von Geschäftsprozessen lassen sich anhand des Vier-Ebenen-Integrationsmodells nach Österle (vgl. Österle, 1996, S. 3; siehe auch Wollnik, 1988), das die unterschiedlichen Abstraktionsbereiche eines Unternehmens in einem Ebenenmodell konzeptionalisiert, verorten (siehe Abbildung 2.2/1).

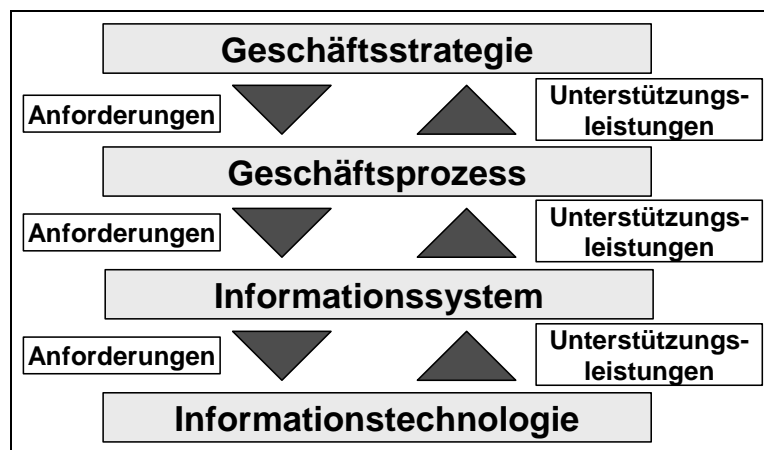


Abb. 2.2/1: Vier-Ebenen-Integrationsmodell nach Österle

Während auf der obersten Ebene die zentralen Konzepte der Geschäftsstrategie eines Unternehmens entwickelt und auf der zweiten Ebene auf konkrete Prozessstrukturen abgebildet werden, werden zur Umsetzung bzw. Unterstützung der Geschäftsprozesse auf der Ebene der Informationssysteme anwendungsdomänenspezifische, auf der Informationstechnologieebene basisinfrastrukturelle Anforderungen spezifiziert (vgl. Krcmar, 2003, S. 40ff.). An dieser Systematik orientiert sich auch die Ableitung der Anforderungen an eine Mehrfachverwertung von Inhalten. Während sich anwendungsdomänenspezifische (hier im Speziellen medieninhaltebezogene) Anforderungen aus dem Content-Workflow einer Mehrfachverwertung von Inhalten ableiten lassen (Prozessebene), orientieren sich basisinfrastrukturelle Anforderungen zum einen an den funktionalen Anforderungen an eine Mehrfachnutzung von Medieninhalten, zum anderen an den technischen Charakteristika der gegebenen Informationssysteme, -netzwerke und -güter (Systemebene).

2.2.1 Medieninhaltebezogene Anforderungen

Medieninhaltebezogene Anforderungen an die Mehrfachverwertung von Medieninhalten lassen sich entlang der Medienwertschöpfungskette und spezifischer, entlang eines abstrakten Content-Workflows, ableiten, der eine modularisierte Mehrfachverwertung unterstützt. Durch die Beschaffung bzw. Erstellung von Medieninhalten gelangen neue Content-Objekte unterschiedlichen Medientyps in das Unternehmen. In einem zweiten Schritt werden Medieninhalte indiziert, kategorisiert und ggf. zielmedienneutralisiert, während die ursprünglich monolithisch strukturierten Medieninhalte modularisiert werden. Nachfragebedingt greifen Redakteure auf gespeicherte Inhaltmodule zurück, wobei sie vor der eigentlichen Bündelung geeignete Inhaltmodule finden müssen. Im letzten Schritt werden die geschnürten Content-Bündel über gewünschte (mehrfache) Medienkanäle distribuiert (vgl. Pankratz/Benlian, 2004, S. 136ff.). Abbildung 3 bildet den Content-Workflow in UML-Notation schematisch ab.

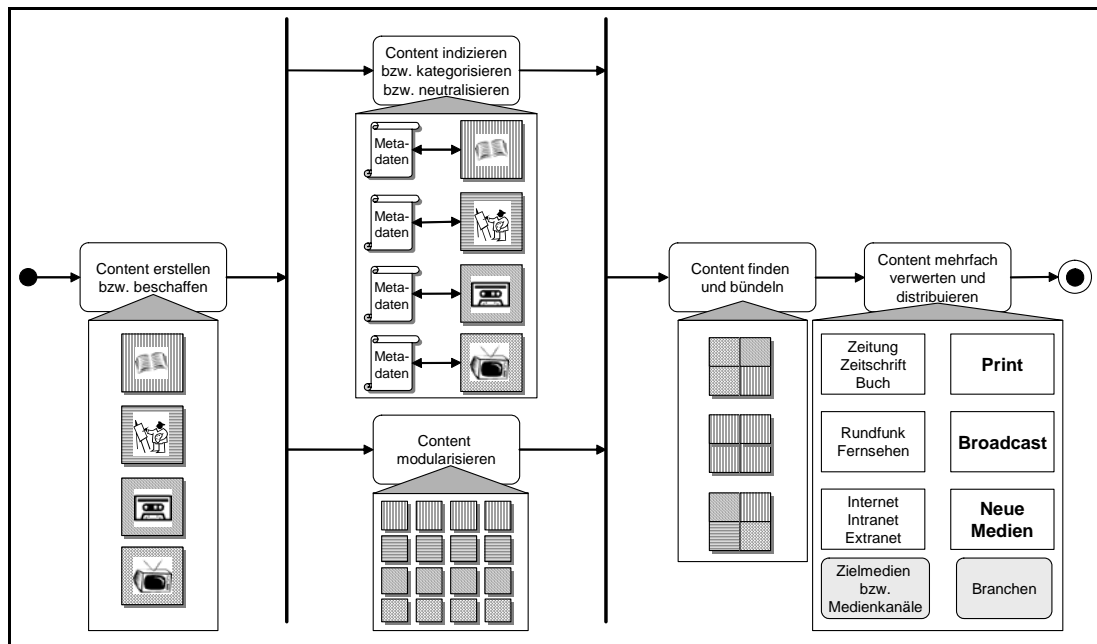


Abb. 2.3.1/1: Content- und Metadatenproduktion und -bündelung im Content Management Workflow

Auf Systemebene werden in den Phasen der Erstellung bzw. Beschaffung und Bündelung von Content insbesondere Integrations- bzw. Editorensysteme, in der Distributionsphase hauptsächlich Übertragungssysteme herangezogen. Entlang des gesamten Content-Workflows werden Datenhaltungs- und -managementsysteme zur Speicherung, Indizierung, Archivierung und Suche von Content verwendet. Die Kombination dieser Systeme wird je nach Schwerpunktsetzung und Funktionsumfang als Content Management System (vgl. Rawolle, 2002) bzw. als Content Produkt Plattform (vgl. Anding et al., 2003; Köhler et al., 2003) bezeichnet. Damit die Medieninhalteflut in Medienunternehmen nicht überhand nimmt, müssen zusätzliche Managementmechanismen eingeführt werden, die den erstellten bzw. den in das Unterneh-

men strömenden Content einerseits kontinuierlich integrieren, andererseits (automatisch bzw. semi-automatisch) in einer Form organisieren, in der Verknüpfungspotenziale zwischen Content-Modulen aufgedeckt und nachfragebedingt ausgeschöpft werden (insbesondere Modularisierungs-, Auszeichnungs-, Matching-, Retrieval- und Bündelungsfunktionalitäten). Aus dem Prozess der Content-Verarbeitung (fachkonzeptionelle Sicht) und der Systemebene (technische Sicht) lassen sich somit folgende medieninhaltebezogenen Anforderungen ableiten (vgl. Reinhard/Rauchfuß, 2004, S. 168).

- Modularisierung
- Zielmedienneutralisierung
- Strukturierung (für eine schnelle Zugänglich- und Auffindbarkeit)
- Komprimierung

Eine Mehrfachnutzung von Medieninhalten bringt zunächst eine Komplexitätserweiterung mit sich, da unterschiedliche Medienkanäle mit den unterschiedlichsten Datenformaten und Medientypen bedient werden müssen. Aus der ökonomischen Sinnhaftigkeit, Kostendegressions- durch Synergieeffekte zu realisieren, in dem ein Medieninhalt (in potenziell abgewandelter) Form wieder verwendet wird, ist es erstrebenswert, den Medieninhalt in Form von Content-Modulen zu organisieren, die in unterschiedlichen Kombinationen für unterschiedliche Zielmedien rekonfiguriert werden können (vgl. Schulze/Hess, 2003). Neben der *Modularisierung* ist im Kontext der Mehrfachnutzung von Inhalten die *zielmedienneutrale Speicherung* der Medieninhalte von zentraler Bedeutung. Erst diese Speicherungsweise ermöglicht eine flexible Konvertierung der Medieninhalte in gewünschte Zielformate, ohne aufwändige Vor-konvertierungen aus anderen proprietären Formaten vornehmen zu müssen. Hinzu kommt die durch eine zielmedienneutrale Datenhaltung ermöglichte Trennung von Inhalt, Struktur und Layout (vgl. Rawolle, 2002).

