

Interoperabilität heterogener Informationsquellen im Gesundheitswesen auf Grundlage von Standards für die medizinische Kommunikation und Dokumentation

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

an der Fakultät II - Informatik, Wirtschafts- und Rechtswissenschaften,
Department für Informatik
der



Dipl.-Math.oec. **Susanne Pedersen**

Erstgutachter: Prof. Dr. Wilhelm Hasselbring
Zweitgutachter: Prof. Dr. Hans-Jürgen Appelrath

Disputation 3.11.2005

Danksagung

An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mir mit unterschiedlichster Unterstützung ermöglicht haben, diese Arbeit zu erstellen.

Ganz besonders möchte ich mich dabei bei Herrn Prof. Dr. Wilhelm Hasselbring bedanken, der mich unermüdlich mit Hinweisen und konstruktiver Kritik durch die Jahre hindurch begleitet hat. Danke für eine tolle Betreuung mit zahlreichen Gesprächen sowie vielen E-Mails und die stets umgehende Reaktion.

Herrn Prof. Hans-Jürgen Appelrath danke ich für sein Interesse und vor allem für seinen Einsatz, mit dem er mir die Möglichkeit an der Universität Oldenburg als externer Doktorand angenommen zu werden und anzufangen, erst eröffnet hat.

Auch Juniorprofessor Dr. Ralf Reussner hat mit diplomatischem Geschick Kritik stets wie Lob wirken lassen und damit sehr ermutigt.

Ferner möchte ich allen meinen Kollegen der Abteilung Software Engineering an der Universität bzw. im OFFIS danken für Ihre konstruktive Kritik und Ihre zahlreichen Anregungen.

Dank an meine Familie, die mich kontinuierlich bei der Stange gehalten hat.

Vor allem aber möchte ich meinem Mann Jürgen für seine endlose Geduld, seine zahlreichen Ermunterungen und seine einfach großartige Unterstützung danken. Ohne ihn wäre diese Arbeit einfach nicht möglich gewesen.

Kurzfassung

**„Wenn die Begriffe sich verwirren,
ist die Welt in Unordnung.“**

Konfuzius

Ein Informationsaustausch zwischen unterschiedlichen Informationssystemen wird immer wieder benötigt. Überspitzt formuliert, möchte jeder jede Information jederzeit überall zur Verfügung haben. Es gibt jedoch, aufgrund verschiedenartiger Heterogenitäten, fast immer Probleme mit der so genannten *Interoperabilität*, womit die korrekte Kommunikation von beteiligten Softwaresystemen verschiedener Herkunft gemeint ist. Da wären z.B. die verschiedenen Hardware-, Betriebssystem- und Netzwerkausstattungen, verschiedene Anwendungsprogramme, unterschiedliche Datenmodelle und Datenstrukturen in semantischer und struktureller Hinsicht innerhalb der Anwendungssysteme sowie unterschiedliche Dokumentationsweisen, die eine Interoperabilität erschweren. Das noch relativ neue Gebiet der Modelltransformationen liefert Anregungen und Lösungsansätze. Integrationstechniken aus dem Gebiet der so genannten föderierten Informationssysteme (FIS) sind ebenfalls wesentliche Hilfsmittel bei der Überwindung der Interoperabilitäts-Herausforderungen.

Bespielsweise wird die Vernetzung im Gesundheitswesen immer wichtiger. Durch eine verbesserte Kommunikation und Koordination im Gesundheitswesen sollen zum einen die Qualität der Patientenversorgung gesteigert werden, zum anderen muss mehr Effizienz Kosten sparen helfen. Mit Hilfe von Standards und Integrationstechniken bemühen sich verschiedene Organisationen die genannten Heterogenitäten zu überwinden. Da wären für das Gesundheitswesen Standards für die medizinische Kommunikation und für die medizinische Dokumentation zu nennen.

Schwerpunkt dieser Arbeit ist die Unterstützung der institutionsübergreifenden Interoperabilität im Gesundheitswesen. Ziel ist es, einen möglichst umfassenden Ansatz für den Informationsaustausch vorzustellen, welcher eine geeignete Integrationstechnik der FIS sowie einen geeigneten Modelltransformationsansatz mit den Vorzügen von Standards der Dokumentation und Kommunikation verbindet. Dabei fokussiert diese Arbeit auf die Ebene der Anwendungsarchitekturen.

Die Frage, die mit dieser Arbeit beantwortet werden soll ist, wie relevante Standards im Gesundheitswesen sinnvoll kombiniert werden können bzw. zwischen ihnen vermittelt werden kann. Es werden die Grundzüge einer Architektur vorgestellt, die auf der Basis von Standards eine institutionsübergreifende Interoperabilität im Gesundheitswesen ermöglichen soll.

Die Strukturen der relevanten Standards werden einheitlich in der Unified Modeling Language (UML) spezifiziert und dann geeignet zueinander in Beziehung gesetzt. Die resultierenden Metamodelle stellen jeweils geeignete Korrespondenzen zur Verfügung, um

bei den Transformationen zwischen den verschiedenen Standards zu helfen. Die notwendigen Modelltransformationen sollten durch einen geeigneten Ansatz unterstützt werden können. Die vorgestellten Metamodelle orientieren sich dabei an der Schichtenstruktur des Meta Object Facility Standard der Object Management Group (OMG) und stellen einen Leitfaden zur Strukturanalyse der Standards dar.

Architektur und Metamodelle dienen als Grundlage für eine prototypische Implementierung am Beispiel des Epidemiologischen Krebsregisters Niedersachsen. Hiermit erfolgt eine Evaluation des vorgestellten Konzeptes.

Abstract

Communication among different information systems is a common requirement. Exaggerated formulated, anybody wants to have any information anytime and anywhere to disposition. However, most often, there are problems with interoperability, which means correct communication between involved software systems of various origin caused by various forms of heterogeneity. There are for example various types of hardware, operating systems and network components, diverse application programs, different datamodels and data structures on semantical and textual levels within the application systems as well as different kinds of documentation complicate interoperability. The new area of the model transformations provides solution approaches. Integration techniques from the area of the so named federated information systems (FIS) are also essential expedients by overcoming the problems with interoperability.

Connecting the various service providers in health care is essential for increasing the efficiency of health services. A great challenge for the involved software systems of dissimilar origin is achieving *semantic interoperability*, which means correct communication among those systems. Various initiatives aim at solving the resulting problems by employing *domain-specific standards* and *integration techniques* for managing heterogeneity. In health care, these are standards for the medical communication and medical documentation.

Main focus of this work is the support of interoperability between various service providers in health care. The aim of this work is, as far as possible, to introduce a comprehensive base for communication, which combines a qualified integration technique of the FIS as well as a qualified model transformation approach with the advantages from standards of the documentation and communication.

The main question to be answered by this work is, how to meaningfully combine relevant standards and how to mediate between them. This approach combines and mediates among domain-specific standards in health care to solve interoperability problems in that

domain. A mediator-based architecture is presented that employs domain-specific standards and enables interoperability in shared care. This work addresses the problems of interoperability on the application level.

The structures of relevant standards are uniformly specified in Unified Modeling Language (UML) and then appropriately relates them to each other. The resulting models and meta models are guided by the Meta Object Facility Standard (MOF) of the Object Management Group (OMG) and are meant as a guideline for structural analysis of candidate standards. A qualified approach shall support the specification of necessary transformations among the standards. The developed architecture and meta models are evaluated in a prototype implementation of the epidemiologic cancer registry of Lower Saxony.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problemstellung und Zielsetzung der Arbeit	1
1.2	Aufbau der Arbeit	3
I	Grundlagen	5
2	Interoperabilität	7
2.1	Grundlagen	7
2.2	Wissensrepräsentation	12
2.3	Enterprise Application Integration-Architekturmuster	12
3	Föderierte Informationssysteme	15
3.1	Definition föderierter Informationssysteme (FIS)	15
3.2	Metadaten für föderierte Informationssysteme	17
3.3	Typen von Föderierten Informationssystemen	18
3.4	Mediatorbasierte Informationssysteme	19
3.5	Entwurfsmuster zur Integration von Informationssystemen	20
3.6	Zusammenfassung	24
4	Metadaten und ihre Modellierung	25
4.1	Metadaten-Standards	25
4.1.1	Framework-/Architekturstandards	25
4.1.2	Metamodellstandards	26
4.1.3	Daten- und Nachrichtendefinitionsstandards	28
4.2	Mehrebenen-Modellierungsframework	30
4.2.1	Metamodellierung über mehr als zwei Ebenen	30
4.2.2	Probleme bei flacher Instanziierung	30
4.2.3	Lösung: Tiefe Instanziierung	31
4.3	Zusammenfassung	34
5	Modelltransformation	35
5.1	Entwurfs-Eigenschaften von Modelltransformationsansätzen	35
5.2	Modelltransformationsansätze	39
5.2.1	Übersicht	39
5.2.2	Modell-zu-Code Ansätze	40

5.2.3	Modell-zu-Modell Ansätze	40
5.2.4	Andere Ansätze	40
5.3	Qualität einer Abbildung	42
5.4	Auswahl Modelltransformationsansatz	42
II	Metamodell-basierte Architektur zur institutionsübergreifenden Interoperabilität	45
6	Analyse domänenspezifischer Standards am Beispiel Medizin	47
6.1	Bedeutung von Standards am Beispiel Medizin	47
6.1.1	Erwartungen an die Telemedizin	47
6.1.2	Die Rolle von Standards	48
6.2	Standards für die medizinische Kommunikation	48
6.2.1	Behandlungsdatenträger (BDT)	49
6.2.2	Health Level Seven (HL7)	51
6.2.3	Digital Imaging and Communications in Medicine (DICOM)	53
6.3	Standards für die medizinische Dokumentation	57
6.3.1	Standards für Dokumente	57
6.3.2	Standards für Begriffssysteme	62
6.3.3	Ontologien	67
6.4	Auswahl von Kombinationen der Standards	67
6.4.1	Zusammenhang der Standards	67
6.4.2	Auswahlschema	68
6.4.3	Vorgehensweise bei der Auswahl	68
6.5	Andere Standardisierungsbemühungen	69
6.5.1	Standards	70
6.5.2	Übertragungstechniken	72
6.5.3	Leitlinien und Klinische Pfade	73
7	Architektur einer institutionsübergreifenden Interoperabilität	75
7.1	Erweiterte Mediator-basierte Architektur	75
7.1.1	Anforderungen an die Architektur	75
7.1.2	Allgemeine Struktur der Architektur	76
7.1.3	Zusammenfassung	76
7.2	Mediator-basierte Architektur für das Gesundheitswesen	78
7.2.1	Anwendung der Architektur auf das Gesundheitswesen	78
7.2.2	Aufgaben der einzelnen Architekturteile	78
7.3	Kombination der Metadaten	83
7.3.1	Zusammenhang der Metadaten	83
7.3.2	Metadaten Kommunikation	85
7.3.3	Metadaten Dokumentation	89
7.3.4	Metadaten Begriffssysteme	91
7.4	Transformationen zwischen den Standards	95
7.4.1	Top-Down-Integration	95
7.4.2	Transformationsregeln	96

III	Evaluation am Beispiel des Epidemiologischen Krebsregisters Niedersachsen	99
8	Architektur für das Epidemiologische Krebsregister Niedersachsen	101
8.1	Problemstellung und Zielsetzung der Evaluation	101
8.1.1	Problemstellung	101
8.1.2	Zielsetzung	102
8.2	Auswahl der geeigneten Standards	102
8.2.1	Dimensionenschema	102
8.2.2	Einordnung des EKN in das Dimensionenschema	104
8.2.3	Auswahl der Standards für das EKN	104
8.2.4	Auswahl der Kombinationen der Standards für das EKN	104
8.2.5	Beurteilung der ausgewählten Kombinationen von Standards	105
8.3	Mediator-basierte Architektur für das EKN	106
8.3.1	Schemaarchitektur für das EKN	106
8.3.2	UML-Objektdiagramm der Architektur für das EKN	109
8.3.3	Facilitatoren	109
8.3.4	Domänenmodelle	109
8.3.5	Komponentenmediatoren	112
8.3.6	Einsatz von RDF und RDF-Schema	114
9	Prototypische Implementierung	121
9.1	Integrationsplattform SeeBeyond e*Gate	121
9.1.1	e*Gate Integrator	122
9.1.2	MapForce 2005 Enterprise Edition	123
9.2	Abbildung Kommunikation	124
9.2.1	Modellkorrespondenzen von HL7-Segment und BDT-Satz	126
9.2.2	Modellkorrespondenzen von HL7-Feld und BDT-Datenfeld	126
9.2.3	Übrige Abbildungen	126
9.2.4	XSLT mit MapForce	126
9.2.5	Beurteilung	128
9.3	Abbildung Dokumentation	132
9.3.1	Modellkorrespondenzen von Strukturierungselementen	132
9.3.2	Übrige Abbildungen	133
9.3.3	XSLT mit MapForce	133
9.3.4	Beurteilung	136
9.4	Ergebnisse der Evaluation	137
9.4.1	Bijektivität der Abbildungen	137
IV	Verwandte Ansätze	139
10	Verwandte Ansätze aus der Informatik	141
10.1	Datenbank-Integration	141
10.1.1	Schemaintegration	141
10.2	Ontologien und Semantic Web	142

10.2.1	Repräsentations- und Anfragesprachen für Ontologien	142
10.2.2	A Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information Systems (ARTEMIS)	142
10.2.3	Semantic Health Records	143
10.3	Semantische Suche	143
10.3.1	Semantische Komponentensuche auf Basis von Geschäftsprozessmodellen	143
10.4	Matching	143
10.4.1	Metamodell-basiertes Matching	143
10.5	Tabellarische Zusammenfassung	144
11	Verwandte Ansätze aus der Medizininformatik	145
11.1	Clinical Document Architecture	145
11.1.1	Standardisation of Communication between Information Systems in Physician Offices and Hospitals using XML (SCIPHOX)	145
11.2	Health Level Seven (HL7)	146
11.2.1	MIRApplus (Universität Freiburg)	146
11.3	Digital Imaging and Communications in Medicine Structured Reporting (DICOM SR)	146
11.3.1	Ein Verfahren zur Bewertung der Interoperabilität medizinischer Bildkommunikationssysteme	146
11.3.2	Ein generisches Verfahren zur adaptiven Visualisierung von strukturierten medizinischen Befundberichten	147
11.4	Tumordokumentation	147
11.4.1	Schnittstellenkonzepte in Tumordokumentationssystemen	147
11.4.2	Patientenzentrierte Dokumentation onkologischer Erkrankungen	148
11.4.3	eMamma-Akte (Fachhochschule Dortmund)	149
11.5	Integrations-Architekturen	149
11.5.1	Modellierung von Integration in Krankenhausinformationssystemen	149
11.6	Tabellarische Zusammenfassung	150
V	Zusammenfassung und Ausblick	151
12	Zusammenfassung	153
13	Ausblick	157
VI	Anhänge	159
A	Föderierte Informationssysteme	161
A.1	Einleitung	161
A.2	Dimensionen für die Klassifikation von Informationssystemen	161
A.2.1	Autonomie	161
A.2.2	Heterogenität	162

A.2.3	Syntaktische Heterogenität	162
A.2.4	Datenmodellbasierte Heterogenität	162
A.2.5	Logische Heterogenität	163
A.2.6	Verteilung	163
A.3	Klassifikation von Informationssystemen	163
A.3.1	Metadaten für föderierte Informationssysteme	166
A.4	Klassifikationskriterien für FIS	166
A.4.1	Arten von Komponenten: strukturiert, semi-strukturiert, unstrukturiert	168
A.4.2	Enge versus lose Föderation	168
A.4.3	Enge Föderation	168
A.4.4	Lose Föderationen	169
A.4.5	Modellierungssprache des FIS	170
A.4.6	Arten von semantischer Integration	171
A.4.7	Transparenz	172
A.4.8	Anfrage-Paradigmen	172
A.4.9	Bottom-up versus Top-down	173
A.4.10	Virtuelle versus materialisierte Integration	174
A.4.11	Nur Lese- oder Lese- und Schreib-Zugriff	175
A.4.12	Erforderliche Zugriffsmethoden	175
A.5	Typen von Föderierten Informationssystemen	176
A.5.1	Lose gekoppelte Informationssysteme	177
A.5.2	Föderierte Datenbanksysteme (FDBS)	177
A.5.3	Mediator-basierte Informationssysteme	179
B	Resource Description Framework und Semantic Web	183
B.1	Grundlagen RDF	183
B.2	Notation 3	184
B.3	Erstellung eines Vokabulars	186
B.4	Semantic Web und Ontology Web Language (OWL)	186
C	Medizinische Dokumentation	189
C.1	Einleitung	189
C.2	Ziele der medizinischen Dokumentation	189
C.3	Grundlagen der Terminologielehre	190
C.3.1	Semiotische Triade	190
C.3.2	Merkmale	192
C.3.3	Merkmalarart	192
C.3.4	Zusammenhänge Begriff und Benennung	193
C.4	Grundlagen medizinischer Begriffssysteme	193
C.4.1	Allgemeines	193
C.4.2	Art der Begriffsbeziehungen	194
C.4.3	Anzahl der Achsen	195
C.4.4	Art der hierarchischen Struktur	196
C.4.5	Codierprinzipien	197
C.4.6	Klassifikation, Thesaurus und Nomenklatur	199

C.5	Medizinische Begriffssysteme	200
C.5.1	Übersicht	200
C.5.2	TNM-Klassifikation und Tumorlokalisierungsschlüssel	203
C.5.3	ICD	203
C.5.4	SNOMED	204
C.5.5	RCC, MeSH, UMLS und GALEN	204
D	Epidemiologisches Krebsregister Niedersachsen	205
D.1	Einleitung	205
D.2	Epidemiologische Krebsregister	205
D.2.1	Allgemeines	206
D.3	Epidemiologisches Krebsregister Niedersachsen (EKN)	208
D.4	Meldungsarten an das EKN	208
D.5	Problemstellung und Zielsetzung der Evaluation	210
D.6	Grundlagen zum Epidemiologischen Krebsregister	211
D.6.1	Anwendungsszenario	211
D.6.2	Technischer Stand der Melder	219
E	Inhaltsverzeichnis der CD	221
	Abbildungsverzeichnis	222
	Tabellenverzeichnis	225
	Literaturverzeichnis	227
	Index	242

1 Einleitung

1.1 Problemstellung und Zielsetzung der Arbeit

Ein Informationsaustausch zwischen unterschiedlichen Informationssystemen wird in heutiger Zeit immer mehr nachgefragt. So sollen beispielsweise Informationen zwischen Produzenten und Händlern, zwischen Behörden und Bürgern, innerhalb großer Konzerne usw. ausgetauscht und weiterverwendet werden. Es gibt jedoch Probleme bei der so genannten *Interoperabilität*, womit die korrekte Kommunikation von beteiligten Software-systemen verschiedener Herkunft gemeint ist, aufgrund verschiedenartiger *Heterogenitäten*. Da wären z.B. die verschiedenen Hardware-, Betriebssystem- und Netzwerkausstattungen, verschiedene Anwendungsprogramme, unterschiedliche Datenmodelle und Datenstrukturen in semantischer und struktureller Hinsicht innerhalb der Anwendungssysteme sowie unterschiedliche Dokumentationsweisen, die eine Interoperabilität erschweren.

Auch die Vernetzung im Gesundheitswesen wird immer wichtiger. Durch eine verbesserte Kommunikation und Koordination im Gesundheitswesen soll zum einen die Qualität der Patientenversorgung gesteigert werden, zum anderen muss mehr Effizienz Kosten sparen helfen. Mit Hilfe von *Standards* und *Integrationstechniken* bemühen sich verschiedene Institutionen die genannten Heterogenitäten zu überwinden. Da wären für das Gesundheitswesen Standards für die medizinische Kommunikation und für die medizinische Dokumentation zu nennen. Das noch relativ neue Gebiet der Modelltransformationen liefert zahlreiche Anregungen und Lösungsansätze. Integrationstechniken aus dem Gebiet der so genannten *föderierten Informationssysteme* (FIS) sind ebenfalls wesentliche Hilfsmittel bei der Überwindung der Interoperabilitätsschwierigkeiten.

Unter Informationsquellen sollen hier diverse Anwendungssysteme von Institutionen im Gesundheitswesen zu verstehen sein, also im wesentlichen Anwendungssysteme in den Krankenhäusern, in den Praxen der niedergelassenen Ärzte, in Laboren und in Pathologien. Ferner kommen auch Anwendungssysteme von Krebsregistern, Gesundheitsämtern, Krankenkassen u.ä. in Frage.

Als Standards in der medizinischen Dokumentation sind die unterschiedlichen medizinischen Begriffssysteme anzusehen, wie z.B. der ICD (International Classification of Diseases). Standards in der medizinischen Kommunikation sind z.B. HL7 (Health Level

Seven) für den stationären Bereich und BDT (Behandlungsdatenträger) für den niedergelassenen Bereich. Bei Electronic Healthcare Records wird an einer Harmonisierung der bestehenden Ansätze gearbeitet.

Täglich müssen im Gesundheitswesen große Mengen von patientenbezogenen Daten bewegt werden. Dazu zählen insbesondere Patientenstammdaten, Abrechnungsdaten, Untersuchungsergebnisse, Labordaten, digitale Bildinformationen, Befunde etc., die für die Abrechnung, die Einweisung ins Krankenhaus, für die Überweisung zum Facharzt oder für die Weiter- bzw. Nachbehandlung wichtig sind.

Für die verschiedenen Bereiche im Gesundheitswesen sind in der Vergangenheit bereits eine Vielzahl von Anwendungssystemen entwickelt worden. So gibt es Anwendungssysteme für den Krankenhausbereich, für die niedergelassenen Ärzte und Zahnärzte sowie für spezielle Anwendungsgebiete wie Pathologie, Labormedizin etc. Durch ihre unterschiedliche Herkunft und der damit verbundenen Heterogenität erschweren sie eine umfassende Interoperabilität. Es werden Standards für die Kommunikation und Dokumentation benötigt, um zu ermöglichen, dass Gesundheitsinformationen zwischen mehreren Partnern im Gesundheitswesen konsistent in die jeweiligen elektronischen Patientenakten integriert werden können sowie eine einheitliche Interpretation der Inhalte zu gewährleisten [PEH⁺01].

Eine Kommunikation der verschiedenen Anwendungssysteme würde die Möglichkeit für einen effizienteren Informationsfluss zwischen den Beteiligten schaffen und könnte so beispielsweise unnötige Doppeluntersuchungen vermeiden helfen, somit dem Patienten Belastungen ersparen und einen Beitrag zur Kostensenkung leisten.

Schwerpunkt dieser Arbeit ist die Unterstützung der institutionsübergreifenden Interoperabilität im Gesundheitswesen. Ziel ist es, einen umfassenden Ansatz für den Informationsaustausch vorzustellen, welcher eine geeignete Integrationstechnik der FIS sowie einen geeigneten Modelltransformationsansatz mit den Vorzügen von Standards der Dokumentation und Kommunikation verbindet. Abbildung 1.1 zeigt eine Drei-Ebenen-Architektur nach [Has00a], wonach eine Organisationseinheit in Geschäftsarchitektur, Anwendungsarchitektur und Technologiearchitektur untergliedert wird. Für eine effiziente Unterstützung der betrieblichen Abläufe müssen alle drei Ebenen jeweils horizontal integriert werden. Diese Arbeit beschränkt sich auf die Ebene der Anwendungsarchitekturen und bezieht sich somit auf das Gebiet der Enterprise Application Integration (EAI).

Die Frage, die mit dieser Arbeit beantwortet werden soll ist, wie relevante Standards im Gesundheitswesen sinnvoll kombiniert werden können bzw. zwischen ihnen vermittelt werden kann. Es werden die Grundzüge einer Architektur vorgestellt, die auf der Basis von Standards eine institutionsübergreifende Interoperabilität im Gesundheitswesen ermöglichen soll.

Die Strukturen der relevanten Standards werden einheitlich in der Unified Modeling Language (UML) [Oes01] spezifiziert und dann geeignet zueinander in Beziehung gesetzt.

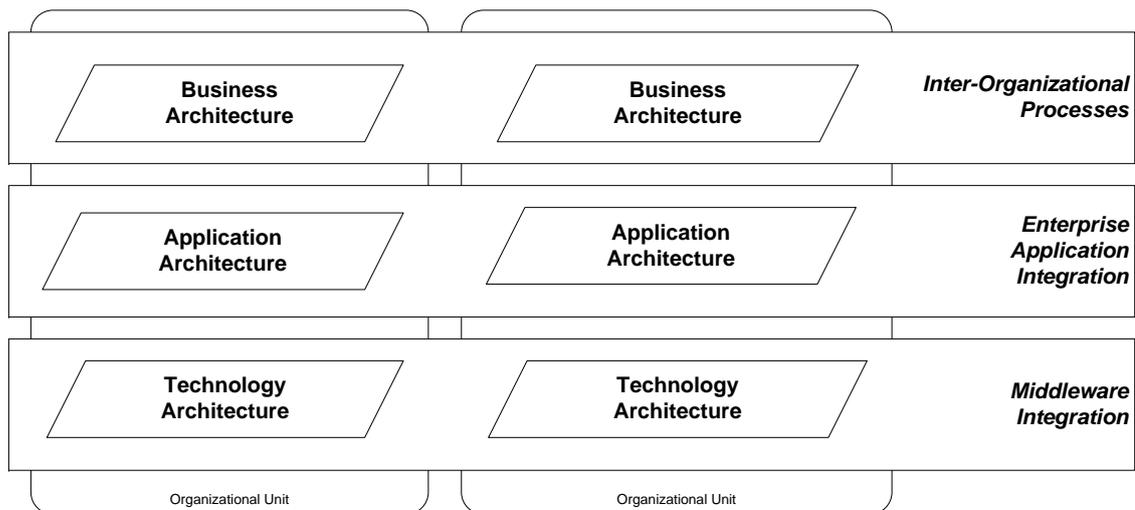


Abbildung 1.1: Architekturebenen von Organisationseinheiten nach [Has00a]

Die resultierenden Metamodelle stellen jeweils geeignete Korrespondenzen zur Verfügung, um bei den Transformationen zwischen den verschiedenen Standards zu helfen. Sie bilden damit einen Leitfaden zur Strukturanalyse der Standards und deren Abbildungen zwischen verschiedenen Kommunikations- und Dokumentationsstandards. Die notwendigen Modelltransformationen sollten durch einen geeigneten Ansatz unterstützt werden können. Die vorgestellten Metamodelle sind angelehnt an die Instanz eines Meta-Metamodells im Meta Object Facility Standard der Object Management Group (OMG) [MOF03]. Unterstützend wird dabei tiefe Instanziierung [AK01] angewandt.

Architektur und Metamodelle dienen als Grundlage für eine prototypische Implementierung am Beispiel des Epidemiologischen Krebsregisters Niedersachsen (EKN). Hiermit erfolgt eine Evaluation des vorgestellten Konzeptes.

1.2 Aufbau der Arbeit

Um die eben vorgestellten Ziele zu erreichen, finden sich in dieser Arbeit folgende Teile:

In Teil I sind die Grundlagen für diese Arbeit zusammengestellt. Nach einem Überblick über Interoperabilität und föderierte Informationssysteme folgen jeweils ein Überblick über wesentliche Metadaten-Standards, ein Mehrebenen-Modellierungsframework sowie den Themenbereich der Modelltransformation.

Der Teil II bildet den Kern dieser Arbeit. Er befasst sich zunächst mit der Analyse domänenspezifischer Standards. Vorgestellt werden einheitlich spezifizierte UML-Modelle

medizinischer Kommunikations- und Dokumentationsstandards sowie von Electronic Healthcare Records. Vertieft wird nicht die Thematik der Middleware-Architekturen, wie z.B. CORBA, da dies die Integration auf Technologieebene betrifft. Ebenso wird nicht die Ebene der Geschäftsarchitekturen betrachtet.

Danach werden die Anforderungen an eine institutionsübergreifende Interoperabilität im Gesundheitswesen formuliert und eine entsprechende Architektur vorgestellt, die die wesentlichen Standards auf Meta-Modellebene kombiniert und, ausgehend von diesen Standards, die Anwendungssysteme von Gesundheitsdienstleistern top-down integriert.

Die zuvor vorgestellte Architektur wird in Teil III anhand eines Beispielmeldeweges umgesetzt und in der Anwendungsumgebung des Epidemiologischen Krebsregisters Niedersachsen (EKN) evaluiert. Zudem wird ein kurzer Überblick über das Szenario der Melder, Meldungsarten, Krebsregister sowie Tumordokumentation gegeben.

Teil IV beschreibt verwandte Arbeiten und Initiativen aus den Bereichen der Informatik sowie Medizininformatik und stellt Bezug zum eigenen Ansatz her.

In Teil V bilden eine Zusammenfassung der hier vorgestellten Forschungsarbeit und ein Ausblick auf mögliche Weiterentwicklungen bzw. Ergänzungen den Abschluss dieser Arbeit.

In Anhängen im Teil VI sind relevante Grundlagen zu den in dieser Arbeit bearbeiteten Themenbereichen Föderierten Informationssysteme, Medizinische Dokumentation und Epidemiologisches Krebsregister Niedersachsen zusammengestellt. Außerdem gibt ein Inhaltsverzeichnis Auskunft über den Inhalt der beigefügten Dissertations-CD (siehe Anhang E auf Seite 221).

Teil I

Grundlagen

2 Interoperabilität

Die Verbindung von heterogenen Informationssystemen in der Medizin verstärkt das Problem von semantischer Interoperabilität, ausgehend von traditionellen Ansätzen terminologischer Standardisierung in drei grundlegenden Aspekten [IRS01]:

1. Die Vielfalt der medizinischen Wortschätze, die aktuell koexistieren, ist eine beachtliche Barriere für die Integration von jeweils eigenständig entwickelten Anwendungen.
2. Die ausgetauschten Daten müssen maschinenverarbeitbar sein für unterschiedliche Zwecke wie Integration von Patientendaten, Zugang zu Literatur und Wissensgrundlagen, klinische Revision und Forschung.
3. Terminologische Dienste sollten konstant und als Komponente mit standardisierten Schnittstellen verfügbar sein.

2.1 Grundlagen

Interoperabilität zwischen Vertretern des Gesundheitswesens einschließlich heterogener Softwareanwendungen ist eine Schlüsseleigenschaft in verteilten Shared-Care Umgebungen. Die folgende Definition der IEEE unterscheidet zwei wichtige Aspekte: Interoperabilität ist die Fähigkeit von zwei oder mehr Systemen oder Komponenten Informationen auszutauschen und die Information, die ausgetauscht wurde, zu nutzen [IEE90, IRS01].

Das grundsätzliche Kommunikationsprinzip wird in Abbildung 2.1 schematisch nach [DJK04] dargestellt: Ein Sender kodiert seine Informationen. Damit werden aus interpretierten Informationen interpretationsbedürftige Daten. Für eine Übertragung als Nachricht werden die Daten serialisiert. Nur wenn der Empfänger entsprechend deserialisiert und dekodiert, ist die Information angekommen und eine korrekte Kommunikation hat stattgefunden. Zum erfolgreichen Austausch der Information ist es notwendig, dass beide Seiten mit demselben Vokabular und derselben Interpretation der Vokabeln arbeiten. Semantik ergibt sich aus der Verknüpfung einer Bezeichnung mit einem Begriff bzw. Konzept, das etwas beschreibt. Unter einem Begriff oder Konzept (engl. concept) versteht man eine Denkeinheit, die aus einer Menge von Gegenständen unter Ermittlung der diesen

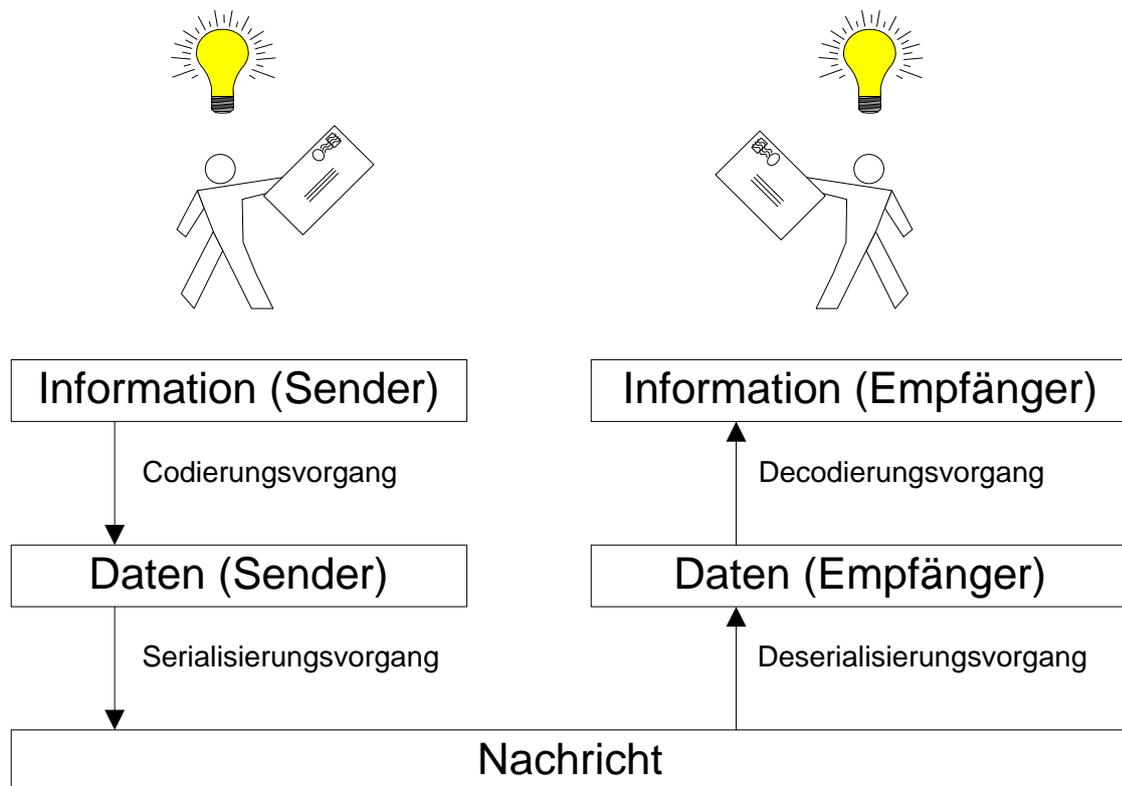


Abbildung 2.1: Grundsätzliches Kommunikationsprinzip

Gegenständen gemeinsamen Eigenschaften mittels Abstraktion gebildet wird [DIN2330]. Die Herausforderung ist, Semantik so eindeutig formulieren zu können, dass menschliche Interpretationsvorgänge verzichtbar werden, ohne Semantik zu verlieren. Mit dieser Aussage befindet man sich nach [DJK04] im Dilemma, da Semantik nicht niedergelegt, sondern nur verstanden werden kann. Auch Menschen können lediglich eine bekannte Bezeichnung mit einem erlernten Begriff (Konzept) verknüpfen, aber nicht direkt aus der Bezeichnung auf eine zugeordnete, aber möglicherweise unbekannte, Semantik schließen. Zudem stellen natürliche Sprachen keinen hinreichenden Bezeichnungsvorrat zur Verfügung, um alle denkbaren Begriffe so eindeutig zu bezeichnen, dass alle Interpretationslücken geschlossen werden können.

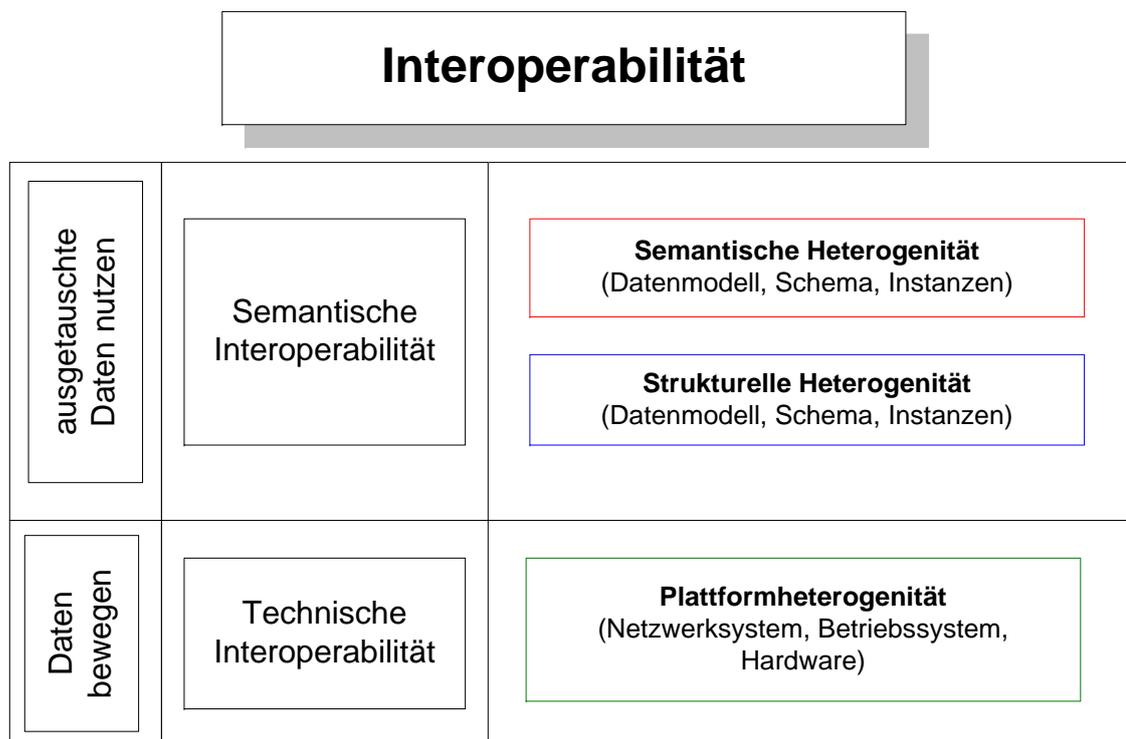


Abbildung 2.2: Grundzüge Interoperabilität

Interoperabilität ist die Fähigkeit von Softwaresystemen unterschiedlicher Herkunft, miteinander kommunizieren zu können [Bun02]. Einerseits gibt es den Aspekt der *technischen Interoperabilität*, z.B. die Übertragung der Daten und die Integration der Dienste. Andererseits gibt es den Aspekt der *semantischen Interoperabilität*, z.B. der Gebrauch von Information. Diese Grundzüge von Interoperabilität sind in Abbildung 2.2 dargestellt.

Die Problematik im Bereich der technischen Interoperabilität betrifft die Heterogenität verursacht durch Netzwerke, Betriebssysteme und Hardware. Diese Art der Heterogenität bezeichnet [Sau98] als Plattformheterogenität. Der Aspekt der technischen Interoperabilität wird hier nicht näher beleuchtet.

Es soll auf Konflikte eingegangen werden, die durch verschiedene Datenmodelle entstehen, durch verschiedene Schemata, zu denen die Modellierungen desselben Sachverhaltes geführt haben können, oder durch Daten, weil sie fehlen oder sich widersprechen. Die Arten von Heterogenität, die in diesen drei Ebenen entstehen können, bezeichnet [Sau98] als semantische Heterogenität und strukturelle Heterogenität.

Die semantische Heterogenität ist weniger ein Problem auf der Ebene der Datenmodelle, als ein Problem der Schema- und Datenebene [Sau98]. Bei der semantischen Interoperabilität ist es wichtig, Information als interpretierte Daten zu definieren unter Berücksichtigung des Zusammenhangs mit dem Gebrauch (Kontext). Semantische Konflikte sind Konflikte zwischen dem Inhalt und der beabsichtigten Bedeutung von heterogenen Anwendungssystemen [Pol01]. Diese treten in drei Formen auf:

1. Verwechslungskonflikte, wo es so erscheint, als wäre dieselbe Bedeutung vorhanden
2. Skalierungskonflikte, wo unterschiedliche Referenzsysteme denselben Wert messen
3. Namenskonflikte, wo Namensschemen sich signifikant unterscheiden, z.B. Synonyme und Homonyme

Diese uneinheitliche Verwendung von Begriffen sorgt in vielen Bereichen oft für Unklarheiten, wie beispielsweise die gleichen Bezeichnungen für verschiedene Begriffe (Homonyme) oder mehrere Bezeichnungen für einen Begriff (Synonyme). Es entstehen Probleme mit der verwendeten Terminologie, der man mit zahlreichen Standardisierungen zu begegnen versucht [PH02].

Strukturelle Unterschiede bei Datenmodellen führen zu strukturellen Heterogenitäten bei den Schemata bis hin zu den Instanzen. Je größer dabei die Unterschiede zwischen den verwendeten Datenmodellen sind, desto größer sind dann auch die Unterschiede bei den Schemata und Instanzen [Sau98]. Dokumentenstrukturen können z.B. in der Medizin sehr vielfältig sein, auch hier versucht man durch Standardisierung die Heterogenitäten zu verringern.

Abbildung 2.3 definiert Applikationsinteroperabilität nach [Pol01] als Summe von Applikations- und Informationsintegration. Applikationsintegration ist dabei der Prozess Softwaresysteme zu verknüpfen, bezieht sich auf die technologische Lösung für die Applikationsinteroperabilität und ist damit mit der technischen Interoperabilität gleichzusetzen. Die Informationsintegration befasst sich dagegen mit dem linguistischen, sozialen und philosophischen Teil, betrifft also die semantische Interoperabilität. Informationsintegrationstechniken fokussieren daher auf den Transport von Bedeutung, die Semantik soll erhalten bleiben. Eine vollständig unterstützte Interoperabilitätslösung sollte nach [Pol01] beinhalten:

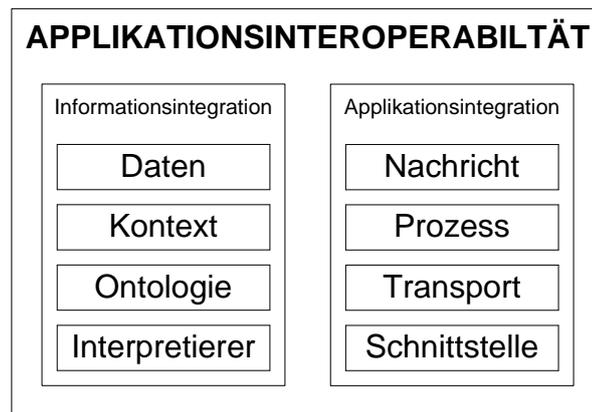


Abbildung 2.3: Applikationsinteroperabilität nach [Po101]

- Bereich Applikationsintegration:

Transport-Service: Ein Mechanismus oder Protokoll, um Nachrichten von einem Ort zu einem anderen zu bewegen

Nachrichten-Container: Ein kodierter Wrapper und Struktur für die Daten und Metadaten während des Austausches (z.B. XML, CORBA, etc.)

Integrationsschnittstelle (Transport oder Nachricht): Ein Weg die Daten in den Container zu bringen und über das Protokoll zu versenden

Prozess-Controller: (Ein Koordinator, der definiert und Multistep, Batch, Publish/Subscribe etc. koordiniert

- Bereich Informationsintegration:

Datenkapselung : Kapselung von groben Elementen eines Informationsaustausches

Ontologische Informationsrepräsentation: Eine Spezifikation eines konzeptuellen Modells, durch das die Datenquellen repräsentiert werden. Diese Repräsentation kann erreicht werden durch direkte Abbildung oder einer dritten konzeptuellen Spezifikation auf höherer Ebene als Quell- und Zielspezifikation

Kontext und Metadaten: Metadaten sind quasi die Daten über die Daten aus Applikationssicht, aus der Organisationsperspektive sind sie Kontext

Informationstransformation: Die Fähigkeit ganze Mengen von Daten, Beziehungen und allem, basierend auf Informationen zu manipulieren, die in Metadaten und vom Nutzer gelieferte Kontextinformation stecken. Dies erfordert typischerweise ein konzeptuelles Modell, welches ontologische Abbildungstechniken nutzt

Datenübersetzung: Eine Menge von Algorithmen und vom Nutzer modifizierbarer Skripts, um auf der Basis von vordefinierten Regeln zu übersetzen

2.2 Wissensrepräsentation

Die Umsetzung eines Denkmodells für ein Anwendungssystem erfordert immer die formale Repräsentation deklarativen und prozeduralen Wissens [Len05]. Nach [Pol01] umfasst der Bereich Informationsintegration, also die hier im Zentrum stehende semantische Interoperabilität, eine ontologische Informationsrepräsentation. Eine Ontologie ist eine Art der formalen Spezifikation deklarativen Wissens. Bei einer Ontologie geht es um die Spezifikation der Semantik der definierten Begriffe, mit dem Ziel, ein gemeinsames Verständnis zwischen kommunizierenden Parteien zu ermöglichen. Der Begriff Ontologie wird zunehmend auch verallgemeinernd im Sinne von Begriffsordnungen verwendet. Detaillierter wird auf Ontologien in den Abschnitten 6.3.3 und B.4 auf den Seiten 67 und 186 eingegangen. Weitere Möglichkeiten von formaler Repräsentation deklarativen Wissens stellen auch Datenbankschemata, Klassenhierarchien und Begriffssysteme dar. Ein Begriffssystem ist eine entsprechend den Begriffsbeziehungen geordnete Menge von Begriffen eines Begriffsfeldes, wobei jeder einzelne Begriff durch seine Position innerhalb des Begriffssystems bestimmt ist (DIN 2342). Ein Begriffssystem hat die Aufgabe, verschiedene Begriffe semantisch klar voneinander abzugrenzen. Auf Begriffssysteme gehen Abschnitt 6.3.2 auf Seite 62 sowie der Anhang C ab Seite 189 näher ein.