Mit Blick auf die verteilte, unternehmensinterne Bearbeitung sowie unternehmensexterne Distribution digitaler Medieninhalte spielen zwei weitere medieninhaltebezogene Anforderungen eine große Rolle. Zum einen sollten Medieninhalte einfach editierbar und leicht zugänglich sein. Durch die technische Vorbedingung der *digitalisierten* Speicherung ist die – im Vergleich zu analog gespeicherten Medieninhalten – einfache Editierbarkeit erfüllt. Unter der leichten *Zugänglichkeit* ist die einfache Auffindbarkeit bzw. die schnelle Identifikation der und der schnelle Zugriff auf die Medieninhalte gemeint. Die alleinige digitale Speicherung der „rohen“ Medieninhalte reicht zur Erfüllung dieser Anforderung nicht aus. Wie oben bereits erwähnt wurde und wie sich bereits in der technischen Ebene gezeigt hat, ist für die Auszeichnung der Rohdaten mit semantischen Textinformationen ein ausgefeiltes Metainformationsmodell von zentraler Bedeutung (vgl. Schmitt, 2002). Schließlich ist für eine prozessoptimale Übertragung von Medieninhalten unter einer Übertragungskapazitäts- und Speicher-

platzrestriktion eine *Komprimierung* der Medieninhalte eine zentrale Anforderung an die Mehrfachnutzung von Inhalten.

Medieninhaltebezogene Anforderungen basieren auf basisinfrastrukturellen Anforderungen, denn erst eine Bereitstellung physikalisch-basistechnischer Vorbedingungen ermöglicht eine an konkreten Prozessen orientierte Umsetzung von Anwendungserfordernissen.

2.2.2 Basisinfrastrukturelle Anforderungen

Basisinfrastrukturelle Anforderungen an Architekturen, die eine Mehrfachnutzung von Medieninhalten unterstützen sollen, basieren alle auf der Erfordernis, eine vollständige Integration der an einer potenziellen Mehrfachnutzung beteiligten Prozesse, Applikationen und Datenquellen zu erzielen (vgl. Österle, 1996). Der Betrachtungsfokus soll in vorliegendem Arbeitsbericht dabei insbesondere auf der Datenebene liegen.

Aus der Heterogenität der oftmals verteilt gespeicherten multimedialen Daten sowie der erhöhten Komplexität verwendeter Technologien in integrierten Systemen ergeben sich für die Basisarchitektur einer Mehrfachnutzung von Inhalten spezifische Anforderungen, die sich aus einer Literaturanalyse zu den Bereichen „Verteilte Systeme“ und „Multimediale Systeme“ übertragen lassen. In der Diskussion um die Frage, welche technischen Anforderungen integrierte Systeme erfüllen müssen, werden in der klassischen Literatur mehrere vergleichbare Systematisierungsansätze verwendet. Als zentrale Anforderungen an (integrierte) verteilte und multimediale Systeme lassen sich die Kategorien *Skalierbarkeit*, *Zuverlässigkeit*, *Performanz*, *Erweiterbarkeit* und *Sicherheit* identifizieren (vgl. Weber, 1998; Dadam, 1996; Grady, 1992; Balzert, 1998, S. 260f.). Mit *Skalierbarkeit* wird eine beliebige Erweiterbarkeit der Rechnerkapazitäten eines Systems gemeint. Ein System, das multimediale Inhalte zur Mehrfachnutzung bündeln und distribuieren muss, sollte in Abhängigkeit der beanspruchten Bandbreiten flexibel skalierbar sein. Die *Zuverlässigkeit* beschreibt die Wahrscheinlichkeit, dass ein SW-Produkt eine gewünschte Funktion erfüllt. Hohe *Zuverlässigkeit* bedeutet, dass Fehler selten auftreten und nur geringe Auswirkungen haben. *Zuverlässige* Systeme lassen sich damit durch ein hohes Maß an Verfügbarkeit und Fehlertoleranz charakterisieren (vgl. Weber, 1998). Hand in Hand mit der *Zuverlässigkeit* geht die *Robustheit* des Systems, die eine sinnvolle Reaktion des Systems bei Fehlern, die in der Umgebung auftreten (z.B. Bedienungsfehler, Fehler anderer Systeme), beschreibt. Die Anforderung einer hohen *Zuverlässigkeit* spielt im Szenario der Mehrfachnutzung von Medieninhalten eine zentrale Rolle, da gerade aufgrund der erhöhten Management- und Verarbeitungskomplexität die Fehleranfälligkeit (*ceteris paribus*) tendenziell zunimmt bzw. die Verfügbarkeit der Systeme (*ceteris paribus*) tendenziell abnimmt. Die *Performanz* kennzeichnet das Antwort-Zeit-

Verhalten (Latenz) eines Systems auf Nutzeranfragen und hängt dabei einerseits von der Programmlogik des Anwendungssystems, andererseits von der Übertragungskapazität der Kommunikationsnetze ab. Im Szenario der Mehrfachnutzung von Medieninhalten müssen vergleichsweise hohe Performanzanforderungen an die Systeme gestellt werden, um im Produktions-, Bündelungs- und Distributionsprozess schnelle Zugriffe auf physisch oftmals verteilte Medieninhalte gewährleisten zu können. Eine weitere Anforderung, die sich insbesondere auf die Anwendungsebene bezieht, ist die *flexible Erweiterbarkeit*. Das Szenario der Mehrfachverwertung von Inhalten muss sich immer offen halten, neue Dienste, Formate und Funktionen integrieren zu können. Die zugrunde liegende Systemarchitektur muss folglich in der Lage sein, laufend angepasst bzw. durch neue Module ergänzt werden zu können (vgl. Stamer, 2002, S. 114). Schließlich sind die Kriterien *Sicherheit* und *Schutz* ein wichtiges Themenfeld. Aufgrund der hohen Menge übertragener Daten über Kommunikationsnetze, sind im Szenario der Mehrfachnutzung von Medieninhalten erhöhte Sicherheitsvorkehrungen (z.B. Verschlüsselung, Authentisierung und Zugriffskontrollmechanismen) gegenüber unrechtmäßigen Attacken und Zugriffen (z.B. Abhören, Intrigieren, Fälschen, Hinzufügen/Generieren) zu treffen (Weber, 1998, S. 263ff.).

Zusammenfassend lassen sich als basisinfrastruktureller Anforderungskatalog folgende fünf Kategorien anführen:

- Skalierbarkeit
- Performanz
- Zuverlässigkeit
- Erweiterbarkeit
- Sicherheit

2.2.3 Zusammenfassung: Technische Anforderungen an eine Mehrfachnutzung von Medieninhalten

In den vorangegangenen Kapiteln wurde zunächst die Integration von Medieninhalten (in der Medienbranche) in den State-of-the-Art der WI-Integrationsliteratur eingeordnet. Auf Basis dieser Verortung wurden anschließend technische (medieninhaltebezogene und basisinfrastrukturelle) Anforderungskategorien anhand eines generischen Content Workflows in Medienunternehmen abgeleitet. Abbildung 2.2.3/1 fasst die zentralen Ergebnisse aus Kapitel 2.2 noch einmal überblicksartig zusammen.

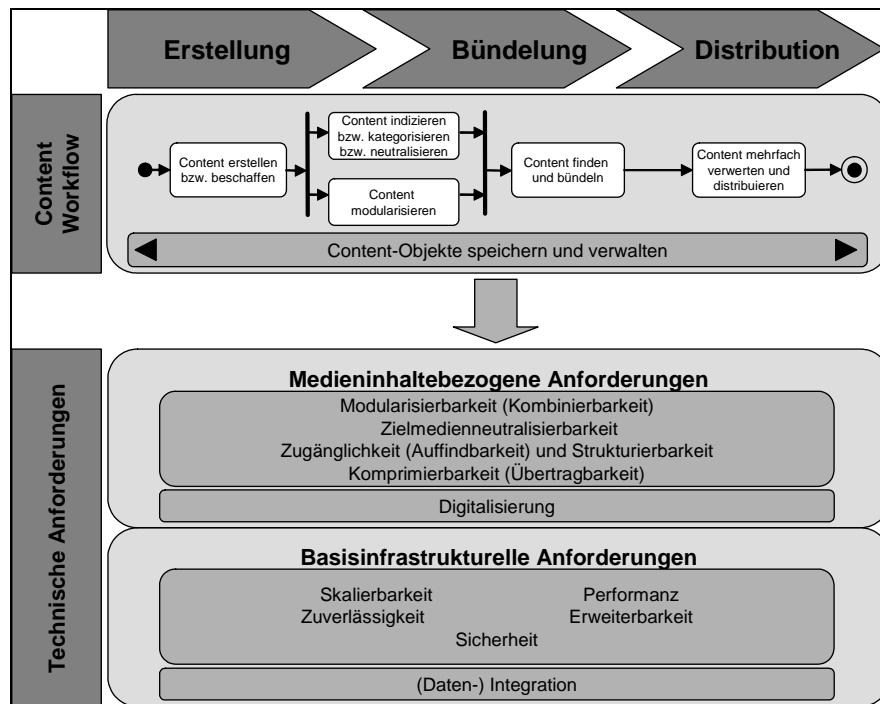


Abb. 2.2.3/1: Technische Anforderungskategorien an eine Mehrfachverwertung von Content

Ziel der Bestimmung technischer Anforderungen ist die Bereitstellung von Kategorien zur Bewertung unterschiedlicher integrierter Datenverteilungsszenarien. Diese sollen im nachfolgenden Kapitel systematisch abgeleitet werden.

3 Integrierte Verteilungsvarianten von Content und deren empirische Ausprägungen

Nach einer kurzen Begriffsklärung zentraler Termini zur Datenverteilung in Kapitel 3.1 werden integrierte Datenverteilungsvarianten zunächst auf Basis einfacher kombinatorischer Überlegungen abgeleitet (Kapitel 3.2). Anschließend werden zur empirischen Untermauerung der identifizierten Datenverteilungsvarianten in der Unternehmenspraxis eingesetzte (Multimedia) Content Management Systeme herangezogen, die aufgrund ihrer Systemarchitektur entsprechende Datenverteilungsszenarien nahe legen (Kapitel 3.3).

3.1 Datenverteilung: Eine Begriffsklärung

Um den Begriff *Datenverteilung* inhaltlich bestimmen zu können, ist eine Klärung der Unterschiede zwischen physischer und logischer Datenintegration vorzuschalten. Der Unterschied zwischen diesen Formen der Datenintegration liegt in der *Lokalisierung der Daten*. Physische Datenintegration basiert auf einem einheitlichen, zentralen Datenbanksystem, in dem alle relevanten Daten auch physisch vereint gespeichert werden. Logische Datenintegration erfordert zwar ein einheitliches Datenmodell und somit die Verknüpfung aller relevanten Daten. Allerdings müssen die Daten physisch nicht in ein und demselben Datenbanksystem gespeichert, sondern können mithin

auch dezentral gehalten werden (Fischer, 1999, S. 92). Bei verteilten Datenbanken sind somit, im Gegensatz zu zentralen Datenbanken, die logisch zusammengehörenden Daten physisch an verschiedenen, durch ein Netz verbundenen Orten gespeichert (vgl. Schwarze, 2000, S. 253). Unter Datenverteilungsvarianten werden vor diesem Hintergrund die Zentralisierung oder Dezentralisierung von Content (bzw. Metadaten) innerhalb eines organisatorischen Kontextes verstanden (siehe Abbildung 3.1/1).

Zustand	Integrierte Datenhaltung		Isolierte Datenhaltung
Vorgang, der oben genannten Zustand erreicht	Zentrale Integration (Verschmelzung)	Dezentrale Integration (Kopplung)	Dezentrale Isolation (Isolation)
Logische Integration	Ja	Ja	Nein
Physische Integration	Ja	Nein	Nein

Abb. 3.1/1: Unterschiede zwischen integrierter und isolierter Datenhaltung

Auf einem *abstrakten* Niveau meint Dezentralisierung dabei die Ausrichtung oder das Streben von einem Mittelpunkt (Zentrum) weg (zentrifugal), Zentralisierung die Konzentration auf einen Mittelpunkt (zentripetal) hin (vgl. Heinrich, 1999, S. 63). Zentralisierung von Content auf *technischer* Ebene meint die Speicherung von Content an einem zentralen Ort (lokal) innerhalb einer Organisation („physische und logische Datenintegration“). Die Dezentralisierung von Content bezeichnet schließlich die Speicherung und Vorhaltung von Content an mehreren Orten (global) in verteilten Organisationseinheiten (z.B. Fachabteilungen, Tochterunternehmen, Netzwerkpartner, etc.) („logische, aber nicht physische Datenintegration“). Dezentrale Contenthaltung erfordert somit implizit eine Kommunikationsschnittstelle, über welche eine logische Verknüpfung des Content erfolgen kann. Eine Schnittstelle bezeichnet dabei den Übergang an der Grenze von zwei Systemen mit den vereinbarten Regeln für die Übergabe von Daten (vgl. DIN 44 300, S. 141). Physische und geographische Zentralisierung bzw. Dezentralisierung können, müssen dabei nicht zusammenfallen (vgl. Dadam, 1996, S. 7f.; Mertens, 1985, S. 20). Zentralisierte Daten können auch an einem geographischen Ort physisch separat (in zwei oder mehreren Datenbanken) oder vereint (in einer Datenbank) abgespeichert werden.

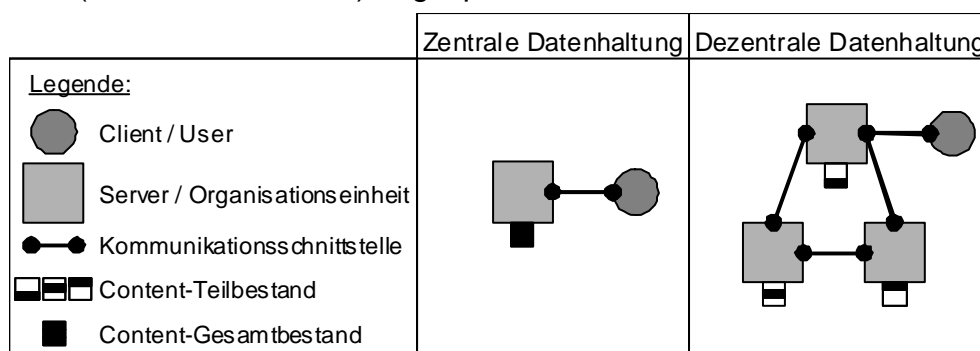


Abb. 3.1/2: Zentrale und dezentrale Datenhaltung

Die Dezentralisierung von Content lässt sich weiter in drei Dezentralisierungsarten untergliedern. Bei der *Dispersion* werden die Daten redundanzfrei auf verschiedene Rechner verteilt, bei der *Replikation* dagegen exakte Kopien von Content an anderen Netzknoten abgelegt. Bei der *Fragmentierung* oder *Partitionierung* werden disjunkte Teile von Content auf verschiedene Netzknoten verteilt (vgl. Rautenstrauch/Schulze, 2003, S. 159f.).

In nachfolgenden Ausführungen wird weiterhin immer von bereits integrierten Datenverteilungsszenarien ausgegangen.

3.2 Integrierte Datenverteilungsvarianten

Zur Ableitung möglicher Datenverteilungsszenarien in (Medien-)Unternehmen werden die Terminologie und der Systematisierungsrahmen der Entscheidungstheorie herangezogen. Die Entscheidungstheorie erschließt die Realität anhand von Entscheidungssituationen. Jede Art von Problem kann mittels eines Entscheidungsmodells erfasst werden (vgl. Laux, 2003). Bezüglich der Fragestellung, welche Datenhaltungsvarianten in (Medien-)Unternehmen bestehen, lassen sich als Entscheidungsobjekte Content-Objekte und Metadaten identifizieren. Als Entscheidungsalternativen bei der Datenhaltung kommen die beiden Antipoden „Zentralisierung und Dezentralisierung“ in Frage, die auf beide Entscheidungsobjekte angewandt werden können. Auf Basis der Kombinatorik sind die beiden Entscheidungsobjekte Content und Metadaten jeweils auf zwei Architekturausprägungen anzuwenden, so dass sich $2^2=4$ Möglichkeiten ergeben. Nachfolgende Tabelle illustriert die möglichen Varianten der Datenhaltung:

Alternative	Content	Metadaten	Bezeichnung der Entscheidungsalternative
A1	Z	Z	Zentralisierte Datenhaltung
A2	Z	D	Hybride, Content-zentralisierte DH
A3	D	Z	Hybride, Metadaten-zentralisierte DH
A4	D	D	Dezentralisierte Datenhaltung

Z = Zentralisierung; D = Dezentralisierung

Tab. 3.2/1: Varianten der Datenverteilung

Anhand der Dimensionen „Content-Verteilungsgrad“ und „Metadaten-Verteilungsgrad“ lassen sich die einzelnen Datenhaltungsvarianten positionieren (siehe Abbildung 3.2/1). Zwischen den vier Extremausprägungen finden sich zahlreiche komplexere Mischformen der Datenverteilung, von denen in den folgenden Ausführungen abstrahiert werden soll.

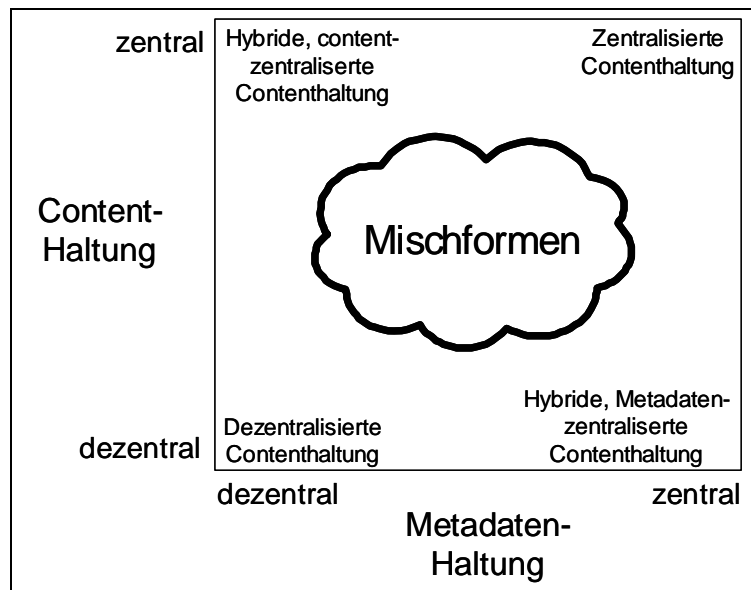


Abb. 3.2/1: Positionierung der Datenhaltungsszenarien

Bezieht man Organisationseinheiten und Anwendungssysteme in die Betrachtung mit ein⁹ und geht man von einer redundanzfreien Speicherung des Content aus (eine Dispersion des Content liegt vor¹⁰), lassen sich die vier identifizierten Datenverteilungsszenarien entlang eines Dezentralisierungskontinuums von Content und Metadaten ausrichten (siehe Abbildung 3.2/2). Datenverteilungsszenarien, die am Anfang des Dezentralisierungskontinuums stehen, lassen sich dabei einer Client-Server-Architektur zuordnen, Datenverteilungsszenarien am Ende des Dezentralisierungskontinuums entsprechen Peer-to-Peer-Architekturen. Während Client-Server-Architekturen also von einem hierarchischen Verhältnis zwischen den Netzwerkknoten bezüglich der Datenverteilung-, Funktions- und Aufgabenverteilung ausgehen (Zentralisierung der Daten auf dem Server), heben hybride und pure Peer-to-Peer-Architekturen diese strikte Rollenunterscheidung der einzelnen Netzwerkknoten (Peers) schrittweise auf (vgl. Minar, 2001). Da eine klare Abgrenzung zwischen Client-Server- und Peer-to-Peer-Architekturen nicht besteht, existiert ein Überlappungsbereich.