In dieser Arbeit sollen nach Analyse von verschiedenen Standards unter anderem Begriffssysteme angeboten werden, die die in den Standards verwendeten Begriffe semantisch deutlich abgrenzen und als Metamodelle bei der Transformation zwischen verschiedenen Standards hilfreich sein können. Mit der Unified Modeling Language (UML) kann deklaratives wie prozedurales Wissen spezifiziert werden [Len05]. UML soll zur Modellierung der Begriffssysteme eingesetzt werden, da sie geeignet visualisieren kann und unter Systementwicklern weit verbreitet ist.

2.3 Enterprise Application Integration-Architekturmuster

Eine Integration auf der Ebene der Anwendungsarchitekturen, die hier betrachtet werden soll, wird als Enterprise Application Integration (EAI) bezeichnet. Architekturmuster sind eine wichtige Hilfe des IT-Fachmannes bei den Herausforderungen bezüglich Integrationskomplexität und -verwirrung [Lut00]. EAI-Architekturmuster bieten einen Architekturblick auf die Thematik EAI und sind eine solide technische Basis für erfolgreiches EAI. EAI-Architekturmuster fokussieren auf die gesamte Struktur eines Systems oder einer Applikation. In [BMR⁺01] wird ein Architekturmuster wie folgt definiert: „Ein Architekturmuster stellt ein fundamentales, strukturelles Organisationsschema für Softwaresysteme dar. Es bietet eine Menge von vordefinierten Subsystemen an, spezifiziert deren Beziehungen und beinhalten Regeln und Leitlinien für die Organisation der Beziehungen zwischen ihnen.“ So ist auch die Definition von EAI-Architekturmuster zu verstehen. Auf verschiedene Entwurfsmuster zur Integration wird später in Abschnitt 3.5 auf Seite 20 eingegangen. Während EAI-Architekturmuster auf die gesamte Struktur eines Systems oder

einer Applikation fokussieren, befassen sich die Entwurfsmuster mit der Art und Weise der Integration.

Ein gemeinsames Charakteristikum aller EAI-Architekturmuster ist der hohe Grad an Entkopplung — eine Eigenschaft, wo die Abhängigkeiten zwischen Elementen, z.B. Applikationen, minimiert ist. Eine Entkopplung auf Architekturebene hat den immensen Vorteil von IT-System-Flexibilität. Eine gelungene Entkopplung erlaubt Entwicklern das Ändern von Applikationen mit minimalem oder wegfallendem Änderungsaufwand für abhängige Anwendungen.

Fünf wichtige EAI-Architekturmuster werden im Folgenden kurz vorgestellt [Lut00]:

1. Integrationsadapter: konvertiert eine bestehende Anwendungsschnittstelle in eine gewünschte Schnittstelle
2. Integrationsbote: beschreibt einen Ansatz der Minimierung der Abhängigkeiten in der Kommunikation zwischen Applikationen
3. Integrationsfassade: bietet eine einfache Schnittstelle für Back-end-Applikationen, die die Abhängigkeiten zwischen Client- und Serverapplikationen minimiert
4. Integrationsmediator: kapselt Anwendungsinteraktionslogik, um Applikationsabhängigkeiten zu minimieren
5. Prozessautomatisierung: beschreibt einen Architekturansatz, der Abhängigkeiten zwischen Prozessautomationslogik und Anwendungen minimiert.

Die EAI-Architekturmuster können in zwei wichtigen Zusammenhängen gesehen werden. Der erste ist die Enterprise Integration Architekturebene. Der Zweck von Enterprise Integration Architektur ist es, Ordnung und Konsistenz in die vielen Applikationen im Unternehmen zu bringen. Die EAI-Architekturmuster können als Rahmenwerk für die Spezifikation von EAI-Standards genutzt werden. Der zweite Zusammenhang ist eine gegebene Applikation. EAI-Architekturmuster sind nützlich bei der Spezifikation der Architektur für eine vorgegebene zu integrierende Applikation.

Im Rahmen dieser Arbeit ist das Architekturmuster Integrationsmediator durch die daraus folgende Kapselung von Interesse, da Drei-Ebenen-Architekturen bereits in vielen Gebieten der Informatik etabliert sind. Mediatoren ermöglichen einen effizienten Zugriff auf verteilte Daten sowie deren Kombination [SCS03]. Daher sollen nun die Grundlagen Förderierter Informationssysteme näher gebracht werden.

3 Föderierte Informationssysteme

Eine wichtige Klasse von Softwaresystemen sind die so genannten *föderierten Informationssysteme* [BKLW99]. Ihre Hauptcharakteristik ist, dass sie als integrierende Schicht über existierende Alt-Anwendungen und -Datenbanken konstruiert werden. Allgemein können sie nach drei Dimensionen klassifiziert werden, dem Grad der Autonomie der integrierten Komponenten, dem Grad der Heterogenität zwischen den Komponenten und der Verteilung.

Autonomie

Wenn ein Informationssystem auf mehreren Komponenten basiert, kann die Autonomie, die integrierte Komponenten behalten können, ein kritischer Punkt sein. [BKLW99] unterscheidet dabei Entwurfs-, Kommunikations- und Ausführungs-Autonomie.

Heterogenität

Heterogenität ist natürlich auf die Tatsache zurückzuführen, dass eine autonome Entwicklung von Systemen immer unterschiedliche Lösungen ergibt, z.B. aufgrund eines anderen Verständnisses und Modellierung der Objekte der realen Welt. [BKLW99] unterscheidet syntaktische, datenmodellbasierte und logische Heterogenität mit verschiedenen Untertypen. Zur syntaktischen Heterogenität gehören die technische und die Schnittstellen-Heterogenität, die datenmodellbasierte Heterogenität betrifft die Tatsache, dass unterschiedliche Datenmodelle eine unterschiedliche Semantik für ihre Konzepte haben. Die logische Heterogenität wird in semantische, schemabasierte und strukturelle Heterogenität unterteilt.

Verteilung

Ein drittes Problem ist die physische Verteilung von Datenquellen. Verteilung wird von einer Reihe von Techniken überwunden, wie z.B. CORBA.

3.1 Definition föderierter Informationssysteme (FIS)

Ein FIS enthält eine Menge von heterogenen und autonomen Informationssystem-Komponenten, die Teilnehmer dieser Föderation sind.

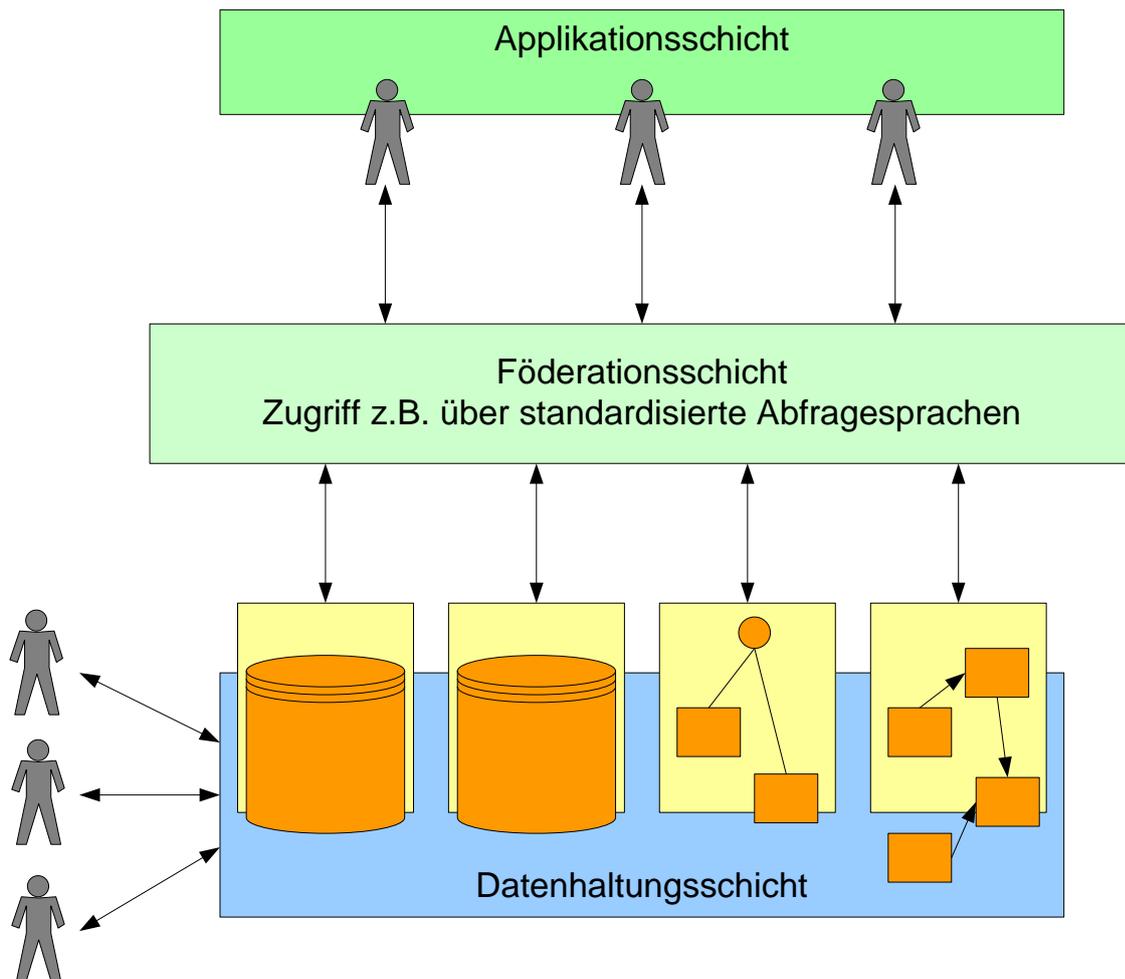


Abbildung 3.1: Grundzüge eines föderierten Informationssystems von [Wil02]

Anwendungen greifen auf eine Menge von heterogenen Datenquellen über eine Föderationsschicht bzw. Integrationsschicht zu, die einen einheitlichen Zugang zu den gespeicherten Daten in den Datenquellen bietet.

Zudem müssen zukünftig entwickelte Systeme flexibel dem stetigen Wechsel und der Weiterentwicklung der Komponenten gerecht werden. Ein wichtiges Konzept, dies zu erreichen sowie Heterogenität und Verteilung zu begegnen, ist das Konzept der Metadaten. Die Datenbankforschungsgruppe an der ETH Zürich verfolgt das Konzept der so genannten *Hyperdatabase* [Sch96], die Datenbankfunktionalität auf einer höheren Ebene der Abstraktion anbietet, z.B. auf der Ebene kompletter Informationskomponenten in einer n-schichtigen Architektur. In Analogie zu traditionellen Datenbanksystemen, die verteilte Daten und Transaktionen verwalten, verwaltet eine Hyperdatabase verteilte Informationskomponenten und Transaktionsprozesse. Entsprechend [ETH02] wird eine Hyperdatabase die Schlüsselinfrastruktur für das Entwickeln und Verwalten zukünftiger Informationssysteme werden. Sie wird die Komponenten- und Dienstedefinition unterstützen sowie den Einsatz dieser, ebenso die Spezifikation von Transaktionsprozessen einschließlich vielfacher Anwendungsdienstaufrufe. Anwendungen wären nicht nur entkoppelt gegen Änderungen bei der Datenhaltung und den Zugriffsstrukturen, sondern auch gegenüber Änderungen in den Anwendungskomponenten und -diensten.

3.2 Metadaten für föderierte Informationssysteme

[BKLW99] unterscheidet folgende Arten von Metadaten bezüglich föderierter Informationssysteme, speziell der Föderationsschicht:

1. **Technische Metadaten** beschreiben Informationen bezüglich der technischen Zugriffsmechanismen von Komponenten, wie z.B. das eingesetzte Protokoll etc. Sie werden eingesetzt, um technische und Schnittstellen-Heterogenitäten zu überbrücken.
2. **Logische Metadaten** beziehen sich auf Schemata und deren logische Beziehungen. Logische Metadaten sind z.B. erreichbar durch die Data Dictionaries in RDBMS oder als Klassendiagramme in OODBMS. Insbesondere sind die Beziehungen und Abhängigkeiten zwischen verschiedenen Schemata eines Datenmodells eine wichtige Art von logischen Metadaten in FIS.
3. **Metamodelle** als Metadaten unterstützen die Interoperabilität von Schemata in verschiedenen Datenmodellen. Sie beziehen sich hauptsächlich auf Datenmodellheterogenität.
4. **Semantische Metadaten** sind Informationen, die helfen, die Semantik von Begriffen zu beschreiben. Insbesondere Ontologien und Thesauren werden für diesen Zweck eingesetzt. Alle domänenspezifischen Beschreibungen gehören in diese Klasse.

5. **Qualitätsbezogene Metadaten** beschreiben quellenspezifische Eigenschaften von Informationssystemen bezüglich deren Qualität, wie z.B. Zuverlässigkeit, Aktualität usw.
6. **Infrastruktur-Metadaten** helfen Nutzern relevante Daten zu finden, wie z.B. Navigationshilfe, kommentierte Lesezeichen usw.
7. **Nutzerbezogene-Metadaten** beschreiben Verantwortlichkeiten und Präferenzen von Nutzern, z.B. Nutzer-Profile.

Ein FIS nutzt typischerweise Metadaten, die als Daten in einem Repository abgespeichert werden. Nicht näher vertieft werden in dieser Arbeit die technischen, qualitätsbezogenen, infrastruktur- und nutzerbezogenen Metadaten.

3.3 Typen von Föderierten Informationssystemen

Es gibt eine Definition von drei Typen von föderierten Informationssystemen: lose gekoppelte Informationssysteme, föderierte Datenbanksysteme und mediator-basierte Informationssysteme [BKLW99].

1. Lose gekoppelte Informationssysteme

Lose gekoppelte Informationssysteme bieten kein föderiertes Schema, sondern nur eine Multidatenbank-Anfragesprache, um auf die Komponenten zuzugreifen. Dies hat den Vorteil, dass Komponenten nicht ihre Autonomie aufgeben müssen, um an der Föderation teilzunehmen. Aber auf der anderen Seite wird keine Orts- oder Schema-Transparenz erreicht. Wird eine einheitliche Anfragesprache angeboten, wird technische und sprachbasierte Heterogenität durch das FIS überbrückt. Aber die fehlende logische Integration führt zu zahlreichen direkten Abhängigkeiten zwischen Anwendungen und Komponenten-Systemen.

2. Föderierte Datenbanksysteme (FDBS)

Föderierte Datenbanksysteme bieten klassische Datenbanksystem-Funktionalität auf globaler Ebene [Con97, SL90]. Dies schließt einen Lese- und Schreibzugriff für das Datenmanagement mit ein. Föderierte Datenbanksysteme sind eng gekoppelte Informationssysteme. Sie werden üblicherweise Bottom-Up konstruiert durch Anwendung von Schemaintegrationstechniken. Das föderierte Schema muss die Anforderungen an Vollständigkeit, Korrektheit, Minimalität und Verständlichkeit erfüllen, die nur möglich sind mittels Integration per Sammlung oder Fusion. Als eng gekoppeltes FIS bieten FDBS ihren Nutzern volle Orts- und Schema-Transparenz. Aber FDBS haben typischerweise eine statische Architektur mit Problemen bei der Systemweiterentwicklung aufgrund der Abhängigkeit von Schemaintegrationsprozessen, die weder ein einfaches Plug-in oder Plug-out von Komponenten erlauben, noch ein flexibles Änderungsmanagement.

3. Mediatorbasierte Informationssysteme

Eine logische Weiterentwicklung von Client-Server-Architekturen wird von den so genannten *vermittelten Architekturen* oder mediatorbasierten Informationssystem-Architekturen angeboten [Wie92, Wie95]. Bei der Vermittlung ist eine weitere Software-Schicht zwischen Client und Server eingefügt. Genauer geht darauf nun der Abschnitt 3.4 ein.

3.4 Mediatorbasierte Informationssysteme

Abbildung 3.2 zeigt die Drei-Schichten-Architektur nach Wiederhold, bei der die zusätzliche Vermittlungsschicht zwischen Basis- und Nutzerschicht illustriert wird.

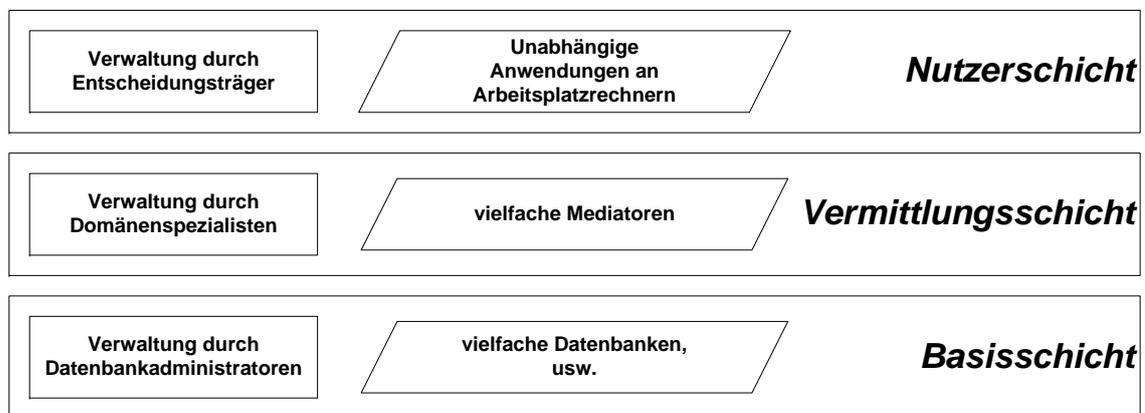


Abbildung 3.2: Die drei Schichten einer Mediator-Architektur nach [Wie92]

Die Vermittlungsschicht entkoppelt, sie macht die Applikationen unabhängig von den Datenquellen. Die vermittelnden Module, die Mediatoren, bringen Quellinformationen in eine gemeinsame Form [Wie95]. Eine allgemeine Definition von Mediator nach Wiederhold: Ein *Mediator* ist ein Software-Modul, das kodiertes Wissen über bestimmte Mengen oder Untermengen von Daten auswertet, um Informationen für eine höhere Schicht von Anwendungen zu kreieren. Mediatoren bieten nach [Wie92] also eine signifikante Funktionalität durch die Transformationen der von einem Server angebotenen Daten in Informationen, die die Applikation benötigt. Solche Transformationen benötigen Wissen, wo die Daten sind sowie Spezifikationen über die Repräsentation der Daten. Mediatoren sind domänenspezifisch, so dass ihre Wartung konzentriert werden kann. Anwendungen, die Informationen aus mehreren Domänen benötigen, würden zwei oder mehr Mediatoren nutzen. Seinerseits kann jeder Mediator auf verschiedene Informationsquellen zugreifen.

Facilitatoren sind spezielle Mediatoren, die Automatisierung, zu allererst bei Quellidentifikation und Datenformat-Umwandlungen, bieten [WG97]. Folgende Metadaten sollte ein Facilitator nutzen, um Daten und Informationen zu konvertieren, zu übersetzen oder weiterzuleiten:

- Metadaten über die Informationsquellen
- Logische Aussagen bezüglich heterogener Konzepte (Abbildungsregeln)
- Formatbeschreibungen

Mediatoren sind Software-Komponenten [Has02]. Abbildung 3.3 zeigt eine beispielhafte mediatorbasierte Integrations-Architektur mit verschiedenen Mediatoren als Ansatz für die Integration von heterogenen Informationssystemen.

Das Konzept der Mediatoren führt zu flexiblen und dynamischen Software-Architekturen. So trennt die mediatorbasierte Integrations-Architektur die Interessen von Domänenfacilitatoren, Komponentenmediatoren, Applikationsmediatoren und Wrappern. Domänenfacilitatoren verwalten die domänenspezifischen Modelle und bieten als Facilitatoren Koordinationsleistungen an. Komponentenmediatoren vollenden die Integration in standardbasierte, domänenspezifische Modelle. Dieser Ansatz erhöht die Skalierbarkeit, indem es eine Trennung der Verwaltung der globalen Modelle und der Integration von Komponentenmodellen in domänenspezifische Modelle erreicht. Ein Wrapper ist eine Komponente, welche den Zugriff auf Daten existierender Systeme ermöglicht, um eine Konformität mit Zugriffsstandards und Konventionen, die die Mediatoren und Facilitatoren nutzen, herzustellen. Auf die Wrapperschicht in Informationssystemen geht [Fri01] näher ein. Optionale Applikationsmediatoren vermitteln zwischen Domänenfacilitatoren und globalen Anwendungen. Der Domänenfacilitator bestimmt die Konditionen, die die Komponentenmediatoren erfüllen müssen, wenn sie an dem integrierten System teilhaben wollen.

Nach Art der *semantischen Integration* unterteilt [Bus02] mediator-basierte Informationssysteme (MBIS) in *metainformationsbasierte MBIS*, die domänenabhängige Metadaten verwenden und *integrierende MBIS*, die andere semantische Integrationsansätze verwenden, wie Schemaintegration oder wissenbasierte Integrationsansätze.

3.5 Entwurfsmuster zur Integration von Informationssystemen

Nachdem verschiedene Architekturmuster in Abschnitt 2.3 auf Seite 12 vorgestellt worden sind, soll nun kurz auf die Integrationsstile eingegangen werden, die grundsätzlich zur Integration von Informationssystemen möglich sind. Dies sind File Transfer, Shared Database, Remote Procedure Invocation und Messaging. Es gibt mehr als einen Ansatz Informationssysteme zu integrieren [CHKT05]. Jeder Ansatz ist für einige Integrationskriterien besser geeignet als andere. Die verschiedenen Ansätze werden in [CHKT05] zu vier Hauptintegrationsstilen zusammengefasst:

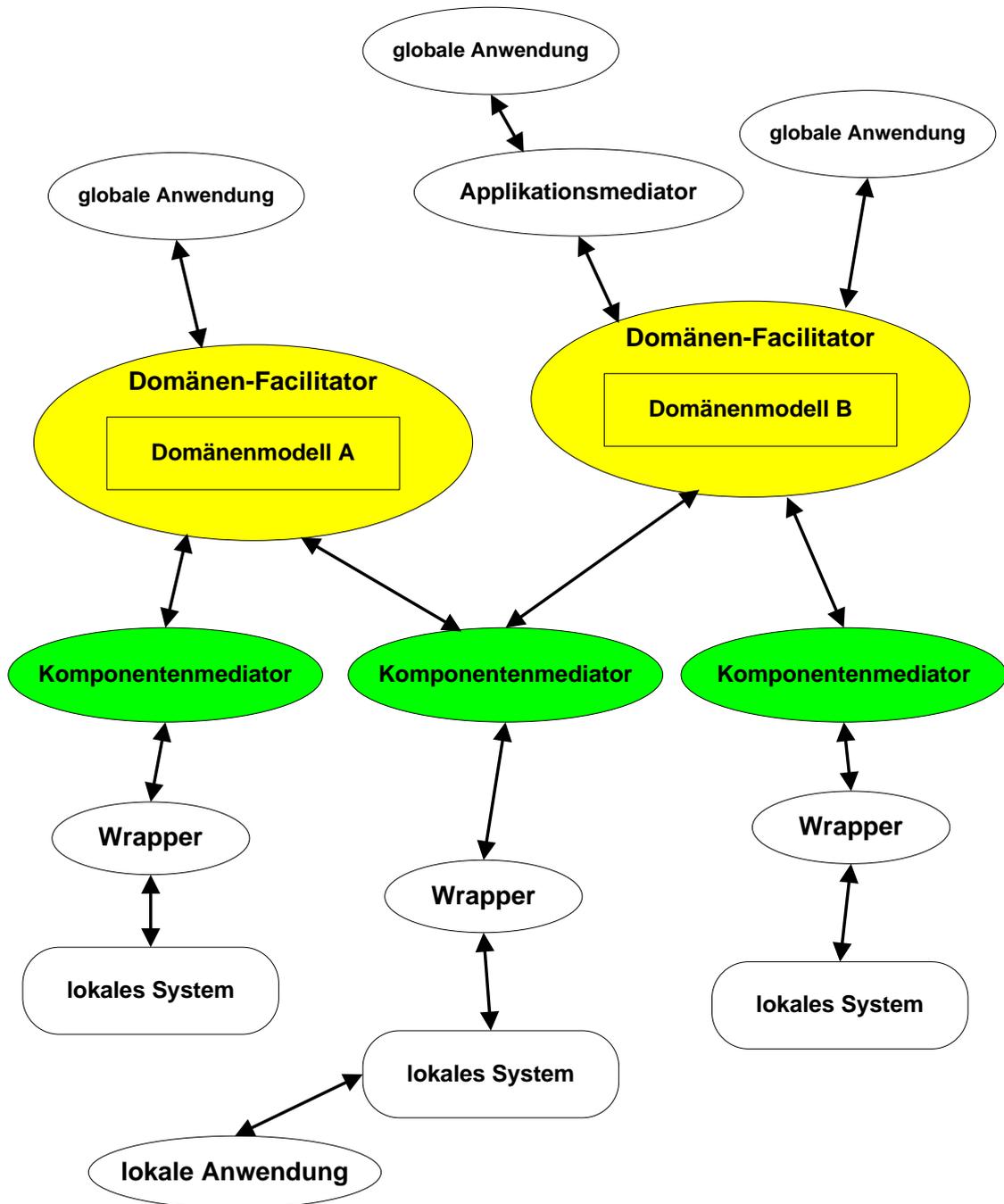


Abbildung 3.3: Eine mediatorbasierte Architektur nach [Has02]

- **File Transfer** — Jedes Informationssystem überträgt Dateien gemeinsamer Daten an andere Informationssysteme und erhält Dateien, die andere Informationssysteme erzeugt haben.
- **Shared Database** — Informationssysteme speichern die Daten, die sie teilen möchten in einer gemeinsamen Datenbank.
- **Remote Procedure Invocation** — Jedes Informationssystem stellt einige seiner Dienste für entfernte Aufrufe bereit, so dass sie von anderen Informationssystemen aufgerufen werden können. Die Informationssysteme rufen diese auf, um Daten auszutauschen und entsprechende Funktionalität zu nutzen.
- **Messaging** — Jedes Informationssystem ist mit einem gemeinsamen Nachrichtenserver verbunden und tauscht Daten und Funktionalität über Nachrichten aus.

File Transfer (Batch Data Exchange) Beim File Transfer bzw. Batch Data Exchange werden die Informationen in ein gemeinsames Dateiformat exportiert und dann im Zielsystem eingelesen (Beispiel: COBOL Flat-Dateien). Die Vorteile liegen in einer guten physischen Entkopplung sowie in der Unabhängigkeit von Sprache und System. Als Nachteil hat der File Transfer, dass kein direkter Datenzugriff möglich ist. Ferner können die Daten der Systeme inkonsistent sein. Probleme bereiten auch große Datenmengen.

Shared Database (Central Database) Die gemeinsame, zentrale Datenbank sorgt dafür, dass alle Anwendungen auf diese gemeinsame Datenbank zugreifen können. Vorteil der Shared Database sind konsistente Daten sowie direkt verfügbare Anfragen auf diese Daten. Allerdings werden nur Daten integriert und keine Geschäftsfunktionen. Schwierig ist es, eine gemeinsame Repräsentation für alle beteiligten Systeme zu finden.

Remote Procedure Invocation (Remote Procedure Call) Bei der Remote Procedure Invocation bzw. Remote Procedure Call ruft eine Anwendung eine andere direkt auf, damit diese eine Funktion ausführt. Die Daten, die für den Aufruf nötig sind, werden als Parameter übertragen. Die Ergebnisse kehren zur aufrufenden Anwendung zurück. Remote Procedure Invocation integriert Geschäftsfunktionen und nicht nur Daten. Die Daten sind zudem konsistent, falls sie bei Bedarf immer neu geholt und nicht repliziert gespeichert werden. Nachteil ist, dass Remote Procedure Invocation nur mit einer geringen Anzahl von Systemen wirklich gut funktioniert. Die entfernten Zugriffe können Performanzprobleme verursachen. Auch erreichen wir eine sehr enge Kopplung durch die synchronen Aufrufe. Aufrufende Systeme sind von der Verfügbarkeit der aufgerufenen Systeme abhängig (Einschränkung der Autonomie).

Messaging Beim Messaging, ähnlich einem Email-System, propagiert ein Message Bus die Ereignisse. Das Messaging hat den Vorteil, dass Daten nur dann ausgetauscht werden, wenn es notwendig ist. Ferner werden hier die Geschäftsfunktionen integriert und nicht nur die Daten. Messaging skaliert auch eine große Zahl von Systemen. Nachteil dieses Integrationsstils ist, dass er programmieretechnisch nicht sehr geläufig ist. Leider sind Tests und Fehlersuche schwierig, da bei diesem asynchronen Kommunikationsstil gelegentlich Synchronisierungen zur Konsistenzsicherung notwendig sind.

Messaging führt zu einer losen Kopplung von Informationssystemen durch asynchrone Kommunikation, die die beteiligten Systeme unabhängiger macht, da die betroffenen Informationssysteme nicht zur gleichen Zeit verfügbar sein müssen. Beim Messaging ist der Messaging-Server für die Übertragung der Daten von einem Informationssystem zu einem anderen verantwortlich. Die folgenden Begriffe sind hier relevant:

1. Channels — Messaging-Server übertragen Daten über einen Nachrichtenkanal, der als virtueller Kanal Sender und Empfänger verbindet.
2. Messages — Eine Applikation teilt Daten in ein oder mehrere Nachrichtenpakete auf, die über Nachrichtenkanäle versandt werden können. Eine Empfänger-Anwendung, die eine Nachricht erhält, kann die Daten aus der Nachricht extrahieren. Der Nachrichtenserver wird versuchen die Nachricht so oft zuzustellen (z.B. durch Übertragung vom Sender zum Empfänger), bis der Nachrichtempfang erfolgreich gewesen ist oder nach einem definierten Timeout abbrechen.
3. Überprüfung und Transformation — Im einfachsten Fall liefert ein Messaging-Server eine Nachricht direkt vom Computer des Senders zum Computer des Empfängers. Häufig werden Aktionen auf der Nachricht ausgeführt, nachdem sie vom ursprünglichen Sender versandt wurde, und bevor sie beim endgültigen Empfänger eingegangen ist. Beispielsweise wird die Nachricht validiert oder transformiert, weil der Empfänger ein anderes Nachrichtenformat benötigt. Eine Architektur mit Kanälen und Filtern beschreibt, wie größere Prozesse zerlegt werden können. Zahlreiche Informationssysteme können für die gleichen konzeptuellen Daten unterschiedliche Formate nutzen. Der Sender formatiert die Nachricht auf eine Weise, wohingegen der Empfänger eine andere Formatierung erwartet. Um dies zu überbrücken, muss die Nachricht in einem Zwischenfilter (Message Translator) von einem Format in ein anderes konvertiert werden.
4. Routing — In einem großen Unternehmen mit zahlreichen Informationssystemen und Kanälen, die diese verbinden, kann eine Nachricht durch verschiedene Kanäle gehen müssen, um ihr endgültiges Ziel zu erreichen. Der Weg, dem eine Nachricht folgen muss, kann so komplex sein, dass der ursprüngliche Sender nicht die Kanäle kennt, die die Nachricht zum endgültigen Empfänger bringt. Stattdessen sendet der ursprüngliche Sender die Nachricht an einen Message Router, eine Komponente und Filter in der Pipes-and-filters-Architektur [SG96]. Diese Architektur wird genau festlegen, wie in der Kanaltopologie navigiert wird und die Nachricht direkt an den endgültigen Empfänger gesendet werden kann, oder mindestens zum

nächsten Router. Bei Messaging Variationen mit einem Message Router und einem Übersetzer sind die Nachrichten vom Empfänger direkt an den Message Router gesendet worden, in dessen Eingangswarteschlange sie auf Weiterleitung warten. Diese eingegangenen Nachrichten werden anschließend vom Message Router, der die Kanaltopologie kennt, über entsprechende Kanäle zum Empfänger navigiert. Ein Übersetzer transformiert eine eingehende Nachricht in ein anderes Format und erzeugt die gewünschte transformierte Nachricht.

5. Adaptoren bzw. Wrapper — Ein Informationssystem hat im Allgemeinen keine eingebauten Möglichkeiten für eine Schnittstelle zu einem Messaging-Server. Vielmehr besitzt es eine Schicht mit Programm-Code, die die Funktionalität sowohl des Informationssystems als auch des Messaging-Servers kennt. Damit können beide Systeme so überbrückt werden, dass sie zusammenarbeiten. Dieser Brückencode ist eine Menge koordinierter Message-Adaptoren, die ein Informationssystem in die Lage versetzen, Nachrichten zu senden und zu empfangen.

3.6 Zusammenfassung

Für diese Arbeit sind mediatorbasierte Informationssysteme von Bedeutung, da entsprechende Mediatoren eine geeignete Problemlösung darstellen. Mediatoren bieten flexible Gestaltungsmöglichkeiten an. Die Standards für Kommunikation und Dokumentation können als domänenabhängige Metadaten dienen.

Als Integrationsstil bietet sich Messaging an, da die lose Kopplung der beteiligten Informationssysteme hohe Flexibilität bietet, weil die beteiligten Systeme nicht zur gleichen Zeit verfügbar sein müssen. Es werden Daten und Abläufe integriert, und die Daten werden nur bei Bedarf ausgetauscht. Gegen den Integrationsstil File Transfer spricht die Gefahr inkonsistenter Daten sowie die Probleme bei großen Datenmengen. Inkonsistente Daten in der Medizin können sogar gefährlich sein. Durch die Vielzahl der Dienstleister im Gesundheitswesen ist eine gemeinsame Datenbank nahezu unmöglich. Da vor allem auch Abläufe effizienter gestaltet werden sollen, ist diese Integrationsmöglichkeit ungeeignet. Remote Procedure Invocation integriert zwar Daten und Abläufe, ist jedoch nur bei einer geringen Anzahl von Systemen gut einsetzbar und wäre damit keine passende Lösung.

4 Metadaten und ihre Modellierung

4.1 Metadaten-Standards

In diesem Abschnitt soll nur kurz und übersichtsartig das Gebiet der Metadaten-Standards gestreift werden. Metadaten sind beschreibende Daten über Daten. Standards für Metadaten verfolgen zwei wesentliche Ziele:

- Verbesserung der Qualität von Metadaten und der Daten, die sie beschreiben
- Verbesserung der Interoperabilität bei Einsatz von Metadaten

Metadaten-Standards helfen dabei, diese Ziele zu erreichen, da sie eine konsistente Repräsentation und eine konsistente Interpretation von Daten und Metadaten erlauben. Es können drei Kategorien von Metadaten-Standards unterschieden werden:

1. Framework-/Architekturstandards – schwerpunktmäßiger Einsatz, um andere Standards zu entwickeln
2. Metamodellstandards – Haupteinsatz bei der Spezifikation von Standardschnittstellen, z.B. für Implementationen
3. Daten- und Nachrichtendefinitionsstandards – Spezifikation von Datenrepräsentation und Semantik

4.1.1 Framework-/Architekturstandards

Dazu zählen beispielsweise die modellgetriebene Architektur (MDA) der OMG und der Standard Information Resource Dictionary Systems (IRDS).

Modellgetriebene Architektur (MDA)

Die modellgetriebene Architektur definiert einen Softwareentwicklungsansatz, der auf Modellierung und automatisierter Abbildung von Modellen auf Implementationen basiert. Das Basis-MDA-Entwurfsmuster umfasst die Definition eines plattformunabhängigen Modells (PIM) und seine automatisierte Abbildung auf eines oder mehrere plattformspezifische Modelle (PSM). Der MDA-Ansatz verspricht eine Reihe von Vorteilen einschließlich einer verbesserten Portabilität durch die Trennung des Anwendungssystemwissens von der Abbildung auf eine spezifische Implementierungstechnologie. Zudem verspricht er eine verbesserte Produktivität durch die Automatisierung der Abbildung, verbesserte Qualität durch die Wiederverwendung gut geprüfter Entwurfsmuster und Verfahren bei der Abbildung und verbesserte Wartbarkeit durch eine bessere „Separation of Concerns“ sowie bessere Konsistenz und Nachvollziehbarkeit zwischen Modellen und Code. Während die aktuellen OMG-Standards wie Meta Object Facility (MOF) und Unified Modeling Language (UML) eine etablierte Basis für die Definition von PIMs und PSMs anbieten, existiert nichts vergleichbares für die Transformation der PIMs in die PSMs. Die OMG hat 2002 in ihren Bemühungen diese Situation zu ändern, einen Standardisierungsprozess initiiert durch die Aufforderung Vorschläge auf dem Gebiet QVT (Query, Views, Transformations) einzureichen. Dieser Prozess führt eventuell zu einem OMG-Standard für die Definition von Modelltransformationen. Dieser Standard wäre nicht nur interessant für PIM-zu-PSM-Transformationen, sondern auch für die Definition von Sichten auf Modelle und Synchronisation zwischen Modellen.

Information Resource Dictionary Systems (IRDS)

Dieser ISO-Standard wurde 1990 veröffentlicht. Zahlreiche Basiskonzepte sind in die Meta Object Facility (MOF) der Object Management Group (OMG) eingeflossen. MOF basiert auf der Unified Modeling Language (UML).

4.1.2 Metamodellstandards

Bekannte Metadaten-Standards der Object Management Group (OMG) sind die Meta Object Facility und das Common Warehouse Metamodel. Bekannte Metadaten-Standards des World Wide Web Consortium sind die Document Type Definition (DTD) bzw. das Schema der Extensible Markup Language (XML) sowie das Resource Description Framework (RDF).

Meta Object Facility (MOF)

MOF überbrückt die Lücken zwischen heterogenen Metamodellen durch eine gemeinsame Basis für Metamodelle. Sind zwei unterschiedliche Metamodelle beide MOF-konform, können Modelle, die auf ihnen basieren, im selben Repository zu Hause sein. MOF ist das Metamodell, das die gesamte Softwareentwicklungsarchitektur (einschließlich sich selbst) beschreibt. Es beschreibt Modellierungssprachen wie die Unified Modeling Language (UML), eine Menge von Technologie-Metamodellen wie das CORBA Komponentenmodell (CCM), Enterprise JavaBeans (EJB) und andere nutzerdefinierte Metamodelle. Die MOF umfasst außerdem eine Menge von Regeln, die Interoperabilitätssemantiken und Austauschformate für all seine Metamodelle spezifizieren.

Common Warehouse Metamodel (CWM)

Im August 2000 vereinigte die Meta Data Coalition (MDC) seine Metadaten-Austauschstandards Open Information Model (OIM) und Meta Data Interchange Specification (MDIS) mit dem CWM-Standard der OMG. CWM ist eine Spezifikation, die Metadatenaustausch beschreibt zwischen Data Warehouses, Geschäftsintelligenz, Wissensmanagement und Portaltechnologien. Die OMG-Spezifikation fußt dabei hauptsächlich auf XML Metadata Interchange (XMI) und MOF.

Extensible Markup Language - Document Type Definition (XML-DTD)

Eine DTD beschreibt den strukturellen Aufbau und die logischen Elemente einer Klasse von Dokumenten, genannt Dokumenttyp [BM00]. Die DTD beschreibt damit die strukturellen Gemeinsamkeiten von gleichartigen Dokumenten.

XML-Schema

Das XML-Schema ist Nachfolger der DTD [BM00]. DTDs können „nur“ die Struktur von Elementen eines Datensatzes festlegen. Sie können aber keine Einschränkungen für den Inhalt von Elementen machen. Dies ist mit dem XML-Schema möglich. Ein XML-Schema definiert Elemente und Attribute, die in einem Dokument auftauchen können. Es definiert, welche Elemente Subelemente sind, wieviel Subelemente es gibt und in welcher Reihenfolge. Es definiert Default- oder feste Werte für Elemente oder Attribute sowie Datentypen für Elemente und Attribute.

Resource Description Framework (RDF)

RDF wird vom W3C (World Web Consortium) empfohlen, um Metadaten von Web-Ressourcen zu modellieren. Die Modell- und Syntax-Spezifikation [RDF04b] beschreibt die syntaktischen Aspekte von RDF. Die Schema-Spezifikation (RDFS) [RDF04c] beschreibt dagegen die Definition von Vokabularen (oft Schemata genannt) und damit die semantische Seite von RDF. Die Ontology Web Language (OWL) wurde entwickelt, um es Anwendungen zu ermöglichen, die Bedeutung von Informationen zu verarbeiten anstatt die Informationen dem Anwender nur zu präsentieren [OWL04]. OWL erleichtert durch zusätzliches Vokabular in Verbindung mit formaler Semantik stärkere Interpretationsmöglichkeiten von Web Inhalten als dies XML, RDF und RDFS allein ermöglichen. OWL ist eine Ontologierepräsentationssprache mit XML-Syntax, die das Ausdrücken z.B. von Klassenäquivalenzen ermöglicht. Wichtiges Ziel dieser W3C-Initiative ist die Beschreibung von Web-Ressourcen mit offenem Format, d.h. jeder kann alles über jedes aussagen in Form eines in XML ausdrückbaren (damit maschinenlesbaren), einfachen Datenmodells. RDF legt als Framework dabei formal die Modellierung und Spezifikation (z.B. XML) fest. Die Inhalte und das Vokabular können für jeden Anwendungsbereich frei definiert werden. So können beispielsweise in [Gra05] Nutzeranfragen an das Web unter anderem mit Hilfe der RDF-Technologie dynamisch verarbeitet werden, um die relevanten Inhalte sehr viel besser zur Verfügung zu stellen. Näheres zu Semantic Web, RDF und OWL ist in Anhang B auf Seite 183 zu finden.

4.1.3 Daten- und Nachrichtendefinitionsstandards

Standards für Datenelemente

Auch für Datenelemente gibt es zahlreiche Standards. Als weit verbreitete Standards seien die Postleitzahlen in der Logistik und die International Standard Book Number (ISBN, <http://www.isbn.org>) genannt. Die ISBN ist eine einmalige maschinenlesbare Identifikationsnummer, die jedes Buch unverwechselbar markiert. Vor 30 Jahren hat die ISBN den internationalen Buchhandel revolutioniert.

Standards für Nachrichten

Es existieren verschiedene Standardisierungsinitiativen im Bereich des E-Commerce [Has02]. Electronic Data Interchange (EDI) ermöglicht einen strukturierten elektronischen Austausch von Informationen zwischen Geschäftspartnern. Verabschiedete Standards für EDI sind ANSI X.12 und UN/EDIFACT (EDIFACT=Electronic Data Interchange For Administration, Commerce and Transport).

Eine der Schwierigkeiten mit diesen EDI-Standards ist die Tatsache, dass sie sehr abstrakt sind. Inzwischen ist die Extensible Markup Language (XML) als Standard für die Definition der Syntax von über das Internet übertragenen Datenstrukturen etabliert. Ziel ist es, mit XML ein einheitliches und universelles Datenformat zu besitzen, das sowohl maschinenlesbar als auch für menschliche Augen verständlich ist [BM00]. Auf diese Weise wird die konkrete Syntax von EDI-Nachrichten festgelegt. Standardisierungsinitiativen, die EDI mit XML kombinieren, sind beispielsweise die XML/EDI-Gruppe (<http://www.XMLEDI.org>) und ebXML (e-business XML, <http://www.ebxml.org>). Electronic Business using eXtensible Markup Language (ebXML) ist Beispiel für modernes EDIFACT. ebXML ist eine internationale Initiative, die 1999 von UN/CEFACT (United Nations Centre for Trade Facilitation and Electronic Business) und OASIS (Organization for the Advancement of Structured Information Standards) etabliert worden ist [eBT03, ebX03, OAS03, HW01]. Das Ziel von ebXML ist es, offenen Handel zwischen Organisationen ohne Rücksicht auf ihre Größe durch konsistenten Einsatz von XML beim Austausch elektronischer Geschäftsdaten zu ermöglichen. Dabei arbeitet UN/CEFACT an der Entwicklung der ebXML-Geschäftsprozesse, OASIS an der Entwicklung der technischen Grundlagen. Beim Einsatz im Gesundheitswesen können HL7-Nachrichten in ebXML-Umschlägen transportiert werden [ebX03]. Damit bewegt sich ebXML auf der Transportebene und spezifiziert nicht die zu transportierenden Inhalte wie die Standards HL7, BDT oder CDA in den Abschnitten 6.2 auf Seite 48 und 6.3 auf Seite 57. Verwandte Techniken zu ebXML wären vielmehr VCS und PaDok (siehe Abschnitt 6.5.2).