⁹ Aus Gründen der Vereinfachung wird angenommen, dass innerhalb einer betrachteten Organisation drei Organisationseinheiten informationstechnisch miteinander vernetzt sind.

¹⁰ Siehe Kapitel 3.1

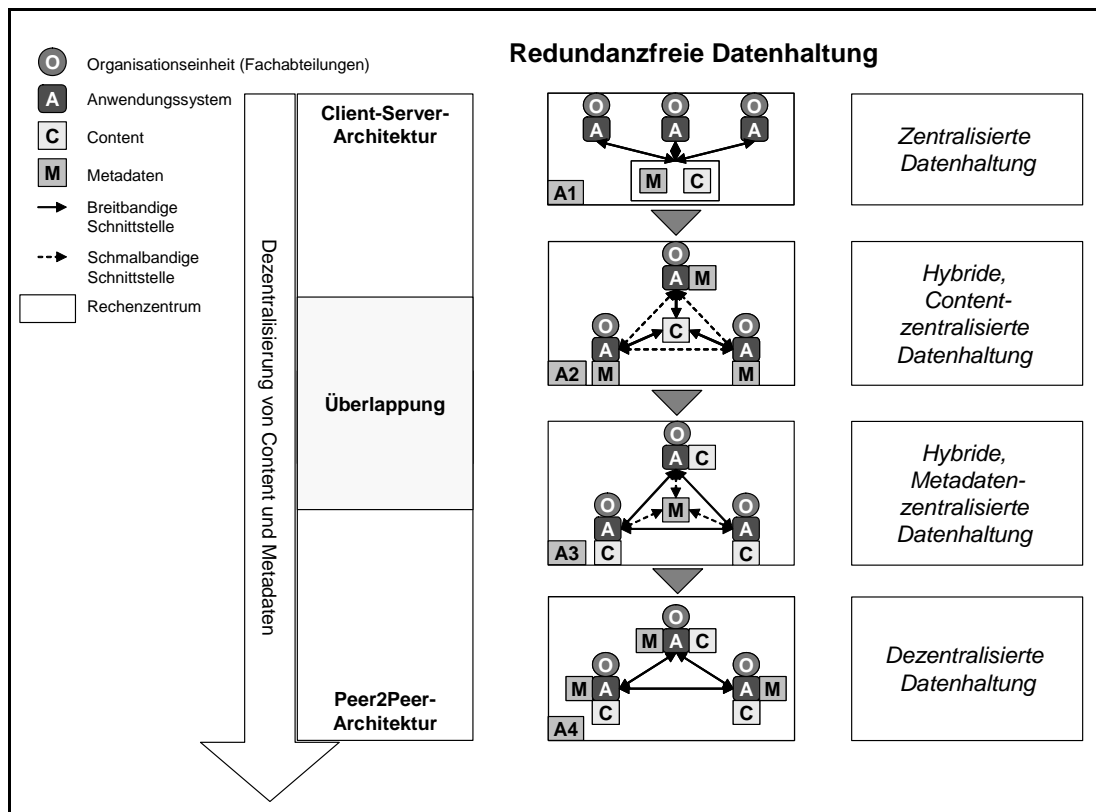


Abb. 3.2/2: Redundanzfreie, integrierte Datenverteilungsvarianten

Empirische Evidenz erhalten die identifizierten Datenverteilungsvarianten durch die Beschreibung von in der Praxis eingesetzten Content Management Systemen. Ihre Architektur kann als Indikator für die realisierte Datenverteilung in Unternehmen interpretiert werden.

3.3 Empirisch beobachtbare Architekturausprägungen

Als empirisch beobachtbare Ausprägungen von (Multimedia) Content Management Systemen mit unterschiedlicher Schwerpunktsetzung in der Daten- und Metadatenverteilung lassen sich exemplarisch drei in der Unternehmenspraxis eingesetzte Architekturen aufzeigen. Eine zentralisierte Contenthaltungsarchitektur bietet CoreMedias Smart Content Technology (vgl. Pakendorf, 2001, S. 726). Content und Metadaten werden während des Produktionsprozesses zentral im sog. Content Server gehalten und erst auf Basis konkreter Publikationsnachfragen auf den Live- und schließlich auf den Delivery-Server übertragen.

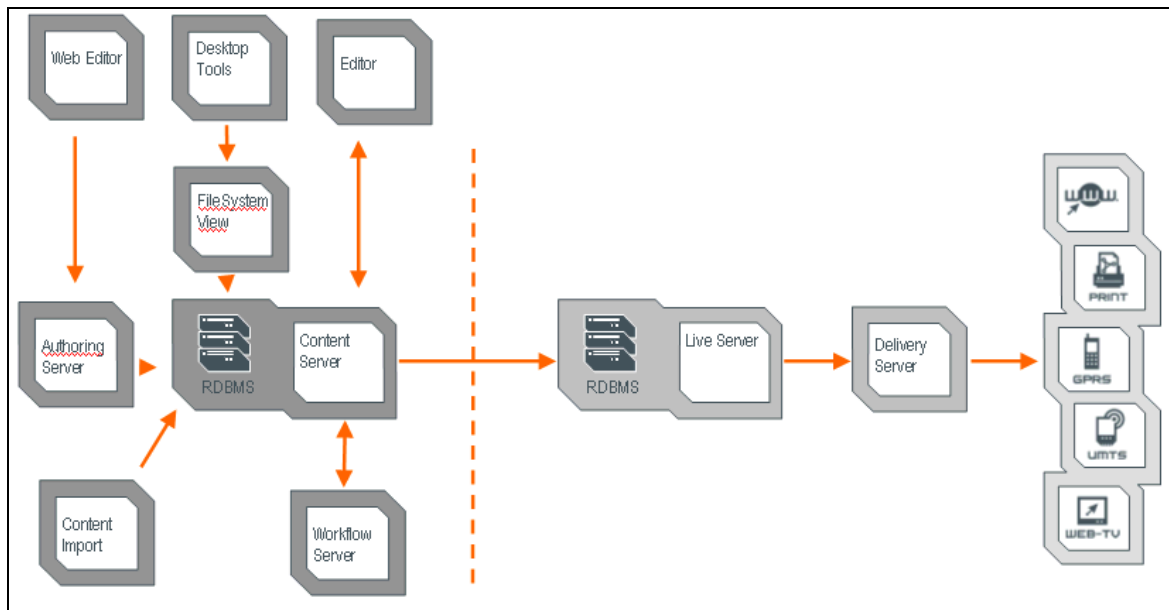


Abb. 3.3/1: Zentralisierte Datenhaltung: CoreMedia's Systemarchitektur

Eine hybride, metadaten-zentrierte Variante der Datenhaltung findet sich beispielsweise in der Dreiecksarchitektur des DB2 Content Manager von IBM wieder (vgl. Reimer, 2002, S. 20). Dabei werden dezentral gespeicherte Content-Objekte in sogenannten Resource Managern verwaltet. Inhaltsbeschreibende, interpretierende Metadaten, die in einem zentralen Library Server gehalten werden, stehen im Zentrum dieser triangularen Architektur (siehe Abbildung 3.3/2). Client-Anfragen werden dabei zuerst per SQL an den Library Server gerichtet (1), der daraufhin Ergebnistrefner samt Referenzen auf den eigentlichen Content zurückgibt (2). Erst nach diesen beiden Schritten kann der Client auf die gewünschten Medieninhalte zugreifen (3+4).