IFX (Interactive Financial Exchange) ist ein domänenspezifischer Standard für das Bankwesen (<http://www.ifxforum.org>). Dieser Standard definiert XML-Nachrichten für Bankdienste. IFX ist Nachfolger des Open Financial Exchange (OFX) Standards. Das Bank Internet Payment System (BIPS) ist ein Projekt des Financial Services Technologie Consortium (FSTC). Die Bankenindustrie möchte elektronischen Zahlungsverkehr über Internet ermöglichen, und BIPS ist ein Weg zu einer entsprechenden Architektur, einer offenen Spezifikation, einer sicheren Vorgehensweise für Zahlungen im Internet und eine Standardschnittstelle zu existierenden Bankzahlungssystemen (<http://www.fstc.org/projects/bips/index.cfm>). Für Standardisierungsbemühungen aus dem Bereich der Versicherungswirtschaft sei ACORD genannt. ACORD (Association for Cooperative Operations Research and Development) mit Sitz in New York ist ein global agierender, gemeinnütziger Zusammenschluss von Versicherern. Diese Vereinigung entwickelt und wartet zentral zahlreiche elektronische Standards für Versicherungen, Rückversicherungen und verwandte Finanzdienstleister (<http://www.acord.org>).

HL7 ist ein Beispiel für einen Nachrichtenstandard im Bereich der Medizin. HL7 wird ausführlich in Abschnitt 6.2.2 beschrieben. Weitere Nachrichtenstandards aus dem Gesundheitswesen werden in Abschnitt 6.2 auf Seite 48 vorgestellt.

4.2 Mehrebenen-Modellierungsframework

Für eine saubere Metamodellierung mit der verwendeten Modellierungssprache UML soll im Folgenden ein Framework vorgestellt werden, das mit dem Konzept der tiefen Instanziierung die Zusammenhänge der Metadaten (siehe Abschnitt 7.3.1) gut darzustellen vermag.

4.2.1 Metamodellierung über mehr als zwei Ebenen

Das Instanzierungsmodell der UML ist nicht für mehrere Modellierungsebenen sauber skalierbar [AK01]. Solange es nur zwei Ebenen gibt — eine mit Klassen (M1), die eine zweite mit Objekten (M0) instanziiert, funktioniert der Instanzierungsmechanismus gut. Wenn jedoch die UML-Konzepte eingesetzt werden, um eine Ebene zu beschreiben, deren Instanz die Klassenebene ist, z.B. die UML Metamodell-Ebene (M2), dann unterstützen die UML-Konzepte keinen natürlichen Modellierungsansatz. Im Einzelnen scheitern sie bei der adäquaten Beschreibung eines Modellelements auf M1-Ebene, das eine Klasse (für weitere Instanziierung) genauso gut wie eben ein Objekt repräsentiert. Leider basieren die aktuell eingesetzten oder vorgeschlagenen Metamodellierungsansätze für UML auf einem Instanzierungsmechanismus, der eine Instanz eines Modellelements gerade wie ein Objekt behandelt. Obwohl eine Instanz auch als Klasse repräsentiert werden kann, sind die Eigenschaften, die es aufgrund des Instanzierungsmechanismus' erhält, nur solche des Objekts. Wenn ein Element der M1-Ebene Attribute und/oder Assoziationen benötigt, müssen diese entweder explizit definiert oder vererbt werden. Die Unfähigkeit des Instanzierungsmechanismus', Informationen betreffend der Attribute und Assoziationen über mehr als eine Ebene bei der Instanziierung zu bringen, ist die Ursache für Probleme in den aktuell vorgeschlagenen UML-Metamodellierungs-Frameworks.

4.2.2 Probleme bei flacher Instanziierung

Eines dieser ernstesten Probleme ist das Problem der *mehrdeutigen Klassifikation*, das andere das Problem der *Replikation der Konzepte*. [AK01] führt einen fundamental anderen Ansatz für Metamodellierung ein, der diese Probleme korrigiert und eine Basis für eine Mehrebenen-Modellierungsarchitektur anbietet. Dieser Ansatz basiert auf der Voraussetzung, dass die Definition von Modellierungselementen inhärent die Möglichkeit von Instanzierungssequenzen erkennen sollte, die mehr als zwei Modellierungsebenen umspannen und sollte erlauben, dass Attribute und Assoziationen nur durch Instanziierung erhalten werden.

Obwohl die dualen Objekt-/Klassen-Eigenschaften vieler Modellierungselemente weitgehend anerkannt sind, werden diese nicht adäquat unterstützt durch das traditionelle flache Instanziierungsmodell. Im Gegenteil, das traditionelle Instanziierungsmodell ist genau genommen deswegen *flach*, weil die Klassenfacette der Modellelemente, die als Typ repräsentiert sind, immer explizit dokumentiert werden muss.

4.2.3 Lösung: Tiefe Instanziierung

Zur Lösung der Probleme wird ein Instanziierungsmechanismus benötigt, durch den die Klassenfeatures der Modellierungselemente automatisch durch den Instanziierungsschritt selbst erworben werden können, als sie immer explizit zu definieren. Dieser Instanziierungsmechanismus wird von [AK01] als „*tiefe Instanziierung*“ bezeichnet. Nur mit einem tiefen Instanziierungsansatz wird es möglich sein, eine Instanz und Klassenfacette von Elementen kohärent einmal und nur einmal zu definieren.

Konzept der Potenz

Der Weg tiefe Instanziierung zu erreichen, ist jedem Modellelement auf jeder Ebene innerhalb des Modellierungsframeworks das Konzept von *Potenz* zu ergänzen. Die Potenz eines Modellelements ist dabei eine ganze Zahl, die die Tiefe definiert, bis zu welcher das Modellelement instanziiert werden kann. Daher korrespondiert ein Modellelement mit der Potenz 0 mit einem Objekt oder jedem Konzept, das nicht für weitere Instanziierung gedacht ist (z.B. eine Schnittstelle oder eine abstrakte Klasse). Ein Element mit der Potenz 1 korrespondiert mit einer Klasse, einem Attribut oder einer Assoziation, die nur einmal instanziiert werden soll. Ein Element mit der Potenz 2 kann eine Instanziierung zweimal unterstützen. Der Vorgang der Instanziierung eines Modellierungselements reduziert seine Potenz um 1. Dieser Ansatz vereinheitlicht daher die Konzepte von Klasse und Objekt in ein einziges Konzept *Modellierungselement*, das Typen (Klassen) und Instanzen (Objekte) durch deren jeweilige Potenz unterscheidet.

Konzept der Ebene

Ein andere wichtige Eigenschaft, die jedes Modellelement besitzt, ist seine *Ebene*. Wie der Name impliziert, ist die Ebene eines Modellelements eine Integerzahl, die die Modellebene repräsentiert, zu der das Element gehört. Daher hat ein M0-Element den Wert 0, ein M1-Element den Wert 1, usw.

Die Notation der Konzepte Potenz und Ebene sind in Abbildung 4.1 dargestellt. Hochgestellt wird die Potenz und tiefgestellt die Ebene eines Modellierungselements angegeben.



Abbildung 4.1: Notation für Potenz und Ebene für ein Modellierungselement

Vorteile dieser beiden neuen Konzepte

Die Ebenen- und Potenzeigenschaft eines Modellelements zusammen erlauben es vielen der Eigenschaften des gewünschten Instanzierungsmechanismus' auf eine sehr kurze und konkrete Weise ausgedrückt zu werden. Instanziierung innerhalb eines strikten Mehrebenen-Modellierungsframework nimmt ein Modellelement mit Ebene e ($e > 0$) und Potenz p ($p > 0$) und erzeugt ein Element mit Ebene $e-1$ und Potenz $p-1$. Da Instanziierung beide Werte um 1 reduziert, kann sie nur angewendet werden auf Modellelemente, deren Potenz und Ebene größer als 0 sind. Elemente, deren Potenz 0 ist, können nicht instanziiert werden, unabhängig von ihrer Ebenennummer. Ein gutes Beispiel für ein Modellelement mit Potenz 0 ist die Vererbungsbeziehung zwischen zwei Elementen, wenn sie nicht weiter instanziiert werden kann. Viele Charakteristiken von heute eingesetzten Metamodellierungs-Architekturen, wie z.B. OMG's MOF- und UML-Architekturen können prägnant erfasst werden durch Bedingungen und Beziehungen zwischen den Ebenen- und Potenzwerten. Die Tatsache, dass M0 als unterste Ebene für alle Elemente vorausgesetzt wird, wird erfasst durch die Regel, dass die Potenz eines Modellelements nicht größer als seine Ebene sein kann. Der Fakt, dass eine Modellierungsarchitektur nur eine bestimmte Anzahl von Ebenen anerkennt, wird durch eine einfache Begrenzung im Ebenenwert erreicht. Die Semantik der flachen Instanziierung schließlich könnte durch die Bedingung erfasst werden, dass die Potenz von Elementen nicht größer als 1 sein kann.

Die Semantik der Instanziierung ist eng verknüpft mit den Eigenschaften der Attribute und Slots. Die Potenz- und Ebenenwerte eines Modellelements zusammen erlauben zahlreiche Vereinfachungen bei Mehrebenen-Modellierung. Eine der signifikantesten ist die Vereinheitlichung der Attribut- und Slotkonzepte. Ein Slot ist nur ein Attribut, welches die Potenz 0 haben soll. Um einen bequemen Weg anzubieten, um über Attribute und Slots allgemein (und einheitlich) zu sprechen und eine Verwirrung bezüglich der Semantik der existierenden Begriffe zu verhindern, wird von [AK01] das Konzept des Feldes eingeführt. Ein Feld mit Potenz 0 repräsentiert einen Slot und muss daher einen Wert besitzen. Ein Feld mit höherer Potenz kann einen oder kann keinen Wert besitzen in Abhängigkeit, ob es sich um ein Einzel- oder ein Dualfeld handelt. Ein Einzelfeld ist ein Feld, welches keinen Wert hat, wenn es die Potenz 0 hat (und korrespondiert so mit einem traditionellen Slot). Ein Dualfeld ist ein Feld, welches einen Wert für Potenzen größer als 0 hat. Ein Dualfeld hat kein Gegenstück in traditionellen Modellierungsansätzen. Die Semantik eines Dualfeldes kann am besten als Menge von Einzelfeldern gedacht werden. Felder haben ihre eigenen Ebenen- und Potenzwerte. Offensichtlich muss die Ebene eines Feldes mit der Ebene des Modellelements übereinstimmen, zu dem es gehört, die Potenz dagegen

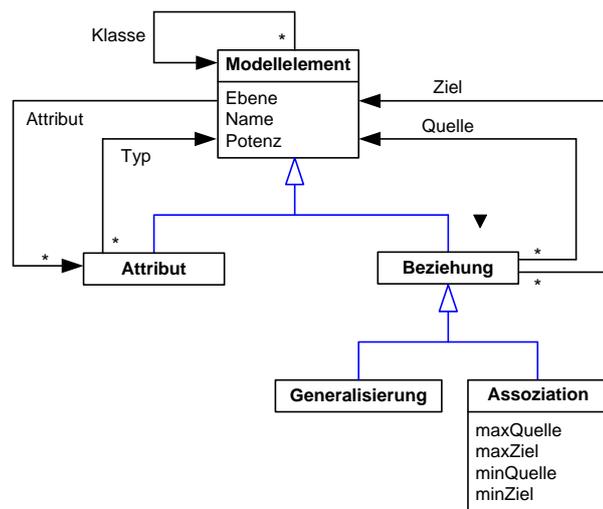


Abbildung 4.2: Das Modell für mehrfache Metaebenen (MoMM) nach [AK01]

muss nicht übereinstimmen. Die Potenz eines Modellelements wird hochgestellt angegeben und die Ebene tiefgestellt. Zusätzlich werden Felder mit der Potenz 0 (die einen Wert besitzen) durch Unterstreichen unterschieden. [AK01] ist bewusst, dass das Unterstreichen zur Zeit genutzt wird, um statische Attribute zu kennzeichnen, argumentiert jedoch dagegen. Felder mit Potenz 0 korrespondieren exakt mit den instanziierten Attributen, wohingegen die Korrespondenz zwischen statischen Attributen und instanziierten Attributen nur teilweise gegeben ist.

Metamodell für mehrfache Metaebenen

Es ist lange das Ziel der Modellierer gewesen, alle fundamentalen Modellierungsideen in einem Mehrebenen-Framework in einem kleinen, kohärenten und einfachen Kern zu erfassen, das üblicherweise als Metametamodell bezeichnet wird. Bisher war eine Umsetzung nicht zufriedenstellend aufgrund der zuvor beschriebenen Probleme. Zum Beispiel dupliziert die MOF, die als Metametamodell für UML vorgeschlagen wird, viele Modellierungskonzepte der UML (aufgrund des Problems der Replikation der Konzepte) und ist nicht eindeutig bezüglich der richtigen Positionierung der Modellelemente.

Eine vorläufige Version des MoMM (Metamodel for Multiple Metalevels) ist in Abbildung 4.2 dargestellt.

Der zu beachtende Kernaspekt dieses Metametamodells ist, dass jedes Modellelement, egal wo in der Metamodellierungs-Architektur, einen Ebenen- und Potenzwert haben muss. Diese müssen als Dualfelder des zugrunde liegenden Basismodellelements repräsentiert werden. Sie müssen Dualfelder sein, da alle Elemente in einer Instanzierungskette bis zu einem Maximum von drei einen Ebenen- und Potenzwert benötigen.

Das grundlegende Problem ist, dass die traditionelle Zwei-Ebenen Semantik der Instanziierung nicht aufwärts skalierbar ist für Metamodellierungs-Hierarchien mit mehr als zwei Ebenen. Die ursprünglichen UML Modellierungskonzepte sind ausreichend um eine Modellierung auf der M1-Ebene zu erlauben, aber scheitern an den Bedürfnissen auf der Ebene M2. Dieses Problem verursacht eine Reihe von Symptomen in existierenden Mehrebenen-Modellierungsarchitekturen, wie UML, einschließlich das Problem der mehrfachen Klassifikation und das Problem der Replikation von Konzepten. Obwohl verschiedene Strategien entwickelt worden sind, um die Auswirkungen dieser Probleme zu mildern, ist keine wirklich komplett erfolgreich bei der Vermeidung all dieser Symptome. [AK01] haben daher eine Reihe von Prinzipien vorgestellt, die eine grundsätzliche Lösung der Probleme bieten und so einen Weg offerieren für eine klare, einfache und kohärente Semantik für Metamodellierung. [AK01] demonstriert wie diese Ideen die Beschreibung realer Modellierungsszenarien, die aktuell in UML zu finden sind, vereinfacht.

4.3 Zusammenfassung

Metadaten sind geeignet die Interoperabilität zwischen Anwendungssystemen zu verbessern. Da die Standards in UML modelliert wurden, ist der Metamodellstandard MOF der naheliegende Standard. An ihn wurde die Metamodellierung angelehnt. Bei den Metadaten-Standards des W3C sind XML-DTD bzw. XML-Schema und RDF von Interesse, weil sie durch den stark gestiegenen Einsatz des Internets in den Mittelpunkt gerückt sind, wie auch [Len05] bemerkt.

Mit dem Mehrebenen-Modellierungsframework kann deutlich gemacht werden, welchen Zusammenhang die Metadaten der zu kombinierenden Standards haben. Auf die Metadaten, die als UML-Modelle erstellt worden sind, wird in Abschnitt 7.3 ab Seite 83 näher eingegangen. Die Konzepte der Ebene und Potenz sind geeignet, um klar zu machen, welcher Ebene die UML-Modelle jeweils zugeordnet sind und wie oft instanziiert werden kann. Der Zusammenhang der Metadaten wird in Abschnitt 7.3.1 auf Seite 83 vorgestellt. Dieser Zusammenhang sowie die einzelnen UML-Modelle haben entscheidenden Einfluss auf die Transformationen zwischen den verschiedenen Standards, da sie als Leitfaden zur Strukturanalyse der Standards und deren Abbildungen zwischen ihnen dienen.

5 Modelltransformation

Die modellgetriebene Architektur (MDA=Model Driven Architecture) ist eine Initiative der Object Management Group (OMG), um die Generierung von plattformspezifischen Modellen aus plattformunabhängigen Modellen zu automatisieren. Während es inzwischen einige etablierte Standards für die Modellierung plattformabhängiger Modelle gibt, gibt es zur Zeit keine ausgereifte Grundlage für die Spezifikation von Transformationen zwischen solchen Modellen. Es wird in [CH03] eine mögliche Taxonomie für die Klassifikation der verschiedenen existierenden und vorgeschlagenen Modelltransformations-Ansätze vorgestellt. Die Taxonomie wird dabei mit einem Feature-Modell beschrieben, das die unterschiedlichen Entwurfswahlmöglichkeiten für Modelltransformationen explizit aufzeigt. Basierend auf dieser Analyse werden Hauptkategorien vorgeschlagen, in die sich die meisten Modelltransformations-Ansätze einordnen lassen. Auf der Basis dieser Taxonomie soll ein geeigneter Transformationsansatz ausgewählt werden. [CH03] stellt ein Feature-Modell vor, welches die verschiedenen Modelltransformationsansätze vergleicht und offeriert einen Überblick und Kategorisierung einer Reihe existierender Ansätze:

1. veröffentlicht in der Literatur
2. eingereicht als Antwort auf die OMG-Aufforderung
3. implementiert in Open-Source MDA-Tools und
4. implementiert in kommerziellen MDA-Tools

Das Featuremodell zeigt die unterschiedlichen Entwurfswahlmöglichkeiten für Modelltransformationen explizit auf.

5.1 Entwurfs-Eigenschaften von Modelltransformationsansätzen

Abbildung 5.1 zeigt das Feature-Diagramm auf oberster Ebene, bei welchem jeder Unterknoten eine wesentliche Variationsmöglichkeit repräsentiert. Es ist das Ergebnis einer

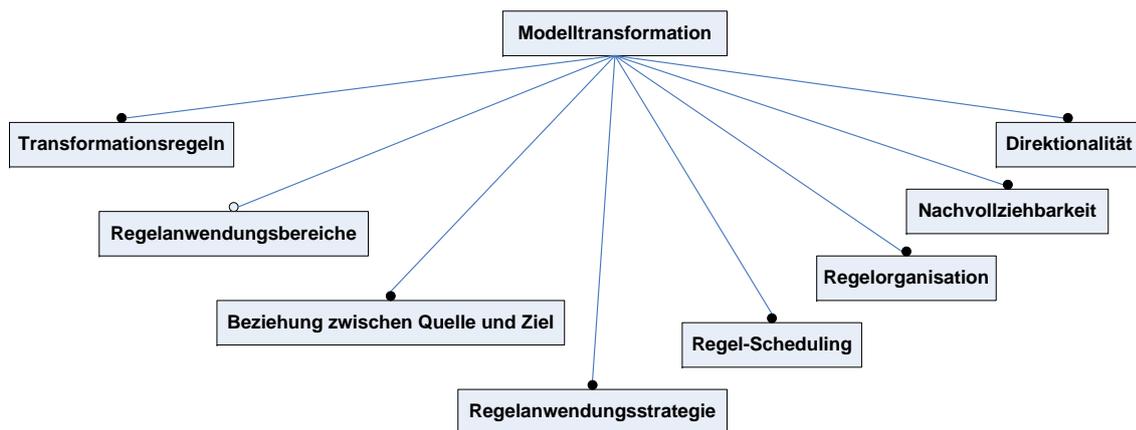


Abbildung 5.1: Ein Feature-Diagramm nach [CH03], das auf oberster Ebene die Hauptvariationensgebiete darstellt

Domänenanalyse auf existierende Ansätze [CH03]. Im wesentlichen definiert ein Feature-Diagramm eine Taxonomie.

Die Verbindungslinien in Abbildung 5.1 mit einem schwarzen kleinen Kreis am Ende symbolisieren dabei obligatorische Eigenschaften, der weiße kleine Kreis am Ende der Linie optionale Eigenschaften.

Transformationsregeln

Eine Transformationsregel besteht aus zwei Teilen: eine linke (LHS) und eine rechte Seite (RHS). Die linke Seite greift auf das Quellmodell zu, wohingegen die rechte Seite sich auf das Zielmodell bezieht. Beide Seiten der Regel können repräsentiert werden durch eine Mischung des folgenden:

1. Transformations-Variablen: Diese Variablen beinhalten Elemente aus dem Quell- und/oder Zielmodell (oder einigen Zwischenelementen).
2. Transformations-Entwurfsmuster: Diese Entwurfsmuster sind Modellfragmente mit keiner oder mehr Variablen. Es gibt String-, Begriffs- und Graphenentwurfsmuster. Stringentwurfsmuster werden in textuellen Vorlagen eingesetzt. Modell-zu-Modelltransformationen nutzen üblicherweise Begriffs- oder Graphenentwurfsmuster.
3. Transformations-Logik: Diese Logik drückt Berechnung und Constraints der Modellelemente aus.

Vier andere Aspekte bei Transformationsregeln sind:

1. Syntaktische Trennung: rechte und linke Seite können oder können nicht syntaktisch getrennt werden.
2. Bidirektionalität: Eine Regel kann in beide Richtungen ausführbar sein.
3. Regelparametrisierung: Transformationsregeln können zusätzliche Kontrollparameter besitzen, die Konfiguration und Tuning erlauben.
4. Zwischenstrukturen: Einige Ansätze erfordern die Konstruktion von Zwischenmodellstrukturen. Dies ist besonders relevant, wenn die Modelltransformation innerhalb eines Modells stattfindet.

Regelanwendungsbereiche

Die Definition von Bereichen für die Regelanwendung erlaubt einer Transformation die Teile eines Modells, die an der Transformation teilnehmen, einzuschränken. Einige Ansätze bieten flexible Bereichswahl beim Quellmodell, wo ein Bereich definiert werden kann, der kleiner ist als das gesamte Quellmodell. Letzteres kann aus Performanzgründen wichtig sein. Der Zielbereich ist derjenige Bereich, in dem die rechte Seite der Regel angewandt wird.

Beziehung zwischen Quelle und Ziel

Einige Ansätze erzwingen die Erstellung eines neuen Zielmodells, welches getrennt von der Quelle sein muss. In einigen anderen Ansätzen sind Quelle und Ziel stets das gleiche Modell, weil Aktualisierungen nur innerhalb eines Modells unterstützt werden. Noch andere Ansätze erlauben, dass das Zielmodell ein neues oder ein existierendes Modell ist, welches das ursprüngliche Quellmodell sein kann.

Regelanwendungsstrategie

Eine Regel muss an einer spezifischen Stelle innerhalb ihres Quellbereichs angewandt werden. Sobald es mehr als einen Treffer für eine Regel innerhalb eines vorgegebenen Quellbereichs geben kann, benötigen wir eine Anwendungsstrategie. Die Strategie kann deterministisch, nicht-deterministisch oder sogar interaktiv sein.

Regel-Scheduling

Scheduling-Mechanismen bestimmen die Reihenfolge, in der einzelne Regeln angewandt werden. Der Scheduling-Mechanismus kann in vier Hauptpunkten variieren:

1. Form: Der Scheduling-Aspekt kann implizit oder explizit ausgedrückt werden. Scheduling implizit bedeutet, dass der Nutzer keine explizite Kontrolle über den Scheduling-Algorithmus durch das Werkzeug besitzt. Der einzige Weg, über den der Nutzer den systemdefinierten Scheduling-Algorithmus beeinflussen kann, ist durch den Entwurf mit den Transformations-Entwurfsmustern und die Logik der Regeln, um bestimmte Ausführungsreihenfolgen zu garantieren. Explizites Scheduling hat spezielle Konstrukte, um explizit die Ausführungsreihenfolge zu kontrollieren.
2. Regelauswahl: Regeln können durch eine explizite Bedingung ausgewählt werden. Einige Ansätze erlauben eine nicht-deterministische Auswahl. Alternativ könnte ein konfliktauflösender Mechanismus, basierend auf Prioritäten, angeboten werden (obwohl keiner der untersuchten Ansätze eine Konfliktauflösung implementiert hat). Interaktive Regelauswahl ist ebenfalls möglich.
3. Regeliteration: Regeliterationsmechanismen schließen Rekursion, Schleifen und Fixpunkt-Iteration ein (z.B. wiederholte Anwendung bis keine Änderungen mehr entdeckt werden)
4. Phasen: Der Transformationsprozess kann in verschiedenen Phasen organisiert sein, wobei jede Phase einen spezifischen Zweck hat und nur bestimmte Regeln in einer vorgegebenen Phase aufgerufen werden können. Zum Beispiel haben manche struktur-orientierte Ansätze eine separate Phase, um die Inhaltshierarchie des Zielmodells zu erstellen und eine separate Phase, um die Attribute und Referenzen im Ziel festzulegen.

Regelorganisation

Regelorganisation ist damit verbunden, vielfache Transformationsregeln zusammen zu stellen und zu strukturieren. Wir betrachten drei Gebiete der Variationen in diesem Zusammenhang:

- Modularitätsmechanismen
- Wiederverwendungsmechanismen
- Organisationsstruktur

Nachvollziehbarkeitsverbindungen

Transformationen können Verbindungen zwischen ihren Quell- und Zielelementen aufzeichnen. Diese Verbindungen können nützlich sein bei der Durchführung einer Wirkungs-/Einfluss-Analyse (z.B. bei der Analyse, wie eine Änderung in einem Modell sich in anderen Modellen auswirkt), bei der Synchronisation zwischen Modellen, modellbasierter Fehlersuche und bei der Festlegung des Ziels der Transformation.

Direktionalität

Transformationen können unidirektional oder bidirektional sein. Unidirektionale Transformationen können nur in einer Richtung ausgeführt werden, bei denen ein Zielmodell basierend auf einem Quellmodell generiert (oder upgedated) wird. Bidirektionale Transformationen können in beiden Richtungen ausgeführt werden, welches nützlich ist im Zusammenhang mit der Synchronisation zweier Modelle. Bidirektionale Transformationen können erreicht werden durch den Einsatz bidirektionaler Regeln oder durch die Definition zweier getrennter komplementärer unidirektionaler Regeln, eine für jede Richtung. Die meisten der untersuchten Ansätze bieten keine Bidirektionalität.

5.2 Modelltransformationsansätze

5.2.1 Übersicht

Auf der obersten Ebene unterscheiden wir Modell-zu-Code- und Modell-zu-Modell-Transformationsansätze. Im Allgemeinen kann man die Transformation von Modellen zu Code als Spezialfall der Modell-zu-Modell-Transformationen betrachten, falls ein Metamodell für die Zielprogrammiersprache angeboten wird. Code wird aus praktischen Erwägungen, nämlich der Wiederverwendung existierender Compiler-Technologie, häufig einfach als Text generiert. Aus diesem Grund unterscheiden wir zwischen Modell-zu-Code-Transformation und Modell-zu-Modell-Transformation. Verschiedene Werkzeuge bieten sowohl Modell-zu-Modell- als auch Modell-zu-Code-Transformationen an. In der Modell-zu-Code-Kategorie unterscheiden wir zwischen traversierenden und Schablonenbasierten Ansätzen. In der Modell-zu-Modell-Kategorie unterscheiden wir zwischen Direktmanipulationsansätzen, relationalen Ansätzen, Graphentransformationsansätzen, strukturgetriebenen Ansätzen und hybriden Ansätzen.

5.2.2 Modell-zu-Code Ansätze

Die Mehrheit der aktuell verfügbaren MDA-Werkzeuge unterstützt Schablonen-basierte Modell-zu-Code Generierung. Eine Schablone besteht üblicherweise aus dem Zieltext, der Metacode enthält, um auf Informationen der Quelle zuzugreifen und um Codeauswahl zu treffen sowie iterative Erweiterung. Schablonen-basierte Ansätze bieten häufig ein nutzerdefiniertes Scheduling in der Form an, dass eine Schablone innerhalb einer anderen aufgerufen wird.

Verglichen mit den traversierenden Transformationen, ähnelt die Struktur einer Schablone mehr dem Code, der generiert werden soll. Schablonen eignen sich selbst für eine iterative Entwicklung, weil sie leicht von Beispielen abgeleitet werden können. Während die in diesem Abschnitt diskutierten Schablonen-basierten Ansätze auf Text operieren, sind die Entwurfsmuster, die sie enthalten, untypisiert und können syntaktisch oder semantisch unkorrekte Code-Fragmente repräsentieren. Auf der anderen Seite sind textuelle Schablonen unabhängig von der Zielsprache und vereinfachen die Generierung textueller Artefakte, einschließlich der Dokumentation. Eine verwandte Technologie ist „Frame Processing“, das Schablonen erweitert mit hochentwickelter Adaption und Strukturierungsmechanismen.

5.2.3 Modell-zu-Modell Ansätze

Modell-zu-Modell Transformationen übersetzen zwischen Quell- und Zielmodellen, welche Instanzen desselben oder unterschiedlicher Metamodelle sein können. Alle diese Ansätze unterstützen syntaktisches Schreiben von Variablen und Entwurfsmustern. Die meisten existierenden MDA-Werkzeuge bieten nur Modell-zu-Code Transformationen an, die sie nutzen, um PSMs aus PIMs zu generieren (in diesem Fall den Implementationscode). Warum werden Modell-zu-Modell Transformationen gebraucht? Dann, wenn große Abstraktionslücken zwischen PIMs und PSMs überbrückt werden müssen. Es ist dann einfacher Zwischenmodelle zu generieren als direkt auf das Ziel-PSM zuzugehen. Dies macht die Transformationen modularer und wartbarer. Zwischenmodelle können auch für Optimierung und Tuning notwendig sein, oder wenigstens für Debugging-Zwecke. Zusätzlich zu PIM-zu-PSM Transformationen sind Modell-zu-Modell Transformationen nützlich, um unterschiedliche Sichten auf ein Systemmodell zu bekommen und sie zu synchronisieren.

5.2.4 Andere Ansätze

Wenigstens zwei weitere Ansätze sollten für die Vollständigkeit unter der Rubrik „Andere Modell-zu-Modell Ansätze“ erwähnt werden: das Transformationsframework definiert in

der OMG Common Warehouse Metamodel (CWM) Spezifikation und Transformationen implementiert mittels Extensible Stylesheet Language Transformations (XSLT).

Das CWM Transformationsframework bietet einen Mechanismus Quell- und Zielelemente miteinander zu verbinden. Jedoch muss die Ableitung der Zielelemente in irgendeiner konkreten Sprache implementiert werden, die nicht durch CWM vorgeschrieben ist. Tatsächlich bietet CWM ein generelles Modell, aber keine wirklichen Mechanismen, um Modelltransformationen zu implementieren.

Seit Modelle in XML mittels XML Metadata Interchange (XMI) serialisiert werden können, ist die Implementation von Modelltransformationen mit XSLT sehr attraktiv. XSLT ist eine Standard-Technologie für die Transformation von XML. XSLT bildet unidirektional ab, um eine bidirektionale Abbildung zu erstellen, müssen beide Richtungen der Abbildung getrennt beschrieben werden. Eine Abbildung in XSLT besteht aus Abbildungsregeln. Jede Regel gibt an, wie ein Element der Eingabedatei für die Ausgabedatei transformiert werden soll.

Modelltransformation ist ein relativ junges Gebiet. Obwohl es in Bezug steht bzw. auf den etablierteren Gebieten wie Programmtransformation und Metaprogrammierung aufbaut, ergibt der Einsatz graphischer Modellierungssprachen und die Anwendung objektorientierter Metamodellierung auf Sprachdefinition einen neuen Kontext. Während es zufrieden stellende Lösungen für die Transformation von Modellen zu Text (wie Schablonenbasierte Ansätze) gibt, ist dies nicht der Fall für die Transformation von Modellen zu Modellen. Viele neue Ansätze für Modell-zu-Modell-Transformationen sind in den letzten Jahren vorgeschlagen worden, aber es gibt nur wenig Erfahrungen für die Bewertung ihrer Effektivität in praktischen Anwendungen.

[CH03] bietet einige Bemerkungen an bezüglich der praktischen Anwendbarkeit der verschiedenen Arten von Modelltransformation. Diese Kommentare basieren auf Intuition und den Anwendungsbeispielen, die zusammen mit den Ansätzen veröffentlicht sind. Wegen des Fehlens kontrollierter Experimente und extensiver praktischer Erfahrung, sind diese Kommentare nicht vollständig validiert.

1. Direktmanipulation ist offensichtlich der Ansatz auf unterster Ebene. Er bietet dem Nutzer wenig oder keine Unterstützung oder Anleitung bei der Implementierung der Transformationen. Die gesamte grundlegende Arbeit muss der Nutzer selbst erledigen. Auf längere Sicht, wird dieser Ansatz nicht praktikabel sein.
2. Die strukturgetriebene Kategorie gruppiert pragmatische Ansätze, die im Kontext bestimmter Arten von Applikationen entwickelt wurden, z.B. das Generieren von EJB-Implementationen und Datenbankschemata aus UML-Modellen. Diese Anwendungen erfordern eine starke Unterstützung für die Modelltransformation mit einer 1-zu-1 und 1-zu-n (und manchmal n-zu-1) Korrespondenz zwischen Quell- und Zielelementen. Auch gibt es in diesem Anwendungskontext typischerweise keine Notwendigkeit für Iteration beim Scheduling. Das Scheduling kann systemdefiniert sein. Es ist unklar wie gut diese Ansätze andere Arten von Anwendungen unterstützen können.

3. Graphentransformationsbasierte Ansätze sind inspiriert von theoretischer Arbeit auf dem Gebiet der Graphentransformationen. Diese Ansätze sind mächtig und deklarativ, aber auch komplex. Die Erfahrung mit praktischen Anwendungen dieser Ansätze ist immer noch begrenzt. Es bleibt abzuwarten, wie gut die Komplexität dieser Ansätze in der Praxis aufgenommen wird.
4. Relationale Ansätze scheinen eine gute Balance zwischen Flexibilität und deklarativer Ausdrucksweise zu erreichen. Sie bieten flexibles Scheduling und eine gute Kontrolle des Nicht-Determinismus. Drei der fünf aktuellen Query Views Transformations-Einreichungen passen in diese Kategorie.
5. Hybride Ansätze erlauben dem Nutzer verschiedene Konzepte und Paradigmen abhängig von der Anwendung zu mischen. Praktische Ansätze werden sehr wahrscheinlich einen hybriden Charakter besitzen.

Die Evaluation der verschiedenen Entwurfsoptionen für einen Modelltransformationseinsatz werden mehr Experimente und praktische Erfahrung erfordern. Die Einführung einer umfassenden Sammlung von Benchmark-Problemen würde ein bewertbarer nächster Schritt in diese Richtung sein.

5.3 Qualität einer Abbildung

Aus einem Vergleich der Informationskapazität von Quell- und Zielschema ergibt sich ein Kriterium für die Informationserhaltung [MIR93]. Bei Abbildungen sollten idealerweise überhaupt keine Informationen verloren werden, sondern komplett erhalten bleiben. Dies ist gleichzeitig eine Maßzahl für die Kongruenz zweier Standards. Je kompletter und informationserhaltender eine Abbildung gelingt, um so besser ist die Qualität der Abbildung. Darauf soll in dieser Arbeit nicht näher eingegangen werden.

5.4 Auswahl Modelltransformationsansatz

Wie in Abschnitt 5.1 bereits angedeutet, fehlen den meisten Ansätzen die Möglichkeiten für bidirektionale Transformationen. Um mit einer Vermittlung zwischen verschiedenen domänenspezifischen Standards institutionsübergreifende Interoperabilität zu erreichen, ist die Bidirektionalität der Abbildungen zwingend. Diejenigen Ansätze, die diese Möglichkeit bieten, sind in der Entwicklungsphase. Sicherlich wäre interessant die in dieser Arbeit vorgestellten Metamodelle der domänenspezifischen Standards mit den neuesten Modelltransformationsansätzen zu testen. Dies würde aber den Rahmen dieser Arbeit übersteigen.

Im Rahmen dieser Arbeit haben wir uns für die Evaluation für Modelltransformationen mittels XSLT entschieden. Die Vorteile von XML liegen in der weiten Verbreitung sowie des „Lingua franca“-Charakters, den XML für Datenaustausch zwischen Anwendungen erworben hat und weiter erwirbt. Lingua franca ist eine Sprache oder ein Weg der Kommunikation, der genutzt wird, wenn nicht die gleiche „Muttersprache“ gesprochen wird.

Teil II

Metamodell-basierte Architektur zur institutionsübergreifenden Interoperabilität

6 Analyse domänenspezifischer Standards am Beispiel Medizin

6.1 Bedeutung von Standards am Beispiel Medizin

Standardisierungen sind ein probates Mittel mit semantischen Heterogenitäten umgehen zu können. Dadurch ist im Idealfall ein gemeinsames Verständnis aller Beteiligten möglich. Dies ist Grundvoraussetzung für eine erfolgreiche Kommunikation. Eine Bereitstellung der notwendigen Informationen am richtigen Ort zur richtigen Zeit ist der wesentliche Aspekt im Bereich der Telemedizin [LBM⁺05].

6.1.1 Erwartungen an die Telemedizin

Da elektronische Kommunikation schnelle und kostengünstige Datenübertragung möglich macht, können die Kosten im Gesundheitswesen durch die Telemedizin sinken. Zur Zeit entstehen nach Untersuchungen von Kostenträgern durch Doppeluntersuchungen und Behandlungsverzögerungen bis zu 20% der Kosten [MED98]. Diese mangelnde und langsame Verfügbarkeit von Informationen lässt sich zum Teil durch technische Lösungen der Telemedizin reduzieren. Die niedergelassenen Ärzte sind als zahlenmäßig weitaus größte Gruppe im Gesundheitswesen der begrenzende Faktor für die Einführung modernster telematischer Lösungen. Zwar sind 70% der Arztpraxen mit Computern ausgerüstet, jedoch sind diese bis zu zehn Jahre alt, und eine generelle Modernisierung in den nächsten Jahren ist nicht zu erwarten [MED98]. Die grundlegende Voraussetzung für eine verlässliche Mitwirkung und Kommunikation zwischen allen beteiligten Akteuren ist die gegenseitige Vereinbarung über Syntax, Semantik und Pragmatik der genutzten Repräsentation medizinischer Daten. Die Syntax bezieht sich auf die Struktur eines Ausdrucks, die Semantik auf die Bedeutung eines Ausdrucks und die Pragmatik definiert die Bedeutung in einem Zusammenhang [IRS01, OS97].

In den Vereinigten Staaten hat die Einrichtung eines nationalen Gesundheitsdatennetzes höchste Priorität [Nat05]. Namhafte IT-Unternehmen wollen gemeinsam offene, nicht-proprietäre Standards für den Datenaustausch im Gesundheitswesen unterstützen. Ziel ist es, durch das Gesundheitsdatennetz die Behandlungsmöglichkeiten zu verbessern und

Kosten einzusparen, indem Daten nicht mehr auf Papier, sondern in digitaler Form vorliegen. Durch einheitliche Formate sollen Informationen ohne Hürden zwischen den verschiedenen Institutionen des Gesundheitssystems austauschbar sein.

In Deutschland wurde im Oktober 2003 mit Verabschiedung des GKV-Modernisierungsgesetzes (GKV=Gesetzliche Krankenversicherung) die Einführung einer elektronischen Gesundheitskarte zum 1.1.2006 beschlossen [bIT05a]. Das Projekt „bIT4health — bessere IT für bessere Gesundheit“, das an ein Konsortium unter Leitung der IBM Deutschland GmbH vergeben wurde, soll die Einführung der Karte unterstützen [bIT05b]. Mit der Telematikrahmenarchitektur, der Solution Outline (basierend auf der Rahmenarchitektur, werden hier Lösungsansätze skizziert, die in der nachfolgenden Lösungsarchitektur konkretisiert werden sollen), und einer ersten Spezifikation der elektronischen Gesundheitskarte sind wesentliche technische Grundlagen für den Aufbau einer Telematikinfrastruktur für das deutsche Gesundheitswesen vom Konsortium erarbeitet worden.

6.1.2 Die Rolle von Standards

Standards spielen eine wichtige Rolle, wenn es darum geht, beim Datenaustausch ein gemeinsames Verständnis der zwischen heterogenen Anwendungssystemen übertragenen Daten sicherzustellen [Has00b]. Ferner können durch eine Top-Down-Integration skalierbare, flexiblere und wiederverwendbare Softwarearchitekturen für heterogene Informationssysteme erreicht werden [Has02]. Dabei sollte das gemeinsame Datenmodell auf einem domänenspezifischen Standard basieren, um eine Grundlage für die semantische Interoperabilität zu besitzen und eine gute Strukturierung der integrierten, gemeinsamen Modelle zu erreichen.

Im Bereich der Medizin existieren nun diverse Standards für die Kommunikation und Dokumentation, die im folgenden vorgestellt werden. Die Strukturen der Standards werden einheitlich in der Unified Modeling Language (UML) spezifiziert [PH03].

6.2 Standards für die medizinische Kommunikation

Für die Architektur von Kommunikationssystemen hat sich als Entwurfsrichtlinie das Open System Interconnection-Basisreferenzmodell der internationalen Standardisierungsorganisation (ISO) durchgesetzt [WK94]. Danach wird ein Kommunikationssystem in sieben Schichten unterteilt. Die im folgenden vorgestellten Kommunikationsstandards sind der Anwendungsschicht zuzuordnen. Die bekannten Protokolle TCP/IP dagegen gehören zur Transport- bzw. zur Vermittlungsschicht.

6.2.1 Behandlungsdatenträger (BDT)

Bei der Kommunikation zwischen niedergelassenen Ärzten hat sich der BDT-Standard durchgesetzt [Hae99]. Das Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland (ZI) ist die für den BDT-Standard verantwortliche Institution [ZI 02]. Seit 1998 liegen Datensatzbeschreibungen (sog. Profile) für Diabetes mellitus und die Tumordokumentation mittels BDT vor. Inzwischen sind BDT-Profile für Hypertonie und Asthma hinzu gekommen. Damit ist das ZI dem Wunsch nach einheitlichen Formaten für die Erfassung von Daten bei einigen chronischen Erkrankungen nachgekommen. Arztpraxis-Softwarehersteller können ihre Produkte beim ZI bezüglich der Umsetzung der BDT-Schnittstellen testen lassen. Der BDT wurde für einen Datenaustausch des gesamten in der Arztpraxis gesammelten Datenumfangs entwickelt. Er beschreibt und kategorisiert die Informationen durch Felddescriptions und durch eindeutige Zuordnung zu definierten Sätzen. Abbildung 6.1 zeigt als Diagramm in der Unified Modeling Language (UML) die Zusammenhänge der wesentlichen Komponenten des BDT-Standards.