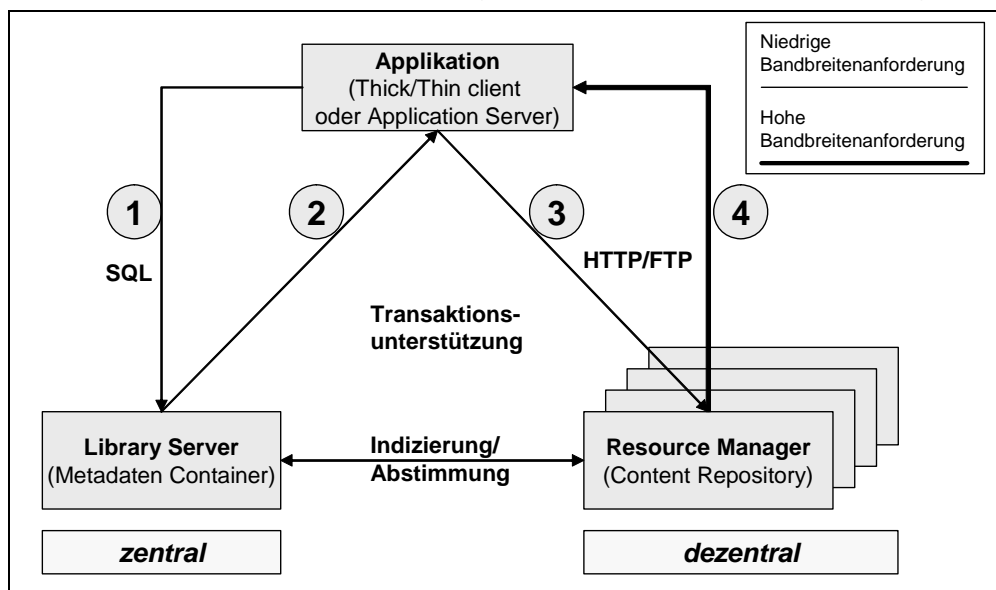


Abb. 3.3/2: Hybrid, metadatenzentrierte Datenhaltung: IBM DB2 Content Manager

Während die zuvor illustrierte Systemarchitektur auf einem hybriden Peer-to-Peer-Prinzip basiert, unterstützt das Dokumentenmanagementsystem „NextPage-NXT 4“

eine pure Peer-to-Peer-gestützte, dezentralisierte Content- und Metadatenhaltung (Schoder/Fischbach, 2003, S. 316). Die Indizierung und Kategorisierung der Daten wird dabei von jedem Peer nach individuellen Kriterien selbständig durchgeführt. Die Suche erfolgt dabei nach dem Schnellballprinzip (vgl. Ripeanu et al., 2002). Eine Suchanfrage wird an eine bestimmte Anzahl von Peers weitergeleitet (siehe Abbildung 3.3/3). Falls diese die Anfrage nicht selbst bedienen können, leiten sie diese wiederum an verschiedene Knoten weiter, bis eine zuvor bestimmte Suchtiefe erreicht oder die gewünschte Datei gefunden wird. Positive Suchergebnisse werden dann an den Nachfrager gesendet, sodass dieser die gewünschte Datei direkt von dem Anbieter herunterladen kann.

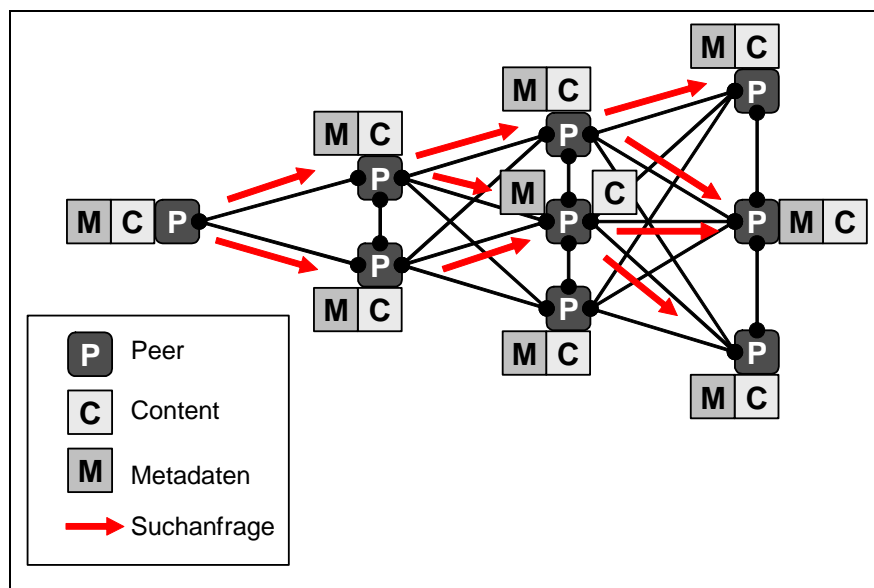


Abb. 3.3/3: Dezentralisierte Datenhaltung: NXT 4-Plattform von NextPage

Nachfolgend sollen die in Kapitel 3 identifizierten Extremvarianten zentralisierter und dezentralisierter Datenhaltung aus technischer Sicht gegenüber gestellt und bewertet werden.

4 Bewertung von Datenverteilungsvarianten aus technischer Sicht

Die Bewertung der in Kapitel 3.2 identifizierten Datenverteilungsvarianten lässt sich aus technischer Sicht vor dem Hintergrund der in Kapitel 2 identifizierten medieninhaltebezogenen (4.1) und basisinfrastrukturellen (4.2) Anforderungen an eine Mehrfachnutzung von Medieninhalten beleuchten. Aus Gründen der Vereinfachung beschränkt sich die Diskussion über die Vorteilhaftigkeit auf zentrale und dezentrale Datenhaltungsvarianten.

Der Diskussion voranstellen lassen sich dabei grundlegende Unterschiede zwischen beiden Datenhaltungsvarianten. Mit der Verteilung bzw. Dezentralisierung von Daten geht einerseits eine höhere Komplexität aufgrund der Entstehung von Heterogenität und der Notwendigkeit zum Aufbau und zur Standardisierung von Schnittstellen ein-

her, möchte man die verteilten Daten wieder logisch zusammenführen. Andererseits bietet Verteilung eine größere Flexibilität in der Bereitstellung von Daten, da das Gesamtsystem bei teilredundanter Datenhaltung nicht von einem Zentrum abhängig ist.

4.1 Erfüllung medieninhaltebezogener Anforderungen

4.1.1 Zentrale vs. dezentrale Datenhaltung aus technischer Sicht

Eine *Modularisierung* von Content lässt sich grundsätzlich sowohl in zentralen als auch in dezentralen Datenhaltungsszenarien vornehmen. Aufgrund der höheren Abstimmungs- bzw. Transaktionskosten (z.B. in Form von beanspruchten Übertragungskapazitäten oder Suchkosten) sowie aufgrund multipler Schnittstellen stellt sich die Administration und Koordination der Module (z.B. Bündelung) in der dezentralen Datenhaltungsvariante allerdings als komplexer dar. In verteilten Szenarien müssen bei der Ausnutzung von Mehrfachverwertungspotenzialen Medieninhalte, Metadaten und Lokalisierungsinformationen über das Netz geschickt werden. Bevor eine Modularisierung stattfindet, muss zudem an jedem Netzknoten kontrolliert werden, welche Inhalte mit Inhalten anderer Netzknoten kombiniert werden können. In extremen Situationen müssen Netzknoten $\frac{n(n-1)}{2}$ Kommunikationsverbindungen aufbauen. Werden lokal produzierte Medieninhalte auch nur lokal verwendet (z.B. direkte Distribution an Kunden), d.h. es bedarf keiner weiteren Bündelung, ist eine dezentrale Datenhaltung technisch effizienter, da die Medieninhalte nicht erst von einem zentralen Server auf den lokalen Rechner übertragen werden müssen (vgl. Jain et al., 1998, S. 5).

Auch bezüglich einer *Zielmedienneutralisierung* der Medieninhalte (z.B. Datenhaltung in XML) weisen zentrale Datenhaltungsszenarien vor dem Hintergrund einer Mehrfachverwertung von Medieninhalten Vorteile auf. Die zentrale Speicherung von medienneutralen Daten ermöglicht einerseits einen schnelleren Zugriff auf alle relevanten Daten für eine Bündelung und andererseits eine geringere Netzbelastung. Ferner treten bei einer Zusammenführung verteilter Daten häufig Inkonsistenzen aufgrund einer Nichtbeachtung von Standards (z.B. Dateiformate, Auflösung, etc.) auf, was zu zusätzlichen Konvertierungskosten führen kann.

Dezentrale Datenhaltungsvarianten weisen aufgrund höherer Suchkosten bei einer Verteilung von Content und Metadaten somit auch Nachteile bezüglich der *Zugänglichkeit und Auffindbarkeit* von Content auf. Gerade diese höheren Suchkosten begründen auch die Vorteilhaftigkeit zentraler Datenhaltung bei den ersten beiden medieninhaltebezogenen Anforderungen. Da es in verteilten Datenhaltungsszenarien zudem zu Inkonsistenzen aufgrund von Datenredundanzen kommen kann, ist das

Management verteilter Metadaten besonders komplex. Aufgrund dieses zentralen Nachteils haben sich hybrid, metadaten-zentrierte Systeme entwickelt, welche eine zentrale semantische Integrationskomponente aufweisen und für eine schnellere Zugänglichkeit zu Content-Objekten sorgen.

Die *Komprimierung* von Medieninhalten wird aufgrund der Beanspruchung größerer Übertragungskapazitäten – bei sonst gleichen Bedingungen (z.B. Größe des übertragenen Contents) – insbesondere in verteilten Datenhaltungsszenarien relevant. Damit sind zentrale Systeme unabhängiger von der Komprimierung von Medieninhalten und damit flexibler gegenüber Kapazitätsänderungen bei vergleichbaren Hardwarebedingungen (vgl. Kretschmar/Dreyer, 2004). Da die Leistungsfähigkeit von Kommunikationsnetzwerken in den letzten Jahren im Vergleich zur Leistungsfähigkeit von Prozessen überproportional gestiegen ist, scheint sich das durch Kommunikationsnetzwerke verursachte Performanzproblem von verteilten Systemen zunehmend abzumildern.

4.1.2 Zusammenfassung

Zentrale Datenhaltungsszenarien erfüllen die medieninhaltebezogenen Anforderungen an eine Mehrfachverwertung von Medieninhalten durchwegs besser als dezentrale Datenhaltungsvarianten. Grund hierfür ist, dass sich bei zentraler Datenhaltung schnellere Synergieeffekte zwischen Content-Modulen ergeben können, da weniger Kommunikationsschnittstellen bei der Übertragung des Content-Bestandes notwendig und die Koordination und Administration des gesamten Medieninhaltebestands weniger komplex sind. Ferner können in einer zentralen Datenbank Bündelungspotenziale zwischen Content-Modulen kontinuierlich überprüft werden, was in dezentralen Datenarchitekturen nur mit extrem hohen Übertragungskosten realisierbar wäre.

4.2 Erfüllung basisinfrastruktureller Anforderungen

Eine Bewertung der Erfüllung basisinfrastruktureller Anforderungen an eine Mehrfachnutzung von Medieninhalten durch zentralisierte bzw. dezentralisierte Datenhaltung ist eng verwoben mit der Fragestellung der Verteilung der Systemkomponenten innerhalb einer Informationssystemarchitektur. Die nachfolgende Argumentation stützt sich aus diesem Grund unter anderem auf Vor- und Nachteile zentraler vs. dezentraler Systemarchitekturen (z.B. Vor- und Nachteile von Peer-to-Peer-Systemen bzw. Verteilten Systemen).

4.2.1 Zentrale vs. dezentrale Datenhaltung aus technischer Sicht

Anhand der Diskussion der Vor- und Nachteile dezentraler Systeme lassen sich gleichzeitig die Vor- und Nachteile zentraler Systeme mit anführen.

Vorteile dezentraler Datenhaltung

Verteilte Datenhaltung weist bezüglich der *Skalierbarkeit* gegenüber zentraler Datenhaltung den Vorteil auf, dass eine Lastverteilung auf verteilte Netzwerkknoten ermöglicht wird (vgl. Weber, 1998, S. 2, 3). Die DV-Ressourcen sind modular anpassbar und können somit eine größere Anzahl gleichzeitiger Systemanfragen bearbeiten (vgl. Fischer, 1999, S. 194; Schwarzer/Krcmar, 1999, S. 56).

Ein großes, verteiltes System kann aufgrund des Charakteristikums des Ressourcen- bzw. Betriebsmittel-Sharings zudem mehr aggregierte Rechenleistung und somit eine größere *Verarbeitungsperformanz* besitzen als ein Großrechner bzw. Server in einer zentralistischen Client-Server-Architektur. Da Prozessorlast über ein gesamtes System verteilt werden kann, ist der einzelne Rechner entlastbar (Load Sharing/Balancing). Die Effizienz der Bearbeitung einer Aufgabe in einem verteilten System, im Vergleich zu unabhängigen Einzelrechnern, hängt im Allgemeinen von der Granularität der nebenläufig oder parallel ausgeführten Programmportionen ab. Dabei nimmt der Kommunikationsaufwand mit steigender Feinheit der Granularität stark zu (vgl. Schoder/Fischbach, 2003, S. 313).

Ein zentraler Vorteil verteilter und gegebenenfalls redundanter Datenhaltung ist die größere *Zuverlässigkeit* bezüglich der Systemverfügbarkeit¹¹. Fällt ein Peer-Rechner im Netzwerk aus, können andere Peer-Rechner die Datenversorgung aufrechterhalten. Die lokale Replikation von Content trägt somit zu einer erhöhten *Ausfallsicherheit* und Systemverfügbarkeit und gegebenenfalls kürzeren Systemantwortzeiten bei (vgl. Tanenbaum/van Steen, 2002, S. 5; Fischer, 1999, S. 194; Lee et al., 1995, S. 41).

Schließlich ist im Zusammenhang mit verteilten Systemen eine *flexiblere Erweiter- und Änderbarkeit* der Datenbasis zu erwähnen. Da mit der Verteilung von Daten eine Loslösung von einer zentralen Instanz einhergeht, können Daten an dezentralen Punkten effizienter verändert, angepasst und schließlich auch aktueller gehalten werden (Diekmann/Hagenhoff, 2003, S. 6).

Nachteile dezentraler Datenhaltung

Durch das Charakteristikum der Dezentralität erhöht sich grundsätzlich die *Komplexität* der Administration der Datenhaltung (z.B. Datenkonsistenz- und Qualitätssicherung des Content), da die Speicherorganisation und -zugriffsgeschwindigkeit sowie die Verbindungsstrukturen als auch die Kommunikations- und Synchronisationsmechanismen optimiert werden müssen. Letztliches Ziel muss es dabei sein, die Orts- und Datenmodellunabhängigkeit der Anwendungsprogramme, das sog. „Single-

¹¹ Werden z.B. anstatt einem drei File-Server mit einer Verfügbarkeitswahrscheinlichkeit von 0,95 eingesetzt, erhöht sich die Gesamtsystemverfügbarkeit von 0,95 auf 0,999875, da die Wahrscheinlichkeit sinkt, dass alle vier File-Server nicht verfügbar sind ($0,05^3 = 0,000125$).

System-Image“ zu erhalten (vgl. Dadam, 1996, S. 5). Verteilte Einheiten sind aus diesen Gründen tendenziell störungsanfälliger und wartungsintensiver (vgl. Fischer, 1999, S. 194). Die durch die Verteilung von Daten häufig entstehende *Heterogenität* (z.B. kann sich Heterogenität auf die für die Kommunikation benötigten Netzwerke, auf die Hardware und die Betriebssysteme der Rechner beziehen) erfordert zudem eine Standardisierung der heterogenen Systemelemente, um eine integrierte und transparente¹² Kommunikationsfähigkeit des Systems aufrechterhalten zu können (vgl. Jain et al., 1998, S. 5). Um die Komplexität der Verteilung automatisiert zu reduzieren, werden zur Unterstützung des Systemmanagements zwischen der (verteilten) Anwendungsschicht und den Betriebssystemen in Netzwerken sog. Middleware-dienste eingesetzt (vgl. Riehm/Vogler, 1996, S. 38), welche eine Aufeinander-Abstimmung zwischen Schnittstellen (z.B. Gewährleistung der Datenintegrität, und -sicherheit) im Netzwerk ermöglichen.

Ein erheblicher Nachteil dezentraler Datenhaltung ist die Abhängigkeit der verteilten Gesamtarchitektur von der Verfügbarkeit des zugrundeliegenden Netzwerks. Netze können Nachrichten verlieren oder können überlastet sein, so dass einzelne Netzwerkknoten eventuell nicht verfügbar sind (vgl. Tanenbaum/van Steen, 2002; Coulouris et al., 2000). Dies hätte nicht nur eine geringere *Kommunikationszuverlässigkeit*, sondern auch zur Folge, dass auf der organisatorischen Ablaufebene bei einem Ausfall kritischer Netzknoten der ganze Prozess zum Stocken gerät. Aus diesem Grund nimmt eine Netzwerküberwachung mittels *Performanz-Parametern* (z.B. *Latenz*¹³, *Datentransferrate*¹⁴ und *Netzwerkbandbreite*¹⁵) einen hohen Stellenwert bei der Administration verteilter Systeme ein.

Als ein weiterer Nachteil kann die *Performanz* dezentraler Systeme darunter leiden, dass extrem hohe Netzwerkübertragungskapazitäten in Anspruch genommen werden, da z.B. Suchanfragen an viele verbundene Peers geschickt und hohe Koordinationskosten zur Vermeidung von Inkonsistenzen in Kauf genommen werden müssen (vgl. Gehrke/Schumann, 2003, S. 2). Die Entwicklung von optimierten Suchtechnologien (wie z.B. verteilte Hash-Tabellen) für pure dezentrale Peer-to-Peer-Strukturen weist allerdings daraufhin, dass die Performanz durch eine Minimierung der Netzwerkbelastung kontinuierlich optimiert werden kann (vgl. Steinmetz/Wehrle, 2004).

¹² Durch Transparenzmechanismen soll dem Anwender ein verteiltes System als „ein Ganzes“ erscheinen. In der Literatur wird dabei zwischen Zugriffs-, Lokations-, Replikations-, Nebenläufigkeits-, Parallelitäts- und Fehler-Transparenz unterschieden (vgl. z.B. Tanenbaum/van Steen 2002, S. 23).

¹³ Unter dem Begriff Latenz versteht man die reale oder wahrgenommene Antwortzeit eines Systems auf ein Signal von einem anderen Punkt im Netzwerk.

¹⁴ Die Datentransferrate als Maß für die Geschwindigkeit der Datenübertragung wird in Kbyte oder Mbyte pro Sekunde angegeben.

¹⁵ Die Netzwerkbandbreite gibt das totale Durchsatzvolumen an, das durch ein Netzwerk in einer bestimmten Zeit transferiert werden kann.

Schließlich bieten dezentrale Systeme eine größere Angriffsfläche als zentrale Systeme. Durch die Verteilung der Systemkomponenten und dem erhöhten Kommunikationsaufkommen, stellt sich der *Datenschutz und die Betriebssicherheit* tendenziell schwieriger dar als in zentralisierten Systemen (vgl. Fischer, 1999, S. 194).