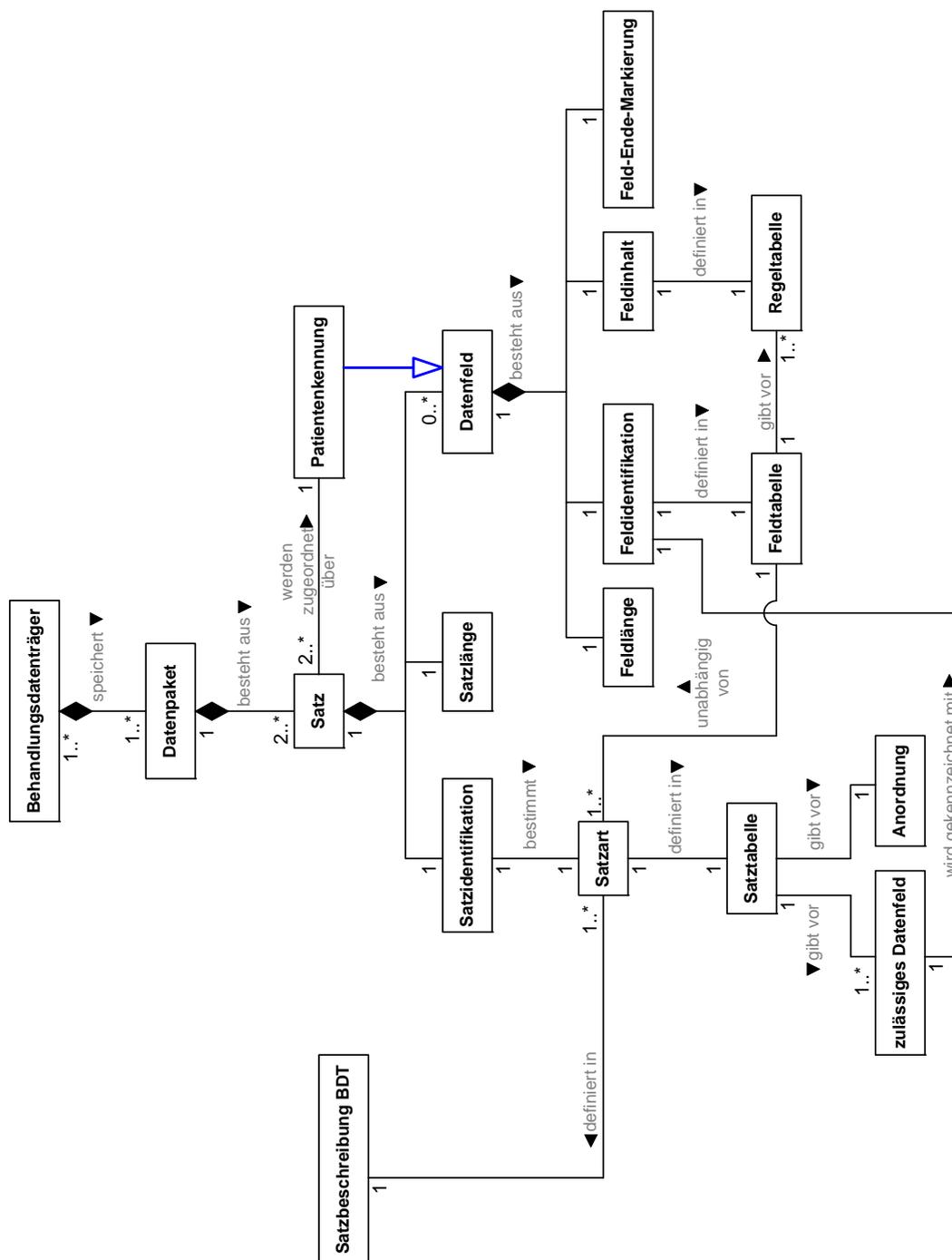


Abbildung 6.1: Ausschnitt BDT-Struktur als UML-Diagramm

Behandlungsdatenträger speichern Datenpakete, die aus Sätzen bestehen. Diese wiederum sind zusammengesetzt aus Satzidentifikation, Satzlänge und eventuell einigen Datenfeldern. Mehrere Sätze können über die eindeutige Patientenkenung zugeordnet werden. Die Satzidentifikation bestimmt genau eine Satzart und die verschiedenen Satzarten sind in der Satzbeschreibung des BDT definiert. Für jede Satzart gibt eine Satztable die zulässigen Datenfelder und eine eindeutige Anordnung vor. Datenfelder bestehen aus Feldlänge, Feldidentifikation, Feldinhalt und der Feld-Ende-Markierung. Für jede Feldidentifikation ist dabei genau eine Feldtable, für jeden Feldinhalt genau eine Regeltabelle definiert. Die Feldtable gibt die Regeltabelle vor und ist unabhängig von der Satzart definiert [Hae99].

Neueste Entwicklungen lassen allerdings ein mittelfristiges Verschwinden des BDT-Standards vermuten, da sich die Kassenärztliche Bundesvereinigung inzwischen aktiv an Umsetzungen bzw. Anpassungen anderer Standards beteiligt und bereits erste Spezifikationen veröffentlicht hat [SCI03, KBV03b].

6.2.2 Health Level Seven (HL7)

HL7 ist ein speziell für das Gesundheitswesen entwickelter Standard, der eine elektronische Kommunikation zwischen allen Gruppen im Gesundheitswesen möglich machen kann [HBD99, HL702a]. Derzeit wird HL7 in erster Linie in Krankenhäusern eingesetzt. In den USA ist HL7 alleiniger offizieller Standard, daneben existieren nur noch einige individuelle Lösungen, wobei dies in der Struktur des amerikanischen Gesundheitswesens begründet ist. In der Bundesrepublik Deutschland sind HL7-Nachrichten Standard für die Kommunikation in Krankenhausinformationssystemen [Dud03].

Die wesentlichen Merkmale und Ziele von HL7 in Version 3 sind nach [Blo02]:

1. Bereitstellung von Formaten und Protokollen zum möglichst weltweiten Austausch bestimmter Datensätze zwischen Computersystemen im Gesundheitswesen, insbesondere im Krankenhausbereich
2. Standardisierung der Nachrichteninhalte
3. HL7 Version 3 strebt Interoperabilität sowohl auf Daten- als auch auf funktioneller und semantischer Ebene an. Dafür stehen insbesondere die Use-Case-Modelle, das Reference Information Model (RIM) und das Vokabular zur Verfügung.
4. Mit Version 3 wird mit dem Message Development Framework eine objektorientierte Entwicklungsmethodologie angeboten, die von der Entwicklung, über die Anpassung bis zur Anwendung der Nachrichten den Entwickler unterstützen soll.

HL7 stellt außerdem nur wenige Vorbedingungen an die Systemarchitektur. So können die Systeme verteilt oder zentral organisiert sein. Der Datenaustausch kann mit unterschiedlichsten Betriebssystemen und Programmiersprachen realisiert sein. HL7 muss auch nicht komplett umgesetzt werden. Meist fängt man mit der Übertragung von Patientenstammdaten an. HL7 standardisiert die Nachrichten-Strukturen (Abstract Message Definition), die Darstellung der Nachrichten für die Übertragung (Encoding Rules) und die nachrichtenauslösenden Anwendungsereignisse (Trigger Events). Ab Version 3 von HL7 sind für alle Nachrichten *Use Cases* definiert. Sie beschreiben, unter welcher realen Bedingung welche HL7-Nachricht bzw. welcher Trigger Event eingesetzt werden sollte. Auch die beteiligten Akteure, Vor- und Nachbedingungen werden genannt sowie der normale Ablauf dargelegt. Dies stellt eine spürbare Erleichterung beim Einsatz von HL7 dar. Um Definitionen in HL7 konsistent zu halten, gibt es ein generisches Reference Information Model (RIM) sowie ein Vokabular. Das Vokabular dient der Identifizierung, Organisation und Wartung kodierter Begriffe aus Wortschätzen, die in HL7-Nachrichten eingesetzt werden [HL702b]. Diese kodierten Begriffe sind definiert als Mitglieder einer kodierten Domäne. Mit Domäne ist hier die Menge aller geeigneten Werte für ein Datenelement gemeint. Es gibt mindestens fünf Typen kodierter Domänen, die der HL7-Standard referenziert:

1. Domänen, deren Inhalt komplett von HL7 Komitees bestimmt wird, also Trigger Events, Nachrichtentypen, Segment-Identifizierer, Formattypen, etc.
2. Domänen, wo HL7 Komitees Listen begonnen haben, wie die Tabellen für Geschlecht, Familienstand, Nationalität, etc.
3. Domänen für spezielle Zwecke, bei denen keine Datenelemente definiert sind und vorausgesetzt wird, dass die kodierten Werte aus Begriffssystemen wie SNOMED, ICD, UMLS, etc. genutzt werden. Dieser Typ beinhaltet z.B. Tabellen für Universal Service IDs, Observation IDs, etc.
4. Domänen, die den Wert für eine spezielle Observation ID ausfüllen, wobei der verschlüsselte Wert aus einem oder mehreren entsprechenden Kodiersystemen kommen muss.
5. Domänen, die der HL7-Standard referenziert, deren Struktur und Inhalte lokal definiert werden müssen. Diese Kategorie umfasst Benutzeridentifikationen, Räume, Betten, etc.

Dieses gemeinsame Verständnis der übertragenen Daten ermöglicht einen Austausch klinischer Informationen.

Im Gesundheitswesen ist XML von der HL7-Gruppe eingeführt worden, welche weiterhin die treibende Kraft für XML darstellt (näheres zu XML siehe Abschnitt 4.1.3 auf Seite 28). Eine Tumordokumentation wurde in Gießen von einer Arbeitsgruppe um Herrn Prof. Dudeck in XML spezifiziert. Diese ist jedoch bisher aus finanziellen Gründen nicht implementiert worden. Im Projekt SCIPHOX dient XML der syntaktischen Integration von BDT und HL7 [SCI02].

Abbildung 6.2 illustriert in UML die wesentlichen Zusammenhänge der Komponenten des HL7-Standards. HL7-Nachrichten bestehen aus Segmenten, die von Segmentseparatoren getrennt werden, jeweils in genau einer Segmenttabelle verwaltet werden, jeweils eine eindeutige Bezeichnung und eine eindeutige Anordnung besitzen. Segmente können Kontroll- oder Daten-Segmente sein. Kontroll-Segmente sind z.B. das Message Header Segment und die Event Description, ein Daten-Segment ist z.B. die Patient Identification. Ein Segment besteht wiederum aus Feldern mit codierten oder uncodierten Datentypen. Der Inhalt der codierten Datentypen wird in diversen Code-Tabellen verwaltet. Diese Code-Tabellen können im HL7-Standard festgeschriebene Tabellen sein, benutzerdefinierte Tabellen, Codierungstabellen oder andere Standards sein. Die Felder werden von Feldseparatoren getrennt und besitzen ebenfalls eine eindeutige Anordnung. Ein Feld wird genau von einer Felddefinition festgelegt. Die Felddefinitionen werden in Vokabularen verwaltet [HBD99].

6.2.3 Digital Imaging and Communications in Medicine (DICOM)

DICOM ist ein etablierter Industrie-Standard, der es erlaubt, digitale Bildinformationen verschiedener Geräte, aus verschiedenen Archiven herstellerübergreifend auszutauschen. Damit soll ein schneller und kostengünstiger Datenaustausch zwischen Krankenhäusern untereinander und mit Arztpraxen erreicht werden. 1993 wurde DICOM das erste Mal verabschiedet und seitdem regelmäßig erweitert [DIC02], in Europa ist DICOM seit 1995 als formaler Standard unter der Bezeichnung MEDICOM akzeptiert. Es handelt sich bei DICOM jedoch nicht um einen internationalen Standard im Sinne von ISO oder CEN. DICOM benutzt eine ganz eigene Sicherheitsinfrastruktur mit einem *secure channel* auf der Basis von SSL.

Abbildung 6.3 veranschaulicht in der UML die relevanten Beziehungen der wesentlichen Komponenten des DICOM-Standards. DICOM definiert Informationsobjekte (IODs), welche aus genau einer Informationsentität als normalisierte IODs oder aus mehreren Informationsentitäten als zusammengesetzte IODs bestehen können. Eine Informationsentität besteht aus Modulen. Modultabellen beschreiben Module, wobei jede Modultabelle alle Module eines Informationsobjektes beschreibt. Eine Modultabelle enthält die zugehörigen Informationsentitäten, die Namen der Module, Referenzen auf den Standard sowie die Typen der Module. Ein Modul beinhaltet mehrere Attribute, wobei ein Attribut auch in mehreren Modulen enthalten sein kann. Jedes Modul besitzt genau eine Moduldefinition, die die enthaltenen Attribute beschreibt. Eine Moduldefinition besteht aus Attributnamen, Markierungen (Tags), Attributtypen, Attributbeschreibungen, Datentypen und Mehrwertigkeiten. Ein Data Dictionary speichert zentral alle Datentypen und Mehrwertigkeiten [DIC02, Eic01].

Ergänzung 23 zu DICOM, das DICOM Structured Reporting (DICOM SR) ist eine Spezifikation, die eine semantische Repräsentation von Bildern und Messkurven unterstützt

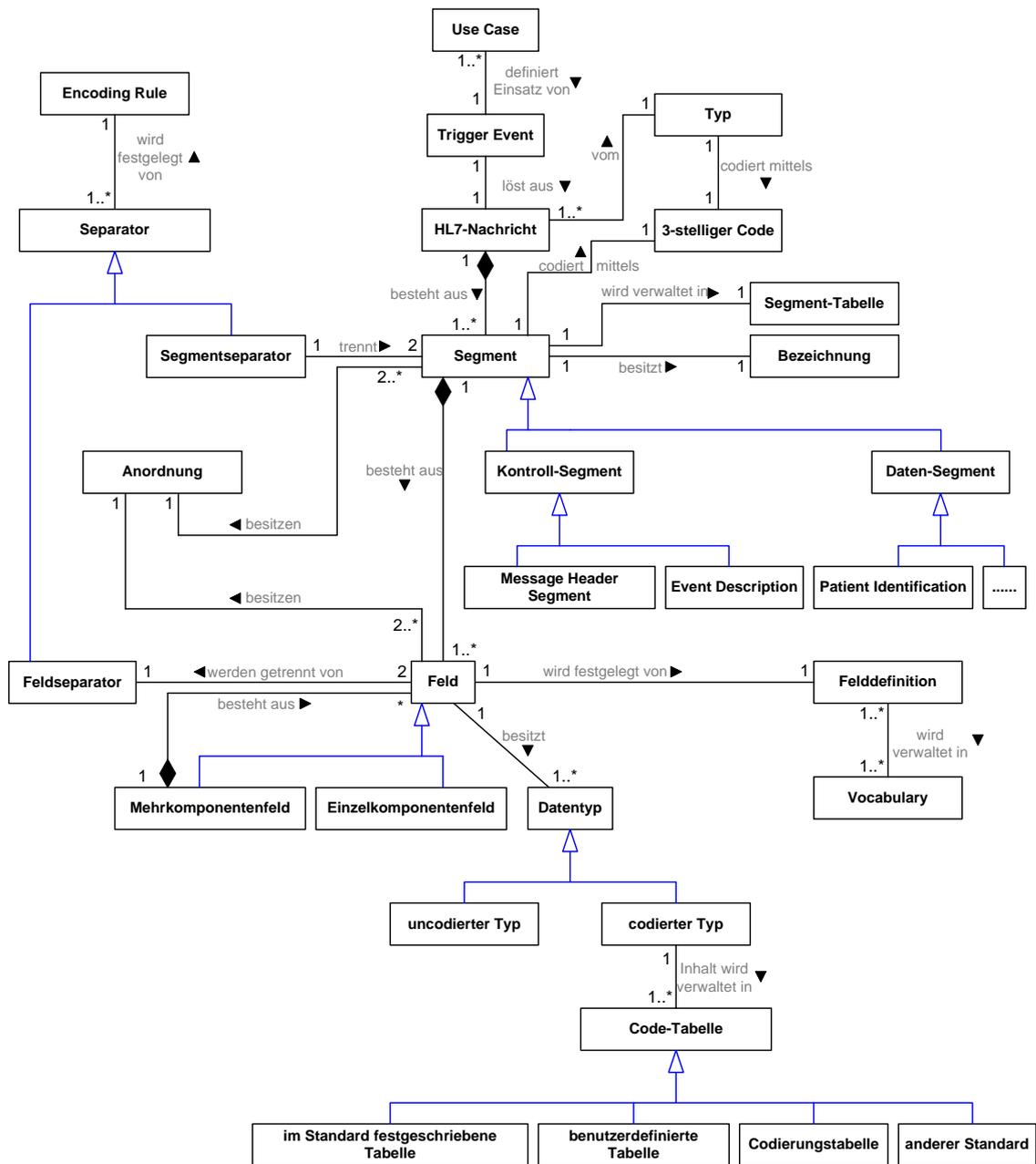


Abbildung 6.2: Ausschnitt HL7 Struktur als UML-Diagramm

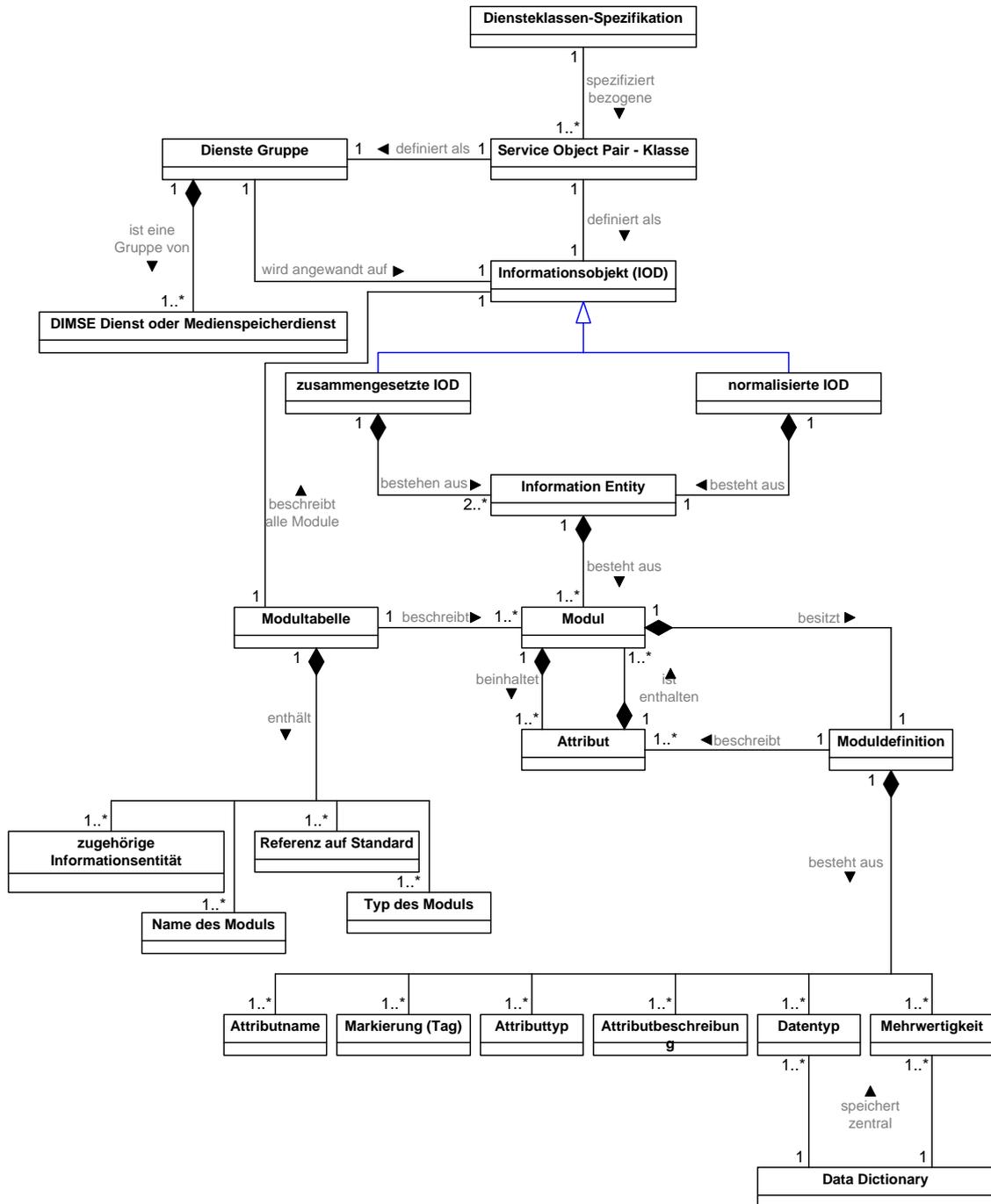


Abbildung 6.3: Ausschnitt DICOM-Struktur als UML-Diagramm

und so die Verteilung von Bildern mit der dazugehörigen Patienteninformation ermöglicht [TRHL02]. DICOM SR unterstützt die Repräsentation von textuellen und kodierten Daten, verknüpft mit Bildern und Messkurven.

DICOM SR strukturiert Daten hierarchisch in eine Baumstruktur mit Knoten wie in Abbildung 6.4 dargestellt. Jeder Knoten besitzt einen Namen für ein Konzept (Name für

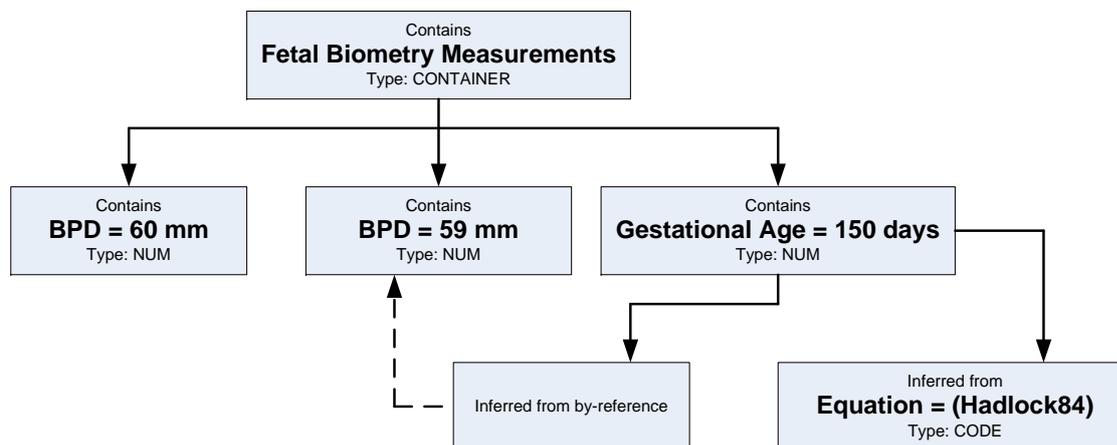


Abbildung 6.4: Struktur von DICOM SR

einen Begriff ist Bezeichnung) mit einem Wert oder Werten (z.B. biparietaler Durchmesser ist 59 mm) [SLM02]. Diese Bezeichnungen werden mit medizinischen Terminologiestandards kodiert, wie beispielsweise SNOMED, LOINC oder DICOM. Der Konzept-Code ist ein sprachunabhängiger, einheitlicher Identifizierer eines Konzeptes. Die Werte der Knoten können verschiedenen Typs sein: Text, Zahlen, kodierter Begriff oder Referenzen zu anderen DICOM-Objekten wie Bildern. Zusätzlich zu den direkten Kindknoten-Beziehungen, erlaubt DICOM SR auch Beziehungen durch Referenzen. Dies ist ein ausdrucks mächtiges Kodierungsschema, dadurch gibt es mehrere Wege, dieselbe Tatsache darzustellen. Interoperabilität ist zu schwierig, wenn es nicht gewisse Einschränkungen bezüglich des Vokabulars und der Struktur gibt. DICOM standardisiert Vorlagen, um eine einheitliche Ausdrucksweise zu forcieren. Dieser Teil des Standards ist die DICOM Content Mapping Resource (DCMR).

Um nun auf einem höheren Niveau einen Überblick über den DICOM Standard und seine komplexen Strukturen zu bekommen, haben [TRHL02] ein objektorientiertes Modell mittels UML erstellt und über XML austauschbare Repräsentationen dieses Modells abgeleitet.

In Abschnitt 7.3 werden diese Kommunikationsstandards in einem gemeinsamen Metamodell integriert.

6.3 Standards für die medizinische Dokumentation

Allgemeines Ziel der medizinischen Dokumentation ist es, berechtigten Personen gezielt Informationen oder Wissen zum richtigen Zeitpunkt, in der richtigen Form und am richtigen Ort zur Verfügung zu stellen. Bei der medizinischen Dokumentation können ganz unterschiedliche Informationen gemeint sein, z.B. Befunde und Therapien einzelner Patienten, Diagnose- und Therapiewissen über spezielle Erkrankungen, ein Verzeichnis medizinischer Veröffentlichungen sowie Resultate aus Arzneimittelstudien [LGHKG99].

Für die Dokumentation sind standardisierte Strukturen für Dokumente notwendig sowie standardisierte Begriffssysteme für die Felder in den Dokumenten.

6.3.1 Standards für Dokumente

In Zukunft soll der Electronic Healthcare Record (EHCR) Mittelpunkt eines medizinischen Informationssystems sein, sozusagen als Kommunikationsdrehscheibe. Dabei ist das Ziel, die Informationen auch korrekt interpretieren zu können. Wichtige Projekte auf diesem Gebiet sind derzeit EHCR (Electronic Healthcare Record) von CEN (Comité Européen de Normalisation), GEHR (Good Electronic Healthcare Record) aus Australien und die CDA (Clinical Document Architecture) der HL7-Gruppe [CEN02, GEH02, HL702a].

Das Projekt openEHR versucht, mit dem Ziel eines globalen EHCR-Modells als generisches Modell, eine Harmonisierung der Ansätze HL7, CEN 13606 und GEHR [HL701b, ope02].

Definitionen

Zunächst einige wichtige Definitionen zum EHCR [HL701b]:

Electronic Healthcare Record: Repository von Informationen über den Gesundheitszustand eines Patienten, die in einem computerlesbaren Format vorliegen.

Electronic Healthcare Record System: Menge von Komponenten, die den Mechanismus zur Erzeugung, Verwendung, Speicherung und zum Wiederauffinden einer Patientenakte realisieren.

Electronic Healthcare Record Architecture: Modell der generischen Eigenschaften, die für jede elektronische Krankenakte erforderlich sind, damit diese Akte eine kommunizierbare, vollständige, nützliche und effektive ethisch-rechtlich verbindliche Krankenakte sein kann, die ihre Integrität unabhängig von Plattformen und Systemen sowie nationalen Besonderheiten über die Zeit bewahrt.

EPA-Stufenmodell

Ein Stufenmodell für die elektronische Patientenakte (EPA) wurde vom Medical Record Institut erstellt [MRI02]. Das Stufenmodell enthält Level 1 bis 5. Level 1 umschreibt die automatisierte medizinische Akte, bei der nur 50% der Information EDV-generiert ist, Level 2 ist die computerisierte Krankenakte, die digital vorliegt, weil zusätzlich Papierdokumente eingescannt werden. Der Aufbau der Krankenakte von Level 1 und 2 entspricht daher der einer papiernen Krankenakte. Level 3 ist die elektronische Krankenhausakte, wobei die Administration nicht mehr Kern-, sondern Subsystem im Haus ist. Level 4 meint die häuserübergreifende Krankenakte und Level 5 beschreibt quasi den gläsernen Bürger [HL701b]. Die Terminologie im Bereich der Electronic Healthcare Records ist zahlreich an Bezeichnungen, und zudem kann sie verwirrend sein [Pro01].

Electronic Healthcare Record (CEN 13606)

Der CEN 13606-Standard ist ein Beispiel für einen so genannten *Ein-Modell-Ansatz* [CEN02]. Der Ein-Modell-Ansatz setzt nur die aktuellen Anforderungen an das System um. Das Hinzufügen neuer Konzepte bedeutet kostenaufwendige Neuentwicklung, Testen und Implementierung. Dadurch, dass alle Anforderungen in einem Schritt umgesetzt werden müssen, ist das Modell äußerst komplex. Besser in dieser Hinsicht als Ein-Modell-Ansätze sind generische Ansätze, wie man sie bei GEHR und CDA findet [HL701b]. Abbildung 6.5 illustriert in UML den Aufbau des EHCR von CEN.

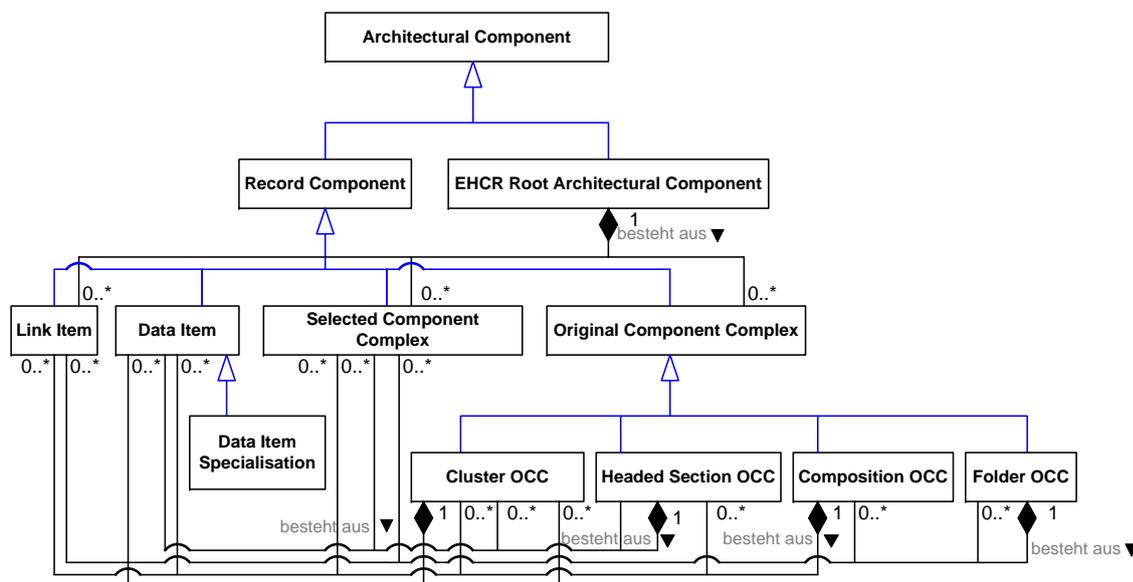


Abbildung 6.5: Ausschnitt EHCR-Struktur des CEN als UML-Diagramm

Die Komponenten der Architektur können zum einen Record Components, zum anderen EHCR Root Architectural Components sein. Record Components sind die Strukturierungselemente Link Item, Data Item, Selected Component Complex und Original Component Complex. Die Original Component Complexes (OCCs) können Cluster OCCs, Headed Section OCCs, Composition OCCs und Folder OCCs sein. Vielfältige Verschachtelungen dieser Elemente ineinander sind möglich. Der Folder OCC ist eine grobe Unterabteilung im EHCR für einen Patienten, für ein spezielles Krankheitsbild oder für Einträge über eine längere Zeitspanne. Ein Composition OCC repräsentiert einen einzelnen identifizierbaren Beitrag in einem EHCR, z.B. den orthopädischen Datensatz. Ein Headed Section OCC enthält Informationen auf einer Ebene unterhalb eines Beitrages, z.B. die aktuelle Medikation. Ein Cluster OCC dient der Aggregation von einzelnen Data Items oder anderen Cluster OCCs, um den Sinnzusammenhang nicht zu verlieren, z.B. Blutdruckwerte [CEN02].

Im Rahmen einer neuen Arbeitsgruppe „EHRcom“ wird der CEN 13606 Standard in Bezug auf die Kommunikation seit 2001 überarbeitet. Der zuvor reinen Strukturbeschreibung sollen Aspekte zur Terminologie hinzugefügt werden. Ziel ist eine Harmonisierung mit anderen Ansätzen durch Anpassen auf deren Paradigmen. So soll der Referenzmodell- und Archetypen-Ansatz vom australischen EHCR (GEHR, jetzt openEHR) ebenso verfolgt werden wie der Einsatz der Clinical Document Architecture mit den HL7 Datentypen [CEN03].

Clinical Document Architecture (CDA)

Die ANSI / HL7 CDA R1.0-2000 ist die erste zertifizierte XML-basierte Spezifikation für das Gesundheitswesen. Vor 1996 hieß die Entwicklung noch Patient Record Architecture (PRA). Die CDA ist eine XML-basierte Dokumentenspezifikation zum Austausch klinischer Informationen. Ein CDA-Dokument ist dabei ein definiertes und komplettes Informationsobjekt, das Texte, Bilder, Klänge und andere multimediale Objekte enthalten kann. Die CDA ist eine Architektur bzw. ein Bauplan für Dokumente mit HL7-Semantik und XML-Syntax. Näheres siehe [HL702a, HL701a]. Abbildung 6.6 zeigt in UML die wichtigsten Elemente der CDA sowie deren Beziehungen.

Die CDA ist eine hierarchische Dokumentenspezifikation, wobei die Dokumente in XML kodiert sind und ihre Semantik aus dem HL7-RIM entnehmen. Die CDA kennt drei Ebenen, die zunehmende Spezialisierung ausdrücken. Das CDA-Dokument der ersten Ebene wird definiert durch eine header DTD, eine level one body DTD und der HL7 V3 data type DTD. Der Header enthält Informationen über das Dokument, das Ereignis, Erbringer und Empfänger der Maßnahmen. Der Body enthält hingegen die eigentliche klinische Information [HL702a, HL701a]. Es gibt derzeit Bemühungen, die HL7 CDA und DICOM SR zu harmonisieren bzw. aufeinander abzubilden [SLM02]. Die CDA trennt domänenunabhängige Anteile (Ebene 1) von domänenabhängigen Anteilen (Ebene 2 und 3) und bietet dadurch nach [LBM⁺05] einen vielversprechenden Lösungsansatz zum Umgang

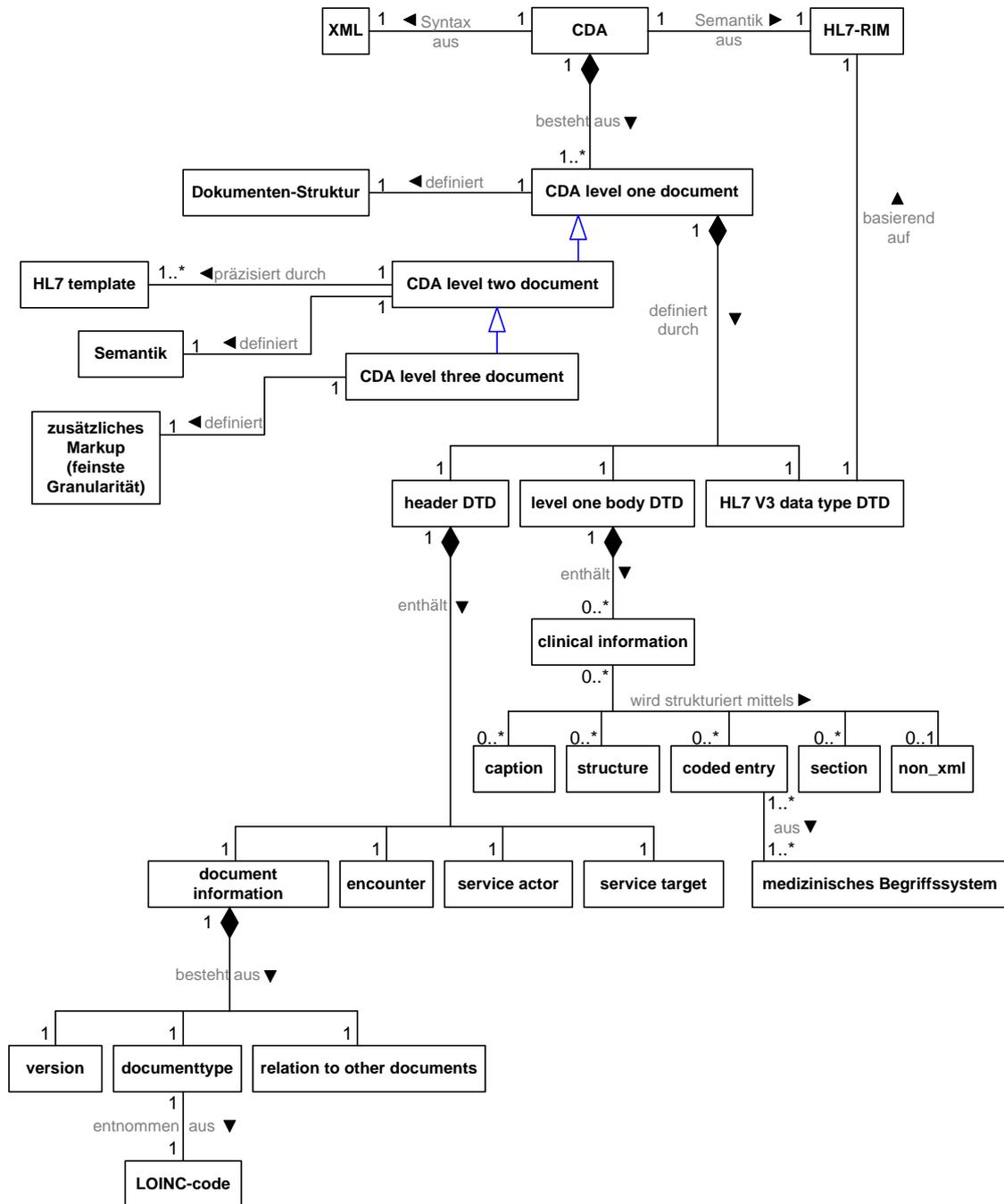


Abbildung 6.6: Ausschnitt CDA-Struktur als UML-Diagramm

mit semantischer Heterogenität. An einer medizinisch-inhaltlichen Standardisierung häufig vorkommender Dokumente arbeitet das Projekt SCIPHOX (siehe Abschnitt 11.1.1 auf Seite 145).

Good Electronic Healthcare Record (GEHR)

Die GEHR - Architektur besteht aus zwei Teilen, dem konkreten Modell *GEHR Object Model* (GOM) und dem Metamodell *Archetype* [GEH02]. Das GOM ist ein generisches Modell für die Wissensrepräsentation, wobei die Komponenten (Objekte) des GOM wie LEGO-Steine verwendet werden. Die Archetypen dagegen beschreiben und standardisieren klinische Inhalte. Archetypen werden wie Entwürfe für LEGO-Strukturen benutzt. Die Pläne beschränken die Nutzung der LEGO-Steine auf sinnvolle Strukturen. Das konkrete Modell kann unabhängig von den Archetypen standardisiert werden.

openEHR Foundation

Das Projekt openEHR, welches aus dem GEHR-Projekt entstanden ist, versucht mit dem Ziel eines globalen EHCR-Modells als generisches Modell, eine Harmonisierung der Ansätze HL7, CEN13606 und GEHR zu erreichen. Dieser Ansatz soll komponentenorientiert, auf der Basis von Modellen sowie plattformunabhängig sein [HL701b, ope02].

Wichtig ist die Trennung von generischem Referenzinformationsmodell und dem anwendungsbereichsspezifischen Domäneninformationsmodell. Dieser Zwei-Modell-Ansatz führt zu einer Trennung von Information und Wissen [Bea02]. Aus dem allgemeinen und schlanken Referenzinformationsmodell erhält man durch Einschränkungen (Constraints) das Archetypen-Modell bzw. Domänenmodell. Die Archetypen stehen für medizinische Konzepte, beispielsweise Blutdruck oder Familiengeschichte. Abbildung 6.7 verdeutlicht graphisch diese Zusammenhänge.

Globaler EHR-Standard

Mit der Revision des CEN ENV 13606, der zeitgleichen Realisierung des nationalen australischen GEHR-Projektes und der internationalen openEHR Foundation wird gegenwärtig ein globaler EHR-Standard beschrieben, der die Strukturmodelle des ursprünglichen CEN ENV 13606 mit dem Referenzmodell und dem Vokabular von HL7 sowie den Domänenmodellen von GEHR und HL7 verknüpft [BP03]. Als Wissensrepräsentationen sind z.B. die Medical Logic Moduls (MLM) der Arden Syntax, die Leitlinien-Repräsentationsmodelle im GuideLine Interchange Format (GLIF) oder die GEHR Archetypen verfügbar, die alle spezifische Constraint-Modelle darstellen. Neben medizinischem Wissen und Konzepten sind auch administrative, prozessbezogene oder rechtliche Constraint-Modelle erforderlich. Die Interoperabilität der resultierenden Lösung

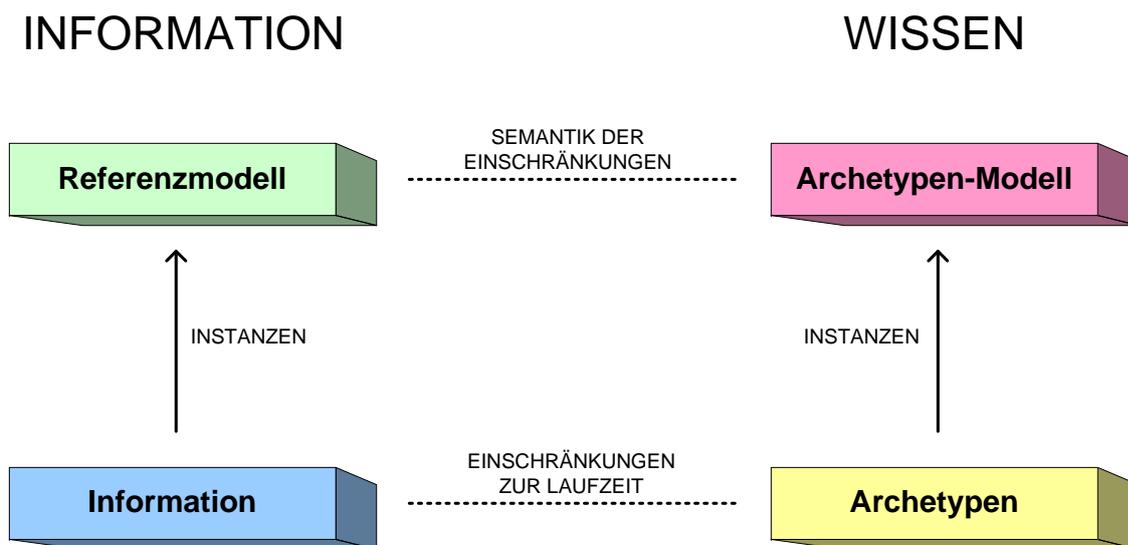


Abbildung 6.7: Trennung von Information und Wissen im openEHR Modell

wird über ein Referenzinformationsmodell sowie ein abgestimmtes Vokabular gesichert. Die Harmonisierung dieser verschiedenen Ansätze wird über Metamodelle und Metasprachen realisiert, die eine Übertragung aus einem Repräsentationsvokabular in ein anderes ermöglichen. Modellierung- und Spezifikationsprachen sind z.B. die UML oder der XML Standard Set. Eingebunden in alle wichtigen globalen Aktivitäten zum EHR, wird in Kooperation mit internationalen Partnern ein generisches System zur Spezifikation und Implementierung einer modellgetriebenen Architektur von EHR-Systemen analog zur MDA (Näheres dazu siehe Abschnitt 4.1.1 auf Seite 26) entwickelt. Basierend auf dem ISO Referenzmodell - Open Distributed Processing, erlaubt die Verknüpfung der verschiedenen Sichten (Enterprise View, Information View, Computational View, Engineering View und Technology View) beschreibenden Constraint-Modelle die plattformunabhängige sowie die plattform-spezifische Modellierung des Systems und seiner Komponenten entsprechend der MDA. Zur Vokabulartransformation wird XML Metadata Interchange (XMI) verwendet.

6.3.2 Standards für Begriffssysteme

Der Inhalt der Felder im EHCR wird durch Begriffssysteme standardisiert. Die uneinheitliche Verwendung von Begriffen sorgt in vielen Bereichen oft für Unklarheiten. So werden häufig die gleichen Bezeichnungen für verschiedene Begriffe verwendet (Homonyme) bzw. verschiedene Bezeichnungen für einen Begriff (Synonyme). Dadurch entstehen Probleme mit der verwendeten Terminologie z.B. für Diagnosen und Maßnahmen [PH02].

Innerhalb eines medizinischen Dokumentes werden medizinische Begriffssysteme eingesetzt, um Maßnahmen und Diagnosen etc. zu verschlüsseln und dadurch die Terminologie

diesbezüglich zu standardisieren.

Zu medizinischen Begriffssystemen zählen Klassifikationen, Nomenklaturen, Thesauern sowie enumerative und kompositionelle Begriffssysteme. Ein Beispiel für

1. eine Klassifikation ist die International Statistical Classification of Diseases and Related Health Problems (ICD) [ICD02a],
2. eine Nomenklatur die Systematized Nomenclature of Human and Veterinary Medicine (SNOMED) [SNO02],
3. einen Thesaurus der Read Clinical Code (RCC) [Rea02],
4. ein enumeratives Begriffssystem UMLS (Unified Medical Language System) [UML02] und
5. ein kompositionelles medizinisches Begriffssystem GALEN (Generalized Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine) [Ber95].

Das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI) wurde vom Bundesministerium für Gesundheit und Soziale Sicherung beauftragt, zum 1. Januar 2003 das Deutsche Zentrum für Medizinische Klassifikation (DZMK) einzurichten. Das Zentrum soll zentral koordiniert medizinische Klassifikationen entwickeln und pflegen, Standards veröffentlichen, Referenzstelle sein und für Fragen von Anwendern zur Verfügung stehen. Der Bedarf für diese zentrale Stelle ergibt sich aus den steigenden Anforderungen an medizinische Klassifikationen, z.B. da sie im Rahmen der Diagnosis Related Groups (DRG) für eine Vergütung der Leistungserbringer herangezogen werden [DZM03].