4.2.2 Zusammenfassung

Zentralisierte und dezentralisierte Datenhaltungsvarianten halten sich bezüglich der Erfüllung basisinfrastruktureller Anforderungen an die Mehrfachnutzung von Medieninhalten tendenziell die Waage. Verteilte Datenhaltungsszenarien zeichnen sich insbesondere durch flexible Skalier- und Erweiterbarkeit aus und bieten Zuverlässigkeit durch Ausfallsicherheit und Fehlertoleranz. Die Vorteile zentralisierter Datenhaltung liegen insbesondere in der weniger komplexen Administration des Contentbestandes, da keine Heterogenität inkompatibler Systemkomponenten aufgrund von Verteilung vorliegt. Eine Folge des Nichtvorhandenseins von Verteilung ist zudem, dass Kommunikationsunterbrechungen zwischen beteiligten Netzknoten aufgrund von Netzüberlastung oder -nichtverfügbarkeit nicht auftreten können.

In Abbildung 4.2.2/1 werden die Gesamtergebnisse der technischen Bewertung zentraler und dezentraler Datenhaltung, welche aufgrund der vorangegangenen abstrakten Argumentation grundsätzlich nur Tendenzaussagen sein können, noch einmal überblicksartig zusammengefasst.

Technische Anforderungen		Idealtypische Verteilungsszenarien	
		Zentralisierte Datenhaltung	Dezentralisierte Datenhaltung
Medieninhaltebezogene Anforderungen	Modularisierung	↑	↓
	Zielmedienneutralisierung	↑	↓
	Strukturierung (Zugänglich- und Auffindbarkeit)	↑	↓
	Komprimierung	↑	↓
Basisinfrastrukturelle Anforderungen	Skalierung	↓	↑
	Zuverlässigkeit	→	→
	Performanz	→	→
	Erweiterbarkeit	↓	↑
	Sicherheit	→	→

↑ = Vorteile überwiegen, → = Vor- und Nachteile halten sich in etwa die Waage, ↓ = Nachteile überwiegen

Tab. 4.2.2/1: Technische Bewertung unterschiedlicher Datenverteilungsvarianten

Resümierend kann festgehalten werden, dass eine zentralisierte Datenhaltung bezüglich einer Mehrfachnutzung von Medieninhalten rein technisch gesehen als vorteilhaft erscheint, da insbesondere die medieninhaltebezogenen Anforderungen effizienter erfüllt werden können. Dies liegt wohl auch darin begründet, dass man im Szenario der Mehrfachnutzung von Medieninhalten danach strebt, organisatorisch verstreut gespeicherte Content-Module zu zentralisieren, um Synergiepotenziale aus der mehrfachen Rekonfiguration und Bündelung auszunutzen. Zu einem ähnlichen Ergebnis kamen auch Jain et al., die in einer empirischen Studie nachweisen konnten, dass Unternehmen mit einer hohen Dateninterdependenz zwischen Organisationseinheiten erfolgreicher bezüglich des Datenmanagements sind, wenn sie IS-Ressourcen (u.a. auch Daten) zentral halten (vgl. Jain et al., 1998, S. 18).

5 Zusammenfassung und Ausblick

In vorliegendem Beitrag wurde versucht, integrierte Daten- bzw. Contentverteilungsvarianten in (Medien-)Unternehmen systematisch abzuleiten und vor dem Hintergrund der Mehrfachnutzung von Content technisch zu bewerten. Aufbauend auf der Abgrenzung der multimedialen von der klassisch-monomedialen Datenintegration wurden zunächst technische Anforderungen an multimediale Informationssysteme zur Datenhaltung und -bereitstellung spezifiziert. Anschließend wurden auf Basis kombinatorischer Überlegungen vier Grundtypen der Datenverteilung (genauer: von Content und Metadaten) abgeleitet, die sich in einem nächsten Schritt in unterschiedliche Datenverteilungsszenarien konkretisieren ließen. Ziel dieser Überlegungen war es, einen Beitrag zur Systematisierung und Konkretisierung von multimedialen Datenhaltungslandschaften zu leisten, die in dieser Form in der einschlägigen Literatur bisher vernachlässigt wurden. Anhand von einzelnen (Multimedia) Content Management Systemen, die gemäß ihrer Architektur spezielle Datenverteilungsszenarien nahe legen, konnten einzelne multimediale Datenverteilungsvarianten empirisch bestätigt werden. Im abschließenden Kapitel 4 wurden die in Kapitel 3 abgeleiteten Datenverteilungsarchitekturen nach ihrem Erfüllungsgrad der in Kapitel 2 identifizierten technischen Anforderungen diskutiert und schließlich bewertet.

Die Gegenüberstellung und Bewertung zentraler und dezentraler Datenhaltungsvarianten haben gezeigt, dass eine Mehrfachnutzung von Medieninhalten von einem rein technischen Standpunkt aus gesehen effizienter durch zentralisierte Systeme unterstützt werden kann. Dies lässt sich letztlich wohl darauf zurückführen, dass eine Mehrfachnutzung produktionsseitig auf eine Bündelung und damit eine Konzentration der Medieninhalte abzielt. Eine verteilte bzw. dezentrale Datenhaltung würde während des Produktionsprozesses aufgrund der höheren Transaktionskosten (Such-, Abstimmungs-, Übertragungs- und Wartungskosten) somit die zweite Wahl bedeuten. Bisherige technische Restriktionen – wie etwa mangelnde Übertragungskapazitäten und Speicherplatz sowie durch Heterogenität bedingte Koordinationskomplexität zwischen verteilten Systemkomponenten – werden durch ökonomische (z.B. Preisverfall) und technische Entwicklungen (z.B. immer effizienter arbeitende Middleware) allerdings immer mehr aufgehoben. Dadurch entstehen größere Freiheitsgrade für eine effiziente technische Verteilung und damit auch Bündelung von Content. Nicht zuletzt diese Entwicklungen können damit Auswirkungen auf die technischen Transaktionskosten der Datenverteilung haben.

Eine Architekturentscheidung auf Basis einer bloßen technischen Bewertung der Verteilung von Daten greift jedoch zu kurz. Neben technischen Gesichtspunkten sind insbesondere organisatorische (z.B. Spezialisierungs- und Motivationseffekte aufgrund von Verteilung) und strategisch-ökonomische Faktoren (z.B. Kosten-Nutzen-

Betrachtung von Content-Outsourcing als Form von Ressourcendekonzentrierung) in die Bewertung miteinzubeziehen. Die Betrachtung der Datenverteilungsproblematik aus organisatorischer und strategisch-ökonomischer Perspektive zur Gewinnung von Erklärungs- und Gestaltungsbeiträgen stellt daher den nächsten Schritt der diesem Beitrag zugrunde liegenden Forschungskonzeption dar. Langfristiges Ziel des Forschungsvorhabens ist es dabei, die Kontingenzfaktoren einer Verteilung von Content aus einer möglichst ganzheitlichen Sicht zu bestimmen, um letztlich situationspezifische (normative) Handlungsempfehlungen für Architekturentscheidungen ableiten zu können.

6 Literaturverzeichnis

- Anding, M. (2004): Online Content Syndication - Theoretische Fundierung und praktische Ausgestaltung eines Geschäftsmodells der Medienindustrie, Gabler, Wiesbaden.
- Anding, M., Hess, T. (2003): Was ist Content? Zur Definition und Systematisierung von Medieninhalten, Arbeitsbericht: 5/2003, Seminar für Wirtschaftsinformatik und Neue Medien der Ludwig-Maximilians-Universität München, Arbeitsbericht Nr. 5/2003, München
- Anding, M., Köhler, L., Hess, T. (2003): Produktplattformen für Medienunternehmen - ein konzeptioneller Rahmen, in: Uhr, W., et al.: Wirtschaftsinformatik 2003 / Band I: Medien - Märkte - Mobilität, Dresden, Physica-Verlag, S. 541-560.
- Balzert, H. (2000): Lehrbuch der Software-Technik. Band I: Software-Entwicklung, Spektrum, Heidelberg, Berlin, Oxford.
- Balzert, H. (1998): Lehrbuch der Software-Technik. Software-Management, Software-Qualitätssicherung, Unternehmensmodellierung, Spektrum Akademischer Verlag, Heidelberg, Berlin.
- Bechtold, S. (2002): Vom Urheber- zum Informationsrecht, Verlag C.H. Beck, München.
- Benlian, A., Wiedemann, F., Hess, T. (2004): KeyTE_x - An integrated prototype for semi-automatic metadata assignment and network-based content retrieval, in: Proceedings of the 10th Americas Conference on Information Systems (AMCIS 2004), New York, NY, (forthcoming).
- Breyer-Mayländer, T., Werner, A. (2003): Handbuch der Medienbetriebslehre, Oldenbourg Verlag, München.
- Bunjes, B., Friebe, J., Götze, R., Harren, A. (2002): Integration von Daten, Anwendungen und Prozessen am Beispiel des Telekommunikationsunternehmens EWE TEL, in: Wirtschaftsinformatik, 44, 5, S. 415-423.
- Chen, P. (1976): The Entity-Relationship Model - Towards a Unified View of Data, in: ACM Transactions on Database Systems, 1, 1, S. 9-36.
- Coulouris, G., Dollimore, J., Kindberg, T. (2000): Distributed Systems, Addison Wesley, Reading, Massachusetts.
- Dadam, P. (1996): Verteilte Datenbanken und Client/Server-Systeme. Grundlagen, Konzepte, Realisierungsformen, Springer, Berlin et al.
- Diekmann, T., Hagenhoff, S. (2003): Verteilte Systeme: State of the Art, Arbeitsbericht: Arbeitsbericht (1/2003) des Instituts für Wirtschaftsinformatik der Georg-August-Universität, Göttingen

- Fetscherin (2002): Present state and emerging scenarios of Digital Rights Management systems, in: *Journal of Media Management*, 4, 3, S. 164-171.
- Fischer, J. (1999): *Informationswirtschaft: Anwendungsmanagement*, Oldenbourg, München, Wien.
- Fluit, C., Sabou, M., van Harmelen, F. (2002): *Ontology-based Information Visualisation*, in: Geroimenko, V.: *Visualising the Semantic Web*, Springer Verlag.
- Gehrke, N., Schumann, M. (2003): *Constructing Electronic Marketplaces using Peer-to-Peer Technology*, in: *Peer-to-Peer Computing Minitrack in the Internet and Digital Economy Track of the Thirty-Sixth Hawaii International Conference on System Sciences (HICSS-36)*, Hawaii.
- Grady, R. B. (1992): *Practical Software Metrics for Project Management and Process Improvement*, P T R Prentice Hall, Englewood Cliffs, NJ.
- Heilmann, H. (1989): *Integration: Ein zentraler Begriff der Wirtschaftsinformatik im Wandel der Zeit*, in: *HMD*, 150, S. 46-58.
- Heine, P. (2000): *Integrierte Datenlogistik bei Energieversorgungsunternehmen*, in: *Wirtschaftsinformatik*, 42, 5, S. 434-442.
- Heinrich, L. J. (1999): *Informationsmanagement. Planung, Überwachung und Steuerung der Informatinsinfrastruktur*, Oldenbourg, München, Wien.
- Holten, R. (2003): *Integration von Informationssystemen*, in: *Wirtschaftsinformatik*, 45, 1, S. 41-52.
- Jain, H., Ramamurthy, K., Ryu, H.-S., Yasai-Ardekani, M. (1998): *Success of data resource management in distributed environments: an empirical investigation*, in: *MIS Quartely*, 22, 1, S. 1-29.
- Köhler, L., Anding, M., Hess, T. (2003): *Exploiting the power of product platforms for the media industry – a conceptual framework for digital goods and its customization for content syndicators*, in: *Proceedings of the third IFIP Conference on e-commerce, e-business and e-government*, Sao Paulo, Kluwer, S. 303 - 313.
- Krcmar, H. (2003): *Informationsmanagement*, Springer Verlag, Berlin et al.
- Krcmar, H. (1991): *Integration in der Wirtschaftsinformatik - Aspekte und Tendenzen*, in: Jacob, H., et al.: *Integrierte Informationssysteme*, Gabler Verlag, Wiesbaden, S. 3-18.
- Kretzschmar, O., Dreyer, R. (2004): *Medien-Datenbank- und Logistik-Systeme. Anforderungen und praktischer Einsatz*, München, Wien, Oldenbourg.
- Laux, H. (2003): *Entscheidungstheorie*, Springer, Berlin et al.

- Lee, A., Cheng, C. H., Chadha, G. S. (1995): Synergism between information technology and organizational structure: a managerial perspective, in: Journal of Information Technology, 10, 1, S. 37-43.
- Martin, J. A., Eisenhardt, K. M. (2001): Exploring cross-business synergies, in: Academy of Management Proceedings, S. H1-7.
- Mertens, P. (1985): Aufbauorganisation der Datenverarbeitung. Zentralisierung - Dezentralisierung - Informationszentrum, Gabler, Wiesbaden.
- Mertens, P. (2001): Integrierte Informationsverarbeitung 1 - Administrations- und Dispositionssysteme in der Industrie, Gabler, Wiesbaden.
- Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M., Hess, T. (2004): Grundzüge der Wirtschaftsinformatik, Berlin et al.
- Minar, N. (2001): Distributed Systems Topologies, in:
http://www.openp2p.com/pub/a/p2p/2001/12/14/topologies_one.html, letzter Zugriff: 13.01.2004
- Österle, H. (1996): Integration: Schlüssel zur Informationsgesellschaft, in: Österle, H., et al.: Middleware. Grundlagen, Produkte und Anwendungsbeispiele für die Integration heterogener Welten, Vieweg, Braunschweig, Wiesbaden, S. 1-23.
- Pakendorf, M. (2001): The CoreMedia Process for Content Applications, in: Bauknecht, K., et al.: Tagungsband der 31. GI/OCG-Jahrestagung - Band 2, Wien, S. 725-732.
- Pankratz, G., Benlian, A. (2004): Konstruktion eines Referenzmodells für das Online Content Syndication auf Basis einer Geschäftsmodellanalyse, in: Tagungsband zur 8. Fachtagung Referenzmodellierung im Rahmen der Multikonferenz Wirtschaftsinformatik, Essen/Duisburg, S. 125-149.
- Rautenstrauch, C., Schulze, T. (2003): Informatik für Wirtschaftswissenschaftler und Wirtschaftsinformatiker, Springer, Berlin et al.
- Rawolle, J. (2002): Content Management integrierter Medienprodukte. Ein XML-basierter Ansatz, Deutscher Universitätsverlag, Wiesbaden.
- Reimer, J. A. (2002): Enterprise Content Management, in: Datenbank Spektrum, 2, 4, S. 17-35.
- Reinhard, K., Rauchfuß, F. (2004): Cross-Media Publishing & Management - Erfolgreich (re-)agieren auf veränderte Anforderungen an die Medienbranche, in: Ottler, S., Radke, P.: Aktuelle Strategien von Medienunternehmen. Ergebnisse der Ravensburger Mediengespräche, Verlag Reinhard Fischer, München, S. 161-174.
- Riehm, R., Vogler, P. (1996): Middleware: Infrastruktur für die Integration, in: Österle, H., et al.: Middleware. Grundlagen, Produkte und Anwendungsbeispiele für die

- Integration heterogener Welten, vieweg, Braunschweig, Wiesbaden, S. 25-136.
- Ripeanu, M., Foster, I., Iamnitchi, A. (2002): Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design, in: IEEE Internet Computing Journal, 6, 1, S. 51-57.
- Schmitt, I. (2002): Retrieval in Multimedia-Datenbanksystemen, in: Datenbank Spektrum, 2, 4, S. 28-35.
- Schoder, D., Fischbach, K. (2003): Peer-to-Peer-Netzwerke für das Ressourcenmanagement, in: Wirtschaftsinformatik, 45, 3, S. 313-323.
- Schulze, B., Hess, T. (2003): Mehrfachnutzung von Inhalten als Synergie-Ansatz in der Medienindustrie: Ökonomische und technologische Grundlagen von derzeit bekannten Varianten, Arbeitsbericht: 4/2003, Institut für Wirtschaftsinformatik und Neue Medien der Ludwig-Maximilians-Universität, München
- Schumann, M. (2003): Möglichkeiten der Mehrfachverwertung von Inhalten bei Medienunternehmen - technische, organisatorische und ökonomische Betrachtung, in: Ehrenberg, D.,H.-J., K.: Herausforderungen der Wirtschaftsinformatik in der Informationsgesellschaft, Leipzig (im Druck).
- Schumann, M., Hess, T. (2002): Grundfragen der Medienwirtschaft, Springer-Verlag, Berlin, Heidelberg, New York.
- Schwarze, J. (2000): Einführung in die Wirtschaftsinformatik, Herne, Berlin.
- Schwarzer, B., Krcmar, H. (1999): Wirtschaftsinformatik. Grundzüge der betrieblichen Datenverarbeitung, Schäffer-Poeschel, Stuttgart.
- Somani, A., Choy, D., Kleewein, J. C. (2002): Bringing together content and data management systems: Challenges and opportunities, in: IBM Systems Journal, 41, 4, S. 686-696.
- Stahlknecht, P. (1995): Einführung in die Wirtschaftsinformatik, Springer, Berlin et al.
- Stamer, S. (2002): Technologie als Enabler für effizientes Cross-Media Publishing, in: Müller-Kalthoff, B.: Cross-Media Management. Content-Strategien erfolgreich umsetzen, Springer-Verlag, S. 89-124.
- Steinmetz, R., Wehrle, K. (2004): Peer-to-Peer-Networking & -Computing, in: Informatik Spektrum, 27, 1, S. 51-54.
- Tanenbaum, A. S., van Steen, M. (2002): Distributed Systems - Principles and Paradigms, Prentice Hall, Upper Saddle River.
- Voigtmann, P., Zeller, T. (2003): Beiträge zur Integrationsproblematik im Kontext von Electronic Business und Elektronischen Marktplätzen, in: Uhr, W., et al.: Wirtschaftsinformatik 2003 / Band I: Medien - Märkte - Mobilität, Dresden, Physica-Verlag, S. 215-237.

-
- Weber, M. (1998): Verteilte Systeme, Spektrum Akademischer Verlag, Heidelberg, Berlin.
- Wildemann, H. (1998): Komplexitätsmanagement durch Prozeß- und Produktgestaltung, in: Adam, D.: Komplexitätsmanagement, Gabler, Wiesbaden, S. 47-69.
- Wirtz, B. W. (2001): Medien- und Internetmanagement, Gabler Verlag, Wiesbaden.
- Wollnik, M. (1988): Ein Referenzmodell des Informationsmanagements, in: Information Management, 1988, 3, S. 34-43.