Über die Zusammenhänge der diversen medizinischen Begriffssysteme gibt [PH02] Auskunft und stellt ein Begriffssystem für Begriffssysteme vor, welches mittels der UML modelliert wurde und im Abschnitt 7.3 mit den Kommunikationsstandards auf Meta-Modellebene kombiniert wird. Die Abbildungen 6.8 und 6.9, modelliert in der UML, geben einen Überblick über die wichtigsten Zusammenhänge in der Welt der medizinischen Begriffssysteme und bilden ein systematisches Begriffssystem für Begriffssysteme.

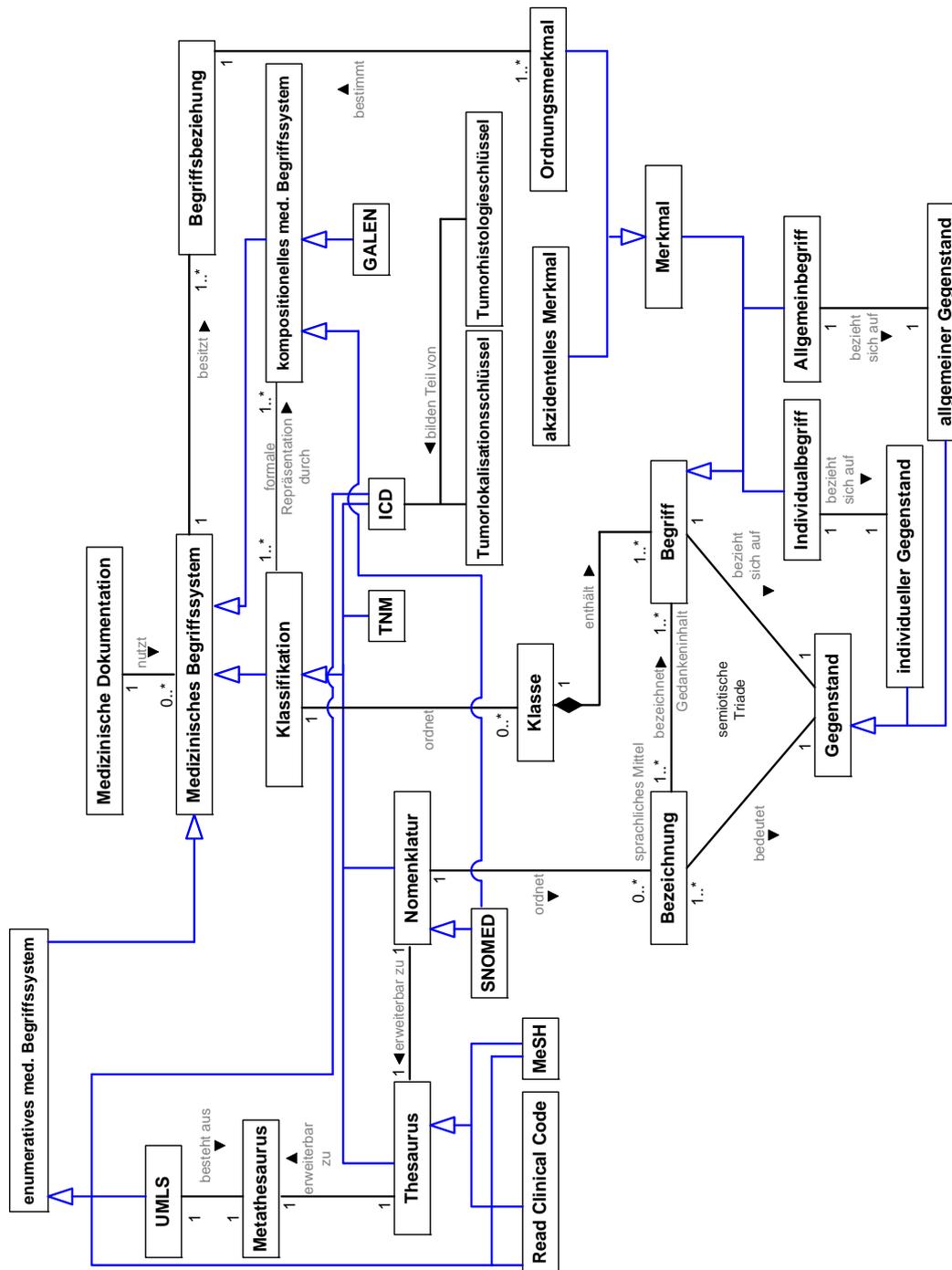


Abbildung 6.8: Ausschnitt medizinische Begriffssysteme als UML-Diagramm

Dargestellt ist in Abbildung 6.8 die *semiotische Triade* [Ber95], die die geltenden Beziehungen der Terminologielehre darstellt: Eine Bezeichnung bezeichnet einen Begriff, eine Bezeichnung bedeutet einen Gegenstand und ein Begriff bezieht sich auf einen Gegenstand. Klassen enthalten nun Begriffe und werden von Klassifikationen geordnet. Klassifikationen sind medizinische Begriffssysteme, welche von der medizinischen Dokumentation genutzt werden. Die TNM-Klassifikation [TNM02] und der ICD [ICD02a] sind Beispiele für eine Klassifikation. Nomenklaturen sind Zusammenstellungen von Bezeichnungen, die dahinterstehende Systematik basiert in der Regel auf einer Begriffsordnung, ein Beispiel ist SNOMED [SNO02]. Eine Nomenklatur kann nun erweitert werden zu einem Thesaurus. Hierzu gehören der Read Clinical Code [Rea02] und die Medical Subject Headings [MeS02]. Wird wiederum der Thesaurus um semantische und linguistische Informationen erweitert, so entsteht ein Metathesaurus, wie ihn der Kern von UMLS [UML02] darstellt. UMLS gehört zu den enumerativen Begriffssystemen und versucht, die wichtigsten medizinischen Begriffssysteme in einem einzigen zu integrieren und dabei möglichst viele Begriffsbeziehungen explizit darzustellen. Dazu abgegrenzt werden kompositionelle Begriffssysteme, wie z.B. GALEN, welches versucht, methodische Grundlagen für die formale Repräsentation medizinischer Begriffssysteme zu entwickeln. GALEN erlaubt die formale Konstruktion von allen anderen fachrelevanten Begriffen durch die Komposition einer Anzahl primitiver Begriffe [Ber95, PH02].

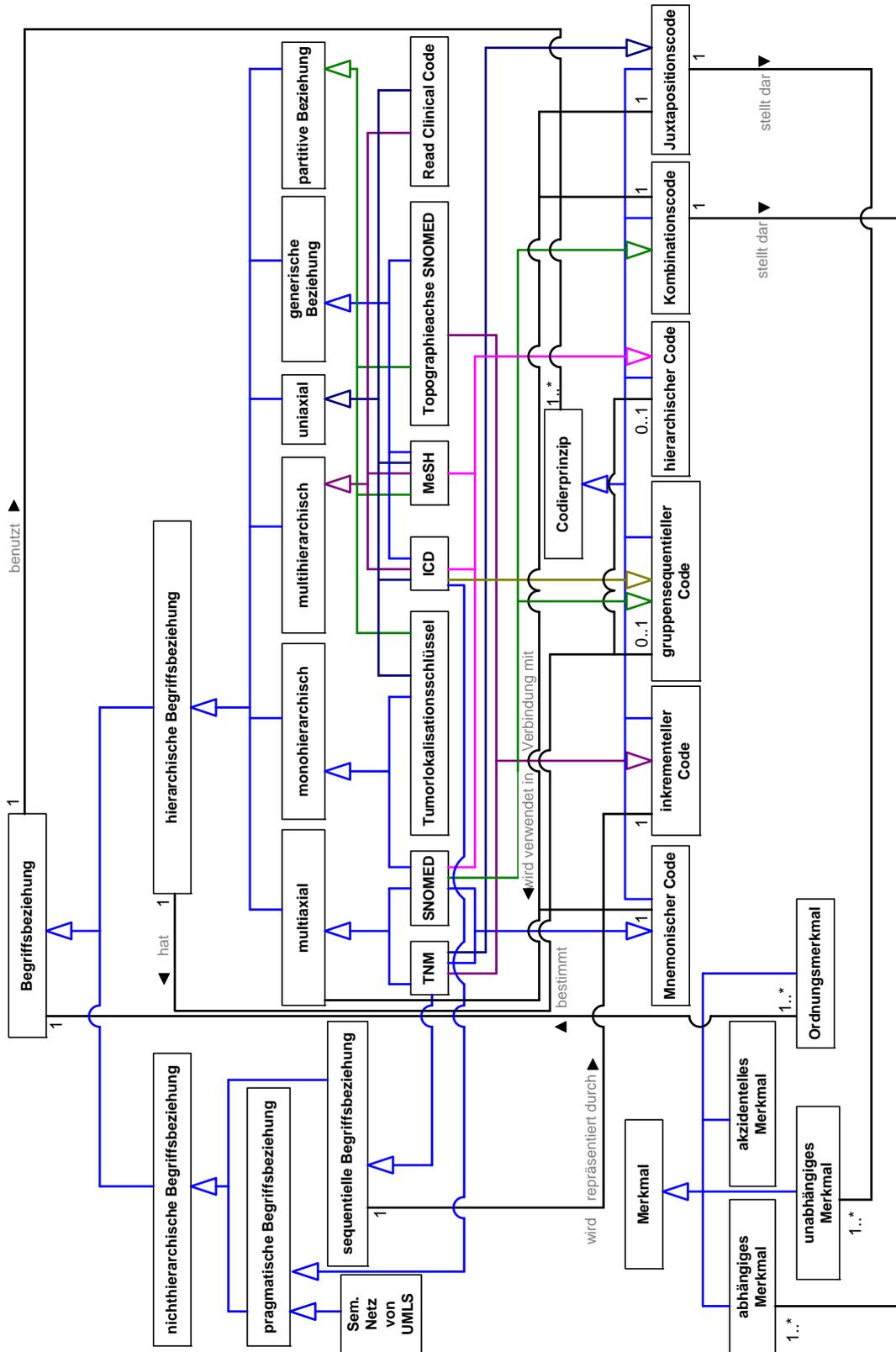


Abbildung 6.9: Ausschnitt Begriffsbeziehungen und Codierprinzipien als UML-Diagramm [Ber95, PH02]

Abbildung 6.9 verdeutlicht die Zusammenhänge zwischen den diversen Begriffsbeziehungen und den Codierprinzipien in den Begriffssystemen. Man unterscheidet in medizinischen Begriffssystemen nichthierarchische und hierarchische Begriffsbeziehungen. Zu ersteren zählen pragmatische und sequentielle Begriffsbeziehungen, zu letzteren generische (Spezialisierung/Generalisierung) und partitive Beziehungen (Teile-Ganzes-Beziehungen). Ferner unterscheidet man bei hierarchischen multi- und uniaxiale Systeme sowie multi- und monohierarchische Begriffssysteme. Begriffssysteme benutzen diverse Codierprinzipien, wobei häufig mehrere Codierprinzipien kombiniert werden. Der ICD ist beispielsweise uniaxial, multihierarchisch mit generischen und pragmatischen Beziehungen. Seine Codierprinzipien sind der gruppensequentielle und der hierarchische Code [Ber95, PH02].

6.3.3 Ontologien

Ontologien sollen den Datenaustausch erleichtern und sind formale Modelle einer Anwendungsdomäne. Eine Ontologie wird ausgedrückt durch eine logische Theorie, die sich zusammensetzt aus einem Vokabular und einer Menge von logischen Aussagen zu der jeweils interessierenden Anwendungsdomäne. Im semiotischen Dreieck schränkt eine Ontologie die Beziehung Bezeichnungen-bedeuten-Gegenstände ein. Idealerweise bleibt genau eine Beziehung übrig [MSS01]. Ontologien sind somit Mechanismen, um eine Repräsentation eines gemeinsamen Verständnisses, z.B. einer Bestellung, zu etablieren. Eine gemeinsame Ontologie für medizinisches Wissen würde das Problem semantischer Heterogenität lösen, leider liegt eine derartige Lösung in der Zukunft [LK01].

In der Medizin bedient man sich seit vielen Jahren zahlreicher medizinischer Begriffssysteme, die jeden Begriff durch die Position innerhalb des Begriffssystems bestimmen. Medizinische Begriffssysteme sorgen so für eine gemeinsame Semantik von Begriffen. Auf Ontologien möchte ich daher im Rahmen dieser Arbeit nicht weiter eingehen.

6.4 Auswahl von Kombinationen der Standards

Prinzipiell ist zunächst jede Kombination von Kommunikations- und Dokumentationsstandards möglich. Bei der Beantwortung der Frage, welche Kombinationen ausgewählter Standards medizinischer Kommunikation und Dokumentation denn nun für einen Anwendungsbereich sinnvoll sind, soll ein Auswahlschema helfen.

6.4.1 Zusammenhang der Standards

Die im Bereich Medizin betrachteten Standards für Kommunikation und Dokumentation sowie die medizinischen Begriffssysteme sind ineinander verschachtelt anzusehen.

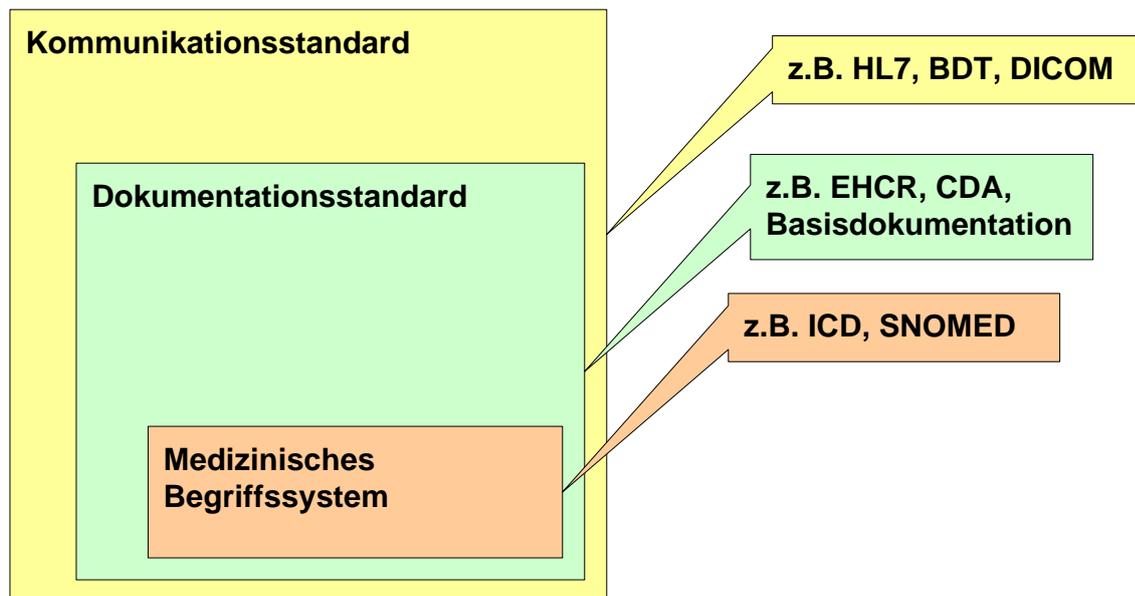


Abbildung 6.10: Zusammenhang der Standards im Gesundheitswesen

Abbildung 6.10 verdeutlicht, dass ein Kommunikationsstandard quasi als Container fungiert. Darin enthalten ist dann beispielsweise ein Dokumentationsstandard als Struktur des Dokuments. In bestimmten Feldern des Dokuments kommen dann medizinische Begriffssysteme zum Zug, z.B. für die Kodierung von Maßnahmen und Diagnosen.

6.4.2 Auswahlschema

Die erste Spalte des in Tabelle 6.1 vorgestellten Auswahlschemas erfasst, ob Abbildungen bereits existieren, die zweite Spalte bewertet bereits im Anwendungsbereich eingesetzte Kombinationen, die dritte und vierte Spalte gehen auf die Vollständigkeit der Abbildung ein.

Mögliche Kombinationen	Existenz der Abbildung	Einsatz der Kombination im Anwendungsbereich	Vollständigkeit der Abbildung	Bewertung der Vollständigkeit
Kommunikationsstandard / Dokumentationsstandard	✓	✓	✓	+++

Tabelle 6.1: Auswahlschema für geeignete Kombinationen von Standards

6.4.3 Vorgehensweise bei der Auswahl

Die Vorgehensweise in folgenden Schritten:

1. Mögliche Kombinationen von zuvor mit dem Dimensionenschema ausgewählten Kommunikations- und Dokumentationsstandards auflisten.
2. Wenn bereits Abbildungen zwischen Kommunikations- und Dokumentationsstandards vorhanden sind, die genutzt werden können, so können diese Abbildungen direkt genutzt werden. Falls enge Verwandtschaften zwischen dem Kommunikations- und Dokumentationsstandard bestehen, wie beispielsweise bei HL7 und CDA, die auf dem gleichen Referenzinformationsmodell basieren, ist eine Abbildung kein Problem.
3. Kombinationen, bei denen keine Abbildung existiert, erhalten keinen Haken, die übrigen erhalten einen bei der Bewertung in der Spalte „Existenz der Abbildung“.
4. Kombinationen, die bereits im Anwendungsbereich eingesetzt werden, bekommen den Haken, die übrigen keinen bei der Bewertung in der Spalte „Einsatz der Kombination“.
5. In der Spalte „Vollständigkeit der Abbildung“ bekommen die Kombinationen, bei denen eine Abbildung vorgenommen wurde, einen Haken, die übrigen keinen.
6. Eine Bewertung der Vollständigkeit der Abbildung wird dann in der Spalte „Bewertung Vollständigkeit“ vorgenommen. Die Bewertungsskala reicht dabei von einem bis zu drei Pluszeichen. Je vollständiger eine Abbildung ist, desto mehr Pluszeichen werden vergeben.
7. Die Anzahl der vergebenen Haken sowie Pluszeichen verdeutlicht, welche Kombinationen am besten geeignet sind.
8. Abschließend muss beurteilt werden, wie gut zwischen zwei Kombinationen vermittelt werden kann, um abzuwägen, ob eine direkte Abbildung besser oder schlechter geeignet ist. Vorzuziehen wäre eine Vermittlung durch den Metafacilitator bzw. Metakomponentenmediator, um größtmögliche Flexibilität zu besitzen. Sonst wären die Vorteile der Architektur nicht vollständig genutzt.
9. Existieren Abbildungen nicht, so ist eine genauere Analyse der Standards nötig, um eine Antwort darauf zu geben, ob eine Abbildung sinnvoll ist oder nicht.

In Abschnitt 7.3 werden detailliertere Angaben zu den benötigten Metadaten und ihren Zusammenhängen gemacht.

6.5 Andere Standardisierungsbemühungen

Neben Standards für die Kommunikation und die Dokumentation im Gesundheitswesen einschließlich Kodierung von Operationen, Diagnosen etc. existieren noch weitere Bemühungen Standards zu entwickeln, die andere Aspekte im Gesundheitswesen tangieren. Im

Folgenden sollen ein paar wichtige Standardisierungsbemühungen kurz vorgestellt werden, um ein etwas umfangreicheres Bild entstehen zu lassen. Diese Standards werden bei dieser Arbeit nicht weiter herangezogen.

6.5.1 Standards

Clinical Data Interchange Standards Consortium (CDISC)

Das CDISC-Konsortium [CDI03a, CDI03b, CDI03c], gegründet 1997, hat ein Operational Data Model (ODM) und ein Submissions Data Model (SDM) ins Leben gerufen, um unabhängige Modelle für den Austausch und die Archivierung klinischer Daten mit Studien-Metadaten basierend auf der XML-Technologie zu definieren. Es geht dabei um Daten klinischer Studien, die zwischen den diversen Beteiligten ausgetauscht und schließlich archiviert werden müssen. Das ODM soll das Zusammenführen der Daten aus verschiedenen Datenquellen (z.B. Laboratorien, Clinical Research Organizations (CRO)), in eine operationale Datenbank erleichtern und das SDM soll bei der Einreichung einer geeigneten Zusammenstellung der Daten bei der entsprechenden staatlichen Stelle (z.B. für Deutschland das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM)) helfen.

Das ODM besteht prinzipiell aus den folgenden vier Abschnitten: Study, AdminData, ReferenceData und ClinicalData. In Study sind enthalten der Studienname, die Studienbeschreibung usw. Das ODM erlaubt die Repräsentation von mehr als einer Studie in einer einzigen Datei. Mit AdminData sind Systembenutzer, beteiligte Kliniken etc. gemeint. Hinter ReferenceData stecken Daten, die relevant sind für die Interpretation, wie z.B. Labor-Normwerte bzw. -bereiche. Aktuelle Dateninhalte finden sich dann im Abschnitt ClinicalData [CDI03c].

Die Bemühungen von CDISC beziehen sich nur auf den Bereich der klinischen Studien und sind daher nicht relevant für diese Arbeit.

Healthcare Information Architecture (HISA)

Hinter HISA verbirgt sich ein Standard der CEN für Krankenhausinformationssystemarchitekturen [Pri97]. Das dabei vorgeschlagene Architekturgerüst für den medizinisch-technischen und klinisch-administrativen Krankenhausbereich lässt sich in drei Schichten strukturieren:

- Bit-Schicht (enthält technologische Infrastruktur für die Netzwerkunterstützung)
- Middleware-Schicht (unterstützt die Kooperation verschiedener Anwendungen)

- Anwendungsschicht (unterstützt die spezifischen Anforderungen der Krankenhausabteilungen)

Distributed Healthcare Environment (DHE) ist eine Implementierung des HISA Standards ENV 12967-1 der CEN [KG01]. DHE stellt Funktionen und Schnittstellen für die KIS-Anwendungen zur Verfügung. Das Ziel ist eine offene Infrastruktur anzubieten, um heterogene Applikationen, die über eine Menge von gemeinsamen medizinspezifischen Komponenten interagieren, zu verteilen und somit zu integrieren. Die vermittelnde DHE-Schicht generischer Dienste ist dabei als Middleware zwischen einer spezifischeren Applikationsschicht und einer technologischen Plattformschicht angesiedelt. DHE ist ein Beispiel, bei dem versucht wird, auf einem vordefinierten zentralen Datenbankschema krankenhausspezifische generische Dienste zu spezifizieren [LK03]. Die Idee, die funktionale Überlappung durch standardisierte domänenspezifische Dienste zu reduzieren, liegt auch der CORBAmed-Initiative der OMG zu Grunde (z.B. Patient Identification Service usw.) [Cor97, Cor03]. Gemeinsam haben HISA und der hier vorgestellte Ansatz die Top-Down-Vorgehensweise bei der Integration. Während sich HISA im Bereich Middleware Integration bewegt, befasst sich das Architekturkonzept dieser Arbeit mit der Ebene der Anwendungsarchitekturen, die darüber liegt (siehe Abbildung 1.1 auf Seite 3).

Integrating the Healthcare Enterprise (IHE)

IHE ist eine Initiative, die 1998 von RSNA (Radiological Society of North America) und HIMSS (Healthcare Information and Management Systems Society) etabliert wurde [IHE03]. IHE hat sich zum Ziel gesetzt, den technischen Informationsfluss zwischen verschiedenen Krankenhausinformationssystemen (KIS), Radiologie Informationssystemen (RIS), Picture Archiving Systems (PACS), etc. zu verbessern. IHE basiert auf bestehenden Standards, zur Zeit HL7 und DICOM, und wendet diese in einem prozessorientierten Ansatz an. Das generische Modell von IHE besteht aus Akteuren und den Transaktionen zwischen ihnen. Das Technical Framework von IHE definiert die relevanten Akteure und beschreibt die erforderlichen Transaktionen, die jeder Akteur unterstützen muss.

Für konkrete Anwendungen wurden verschiedene Integrationsprofile geschaffen. Damit betreibt IHE die Integration auf Ebene der Geschäftsarchitekturen, wobei sich diese Arbeit auf die Ebene der Anwendungsarchitekturen bezieht (siehe Abbildung 1.1 auf Seite 3).

Clinical Context Object Workgroup (CCOW)

CCOW von HL7 ist ein domänenspezifischer ANSI-Standard für das Gesundheitswesen, der auf der Datenlogik-, der Geschäftslogik- und der Präsentationslogik-Ebene eine Rolle spielt [CS04][ANS02]. Ursprünglich ging das Technische Komitee von CCOW

aus einem Industriekonsortium hervor, bevor es sich dann HL7 anschloss [CCO01]. Die Kernarchitektur von CCOW setzt sich aus drei Hauptkomponenten zusammen: den Applikationen, einem Kontext-Manager, der die Applikationen koordiniert und synchronisiert sowie Abbildungsagenten, die die zahlreichen synonymen Bezeichnungen repräsentieren und abbilden können. Die CCOW-Architektur definiert für jede dieser Komponenten Rollen und Zuständigkeiten. Außerdem beschreibt sie präzise die Schnittstellen, die eine Kommunikation zwischen den Komponenten zulässt. Die Architektur schreibt dabei keineswegs eine konkrete Implementation vor. Der CCOW-Standard wird die Möglichkeiten von Kontext-Management stetig weiter entwickeln.

6.5.2 Übertragungstechniken

Im Folgenden werden die zur Zeit populärsten Übertragungstechniken im Bereich der niedergelassenen Ärzte kurz vorgestellt. An der Integration dieser Techniken mit dem Bereich der Krankenhausinformationssysteme wird gearbeitet [SCI02]. Da es in dieser Arbeit im Schwerpunkt um die semantische Interoperabilität geht, spielen diese Techniken eine untergeordnete Rolle.

VDAP Communication Standard (VCS)

Der VDAP e.V. ist der Verband der Deutschen Arztpraxis-Softwarehersteller, in dem die fünf größten Anbieter organisiert sind [VDA02]. Dies sind die Firmen ALBIS, ComputoMED, Data Vital, Medistar und TurboMed. 1999 einigten sich die damaligen Mitglieder des Verbandes auf eine gemeinsame Kommunikationsschnittstelle VCS. Diese soll zum Standard für jede elektronische Kommunikation zwischen Ärzten werden, daher ist auch für andere Hersteller die Spezifikation von VCS offen. In einer ersten Entwicklungsstufe ermöglicht die VCS-Schnittstelle den systemübergreifenden Austausch von elektronischen Arztbriefen zwischen zwei Ärzten. Dabei ist die VCS-Schnittstelle direkt in die Applikation integriert. In einer ersten Entwicklungsstufe, die im Juli 2000 abgeschlossen wurde, wurde die gezielte Kommunikation Arzt zu Arzt, Arzt zu Klinik usw. umgesetzt. In einer zweiten Stufe wurde eine zuweisende (ungezielte) Kommunikation mit der Bereitstellung der Patientendaten in einem virtuellen Postfach möglich gemacht, damit der Patient jederzeit freie Arztwahl hat und der entsprechend autorisierte Arzt die Patientendaten abrufen kann. Diese so genannte VCS-Stufe B wurde zur MEDICA 2001 vorgestellt. Die jeweilige konkrete Integration in die Praxissoftware unterliegt dem entsprechenden Hersteller und wird daher in Details unterschiedlich ausfallen. Der VDAP hat jedoch ein Zertifizierungsverfahren für die Prüfung erarbeitet. Eine Werbung mit VCS-Gütesiegel ist nur nach Zertifizierung erlaubt. Der VDAP wird die Spezifikation bei CEN einreichen, um den Status eines offiziellen Standards zu erreichen. Die Umsetzung der gerichteten Kommunikation mittels VCS entspricht den Vorgaben der so genannten *adressierten Vertraulichkeit* des HCP-Protokolls aus Bayern [HCP00]. VCS ist als DIN-Veröffentlichung verfügbar. Die Mitglieder im VDAP haben die VCS-Schnittstelle den Ärzten kostenlos

zur Verfügung gestellt. Da einige Mitglieder aufgrund von Interessenskonflikten aus dem VDAP ausgetreten sind, ist die weitere Zukunft von VCS ungewiss [LBM⁺05].

Patientenbegleitende Dokumentation (PaDok)

Das Fraunhofer-Institut für Biomedizinische Technik (IBMT) hat zusammen mit der MEDNET AG PaDok als integriertes Kommunikationskonzept für die Vernetzung niedergelassener Allgemein- und Fachärzte entwickelt [D2D02]. Wie bei VCS geht es um eine hohe Datensicherheit bei der Übertragung der Daten, wie auch um eine Integration in die wichtigste Arztpraxissoftware. Auch PaDok konzipiert eine ungezielte Kommunikation mit der zweckbestimmten Bereitstellung von Daten auf einem zentralen Server. Basierend auf PaDok wurde von der Kassenärztlichen Vereinigung Nordrhein (KV Nordrhein) im Rahmen einer Telematik-Initiative [D2D02] die Client-Server-Technologie D2D (Doctor to Doctor) ins Leben gerufen. Nach der Umsetzung eines eArztbriefes, wurden die eÜberweisung und die eKrankenhauseinweisung mittels XML umgesetzt. Einstieg in die ePatientenakte soll die eMamma-Akte darstellen (siehe dazu auch Abschnitt 11.4.3 auf Seite 149). D2D ist laut der KV Nordrhein ein offenes Konzept mit dem Ziel einer umfassenden Standardisierung der Telematik im Gesundheitswesen.

Seit Ende 2003 gibt es eine Kooperationsvereinbarung zwischen dem VDAP und dem IBMT über eine Abstimmung und Abgleichung der Aktivitäten im Bereich der elektronischen Kommunikationslösungen für Ärzte [Koo05]. Erste Zusammenarbeit fand unter Vermittlung der Fachhochschule Dortmund bei der eMamma-Akte statt.

6.5.3 Leitlinien und Klinische Pfade

Auch auf anderen Gebieten der Medizin gibt es vor allem zur Kostensenkung Bemühungen um eine Standardisierung. So ist in der Neufassung des Sozialgesetzbuches V verankert, dass anerkannte Leitlinien zur Qualitätssicherung im Rahmen einer leistungsfähigen Patientenversorgung bereitzustellen sind [Nie01]. Leitlinien sind systematisch entwickelte Entscheidungshilfen für Leistungserbringer und Patienten über die angemessene Vorgehensweise bei speziellen Gesundheitsproblemen [AEZ02]. Verwendet werden die Leitlinien unter anderem im Rahmen der Disease Management-Programme für chronisch Kranke [LS01b]. Krankenkassen müssen in mehreren Schritten bis 2007 chronisch Kranke in solche Programme aufnehmen, um finanzielle Unterstützung über den Risikostrukturausgleich zu bekommen [Lex02]. Damit soll eine qualitativ hochwertige Versorgung chronisch Kranker wieder attraktiv und finanzierbar sein.

Die Klinischen Pfade oder Clinical Pathways sind krankenhausintern verbindlich und erfassen den gesamten patientenbezogenen Handlungsablauf und unterscheiden sich daher von den Leitlinien [hil01]. Auch sie dienen der Qualitätssicherung und Straffung des

Behandlungsprozesses. Bedeutung haben diese Pfade ebenfalls im Zusammenhang mit dem neuen diagnosebasierten Pauschalentgeltsystem (DRG). Anhand der Klinischen Pfade können die Kosten quer durch die Abteilungen eines Krankenhauses für den Behandlungsfall ermittelt werden und so Sicherheit in die Budgetplanungen und -verhandlungen bringen [Küh01]. Ziel ist es, Leitlinien in Dokumentationen als externe Wissensquellen zu integrieren [Wol02].

7 Architektur einer institutionsübergreifenden Interoperabilität

7.1 Erweiterte Mediator-basierte Architektur

In den vorangegangenen Abschnitten sind die wichtigsten Grundlagen und Techniken kurz beschrieben worden, die für eine Konzeption einer institutionsübergreifenden Interoperabilität wichtig sind.

Verschiedene Institutionen setzen häufig unterschiedliche Standards ein, die kombiniert und zwischen denen entsprechend vermittelt werden muss, um Interoperabilität zu erreichen. Dafür soll in den nächsten Abschnitten eine entsprechende Architektur vorgestellt werden, die zuvor formulierte Anforderungen erfüllt. Die vorgestellte Architektur ist unabhängig von domänenspezifischen Standards. In Abschnitt 7.2 wird die Architektur dann auf die Domäne Gesundheitswesen angepasst.

7.1.1 Anforderungen an die Architektur

Es ist nun eine Architektur für eine institutionsübergreifende Interoperabilität zu entwickeln, die:

- prinzipiell für beliebig viele Informationsquellen anwendbar ist (*Skalierbarkeit*)
- gute Weiterentwicklungsmöglichkeiten bietet (*Evolution*)
- flexibel anpassbar bei Entfernen und Hinzufügen von Informationsquellen ist (*Flexibilität*)

7.1.2 Allgemeine Struktur der Architektur

Erfüllt werden die oben genannten Anforderungen durch mediator-basierte Informationssysteme [Wie92, Wie95] mit entsprechenden Metadaten, Facilitatoren und Komponentenmediatoren. Damit können die relevanten Standards im Hinblick auf Interoperabilität kombiniert werden. Ein Grund dafür ist vor allem die Tatsache, dass die zusätzliche Vermittlungsschicht mit den Mediatoren und Facilitatoren die Basisschicht (Datenbanken usw.) und die Nutzerschicht mit den Applikationen in einer Schichtenarchitektur entkoppelt. Drei-Ebenen-Architekturen sind in weiten Gebieten der Informatik etabliert. Als Grundlage eignet sich also die mediator-basierte Architektur nach [Has02] (Abbildung 3.3), die für eine Vermittlung zwischen unterschiedlichen Standards noch erweitert wird mit einem Meta-Facilitator und einem Meta-Komponentenmediator, wie Abbildung 7.1 veranschaulicht.

Eine Vermittlungsschicht, die eine umfassende Interoperabilität auf der Basis von Domänenstandards leisten soll, besteht aus:

1. Wrappern
2. Komponentenmediatoren
3. domänenspezifischen Facilitatoren
4. optionalen Applikationsmediatoren und
5. diversen Metadaten, dargestellt als 

Die diversen Facilitatoren verwalten als spezielle Mediatoren die jeweiligen domänenspezifischen Modelle. Ein Facilitator ist eine Komponente, die Koordinationsleistungen, basierend auf Regeln, anbietet. Der Meta-Facilitator koordiniert mit entsprechenden Abbildungsregeln die einzelnen Sub-Facilitatoren.

Die verschiedenen Komponentenmediatoren existieren je Domänenstandard. Der Meta-Komponentenmediator vermittelt mittels entsprechender Abbildungsspezifikationen zwischen den spezifischen Komponentenmediatoren. Diese Komponentenmediatoren vervollständigen die Integration in standardbasierte, domänenspezifische Modelle. Die Metadaten sollen helfen, zwischen den verschiedenen Domänenstandards zu vermitteln. So kann eine Interoperabilität zwischen diversen Domänenstandards gelingen.

7.1.3 Zusammenfassung

Durch Erweiterung der Mediator-basierten Architektur nach [Has02] durch einen Meta-Facilitator und einen Meta-Komponentenmediator kann nun auch eine Vermittlung zwischen unterschiedlichen Standards erfolgen.

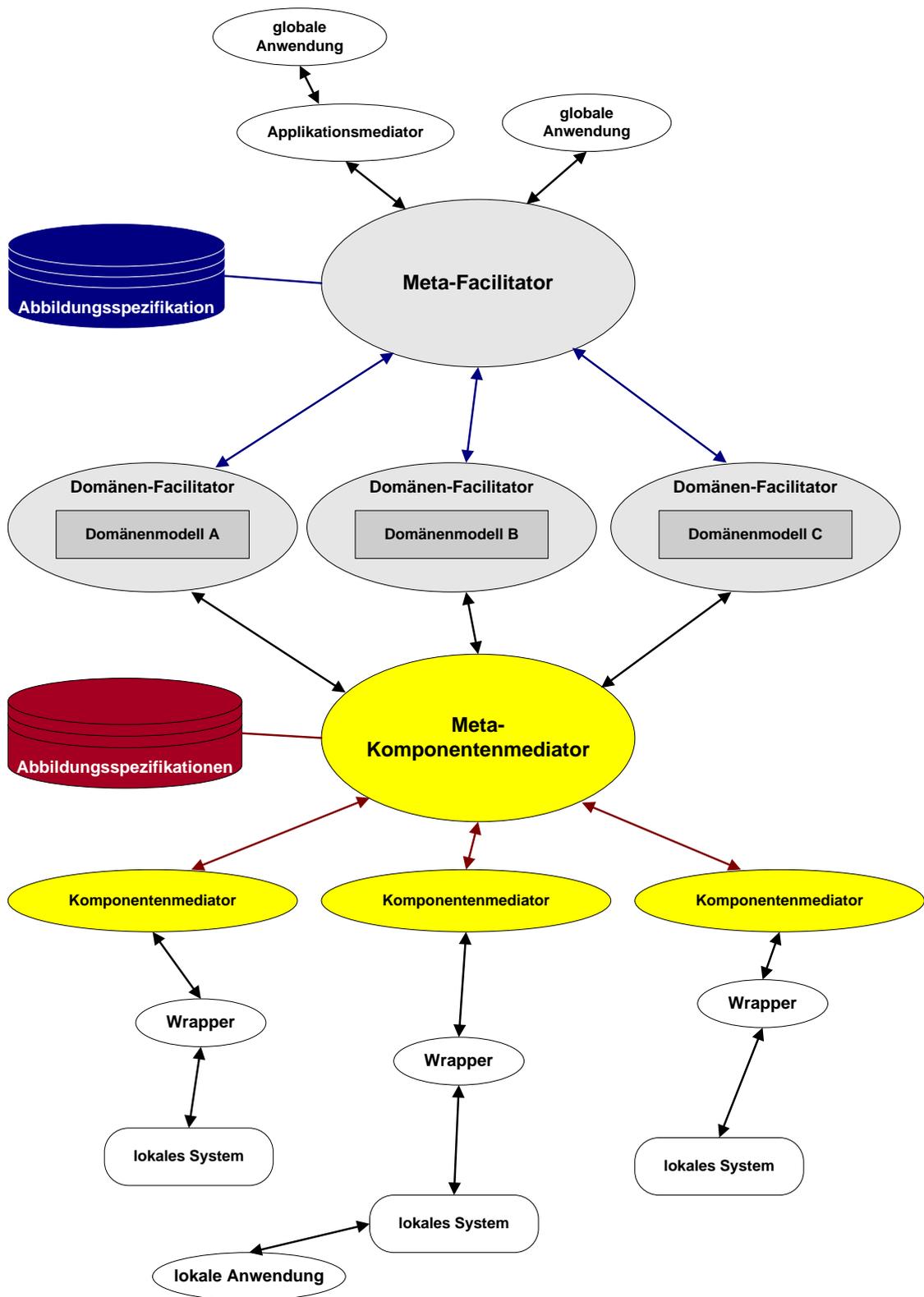


Abbildung 7.1: Allgemeine Mediator-basierte Architektur

7.2 Mediator-basierte Architektur für das Gesundheitswesen

Die erweiterte Mediator-basierte Architektur ist geeignet, um zwischen den verschiedenen Kommunikationsstandards auf der einen Seite und den verschiedenen Dokumentationsstandards auf der anderen Seite in der Domäne Gesundheitswesen zu vermitteln. Dazu wird die allgemeine Architektur zunächst auf die konkrete Domäne Gesundheitswesen angepasst.

7.2.1 Anwendung der Architektur auf das Gesundheitswesen

Zunächst wird die auf die Domäne Gesundheitswesen angepasste mediator-basierte Architektur vorgestellt, bevor in Abschnitt 7.3 genauer auf die Zusammenhänge der Metadaten der Domäne eingegangen wird.

Abbildung 7.2 zeigt nun exemplarisch die mediator-basierte Architektur für eine institutionsübergreifende Interoperabilität im Gesundheitswesen. Später wird die zugrunde liegende allgemeine Architektur noch genauer beschrieben.

7.2.2 Aufgaben der einzelnen Architekturteile

Facilitatoren

Die diversen Facilitatoren verwalten als spezielle Mediatoren die jeweiligen domänenspezifischen Modelle, z.B. HL7-Facilitator für das HL7-DIM (Domain Information Model). Der Kommunikations-Facilitator koordiniert als Meta-Facilitator mit entsprechenden Abbildungsregeln die einzelnen Sub-Facilitatoren. Beim Entwurf der Abbildungsspezifikation hilft das Begriffssystem für Kommunikationsstandards, das die Gemeinsamkeiten der Kommunikationsstandards zeigt, und die einheitlich in der UML modellierten Standards [PH03].

Komponentenmediatoren

Die verschiedenen Komponentenmediatoren existieren je Dokumentenstandard, z.B. CDA-Komponentenmediator. Der Dokumentations-Komponentenmediator vermittelt als Meta-Komponentenmediator mittels entsprechender Abbildungsspezifikationen zwischen den

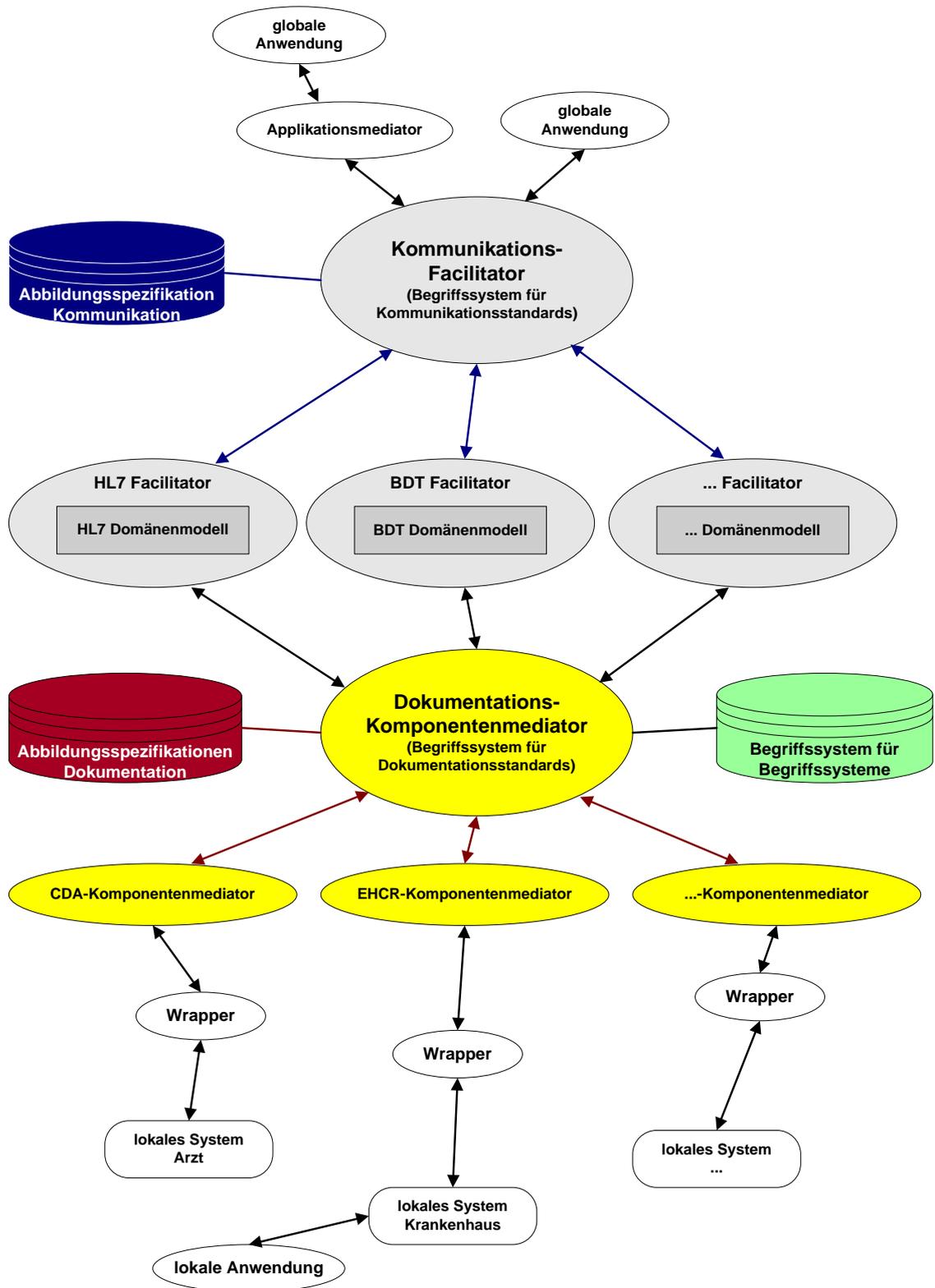


Abbildung 7.2: Mediator-basierte Architektur für das Gesundheitswesen

spezifischen Komponentenmediatoren. Hilfreich beim Entwurf der Abbildungsspezifikationen ist das Begriffssystem für Dokumentationsstandards, welches die Gemeinsamkeiten der Dokumentenstandards aufzeigt. Diese Komponentenmediatoren vervollständigen die Integration in standardbasierte, domänenspezifische Modelle. Die Metadaten „Begriffssystem für Begriffssysteme“ sollen helfen zwischen den verschiedenen medizinischen Begriffssystemen, die innerhalb der Dokumentation verwendet werden, zu vermitteln.

Dieser Ansatz erreicht eine Trennung der Verwaltung der globalen Modelle und der Integration von Komponentenmodellen in domänenspezifische Modelle [Has00b, Has02].

Wrapper

Wrapper sind Komponenten, die die Datenmodellheterogenität kapseln und somit den Zugriff auf Daten existierender Systeme ermöglichen. Optionale Applikationsmediatoren vermitteln zwischen Domänenfacilitatoren und globalen Applikationen.

Wartung der Integration

Durch diese Architektur wird eine dezentralisierte Verantwortung für die Wartung der Integration erreicht. Jeder Komponentenmediator muss eine Sicht seines lokalen Informationssystems in die gemeinsamen Domänenmodelle einbinden, welche durch die Domänenfacilitatoren angeboten werden. Somit diktiert die Domänenfacilitatoren die Konditionen, die die Komponentenmediatoren erfüllen müssen, wenn sie an dem integrierten System teilhaben wollen. Nur dadurch kann eine gute Strukturierung der gemeinsamen Modelle gewährleistet werden.

Prozess der Integration

Der Prozess der Integration soll, top-down beginnend bei den jeweiligen Domänenstandards, erfolgen. Dadurch ist die Struktur des gemeinsamen Modells nicht von den Überlappungen zwischen den Komponentenmodellen abhängig, sondern durch die Anforderungen globaler Applikationen und domänenspezifischer Standards bestimmt. Damit führen die vorgestellten Mediator-basierten Architekturen zusammen mit dem Top-Down-Ansatz zu skalierbaren, weiterentwickelbaren und flexiblen Softwarearchitekturen.

Metamodell der Architektur

Das zur Architektur gehörende Metamodell in Form eines UML-Klassendiagramms zeigt Abbildung 7.3. Abbildung 7.2 stellt eine Instanz dieses Modells dar.

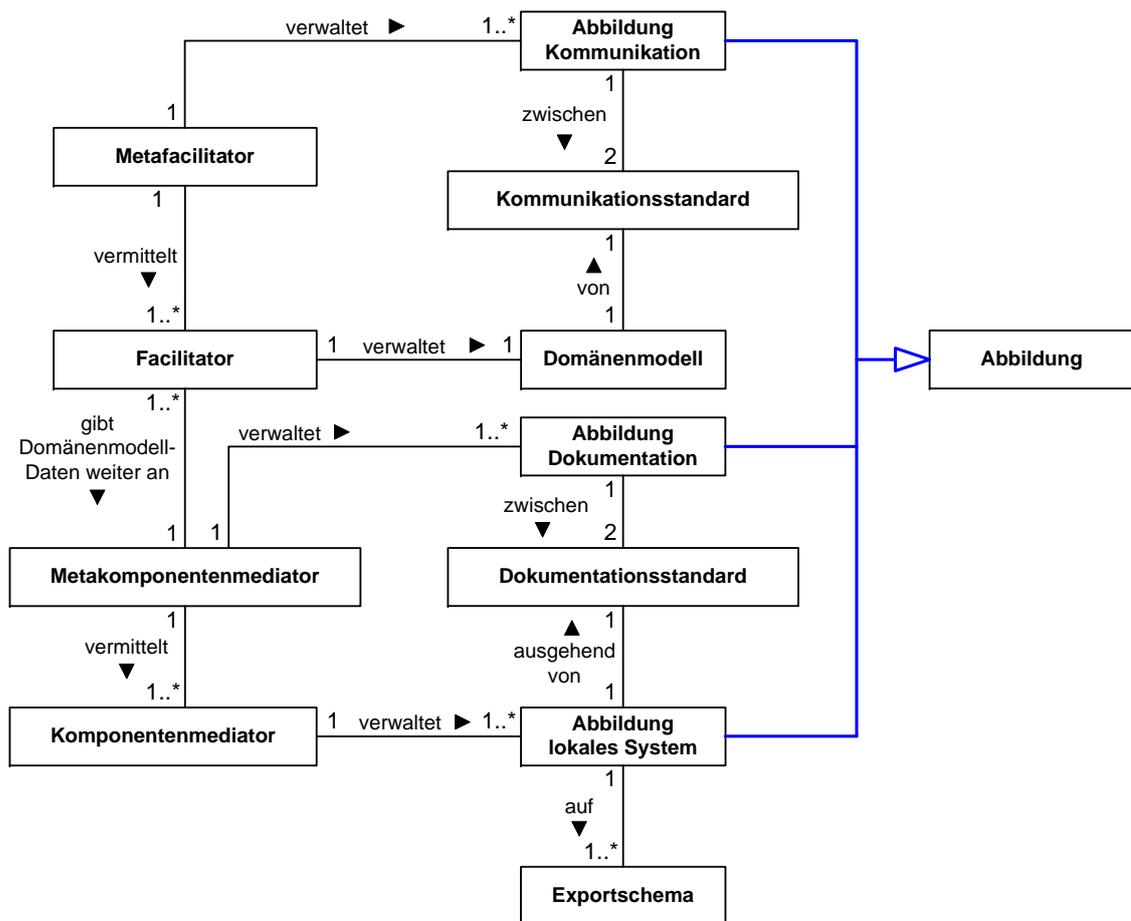


Abbildung 7.3: Metamodell der vorgestellten Mediatorbasierten-Architektur (Abbildung 7.2 stellt eine Instanz dieses Modells dar)

Schemaarchitektur

Abbildung 7.4 verdeutlicht in Form einer Schemaarchitektur die entsprechende statische Sicht für die vorgestellte Softwarearchitektur.

Die lokalen Datenmodelle werden durch Wrapper in Komponentenschemata und umgekehrt transformiert, um die jeweiligen lokalen Schemata in ein gemeinsames kanonisches Datenmodell zu überführen. Damit kapseln Wrapper die Datenmodellheterogenität. Die Exportschemata stellen eine gewünschte Sicht auf die Komponentenschemata dar. Die Informationen über die Abbildungen von den föderierten Schemata der Dokumentationsstandards auf die einzelnen Exportschemata werden von den entsprechenden Komponentenmediatoren verwaltet. So verwaltet z.B. der CDA-Komponentenmediator die Abbildungsinformationen vom CDA-Schema auf Exportschemata. Das CDA-Schema liegt ebenfalls im kanonischen Datenmodell vor. Hierbei findet bereits eine Top-Down-Integration, ausgehend von Dokumentationsstandards, statt. Der Dokumentationskomponentenmediator verwaltet die Abbildungsspezifikationen zwischen den Dokumentationsstandards für eine Transformation z.B. von CDA in EHCR. Die Schemata der Standards

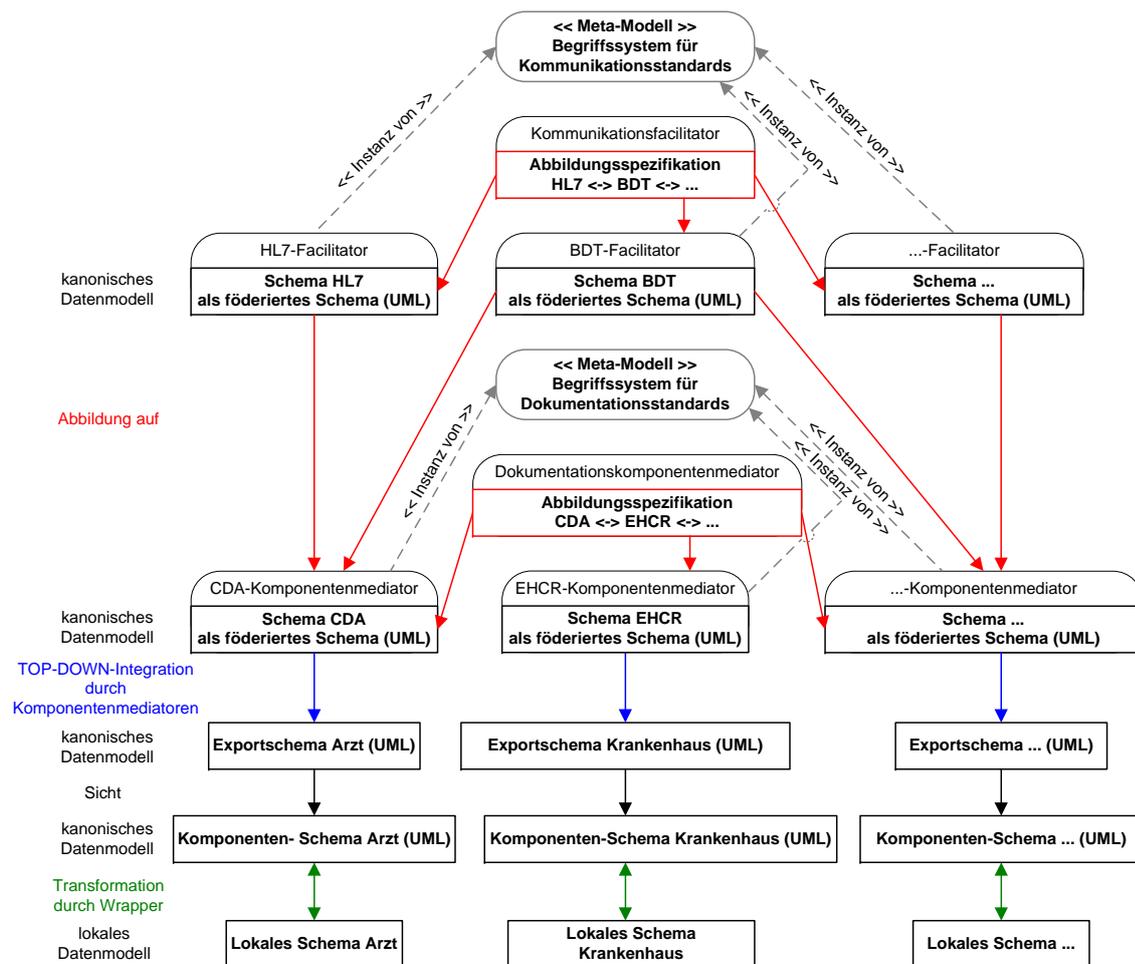


Abbildung 7.4: Schemaarchitektur

sind dabei jeweils Instanz eines gemeinsamen Metamodells bzw. Begriffssystems für Dokumentationsstandards, welches in Abschnitt 7.3 vorgestellt wird.

Die föderierten Schemata der Kommunikationsstandards werden im kanonischen Datenmodell von entsprechenden Facilitatoren verwaltet. Der HL7-Facilitator verwaltet so z.B. das Referenzinformationsmodell (RIM) von HL7. Weitere Komponentenmediatoren verwalten die Abbildungsspezifikationen zwischen den föderierten Schemata der Kommunikationsstandards und denen der Dokumentationsstandards. Ein Komponentenmediator für die Abbildung HL7 auf CDA ist dabei nicht nötig, da die CDA auf HL7 basiert. Die Abbildungsregeln für eine Transformation zwischen den verschiedenen föderierten Schemata der verschiedenen Kommunikationsstandards verwaltet der Kommunikationsfacilitator. Die Schemata der entsprechenden Kommunikationsstandards sind wieder jeweils Instanz des gemeinsamen Metamodells bzw. Begriffssystems für Kommunikationsstandards. Das Metamodell wird ebenfalls in Abschnitt 7.3 vorgestellt.

7.3 Kombination der Metadaten

Metadaten für föderierte Informationssysteme sind ein wichtiges Konzept für das Erreichen von Flexibilität bei der Weiterentwicklung und bei Änderungen sowie für die Überwindung von Heterogenität und Verteilung [BKLW99].

Die Metadaten für die betrachtete Domäne Gesundheitswesen werden im Folgenden auf Modell- und Metamodellebene in einer an die MOF-Architektur angelehnten Architektur vorgestellt werden. Auch ihre Zusammenhänge werden in Form eines Diagrammes aufgezeigt.

7.3.1 Zusammenhang der Metadaten

Abbildung 7.5 zeigt die Zusammenhänge der jeweiligen Metadaten aus den Abschnitten 7.3.2, 7.3.3 und 7.3.4 für Kommunikation, Dokumentation und Begriffssysteme in Form eines UML-Diagrammes. So enthalten die entsprechenden Begriffssysteme jeweils die Metadaten für die Standards bzw. die medizinischen Begriffssysteme. Das Begriffssystem für Kommunikationsstandards enthält die Metadaten für die wichtigsten Kommunikationsstandards HL7, BDT und DICOM. Dokumentationsstandards wie Electronic Healthcare Records, Arztbriefe und Basisdokumentation für Tumorkranke verwenden medizinische Begriffssysteme. Der Electronic Healthcare Record der HL7-Gruppe, die CDA, wird im Standard HL7 formuliert.

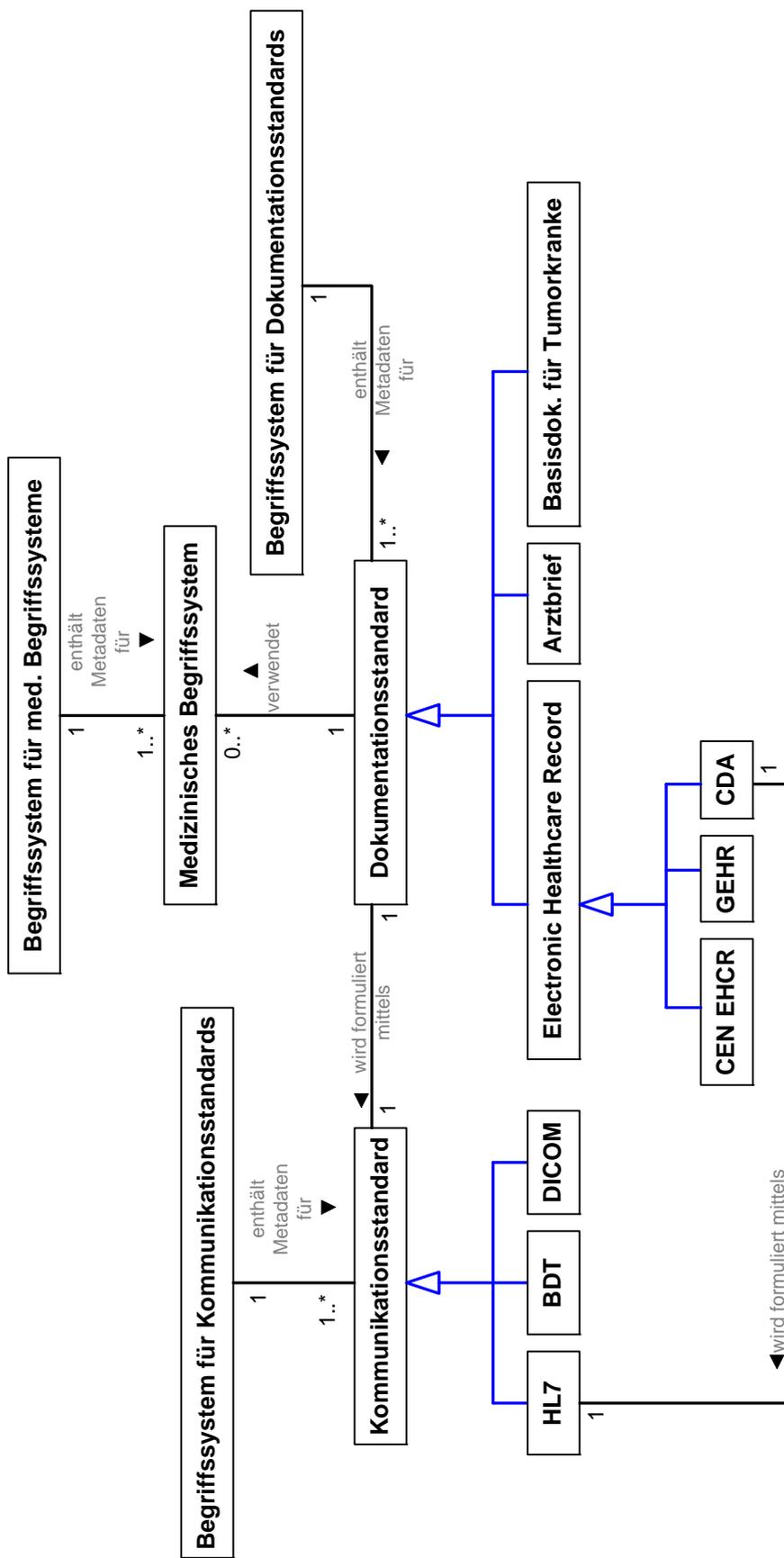


Abbildung 7.5: Zusammenhang der Metadaten als UML-Diagramm

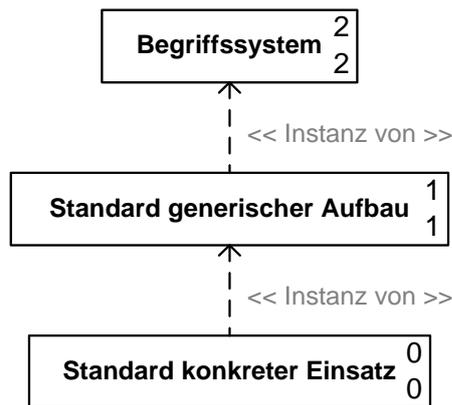


Abbildung 7.6: Tiefe Instanziierung, bezogen auf die Begriffssysteme

Alle Klassen der Begriffssysteme müssen gemäß des Mehrebenen-Modellierungsframeworks nach dem Prinzip der tiefen Instanziierung mit Potenz 2 und Ebene 2 betrachtet werden. Tiefe Instanziierung wird in Abschnitt 4.2.3 auf Seite 31 vorgestellt. Die Begriffssysteme gehören auf die Ebene M2, die Metamodell-Ebene und haben daher die Ebene 2. Potenz 2 bedeutet, dass die Begriffssysteme zweimal instanziiert werden können. Die UML-Modelle der Standards sind der Ebene M1 zuzuordnen. Sie veranschaulichen die generische Struktur des jeweiligen Standards. Jede Klasse wird mit der Potenz 1 und der Ebene 1 ergänzt. Sie können einmal instanziiert werden. Die konkrete Struktur einer Nachricht in einem Standard, z.B. eine HL7 Tumor-DTD gehört zur Ebene M0 und kann nicht weiter instanziiert werden. Abbildung 7.6 veranschaulicht die Ebenen und tiefe Instanziierung ausgehend vom Begriffssystem hin zur konkreten Struktur bei konkretem Einsatz, z.B. bei einem Rezept.

Diese Begriffssysteme erfassen jeweils die Gemeinsamkeiten der Standards. Da dies nur auf einer übergeordneten Ebene stattfinden kann und die Standards zum Teil stark differieren, können diese Zusammenhänge nicht bei den eigentlichen Transformationen helfen. Es existiert kein gemeinsames, mehrere Standards umfassendes Datenschema. Die Begriffssysteme stellen jedoch einen Leitfaden zur Strukturanalyse der Standards und deren Abbildungen dar und unterstützen den Entwickler bei der Erstellung der Abbildungen dadurch, dass er weiß, auf welche Kriterien er bei der Erfassung der Korrespondenzen zu achten hat. Die Metamodelle helfen also beim Entwurf, nicht aber bei der automatischen Vermittlung. Die UML besitzt keine vollständige formale Semantik, da sie nicht mathematisch fundiert ist, sie ist lediglich semiformal. Dies unterscheidet die vorgestellten Metamodelle von einer formalen Ontologie [MM03].

7.3.2 Metadaten Kommunikation

Abbildung 7.7 zeigt als UML-Diagramm die Metadaten für den Kommunikations-Facilitator. Sie stellen ein Begriffssystem für medizinische Kommunikationsstandards dar.

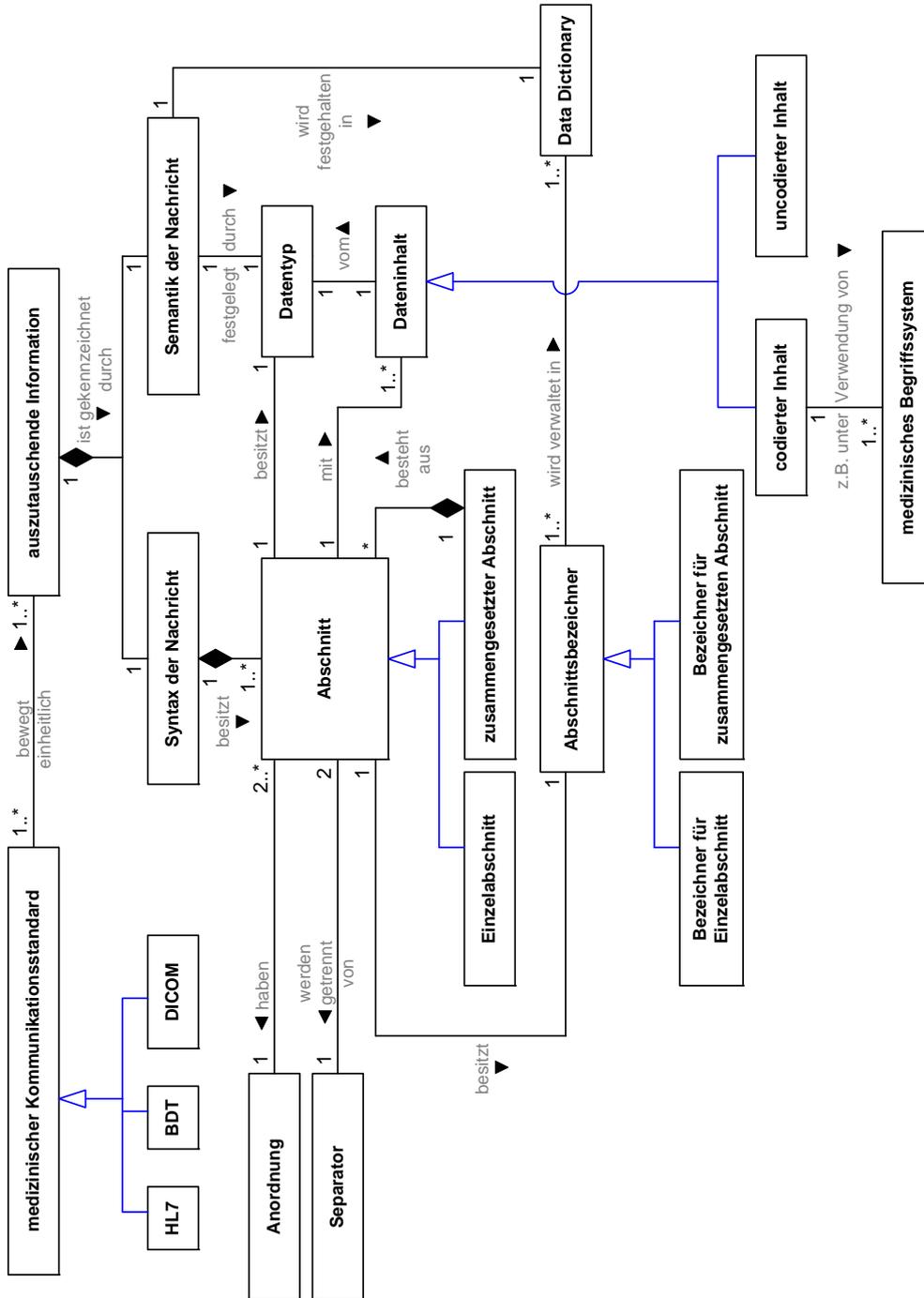


Abbildung 7.7: Ausschnitt aus dem Metamodell der Kommunikationsstandards in der Medizin als UML-Diagramm

Zwei unterteilende Abschnitte besitzen eine Anordnung und werden durch Separatoren getrennt. Auf jeder Ebene besitzt ein unterteilender Abschnitt einen eindeutigen Abschnittsbezeichner, der in den jeweiligen Data Dictionaries verwaltet wird. Die unterteilenden Abschnitte besitzen Dateninhalte von bestimmtem Datentyp, der zusammen mit dem Data Dictionary die Semantik der auszutauschenden Information festlegt. Zum koordinierten Dateninhalt zählt beispielsweise die Verwendung von medizinischen Begriffssystemen.

Die Syntax der Nachricht besitzt allgemein mehrere Ebenen unterteilender Abschnitte. Diese Ebenen zeigt Abbildung 7.8 detaillierter.

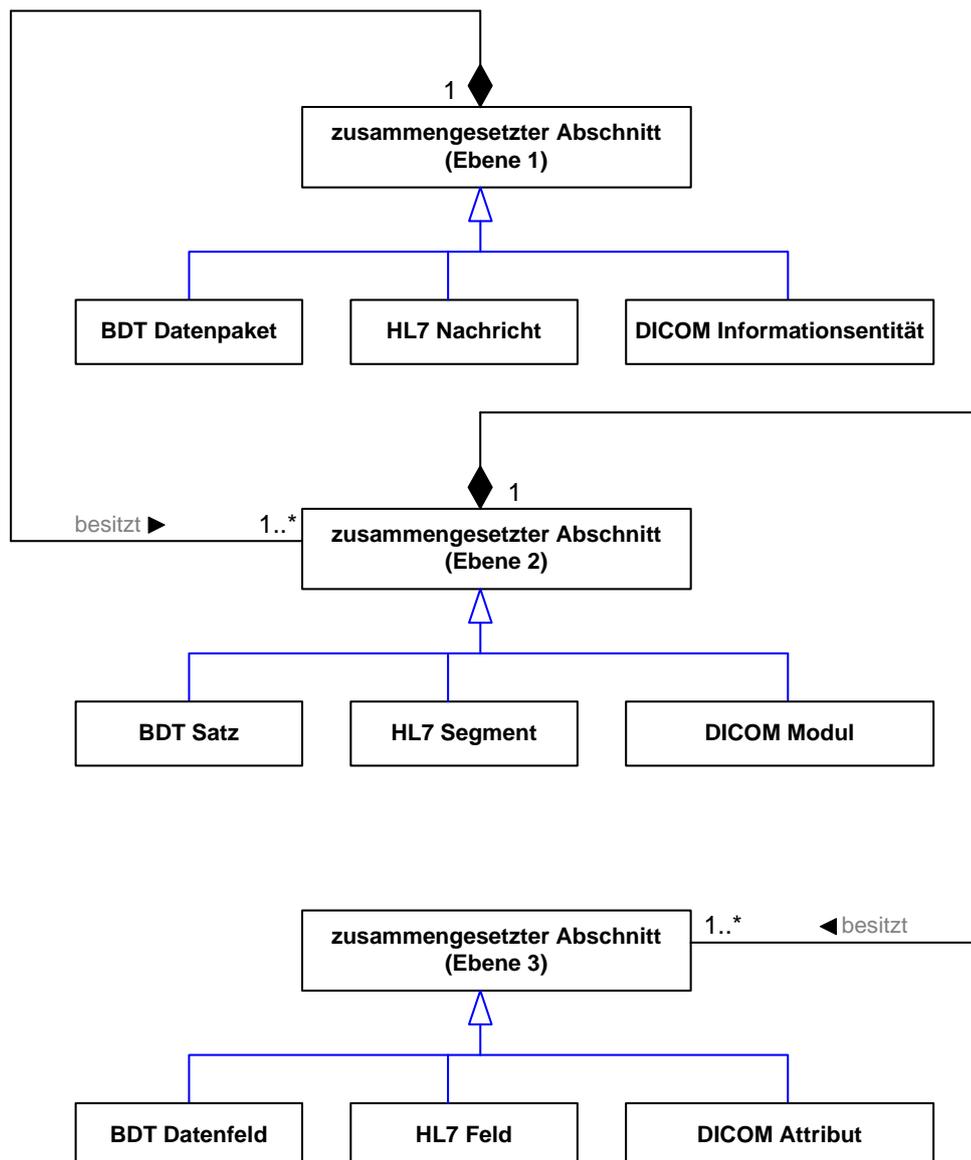


Abbildung 7.8: Die drei Ebenen der unterteilenden Abschnitte als UML-Diagramm

So finden sich auf einer Ebene das BDT-Datenpaket, die HL7-Nachricht und die DI-

COM Informationsentität wieder. Diese wiederum enthalten entsprechend BDT Sätze, HL7 Segmente und DICOM Module auf der nachfolgenden Ebene. Auf einer weiteren Ebene sind das BDT-Datenfeld, das HL7-Feld und das DICOM Attribut zu nennen.

Abbildung 7.9 zeigt die Untertypen zu den Klassen Datentyp, Data Dictionary und eindeutige Abschnittsbezeichner zur Ebene 2 der Abbildung 7.8. a) zeigt die Terminologie

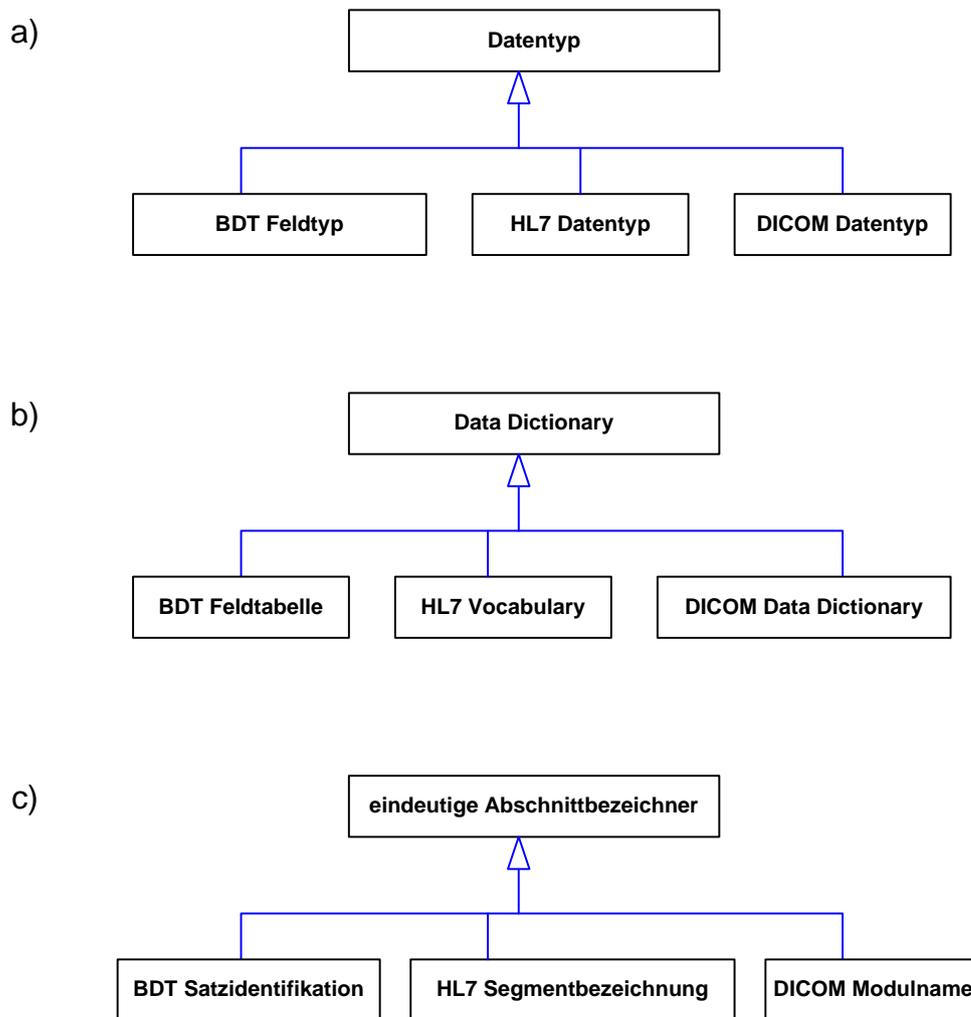


Abbildung 7.9: Details zu drei Klassen des Begriffssystems als UML-Diagramm

in BDT, HL7 und DICOM für Datentyp, b) für Data Dictionary und c) bezeichnet die Abschnittsbezeichner der drei Standards

Durch weitere Standardisierung bzw. Festlegung auf ein bestimmtes Kommunikationsformat oder syntaktische Integration (keine semantische) von HL7 und BDT z.B. mittels XML im Projekt SCIPHOX [SCI02] kommt man einer weniger komplexen Vermittlung näher.

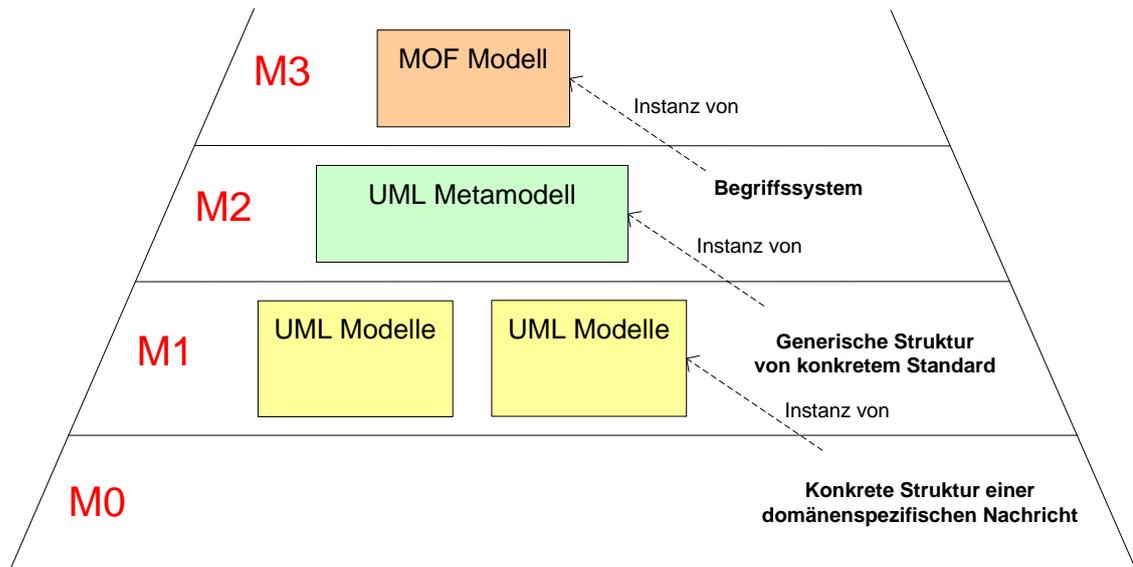


Abbildung 7.10: An die Meta-Object Facility Architektur angelehnte Architektur

Das Begriffssystem ist ein UML-Metamodell und quasi Instanz des MOF-Modells [MOF03]. Instanzen des Begriffssystems sind dann allgemeine Nachrichtenstrukturen in einem Standard (UML-Modelle), beispielsweise HL7 oder BDT. Eine Instanz des UML-Modells ist dann eine konkrete Nachrichtenstruktur, wie beispielsweise eine HL7-DTD für einen ärztlichen Kurzbericht oder eine Tumordokumentation. Abbildung 7.10 zeigt die entsprechenden Zusammenhänge, die auch für die anderen Begriffssysteme gelten.

Grundlage für MOF ist der ISO-Architekturstandard IRDS (siehe Abschnitt 4.1.1 auf Seite 25). Das MOF-Modell als Meta-Metamodell beschreibt die Modellierungselemente, auf die Metamodellebene gehören die verschiedenen Begriffssysteme für Kommunikation, Dokumentation und medizinische Begriffssysteme. Die UML-Modelle auf der untersten Ebene beschreiben dann die allgemeinen Strukturen des Standards.

7.3.3 Metadaten Dokumentation

Bei der Dokumentation spielen für die Metadaten verschiedene Dokumentenstandards für die Medizin eine Rolle. Abbildung 7.11 zeigt die Metadaten für den Dokumentations-Komponentenmediator als UML-Diagramm. Dies stellt ein Begriffssystem für medizinische Dokumentenstandards dar.

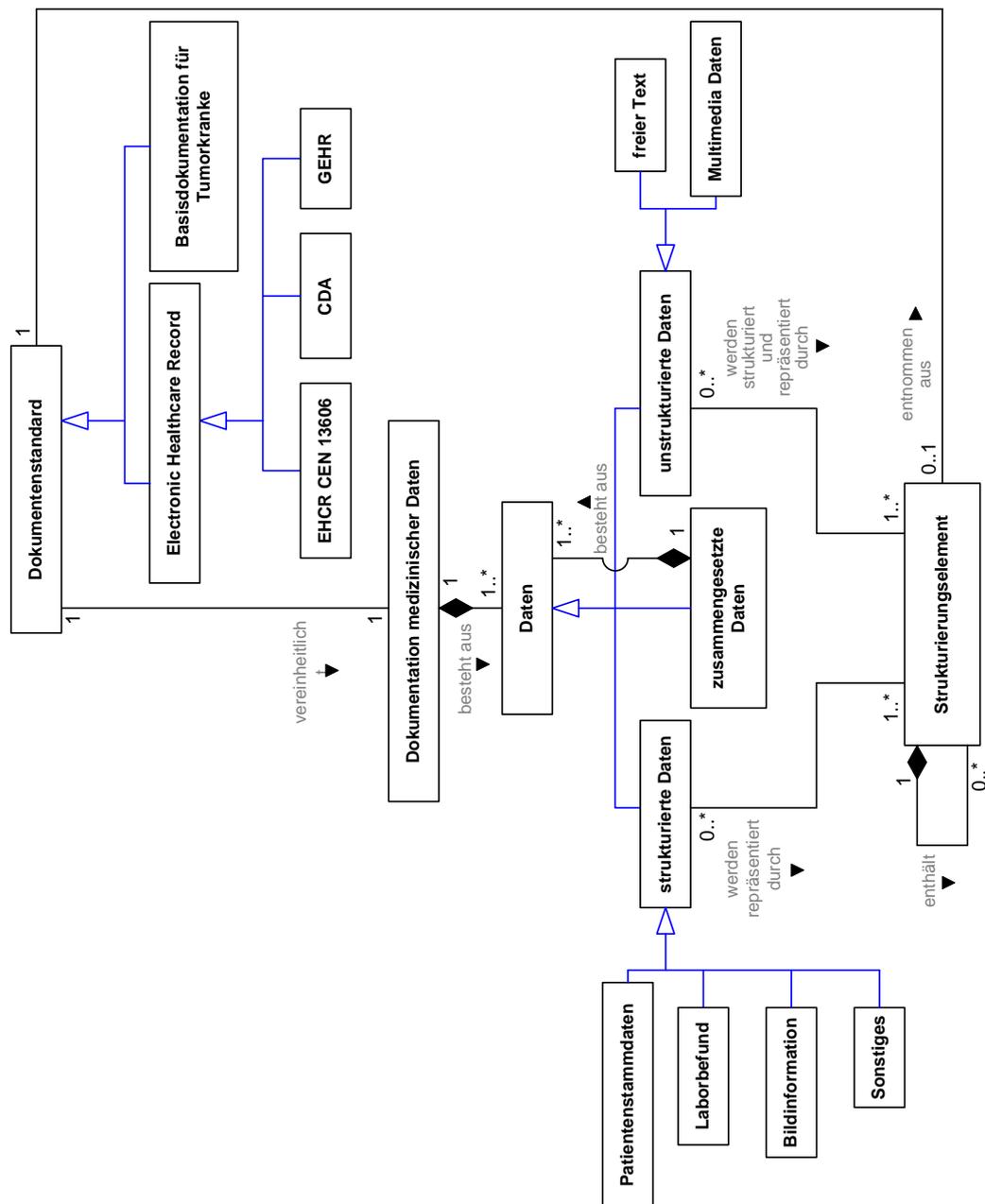


Abbildung 7.11: Ausschnitt aus dem Metamodell der Dokumentenstandards in der Medizin als UML-Diagramm

Ein Dokumentenstandard erhält seine Semantik aus den medizinischen Begriffssystemen und er vereinheitlicht die Dokumentation medizinischer Daten. Diese können strukturiert und unstrukturiert sein. Die Strukturierungselemente, mit denen die Daten strukturiert und repräsentiert werden, kommen aus den Dokumentenstandards, wie z.B. dem EHCR CEN 13606, der CDA oder der Basisdokumentation für Tumorkranke. Speziellere Metamodelle für den Bereich der Dokumentation gibt es für die Tumordokumentation in Form eines generischen XML-basierten OnkoDok-Modells [Wol02]. Für die gesamte medizinische Dokumentation ist ein solches Modell sicher nicht realisierbar.

7.3.4 Metadaten Begriffssysteme

Die Abbildungen 7.12 und 7.13, modelliert in der UML, geben einen Überblick über die wichtigsten Zusammenhänge in der Welt der medizinischen Begriffssysteme und bilden ein systematisches Begriffssystem für die Begriffssysteme. Dargestellt werden in Abbildung 7.12 die Grundelemente der Terminologielehre, Bezeichnung, Begriff und Gegenstand, die Zusammenhänge Klassifikation, Nomenklatur, Thesaurus sowie enumeratives und kompositionelles Begriffssystem. Abbildung 7.13 verdeutlicht die Zusammenhänge zwischen den diversen Begriffsbeziehungen und den Codierprinzipien in den Begriffssystemen.

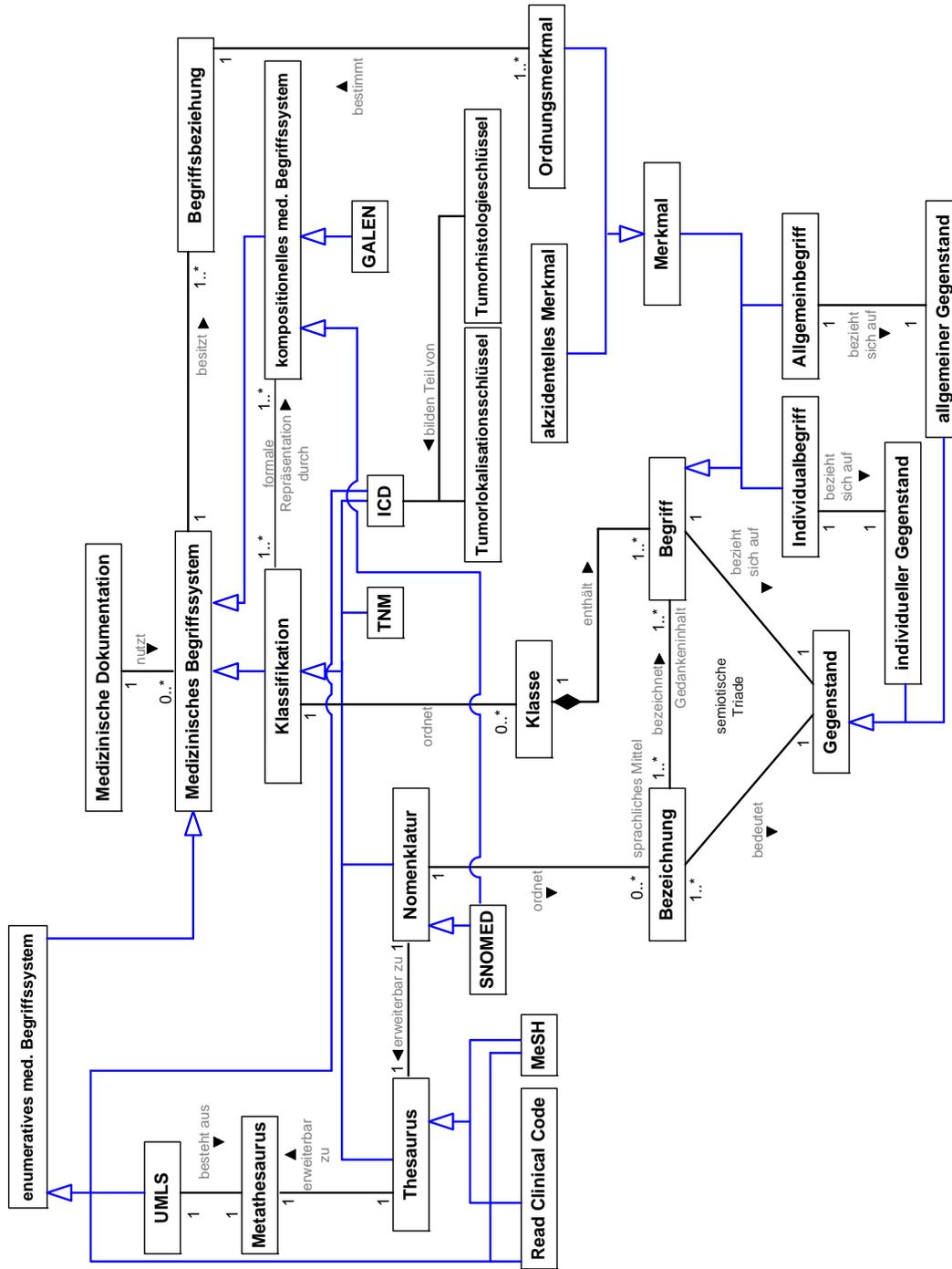


Abbildung 7.12: Ausschnitt medizinische Begriffssysteme als UML-Diagramm

Dargestellt ist die *semiotische Triade* [Ber95], die die geltenden Beziehungen darstellt: Eine Bezeichnung bezeichnet einen Begriff, eine Bezeichnung bedeutet einen Gegenstand und ein Begriff bezieht sich auf einen Gegenstand. Klassen enthalten nun Begriffe und werden von Klassifikationen geordnet. Klassifikationen sind medizinische Begriffssysteme, welche von der medizinischen Dokumentation genutzt werden. Die TNM-Klassifikation [TNM02] und der ICD [ICD02a] sind Beispiele für eine Klassifikation. Nomenklaturen sind Zusammenstellungen von Bezeichnungen, die dahinterstehende Systematik basiert in der Regel auf einer Begriffsordnung, ein Beispiel ist SNOMED [SNO02]. Eine Nomenklatur kann nun erweitert werden zu einem Thesaurus. Hierzu gehören der Read Clinical Code [Rea02] und die Medical Subject Headings [MeS02]. Wird wiederum der Thesaurus um semantische und linguistische Informationen erweitert, so entsteht ein Metathesaurus, wie ihn der Kern von UMLS [UML02] darstellt. UMLS gehört zu den enumerativen Begriffssystemen und versucht die wichtigsten medizinischen Begriffssysteme in einem einzigen zu integrieren und dabei möglichst viele Begriffsbeziehungen explizit darzustellen. Dazu abgegrenzt werden kompositionelle Begriffssysteme, wie z.B. GALEN, welches versucht, methodische Grundlagen für die formale Repräsentation medizinischer Begriffssysteme zu entwickeln. GALEN erlaubt die formale Konstruktion von allen anderen fachrelevanten Begriffen durch die Komposition einer Anzahl primitiver Begriffe [Ber95, PH02].

Man unterscheidet in medizinischen Begriffssystemen nichthierarchische und hierarchische Begriffsbeziehungen. Zu ersteren zählen pragmatische und sequentielle Begriffsbeziehungen, zu letzteren generische (Spezialisierung/Generalisierung) und partitive Beziehungen (Teile-Ganzes-Beziehungen). Ferner unterscheidet man bei hierarchischen multi- und uniaxiale Systeme sowie multi- und monohierarchische Begriffssysteme. Begriffssysteme benutzen diverse Codierprinzipien, wobei häufig mehrere Codierprinzipien kombiniert werden. Der ICD ist beispielsweise uniaxial, multihierarchisch mit generischen und pragmatischen Beziehungen. Seine Codierprinzipien sind der gruppensequentielle und der hierarchische Code [Ber95, PH02].

Eine Vereinfachung ist möglich durch weitere Standardisierung bzw. Festlegung auf ein bestimmtes Begriffssystem bzw. Integration in ein Begriffssystem wie z.B. UMLS. Ein Ansatz, der UMLS nutzt, ist das Projekt Medical UMLS based Terminology Server for Authoring, Navigating and Guiding the Retrieval to Heterogeneous Knowledge Sources (MUSTANG) der Universität Lübeck [IRS01]. Die terminologischen Dienste des MUSTANG-Servers sind dabei über zahlreiche Schnittstellen zugänglich.

7.4 Transformationen zwischen den Standards

Für eine Transformation zwischen den Standards wurden die Standards einheitlich in der UML spezifiziert. Ausgehend vom Begriffssystem, welches die Gemeinsamkeiten der Standards erfasst, sollen top-down bis zu einer konkreten Nachrichtenstruktur, entsprechend des vorgestellten Leitfadens, die Strukturen analysiert werden, um geeignete Abbildungen für Transformationen zu finden. Die entsprechenden Transformationen können auf generischen Strukturen und dann auf konkreten Strukturen definiert werden. Die einheitliche Spezifikation der Standards sowie die Begriffssysteme dienen als Leitfaden zur Strukturanalyse.

7.4.1 Top-Down-Integration

Abbildung 7.14 zeigt graphisch wie ausgehend vom Begriffssystem für Kommunikation, das die gemeinsame Struktur der Kommunikationsstandards wiedergibt, bis zur generischen Struktur der Standards HL7 und BDT die Strukturanalyse für spätere Abbildungen durchgeführt werden kann. Das Begriffssystem ist dabei in der an die MOF-Architektur angelehnten Architektur auf der UML-Metamodell-Ebene angesiedelt. Die Instanzen davon sind auf der Ebene der UML-Modelle zu finden, die die generische Nachrichtenstruktur spezifizieren. Eine konkrete Nachrichtenstruktur, wie beispielsweise eine HL7 Tumormeldung, ist eine Instanz des UML-Modells für HL7 (siehe Abbildung 6.2 auf Seite 54).

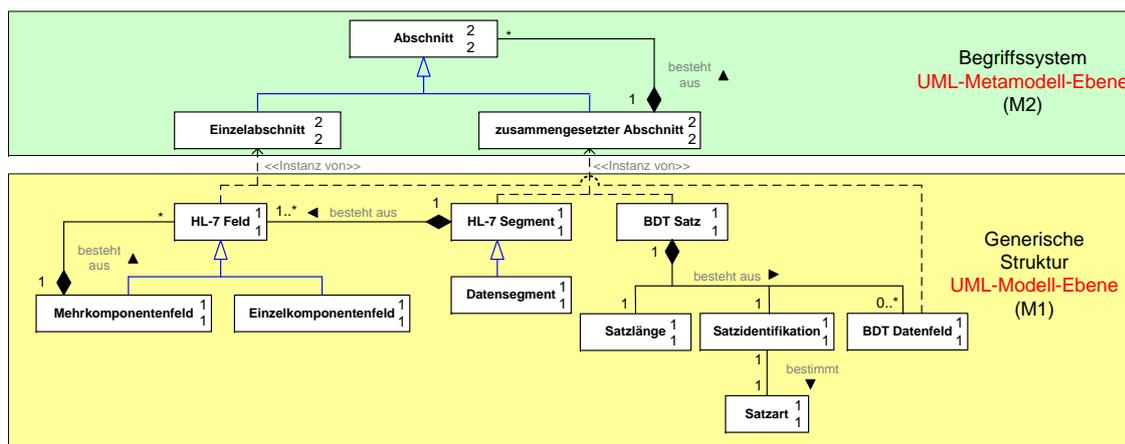


Abbildung 7.14: Strukturanalyse top-down vom Begriffssystem für Kommunikation zur generischen Struktur von HL7 und BDT

7.4.2 Transformationsregeln

Abbildung 7.15 soll verdeutlichen, dass bereits auf der allgemeinen Ebene der UML-Modelle für die generische Struktur der Standards Transformationsregeln gefunden werden können, die dann natürlich entsprechende Konsequenzen für die konkreten Strukturen haben. Dies ist in Abbildung 7.15 durch den Pfeil von „Transformationsregeln“ zu „Transformationsregeln generiert“ dargestellt. Die UML-Metamodell-Ebene entspricht der M2-Ebene, die UML-Modell-Ebene entspricht der M1-Ebene, die konkreten Nachrichtenstrukturen der Standards HL7 und BDT, hier z.B. als DTDs, befinden sich auf der M0-Ebene.

So sind zum einen Transformationsregeln zwischen den UML-Modellen für HL7 und BDT zu spezifizieren, zum anderen zwischen den konkreten Nachrichten-DTDs beider Standards für eine bestimmte Anwendung. Eine Transformationsregel auf UML-Modell-Ebene erfasst die Abbildungen, die zwischen den generischen Strukturen der verschiedenen Standards definiert werden können, beispielsweise HL7-Feld und BDT-Satz. Abbildung 7.16 zeigt diese Abbildungsmöglichkeiten mit einer UML-basierten Abbildungssprache [Wee03] für das Beispiel HL7 und BDT. Die Transformationsregeln zwischen den generischen Strukturen der Standards können, einmal definiert, immer wieder verwendet werden. Diese Abbildung ist Ergebnis der Strukturanalyse nach dem vorgestellten Leitfaden, ausgehend vom UML-Metamodell „Begriffssystem für Kommunikationsstandards“.

Zusätzlich zu den Transformationsregeln, die aus den Regeln aus der M2-Ebene generiert werden können, kommen Transformationsregeln hinzu, die manuell erstellt zwischen den konkreten DTDs die möglichen Transformationen komplettieren. Die Transformationsregeln zwischen den DTDs spezifizieren detailliertere Abbildungen je nach Anwendungsgebiet. Im Evaluationsbeispiel des Epidemiologischen Krebsregisters Niedersachsen (EKN) wird es die Basisdokumentation für Tumorkranke sein.

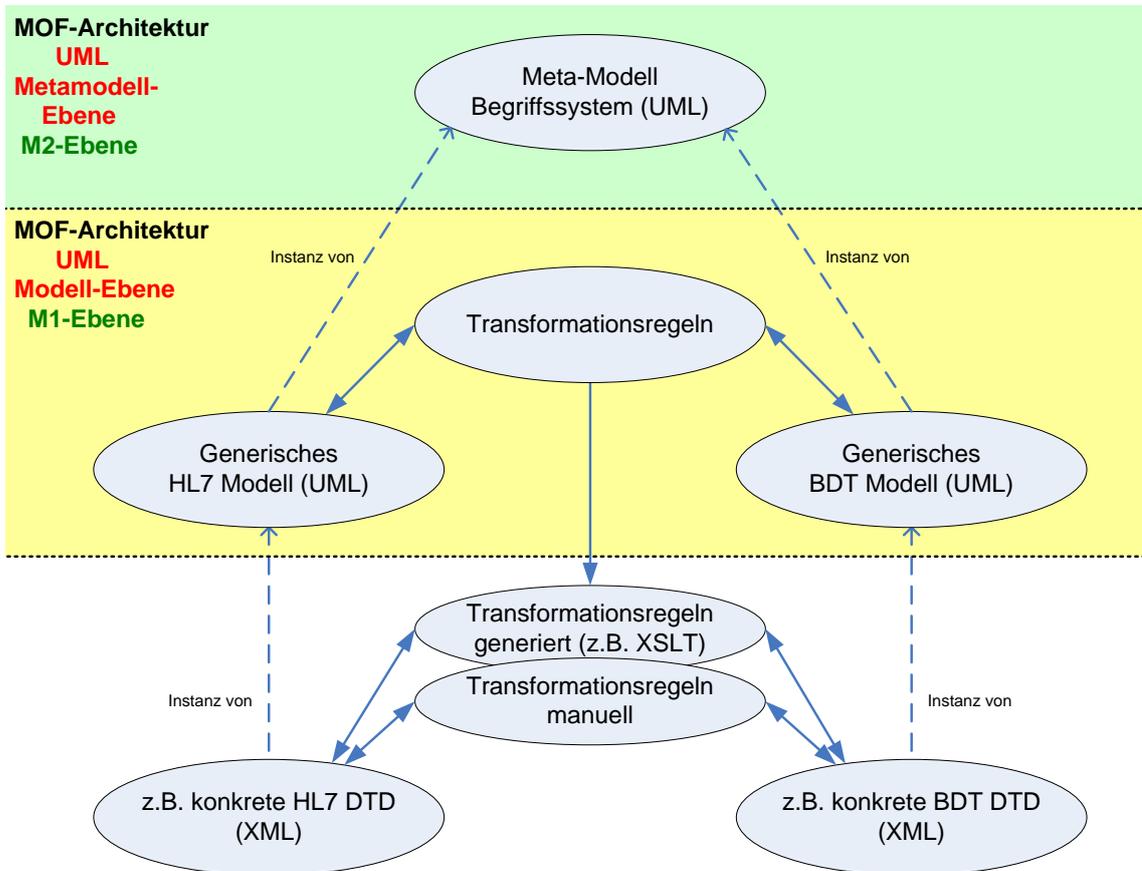


Abbildung 7.15: Transformationen bezogen auf die Ebenen, angepasst auf die analysierte Domäne Medizin

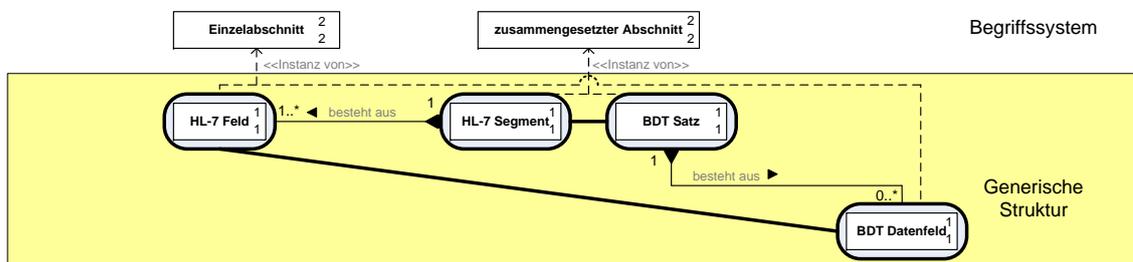


Abbildung 7.16: Abbildung von Klassen aus der generischen Struktur mit einer UML-basierten Abbildungssprache

Teil III

Evaluation am Beispiel des Epidemiologischen Krebsregisters Niedersachsen

8 Architektur für das Epidemiologische Krebsregister Niedersachsen

8.1 Problemstellung und Zielsetzung der Evaluation

Epidemiologische, das heißt bevölkerungsbezogene Krebsregister messen z.B. die Krebsinzidenz, analysieren Häufungen von Krebserkrankungen, errechnen Überlebensraten und mittleres Erkrankungsalter, liefern Aussagen über die Krebsprävalenz und sind Basis für Ursachen- und Risikostudien. Nur eine flächendeckende Registrierung mit über 90% erfasster Krebsfälle führt zu statistisch-epidemiologischen Auswertungen mit ausreichender Qualität.

Um alle benötigten Daten für ein epidemiologisches Krebsregister sammeln zu können, müssen verschiedene Institutionen einbezogen werden, die Daten über Krebserkrankungen erheben. Dies sind in Niedersachsen Nachsorgeleitstellen, klinische Register, Ärzte und Zahnärzte, Pathologen und Gesundheitsämter. Meldende Einrichtungen in einem anderen Sinne sind das Landesvermessungsamt, das Landesamt für Statistik und die Einwohnermeldeämter [RW98]. Die Heterogenität der meldenden Institutionen macht eine flexible elektronische Anbindung der Melder problematisch. Mehr zum Epidemiologischen Krebsregister Niedersachsen (EKN) befindet sich in Anhang D.

8.1.1 Problemstellung

Alle meldenden Institutionen eines epidemiologischen Krebsregisters können einen ganz unterschiedlichen Stand ihrer EDV aufweisen sowie entsprechend ihrer jeweiligen Aufgabengebiete völlig unterschiedlich aufgebaute Softwaresysteme haben. Aufgrund dieser Heterogenität macht die Anbindung der Melder an die Epidemiologischen Krebsregister Probleme. Mühevoll muss analysiert werden, welche Gegebenheiten beim Melder vorliegen und viele individuelle Lösungen zur Anbindung wurden geschaffen. In Abschnitt 6.2 werden verschiedene Kommunikationsstandards vorgestellt, die verschiedene Meldergruppen einsetzen bzw. sich für diese Meldergruppe anbieten. Abschnitt 6.3 weist auf verschiedene Dokumentationsstandards hin. Die Standards, wie auch Metadaten für sie, wurden einheitlich in UML spezifiziert, um einen Leitfaden für eine Transformation zwischen den jeweiligen Standards zu besitzen. Ausgehend von diesen Standards sollen top-down die lokalen Gegebenheiten integriert werden.

8.1.2 Zielsetzung

Es sollte zunächst eine Untersuchung der Schnittstellen zwischen den heterogenen Meldern und dem Krebsregister erfolgen. Entsprechend werden Standards und geeignete Kombinationen von Standards nach einem entwickelten Auswahlverfahren ermittelt und ausgewählt (Abschnitt 8.2). Anschließend müssen Transformationen gemäß der Architekturkonzeption in Abschnitt 7.2 definiert werden, die eine Interoperabilität zwischen ausgesuchten Meldergruppen und Krebsregister ermöglichen. Der Schwerpunkt liegt dabei auf dem Problem der semantischen Interoperabilität. Anhand eines Ausschnittes aus einem Beispielmeldeweg soll das zuvor vorgestellte Konzept prototypisch realisiert werden. Dabei liegt der Schwerpunkt auf den Transformationen zwischen den ausgewählten Standardkombinationen. Es soll geklärt werden, in welcher Weise die einheitliche Spezifikation der Standards und der Begriffssysteme für einen konkreten Anwendungsbereich Hilfestellung geben können. Aus datenschutzrechtlichen Gründen wird dabei selbstverständlich nur mit der Datenstruktur des EKN gearbeitet und nicht mit realen Daten des Krebsregisters.

8.2 Auswahl der geeigneten Standards

Bevor Standards für den Anwendungsbereich kombiniert werden können, müssen zuvor geeignete Standards ausgewählt werden. Ein Dimensionenschema wird helfen dieses Auswahlverfahren zu systematisieren. Anschließend werden nach dem in Abschnitt 6.4 auf Seite 67 vorgestellten Auswahlschema Kombinationen ausgewählt und beurteilt. Erst danach wird die vorgestellte Mediator-basierte Architektur angewendet.

8.2.1 Dimensionenschema

Tabelle 8.1 stellt ein Dimensionenschema dar, mittels dessen eine Auswahl der richtigen Standards gemäß [LS01a] vorgenommen werden kann, um das Procedere zu systematisieren. Tabelle 8.2 zeigt dabei die Einordnung der wichtigsten Standards in das Dimensionenschema. SNOMED ist ein international akzeptierter Standard für medizinische Terminologie, der von Menschen in Institutionen mit stationärer Versorgung verwendet wird. Der ICD wird ebenfalls in der ambulanten Versorgung eingesetzt. BDT dagegen ist ein Standard in Deutschland bei den Institutionen niedergelassene Arztpraxen und dient in der ambulanten Versorgung für das Austauschen von Nachrichten zwischen Maschinen. Der BDT-Standard wurde ergänzt.

DIMENSIONEN				
	Akzeptanz	Anwendungsbereich	Objekt	Nutzer
mögliche Werte	institutionell	institutionell	Nachrichtenstruktur	Mensch
	regional	regional	Nachrichtenintention	Maschine
	national	national	Nachrichtenrepräsentation	
	international	international	Prozedur Terminologie und Vokabular Autorisierung / Authentizität	

Tabelle 8.1: Dimensionen zur Einordnung von Standards im Gesundheitswesen nach [LS01a]

DIMENSIONEN					
	Akzeptanz	Anwendungsbereich	Objekt	Nutzer	Einschränkungen
SNOMED	international	institutionell	Terminologie	Mensch	stationäre Versorgung
HL7	international	institutionell	Nachrichten	Maschine	Krankenhausprozesse
DICOM	international	institutionell	Nachrichten	Maschine	Bilder
ICD	international	institutionell	Terminologie	Mensch	ambulante+stationäre Versorgung
BDT	national	institutionell	Nachrichten	Maschine	ambulante Versorgung

Tabelle 8.2: Beispiele von Standards im Gesundheitswesen nach [LS01a]

DIMENSIONEN				
Notwendige Akzeptanz	Anwendungsbereich	Objekt	Nutzer	Einschränkungen
national	institutionell	Terminologie	Mensch	ambulante+stationäre Versorgung von Tumorpatienten
		Nachrichten	Maschine	Datensammlung für statistische Auswertungen sowie für Forschung

Tabelle 8.3: Das EKN im Dimensionenschema nach [LS01a]

8.2.2 Einordnung des EKN in das Dimensionenschema

Gemäß den Ausführungen von [LS01a] haben wir eine Eingliederung des Anwendungsbereichs EKN in dieses Schema vorgenommen (siehe Tabelle 8.3). Die benötigten Standards im Bereich EKN müssen dabei mindestens national akzeptiert werden, daher spricht man hier nun von notwendiger Akzeptanz.

8.2.3 Auswahl der Standards für das EKN

Für diesen Anwendungsbereich wurden nun systematisch die folgenden Standards ausgewählt. Ausschlaggebend für die Auswahl waren in diesem Fall die Einschränkungen (siehe Hervorhebung):

1. Zur Kommunikation: HL7 und BDT
2. Zur Dokumentation:
 - Für Dokumente: Basisdokumentation für Tumorkranke
 - Als Begriffssystem: ICD-O [ICD02b]

8.2.4 Auswahl der Kombinationen der Standards für das EKN

Die heterogenen Dokumentationen, bezüglich des epidemiologischen Krebsregisters, der verschiedenen Meldergruppen sollen über einen Domänenstandard auf Top-Down-Weise integriert werden. Als Domänenstandard, der ein gemeinsames Verständnis der abzubildenden Daten bietet (Domain Information Model), kann man die Basisdokumentation für Tumorkranke der Deutschen Krebsgesellschaft e.V. und der Arbeitsgemeinschaft

Mögliche Kombinationen	Existenz der Abbildung	Einsatz der Kombination im EKN	Vollständigkeit der Abbildung	Bewertung der Vollständigkeit
HL7 / CDA	✓		✓	+++
HL7 / Basisdoku	✓		✓	++
BDT / CDA	✓		✓	+
BDT / Basisdoku	✓	✓	✓	+++

Tabelle 8.4: Auswahlschema zur Auswahl geeigneter Kombinationen von Kommunikations- und Dokumentationsstandards für den Anwendungsbereich EKN

Deutscher Tumorzentren ansehen. Bei der Beantwortung der Frage, welche Kombinationen dieser ausgewählten Standards für den Anwendungsbereich EKN sinnvoll sind, soll das Auswahlschema, welches Tabelle 8.4 zeigt, helfen. In der ersten Spalte werden alle möglichen Kombinationen von zuvor mit dem Dimensionenschema ausgewählten Kommunikations- und Dokumentationsstandards aufgelistet. Für bestehende Abbildungen bzw. bei engen Verwandtschaften zwischen den Standards wie bei HL7 und CDA, gibt es einen Haken in der Spalte Existenz der Abbildung. Die Spalte „Einsatz der Kombination im EKN“ bekommt einen Haken, wenn Meldungen in dieser Kombination bereits verarbeitet werden. Die Vollständigkeit der Abbildung wird mit einem Haken versehen, wenn die Abbildung vorgenommen wurde sowie geschätzt bewertet, wobei mit wachsender Anzahl von „+“ die Vollständigkeit der Abbildung steigt.

8.2.5 Beurteilung der ausgewählten Kombinationen von Standards

HL7 und CDA basieren auf demselben Referenzinformationsmodell, daher ist eine komplette Abbildung kein Problem. Eine Abbildung HL7 auf die Basisdokumentation gelingt zu einem großen Teil, aber nicht vollständig [Wil02]. Charakteristika des BDT-Standards wurden berücksichtigt bei der Abbildung auf die CDA in der Arbeitsgemeinschaft SCI-PHOX [HSD01, SCI02, GH03, Hei03]. Im EKN eingesetzt wird die vom ZI veröffentlichte Abbildung des BDT-Standards auf die Basisdokumentation [ZI 02]. Insgesamt sind danach die Kombinationen HL7/CDA und BDT/Basisdokumentation zu präferieren. Abschließend muss beurteilt werden, wie gut zwischen HL7/CDA und BDT/Basisdokumentation vermittelt werden kann, um den Grad der Funktionalität beurteilen zu können. Durch die Vermittlung durch den Metafacilitator bzw. Metakomponentenmediator besitzt die Architektur größtmögliche Flexibilität. Existieren Abbildungen nicht, so ist eine genauere Analyse der Standards nötig, um eine Antwort darauf zu geben, ob eine Abbildung sinnvoll ist oder nicht.

8.3 Mediator-basierte Architektur für das EKN

In diesem Kapitel soll die vorgestellte Mediator-basierte Architektur für die Domäne Gesundheitswesen eingesetzt werden für das konkrete Anwendungsbeispiel EKN. Abbildung 8.1 zeigt die Umsetzung der Architektur (Abbildung 7.2) angepasst auf das Beispiel des EKN. Ferner werden auch die angepasste Schemaarchitektur sowie das UML-Objektdiagramm der Architektur für das Anwendungsbeispiel vorgestellt. Die Domänenmodelle sind als DTDs erstellt worden bzw. sind zur Verfügung gestellt worden.

Die notwendigen Transformationen zwischen den ausgewählten Standards werden in Abschnitt 8.3.6 zunächst mit Techniken des Semantic Web angedeutet, in Kapitel 9 wird eine prototypische Implementierung vorgestellt.

8.3.1 Schemaarchitektur für das EKN

Während in Abbildung 8.1 die Sicht der dynamischen Datenflüsse dargestellt wird, zeigt Abbildung 8.2 die zugehörige Fünf-Ebenen-Schemaarchitektur [SL90], auf deren Basis die Mediatoren und Facilitatoren arbeiten und die die statischen Abhängigkeiten zwischen den Modellen zeigt. Das Auflösen von Konflikten zwischen Datenmodellen, welche unter Benutzung verschiedener Modellierungstechniken entstanden sind, erfordert Schemata in einem einheitlichen Datenmodell. Daher werden die lokalen Schemata in ein gemeinsames (kanonisches) Datenmodell transformiert.

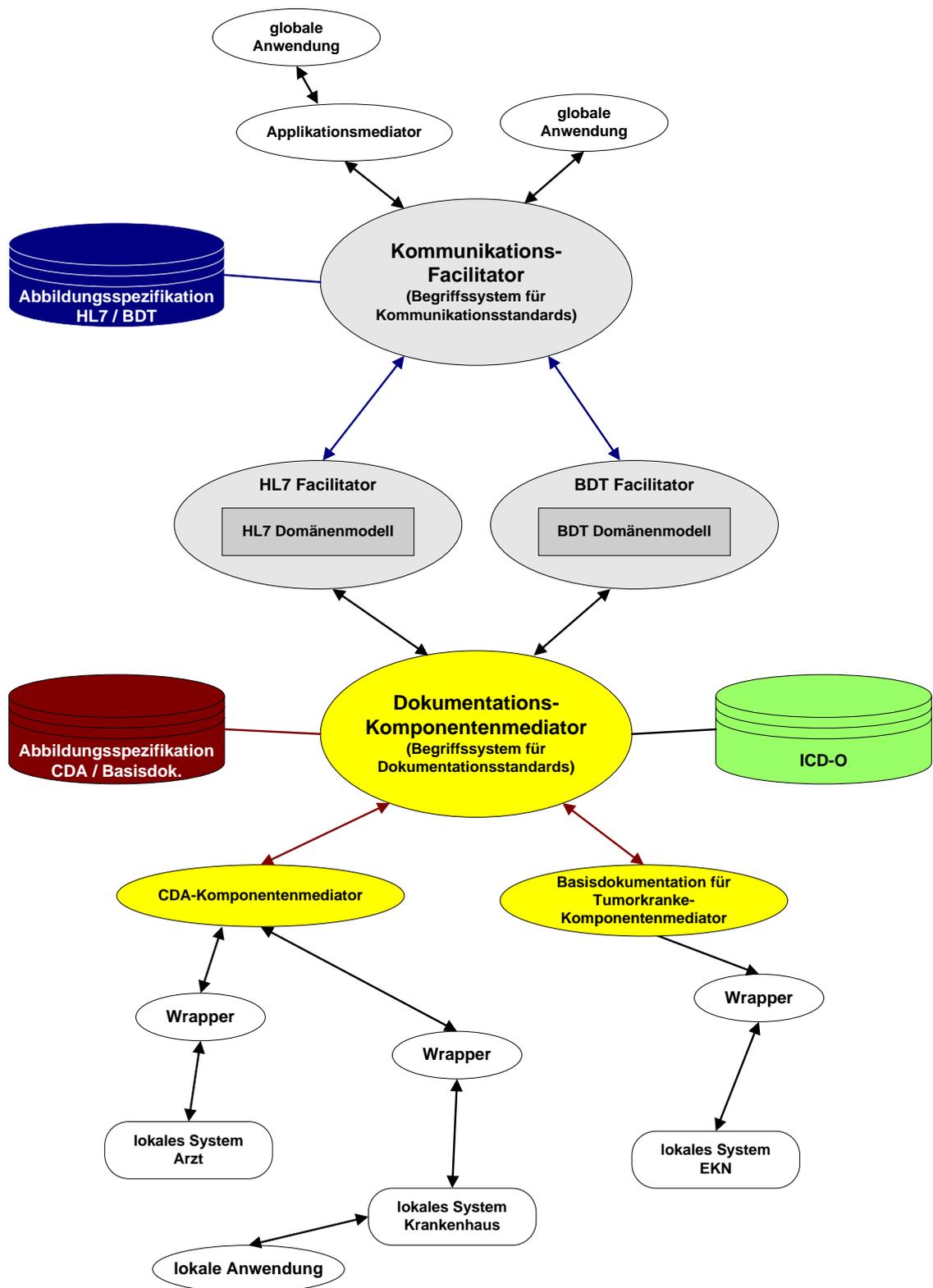


Abbildung 8.1: Evaluation der Architektur am Beispiel des EKN: Spezialisierung der allgemeinen Architektur in Abbildung 7.2 für die Anforderungen des EKN mit den Standards HL7, BDT und ICD-O.

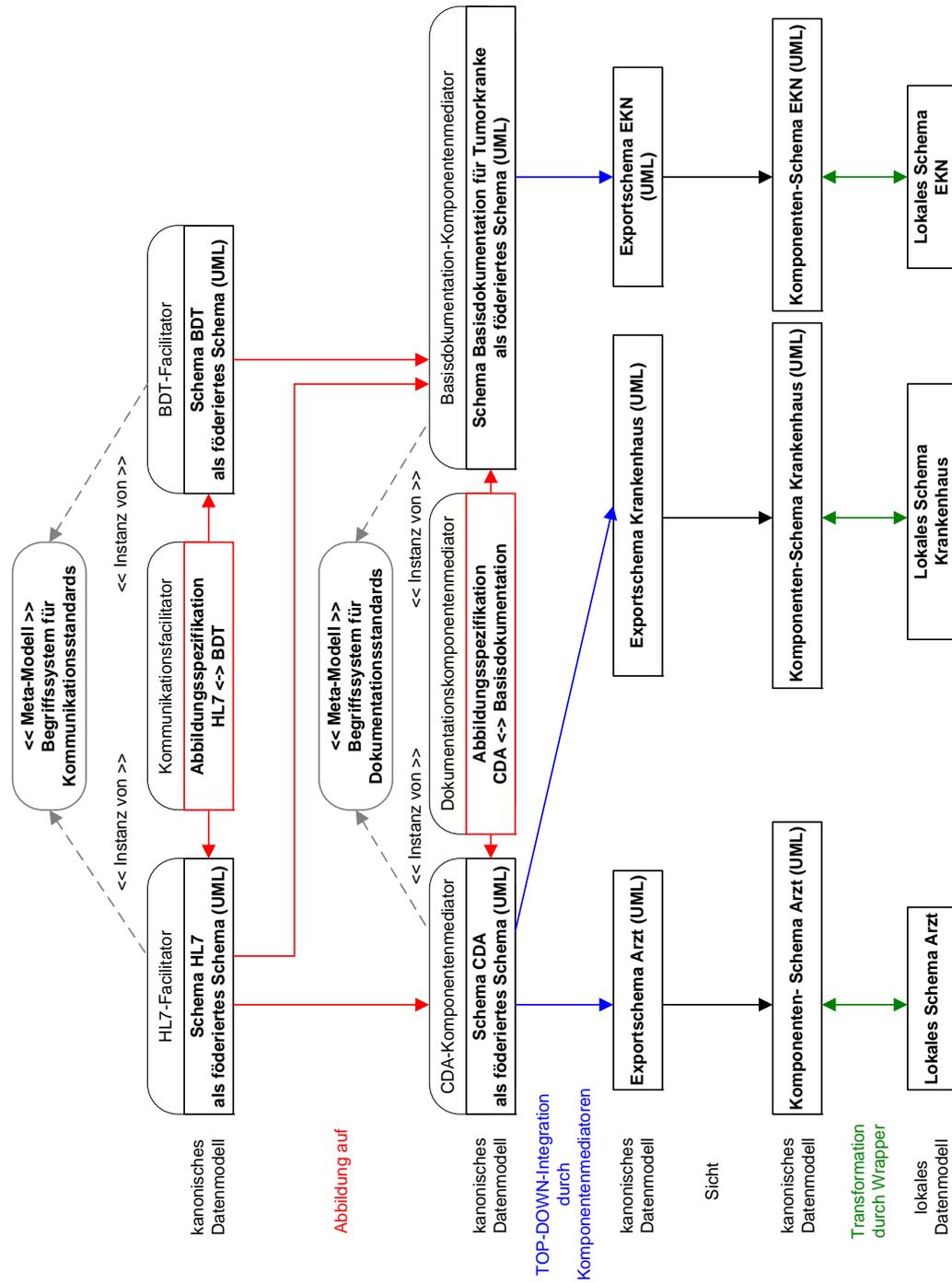


Abbildung 8.2: Schema-Architektur: Eine Sicht der statischen Abhängigkeiten zwischen den Modellen

8.3.2 UML-Objektdiagramm der Architektur für das EKN

Die entsprechende Instanz zum UML-Klassendiagramm der Architektur (Abbildung 7.3 auf Seite 81) zeigt das UML-Objektdiagramm in Abbildung 8.3. Die Abbildungen sind als XML-Transformationen realisiert. Damit dient das XML-Schema als kanonisches Datenmodell für die Integration [CNM02].

8.3.3 Facilitatoren

Die Facilitatoren verwalten die jeweiligen Domänenmodelle HL7 und BDT. Der Kommunikations-Facilitator koordiniert die Transformation zwischen HL7 und BDT über eine entsprechende Abbildungsspezifikation. Beim Entwurf dieser Spezifikation unterstützt das Begriffssystem für Kommunikationsstandards und die in UML spezifizierten Standards HL7 und BDT. Die Komponentenmediatoren verwalten die Abbildungen zu den lokalen Modellen. Der Dokumentations-Komponentenmediator verwaltet die Abbildungsspezifikation zwischen CDA und Basisdokumentation. Beim Entwurf dieser Spezifikation ist das Begriffssystem für Dokumentationsstandards hilfreich.

8.3.4 Domänenmodelle

Die Domänenmodelle kommen, falls sie noch nicht existieren, durch Abbildungen zwischen Kommunikations- und Dokumentationsstandards gemäß den zuvor ausgewählten Kombinationen zustande. So muss für HL7 Version 3 das generische HL7-RIM zunächst auf den speziellen Anwendungsbereich abgebildet werden, um das entsprechende Domänenmodell zu erhalten, bei den Versionen HL7 2.x müssen Z-Segmente hinzugezogen werden. Die Abbildung des HL7-RIM auf die Basisdokumentation für Tumorkranke wurde in [Wil02] prototypisch realisiert. Das BDT-Domänenmodell für die Basisdokumentation für Tumorkranke existiert bereits.

HL7-DTD

Die HL7-DTD für Version 2.3.1 ist erstellt worden, um eine zeitnahe DTD zur Verfügung zu haben, mit der eine Transformation durchgeführt werden kann. Aus [Hae99] wurden dabei zur Erfassung der gesamten Tumordokumentation die Z-Segmente verwendet. Als Editor wird Altova XMLSPY Enterprise Edition version 2004 rel. 4 [XML04] eingesetzt.

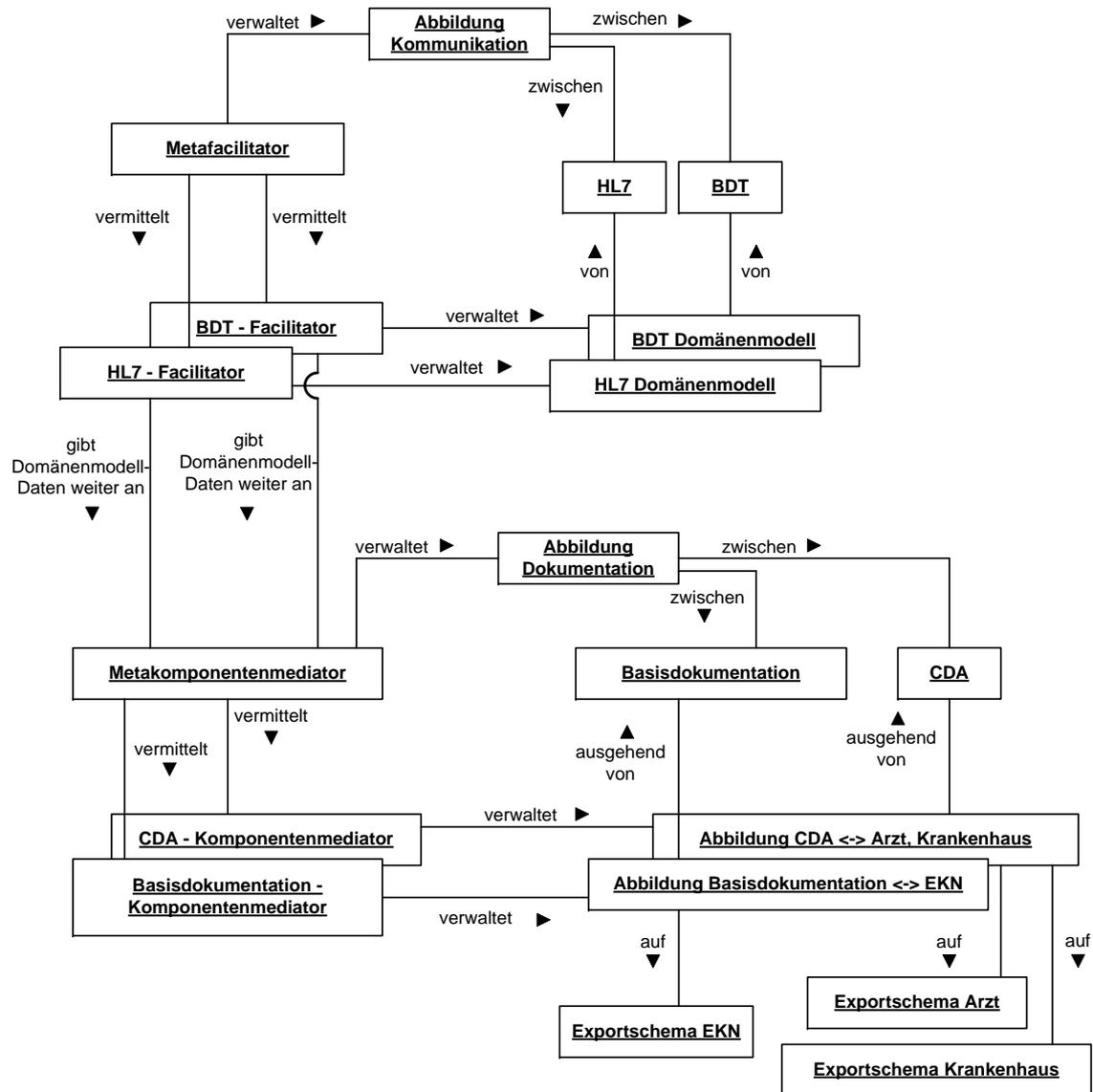


Abbildung 8.3: UML-Objektdiagramm als Instanz des UML-Klassendiagramms aus Abbildung 7.3 auf Seite 81 für das Fallbeispiel EKN

```
<?xml version="1.0" encoding="UTF-8"?>

<!-- edited with XMLSPYv2004 rel. 4 U (http://www.xmlspy.com) by
Susanne Pedersen (University of Oldenburg) -->

<!ELEMENT ENTER_NAME_OF_ROOT_ELEMENT_HERE EMPTY>

<!--HL7 V2.3.1 XML DTD for message TUMELD-->

<!-- Datatype declarations -->
<!ENTITY % HL7V231-datatypes PUBLIC
"-//HL7//DTD HL7 V2.3.1 datatype definitions//EN"
"datatypstumeld.dtd"> %HL7V231-datatypes;

<!-- .LST fuer wiederholbare Elemente -->

<!-- Message declarations-->
<!ENTITY % TUMELD.CONTENT " (MSH, EVN, PID, ZAI, ZPZ, ZST, ZAN, ZLO, ZHI, ZSM,
ZTO, ZTB, ZTI, ZAB, ZME, ZVM, ZNW, ZKO, ZMV, ZFE, OBR, OBX, DG1, NK1, IN1) ">
<!ELEMENT TUMELD %TUMELD.CONTENT;>
```

Abbildung 8.4: Auszug aus HL7-DTD für Tumormeldungen

In Abbildung 8.3.4 ist ein Auszug aus der erstellten HL7-DTD für Tumormeldungen dargestellt. Die Datentyp-Deklarationen für HL7V231-datatypes befinden sich in einer gesonderten DTD mit dem Namen datatypstumeld.dtd. Welche Segmente zu einer Tumormeldung gehören, ist in den Message declarations am Ende definiert. Neben den Segmenten aus dem HL7-Standard wie MSH, EVN, PID, OBR, OBX, DG1, NK1 und IN1 sind sämtliche Z-Segmente aus [Hae99] ergänzt. Die ausführliche Spezifikation der HL7-DTD für Tumormeldungen befindet sich auf der Dissertations-CD (siehe Anhang E auf Seite 221).

Abbildung 8.5 zeigt die Abbildung top-down, ausgehend vom Begriffssystem für Kommunikation über die generische Struktur von HL7 bis hin zum konkreten Einsatz des Standards HL7 für die Basisdokumentation für Tumorkranke.

BDT-DTD

Eine DTD für die Basisdokumentation für Tumorkranke im Format des BDT-Standards ist freundlicherweise vom Epidemiologischen Krebsregister in Version 1.0 zur Verfügung gestellt worden. Auf der Dissertations-CD ist diese Datei ebenfalls zu finden (siehe Anhang E auf Seite 221).

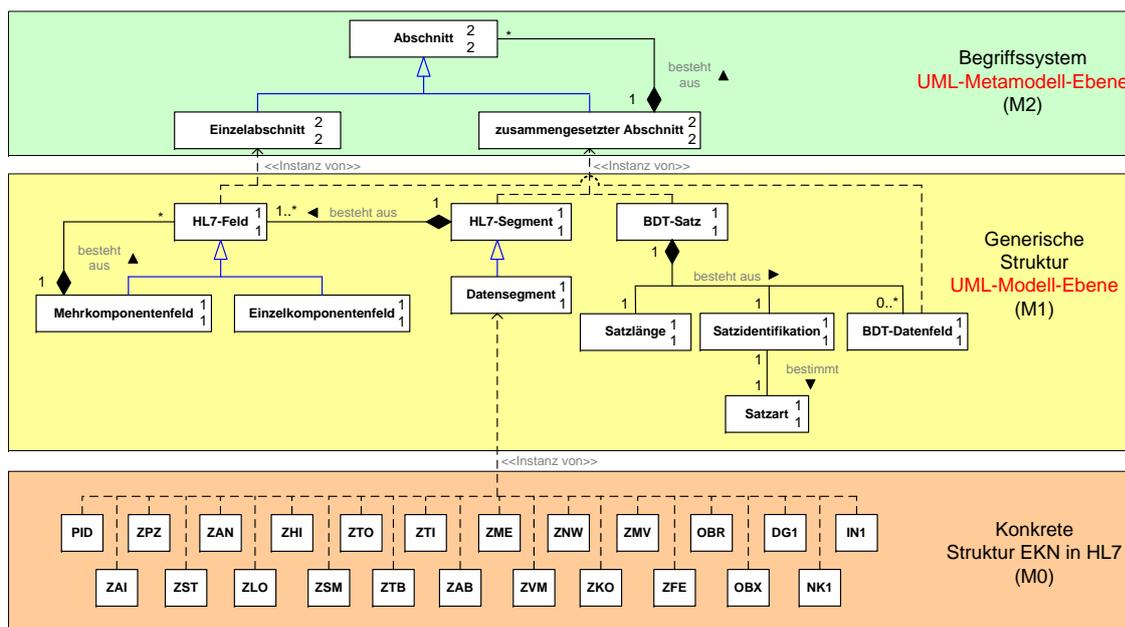


Abbildung 8.5: Top-Down vom Begriffssystem zum konkreten Einsatz des Standards HL7

Abbildung 8.6 veranschaulicht die Abbildung vom Begriffssystem für Kommunikation über die generische Struktur von BDT top-down bis hin zum konkreten Einsatz für die Tumordokumentation.

Abbildung Kommunikation

Ein Metafacilitator verwaltet die Abbildung Kommunikation zwischen den Standards HL7 und BDT und vermittelt so zwischen den beiden einzelnen Facilitatoren, die die HL7- und BDT-Domänenmodelle verwalten. Das Begriffssystem für Kommunikationsstandards ist bei der Erstellung der Abbildungsspezifikation hilfreich.

8.3.5 Komponentenmediatoren

Der CDA-Komponentenmediator verwaltet die Abbildung zwischen der CDA und dem Exportschema des Informationssystems des Arztes oder des Krankenhauses. Der Basisdokumentationskomponentenmediator verwaltet dagegen die Abbildung der Basisdokumentation für Tumorkranke auf das Exportschema des EKN.

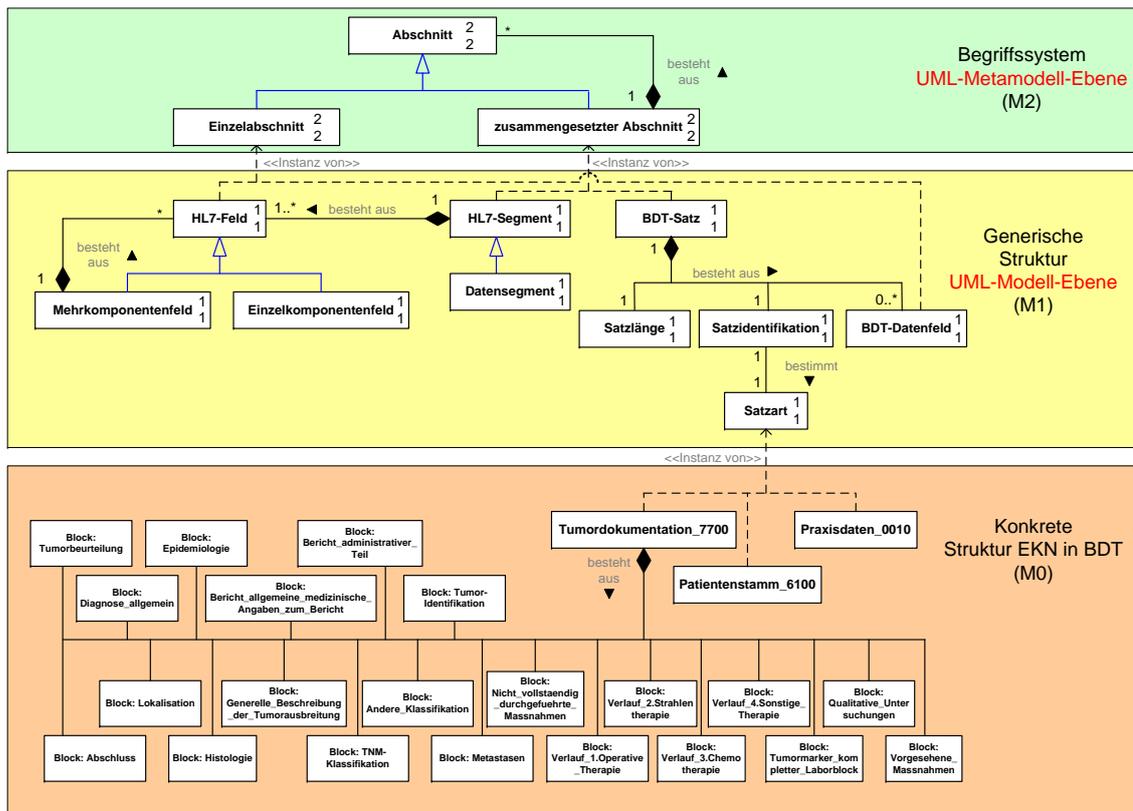


Abbildung 8.6: Top-Down vom Begriffssystem zum konkreten Einsatz des Standards BDT

```
:Datenfeld a rdfs:Class.  
  
:Patientenkennung a rdfs:Class; rdfs:subClassOf :Datenfeld.  
  
:Behandlungsdatenträger a rdfs:Class.  
  
:speichert a rdfs:Class.  
  
:Datenpaket a rdfs:Class.  
  
:Satz a rdfs:Class.
```

Abbildung 8.7: Klassendefinitionen für BDT in N3

Abbildung Dokumentation Für die Evaluation wichtig ist jetzt der Metakomponentenmediator, der zwischen den erwähnten Komponentenmediatoren in diesem Szenario vermitteln soll. Dies soll er mit Hilfe einer Abbildung Dokumentation erreichen, die zwischen Basisdokumentation für Tumorkranke und der CDA (konkretisiert für das Anwendungsszenario) abbildet. Dies soll durch das Begriffssystem für Dokumentationsstandards erleichtert werden.

8.3.6 Einsatz von RDF und RDF-Schema

Für die Evaluation ist der Einsatz von Techniken des Semantic Web eine Möglichkeit, die in UML modellierten Metadaten der Standards und Begriffssysteme zu spezifizieren. Im Folgenden werden die erstellten Spezifikationen zu den Standards und den Begriffssystemen in Auszügen vorgestellt. Bewusst geht es hier um eine prinzipielle Umsetzung, die nur eine Vorstellung von der Machbarkeit geben soll. Für eine konkrete Umsetzung wäre sehr viel mehr Detailaufwand notwendig.

Die Metadaten können mittels RDF-Tripel ausgedrückt und in RDF/XML [RDF04d] spezifiziert werden (siehe Abschnitt B.1). Das konzeptuelle Modell von RDF ist ein Graph. RDF bietet eine XML-Syntax, um RDF-Graphen niederzuschreiben und auszutauschen, genannt RDF/XML. Ein RDF-Schema (RDFS) dient als Vokabular. Wegen der einfacheren Lesbarkeit wird zunächst N3 als Notation genutzt (siehe Abschnitt B.2).

Definitionen der Klassen mit RDFS Jede Klasse der UML-Diagramme „Begriffssystem für Kommunikationsstandards“ und „Begriffssystem für Dokumentationsstandards“ oder z.B. die Struktur von BDT können auf der Basis des entsprechenden UML-Diagramms (Abbildung 6.1 auf Seite 50) als RDFS-Klassen definiert werden (siehe Abschnitt B.3).

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <rdfs:Class rdf:ID="Datenfeld">
    <rdfs:comment>Klasse der Datenfelder</rdfs:comment>
  </rdfs:Class>

  <rdfs:Class rdf:ID="Patientenkennung">
    <rdfs:comment>Klasse der Patientenkennungen</rdfs:comment>
    <rdfs:subClassOf rdf:Resource="http://somesite.org/terms/Datenfeld">
      </rdfs:subClassOf>
    </rdfs:Class>

  <rdfs:Class rdf:ID="Behandlungsdatentraeger">
    <rdfs:comment>Klasse</rdfs:comment>
  </rdfs:Class>
```

Abbildung 8.8: Klassendefinition von BDT in RDF/XML

Abbildung 8.3.6 zeigt einen Auszug aus der Klassendefinition für BDT in N3. Die ausführliche Spezifikation findet sich auf der Dissertations-CD (siehe Anhang E auf Seite 221). Beispielhaft wird in Abbildung 8.3.6 ein Auszug aus den Klassendefinitionen von BDT in RDF/XML gezeigt.

Diese Klassendefinitionen dienen als Vokabular für die Bezeichnungen Datenfeld, Patientenkennung und Behandlungsdatenträger. Die Patientenkennung ist dabei eine Unterklasse der Klasse Datenfeld. Damit ist in RDFS ausgedrückt, dass Patientenkennung ein Datenfeld im BDT-Standard ist. Als URI der Ressource wurde die Dummy-Website <http://somesite.org/terms/> gewählt, da die Spezifikation, wie bereits erwähnt, nur prinzipiell erläutert werden soll.

So können nun für alle erstellten UML-Diagramme der Standards bzw. der Begriffssysteme entsprechende Klassen und damit ein Vokabular definiert werden.

RDF-Statements Alles was über eine Aussage: „Ist Oberklasse von“ hinausgeht, muss mit RDF-Aussagen, sogenannten Statements, ausgedrückt werden. Abbildung 8.3.6 zeigt einen Auszug aus den RDF-Statements für BDT als Beispiel in N3. Anschließend finden sich in Abbildung 8.3.6 dieselben Statements in RDF/XML.

:Behandlungsdatenträger :speichert :Datenpaket.

:Datenpaket :besteht_aus :Satz.

:Satz :wird_zugeordnet_ueber :Patientenkennung.

:Satz :besteht_aus :Satzidentifikation.

:Satz :besteht_aus :Satzlänge.

:Satz :besteht_aus :Datenfeld.

Abbildung 8.9: RDF-Statements für BDT in N3

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:ex="http://somesite.org/terms/">

<rdf:Description
  rdf:about="http://somesite.org/terms/Behandlungsdatenträger">
  <ex:speichert rdf:Resource="http://somesite.org/terms/Datenpaket">
  </ex:speichert>
</rdf:Description>

<rdf:Description rdf:about="http://somesite.org/terms/Datenpaket">
  <ex:besteht_aus rdf:Resource="http://somesite.org/terms/Satz">
  </ex:besteht_aus>
</rdf:Description>

<rdf:Description rdf:about="http://somesite.org/terms/Satz">
  <ex:wird_zugeordnet_ueber rdf:Resource="http://somesite.org/terms/
  Patientenkennung"></ex:wird_zugeordnet_ueber>
  <ex:besteht_aus rdf:Resource="http://somesite.org/terms/Satz-
  identifikation">
  </ex:besteht_aus>
  <ex:besteht_aus rdf:Resource="http://somesite.org/terms/Satzlaenge">
  </ex:besteht_aus>
  <ex:besteht_aus rdf:Resource="http://somesite.org/terms/Datenfeld">
  </ex:besteht_aus>
</rdf:Description>
```

Abbildung 8.10: RDF-Statements für BDT in RDF/XML

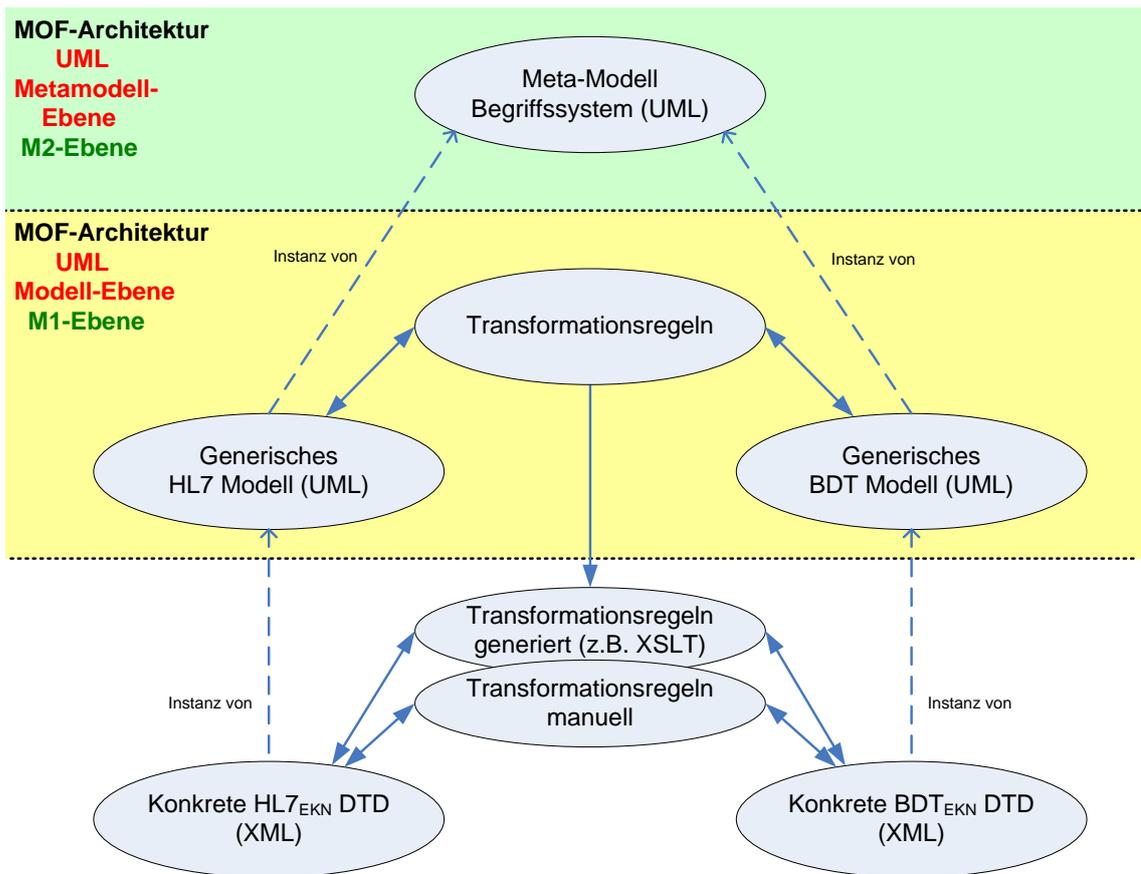


Abbildung 8.11: Transformationen bezogen auf die Ebenen für das Anwendungsbeispiel EKN

Einsatz von OWL

Um Aussagen spezifizieren zu können, die über die Möglichkeiten von RDFS hinausgehen, setzen wir zusätzlich die Ontologierepräsentationssprache OWL ein (siehe Abschnitt B.4). Klassenäquivalenzen sind Ausgangspunkt für Transformationsregeln. Abbildung 8.11 ist die bereits übertragene Abbildung 7.15 (auf Seite 97) angepasst auf das Evaluationsbeispiel EKN.

Äquivalenzen ergeben sich aus den UML-Diagrammen der Standards, hier also von BDT (Abbildung 6.1 auf Seite 50) und HL7 (Abbildung 6.2 auf Seite 54) sowie aus dem UML-Diagramm des Begriffssystems für Kommunikationsstandards (Abbildung 7.7 auf Seite 86). Die abgeleiteten Modellkorrespondenzen, aus denen die Klassenäquivalenzen hervorgehen können, zeigt Abbildung 8.12. Diese Modellkorrespondenzen sind ein Ergebnis der Nutzung des in dieser Arbeit vorgestellten Leitfadens. Die einheitliche Spezifikation von HL7, BDT und des Begriffssystems für Kommunikationsstandards in UML ermöglicht das schnelle Erfassen der Kommunikationsstrukturen und ihrer Zusammenhänge.

Modellkorrespondenzen HL7 und BDT		
HL7 Element	entspricht	BDT Element
Use Case	↔	---
Trigger Event	↔	---
Nachricht	↔	Datenpaket
Nachrichtentyp	↔	---
3-stelliger Code	↔	---
Segment	↔	Satz
Segment-Tabelle	↔	Satztabelle/ Satzbeschreibung
Segmentbezeichnung	↔	Satzidentifikation
Segmentseparator	↔	Prinzip der Satzlänge
Encoding Rule	↔	---
Anordnung	↔	Anordnung (Satztabelle)
Kontroll-Segment	↔	Satzart
Daten-Segment	↔	Satzart
Feldseparator	↔	Feld-Ende-Markierung und Feldlänge
Message Header Segment	↔	Satzart
Event Description	↔	---
Patient Identification	↔	Patientenkennung
Feld	↔	Datenfeld
Felddefinition	↔	Feldidentifikation
Mehrkomponentenfeld	↔	---
Datentyp	↔	Feldtyp
Vocabulary	↔	Feldtabelle
Uncodierter Typ	↔	---
Codierter Typ	↔	---
Codetabelle	↔	---
Im Standard festgeschriebene Tabelle	↔	---
Benutzerdefinierte Tabelle	↔	---
Codierungstabelle	↔	---
Anderer Standard	↔	---
---	↔	Behandlungsdatenträger
---	↔	Zulässiges Datenfeld
---	↔	Feldinhalt
---	↔	Regeltabelle

Abbildung 8.12: Modellkorrespondenzen der UML-Diagramme von HL7 und BDT

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <owl:Class rdf:ID="HL7_Nachricht">
    <owl:equivalentClass rdf:Resource="BDT_Datenpaket">
    </owl:equivalentClass>
  </owl:Class>

  <owl:Class rdf:ID="HL7_Segment">
    <owl:equivalentClass rdf:Resource="BDT_Satz">
    </owl:equivalentClass>
  </owl:Class>
```

Abbildung 8.13: Prinzipielle Klassenäquivalenzen von HL7- und BDT-Klassen in OWL

Ein Beispiel in OWL für die prinzipiellen Klassenäquivalenzen von HL7- und BDT-Klassen zeigt Abbildung 8.3.6. Für die entsprechenden Klassen aus den Standards HL7 und BDT sind jeweils HL7 und BDT den Klassennamen vorangestellt worden, um eine bessere Lesbarkeit zu erzielen. Die ausführlichere Spezifikation der Klassenäquivalenzen von HL7 und BDT sowie für die Dokumentation befindet sich ebenfalls auf der Dissertations-CD (siehe Anhang E auf Seite 221).

Zusammenfassung zur Spezifikation mit RDF, RDFS und OWL

Auf diese Weise sind grundsätzlich alle erstellten UML-Diagramme in RDF bzw. RDFS spezifizierbar. Natürlich muss bei der Definition des Vokabulars ermittelt werden, welche Definitionen schon vorliegen (z.B. Dublin Core [Dub04]). Außerdem müssen etliche Transformationen im Detail eingesetzt werden, um die Spezifikationen konkret einsetzen zu können, beispielsweise Transformationen von Klartexten in Codes und ähnliches bei konkreten Nachrichten, wie beispielsweise Tumormeldungen. Prinzipiell lassen sich aber die erstellten UML-Diagramme des Leitfadens sowie die abgeleiteten Modellkorrespondenzen mit Semantic Web-Technologien umsetzen.

Für die prototypische Umsetzung wurde die gegenwärtig aktuelle Technik des Kommunikationsservers gewählt, die gerade in der Domäne Medizin vielfach eingesetzt wird. Die Ergebnisse aus der rudimentären Semantic Web-Spezifikation sind bei der Realisierung dieses Prototyps jedoch schon hilfreich.

9 Prototypische Implementierung

Die Transformationen zwischen den ausgewählten Kommunikations- und Dokumentationsstandards wurden prototypisch zur Evaluation der Konzepte implementiert. Da die CDA nicht ausreichend definiert ist, um eine Transformation zur Basisdokumentation für Tumorkranke und umgekehrt vorzunehmen, wird auf das der CDA ähnliche OnkoDok-Modell zurückgegriffen. Aus den entsprechenden Semantic Web-Spezifikationen ist aber bereits jetzt klar, dass eine Abbildung nicht annähernd so vollständig erfolgen kann, wie dies auf den ersten Blick zu vermuten wäre. Die in [Wol02] erwähnte Tumorbasisdokumentation, die aus dem OnkoDok-Modell einfach zu extrahieren sei, folgt in der Definition nicht primär der Basisdokumentation für Tumorkranke. Dies ist insofern verständlich, da sich das OnkoDok-Modell vor allem für die Dokumentation in den Kliniken eignet. Die Dokumentation für epidemiologische Krebsregister ist dagegen ein auf diesen Zweck ausgerichtetes Konzentrat aus einer Reihe von Dokumentationen verschiedener Institutionen. Im Falle der hier vorgestellten Evaluation soll das Entwurfsmuster Messaging (siehe Abschnitt 3.5) zum Einsatz kommen, implementiert auf der Integrationsplattform e*Gate der Firma SeeBeyond.

9.1 Integrationsplattform SeeBeyond e*Gate

e*Gate ist eine Weiterentwicklung des klassischen Kommunikationsservers DataGate [LPH99], verbunden mit dem Paradigmenwechsel von einem anwendungsorientierten zu einem unternehmensorientierten Kommunikationsmodell, auf der Basis einer Service-Orientierten Architektur (SOA) [See04, CHKT05]. e*Gate ist im Gegensatz zur traditionellen „Hub and Spoke“-Architektur von Kommunikationsservern eine verteilte, skalierbare Integrationsplattform. Prozesse können mit e*Gate nach Bedarf auch über verschiedene Rechner verteilt und trotzdem zentral administriert werden. Auch spezielle Plattformen bzw. Betriebssysteme können über System-APIs eingebunden werden. e*Gate bietet die Möglichkeit Datenformate, Mapping- und Transformationsregeln komplett in JAVA zu definieren. Es verfügt aber auch über eine interne Scriptsprache MONK, die von LISP abgeleitet ist. Ebenfalls sind C/C++ oder XSLT verwendbar. Bei der Evaluation ist die e*Gate 5.0.4 SRE (Schema-Runtime-Environment)-Version zum Einsatz gekommen.

9.1.1 e*Gate Integrator

Das e*Gate Integrator System (e*Gate) basiert auf einer verteilten und offenen Architektur, die es Komponenten ermöglicht, auf verschiedenen Workstations innerhalb eines globalen Netzwerks zu verbleiben. Basierend auf den Kommunikationsprotokollen und Adaptoren, die ausgewählt werden, kann e*Gate quer über eine Vielzahl unterschiedlicher Betriebssysteme mit zahlreichen Applikationen kommunizieren bzw. diese verbinden.

Schema Organisation

Die e*Gate Systemkomponenten sind in Schemata organisiert. Ein Schema ist dabei ein Konfigurationsformat, das alle Module und Konfigurationsparameter enthält, die den Datenfluss durch das e*Gate-System kontrollieren, routen und transformieren. Ein Schema unterhält außerdem die Beziehungen zwischen den Komponenten, einschließlich der Publish/Subscribe-Information.

Komponenten und Subkomponenten

Für die Evaluation wesentliche Komponenten sind die Graphical user interfaces (GUI) e*Gate Schema Designer, SeeBeyond Java Collaboration Rules Editor und der SeeBeyond Java Event Type Definition (ETD) Editor.

e*Gate Schema Designer

Mit dem Schema Designer können die Komponenten des e*Gate-Systems erstellt und konfiguriert werden. Das Schema Designer Fenster ist geteilt in zwei Fenster. Das linke Fenster ist das Navigator-Fenster. Hier werden alle Systemkomponenten in einer Baumstruktur organisiert, ähnlich dem Windows Explorer. Das rechte Fenster ist das Editor-Fenster. Über links ausgewählte Komponenten werden Details oder die untergeordneten Komponenten angezeigt. Das Navigator-Fenster ermöglicht zwei Sichten: Komponenten-Editor- und Netzwerk-Sicht, wobei Abbildung 9.1 die Komponenten-Editor-Sicht zeigt. Der Schema Designer umfasst ferner verschiedene spezialisierte Editoren. Im Zusammenhang mit dieser Arbeit sind die Collaboration Rules Editoren von Interesse. Collaboration Rules definieren, wie Daten transformiert werden. Wenn die zugehörige Input-Output-Logik definiert wird, können so genannte Event Type Definitions (ETD) eingesetzt werden. Diese ETDs können durch entsprechende e*Gate-Systemwerkzeuge generiert werden oder mit einem ETD-Editor selbst erstellt werden. Den Collaboration Rules Editor wie auch den ETD-Editor gibt es jeweils in einer Java- und in einer MONK-Version. In dieser Arbeit wollen wir uns auf die Java-Versionen konzentrieren. Mit Hilfe von ETD builder Wizards, die der Java ETD Editor anbietet, können aus zahlreichen Quellen ETDs generiert werden, so z.B. aus XML-DTDs und XML-Schemata.

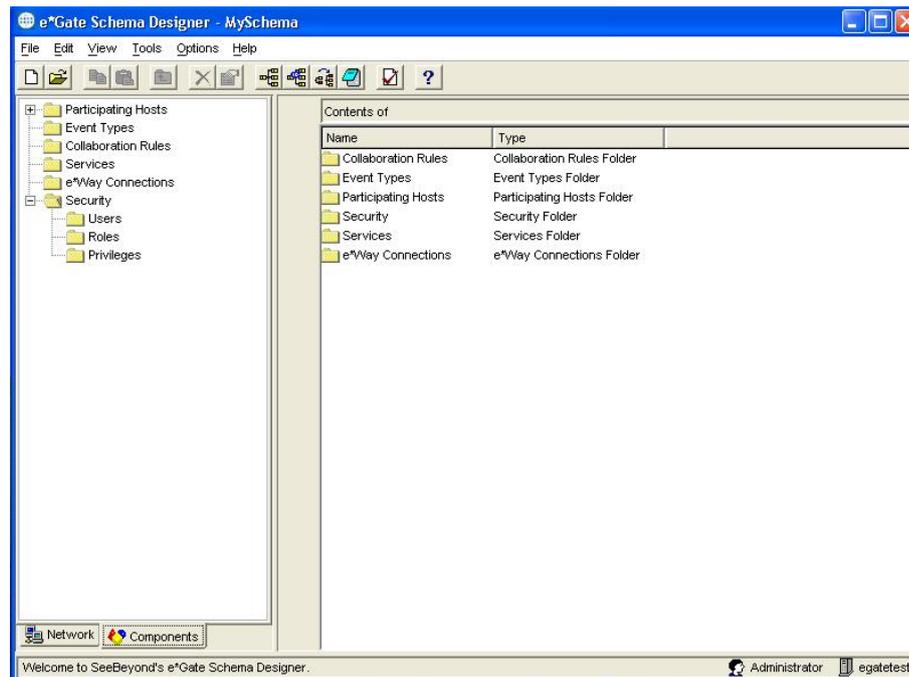


Abbildung 9.1: Komponenten-Editor-Sicht des e*Gate Schema Designers

9.1.2 MapForce 2005 Enterprise Edition

Die Extensible Stylesheet Language Transformationen (XSLT) [XSL99] wurden mit Altova MapForce 2005 Enterprise Edition [Map05] erstellt. [Wee03] stellt eine UML-basierte graphische Abbildungssprache vor, die von einem prototypisch implementierten Diagrammeditor zu einem Teil umgesetzt worden ist. Dieser Diagrammeditor ist geeignet, um Abbildungen in der graphischen Abbildungssprache zu erstellen und zu visualisieren. Der Schwerpunkt liegt allerdings bei Schemaabbildungen föderierter Datenbanksysteme, weswegen eine Unterstützung von XSLT nicht Ziel der Arbeit war. In Abschnitt 7.4.2 auf Seite 96 haben wir diese graphische Abbildungssprache eingesetzt, um die Modellkorrespondenzen vorzustellen, die sich aus dem Begriffssystem für Kommunikationsstandards und den UML-Modellen von HL7 und BDT ergeben haben. Für eine entsprechende prototypische Implementierung mit XSLT ist sie allerdings nicht geeignet.

MapForce ist ein XML- / Datenbank- / Flat File / EDI-Mappingtool für komplexe Integrationsprojekte. Dieses visuelle Datenmapping-Tool generiert automatisch Mapping-Code in XSLT 1.0/2.0, XQuery, Java, C++ und C#. Es können visuelle Mappings zwischen Inhaltsmodellen erstellt werden, indem Verbindungslinien zwischen übereinstimmenden Elementen in Quelle und Ziel gezogen und Datenverarbeitungsregeln eingefügt werden. MapForce 2005 generiert dann automatisch den Code, der benötigt wird, um Daten aus dem Quelldokument auf das Zieldokument abzubilden, um sie dann in serverseitigen Datenintegrationsapplikationen, wie beispielsweise e*Gate zu verwenden. Mit Hilfe des integrierten MapForce-Prozessors kann das Ergebnis jedes Mappings mit einem Mausklick in einer Vorschau angesehen werden.

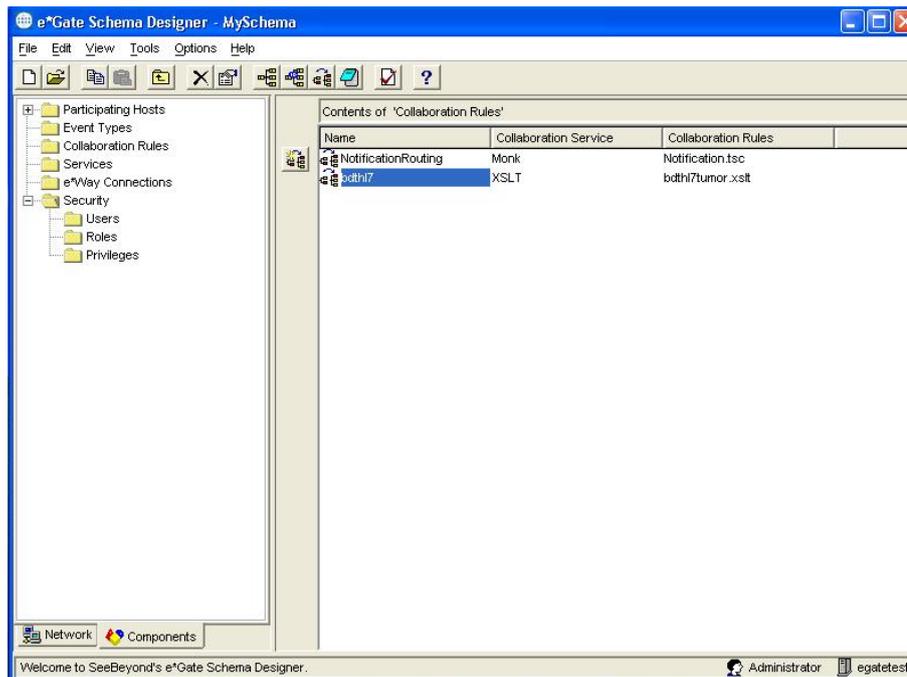


Abbildung 9.2: Übergebene XSL-Datei als Collaboration Rules in e*Gate

MapForce erlaubt eine graphische Verknüpfung zweier XML-Bäume und generiert daraus XSL-Transformationen. Dazu wurden jeweils die DTDs der zu transformierenden Standards mit dem Wurzelement geladen und dann graphisch verbunden. Die XSL-Dateien werden dem e*Gate Collaboration Service übergeben und stehen dann als Collaboration Rules zur Verfügung [See05]. Abbildung 9.2 zeigt die als Collaboration Rules übergebene XSL-Datei *bdth17tumor*. Abbildung 9.3 zeigt die dafür notwendigen Schritte, die so auch durchgeführt worden sind.

9.2 Abbildung Kommunikation

Im Folgenden soll aufgezeigt werden, wie zwischen den Kommunikationsstandards HL7 und BDT für das Beispiel der Basisdokumentation für Tumorkranke vermittelt werden kann. Als Domänenmodelle dienen entsprechende DTDs. Zunächst werden dazu die Strukturen der beiden DTDs analysiert und anhand des vorgestellten Leitfadens die groben Modellkorrespondenzen erstellt. Anschließend können auf einer Detailebene darunter feinere Korrespondenzen spezifiziert werden. Zuordnungen, die dann noch offen bleiben, müssen manuell nach genauerer Prüfung über die gesamte Struktur gesucht werden. Einige Zuordnungen können überhaupt nicht getroffen werden.

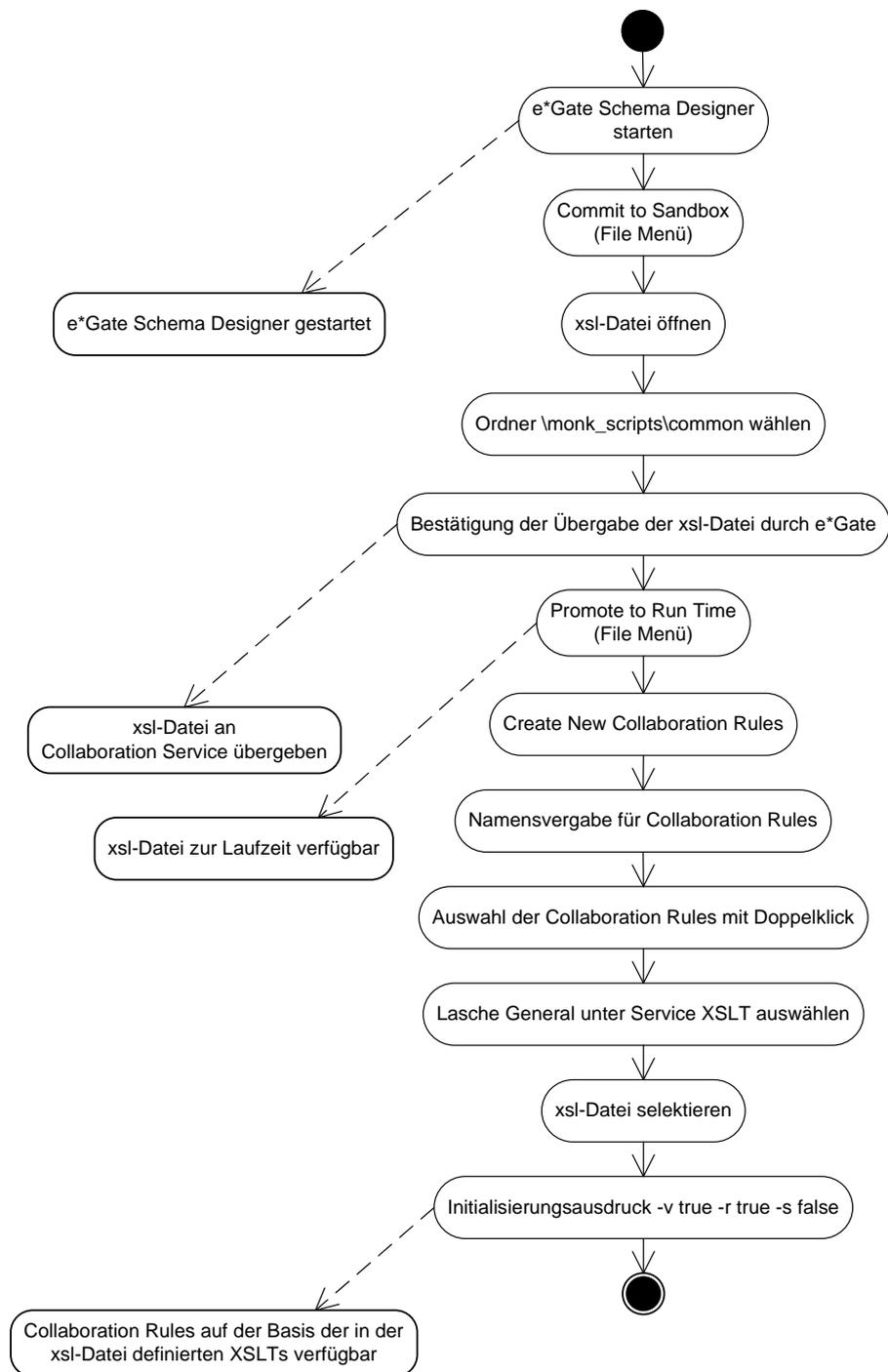


Abbildung 9.3: UML-Aktivitätsdiagramm zur Übergabe einer xsl-Datei an den e*Gate Collaboration Service

9.2.1 Modellkorrespondenzen von HL7-Segment und BDT-Satz

Die aus dem Leitfaden hervorgehenden Modellkorrespondenzen aus Abbildung 8.12 auf Seite 118 und die daraus mit OWL formulierten Klassenäquivalenzen (Abbildung 8.3.6 auf Seite 119) geben die ersten Abbildungen vor. Abbildung 9.4 zeigt die sich quasi automatisch ergebenden Abbildungen zwischen HL7-Segmenten und BDT-Sätzen für die Basisdokumentation für Tumorkranke.

9.2.2 Modellkorrespondenzen von HL7-Feld und BDT-Datenfeld

Nach dieser groben Zuordnung können innerhalb dieser HL7-Segmente und BDT-Blöcke die Zuordnungen der einzelnen HL7-Felder und BDT-Datenfelder zueinander erfolgen. Auch zwischen HL7-Feld und BDT-Datenfeld besteht nach dem Leitfaden eine Modellkorrespondenz. Beispielhaft zeigt Abbildung 9.5 die detaillierten Zuordnungen zwischen dem HL7-Segment `ZST-Status Tumorerkrankung` und dem BDT-Satz `Tumorbeurteilung`. Es gibt auch mehrere HL7-Segmente, die auf einen BDT-Satz abgebildet werden müssen, z.B. HL7-Segmente `PID (Patient Identification)` und `IN1 (Insurance)` auf den BDT-Satz `Patientenstamm`.

9.2.3 Übrige Abbildungen

Leider bleiben bei der Abbildung gemäß des Leitfadens viele HL7-Felder zunächst ohne Korrespondenz. Hier muss versucht werden im Detail eine Äquivalenz zu finden. Da sich die Z-Segmente und damit die HL7-DTD auf die 4. Auflage der Basisdokumentation für Tumorkranke beziehen und die BDT-DTD auf der veränderten 5. Auflage beruht, sind die Abbildungen zum Teil leider sehr lückenhaft. Abbildung 9.6 zeigt die Zuordnungen zwischen den BDT-Datenfeldern und den HL7-Feldern, die für beide Abbildungsrichtungen gelten.

9.2.4 XSLT mit MapForce

In MapForce werden die Tumordokumentations-DTD von HL7 aus Abschnitt 8.3.4 auf Seite 109 und die entsprechende BDT-DTD aus Abschnitt 8.3.4 auf Seite 111 importiert und das Wurzelement angegeben. Abbildung 9.7 zeigt dazu den entsprechenden Bildschirmausdruck.

Die Verbindungslinien zwischen den beiden vorliegenden XML-Bäumen definieren die Abbildungen, aus denen MapForce automatisch die XSL-Transformationen generiert.

Abbildung nach Leitfaden zwischen HL7-Segmenten und BDT-Sätzen

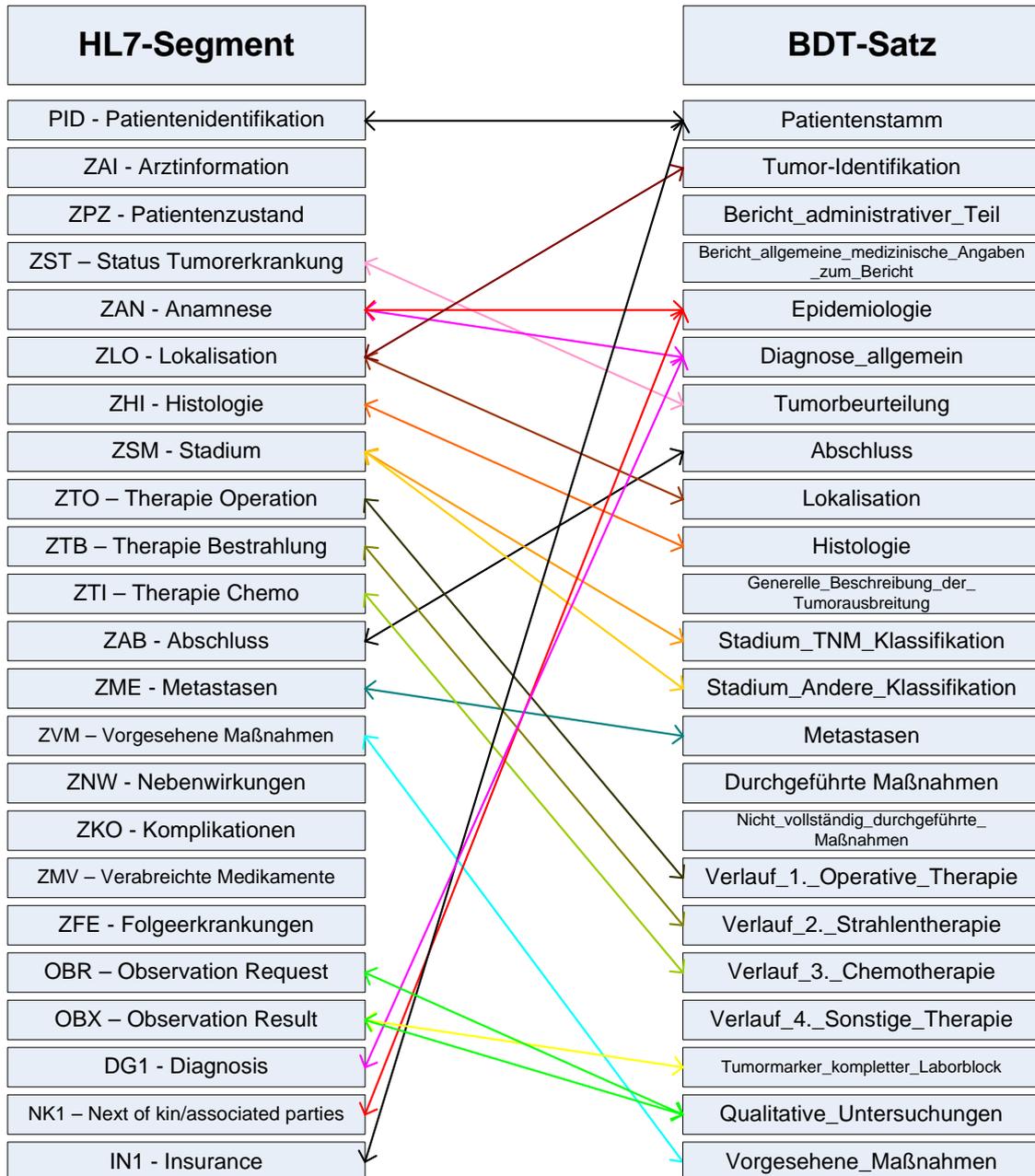


Abbildung 9.4: Abbildungen zwischen HL7-Segmenten und BDT-Sätzen nach dem vorgestellten Leitfaden

Abbildung nach Leitfaden zwischen HL7-Segment ZST und BDT-Satz Tumorbeurteilung

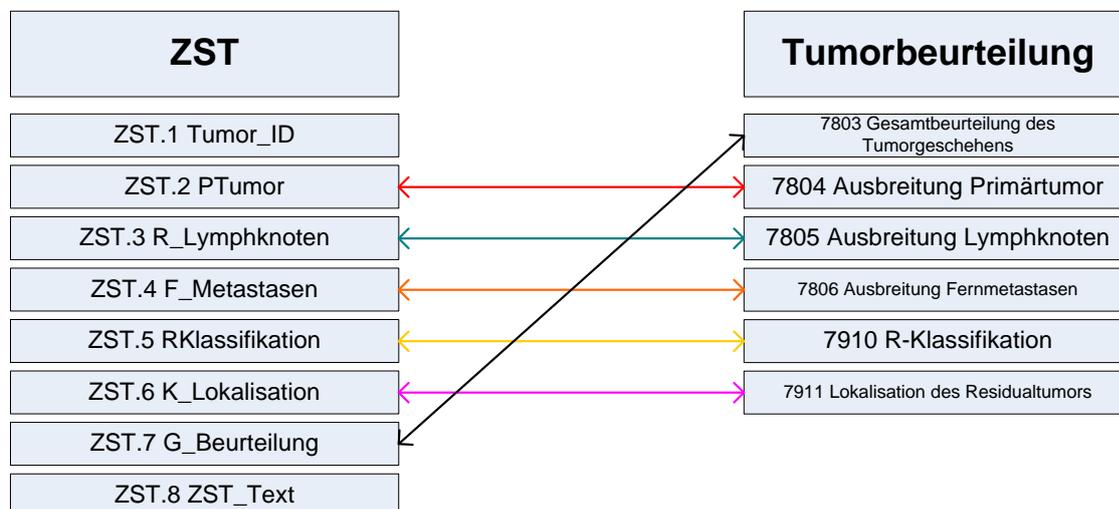


Abbildung 9.5: Abbildungen zwischen HL7-Felder und BDT-Datenfelder, beispielhaft für HL7-Segment ZST und BDT-Satz Tumorbeurteilung

Die Verbindungslinien zeigt Abbildung 9.8, die entsprechende XSLT-Generierung Abbildung 9.9.

Bidirektionale Transformationen können erreicht werden durch den Einsatz bidirektionaler Regeln oder durch die Definition zweier getrennter komplementärer unidirektionaler Regeln, eine für jede Richtung (siehe Abschnitt 5.1 auf Seite 39). XSLT bildet grundsätzlich unidirektional ab. Um bidirektionale Abbildungen zu erreichen, müssen beide Richtungen getrennt beschrieben werden (siehe Abschnitt 5.2.4 auf Seite 40). Die Abbildungen wurden daher einmal zwischen BDT und HL7 und nochmals zwischen HL7 und BDT definiert. Es handelt sich insgesamt um bidirektionale Abbildungen zwischen HL7 und BDT. Abbildung 9.10 zeigt anhand einer BDT-Beispieldatei in der MapForce-Vorschau die Transformation in die entsprechende HL7-Nachricht.

Die mit MapForce erstellten XSL-Dateien für die XSL-Transformationen zwischen der BDT-DTD und der HL7-DTD in der einen, wie in der anderen Richtung, befinden sich auf der Dissertations-CD (siehe Anhang E auf Seite 221).

9.2.5 Beurteilung

Mittels Leitfaden ergeben sich bei der Zuordnung von BDT-Sätzen zu HL7-Segmenten zunächst schnell kleinere Arbeitspäckchen, in denen zügig die detaillierteren Modell-

Zuordnung zwischen BDT-Datenfeldern und HL7-Feldern					
BDT-Datenfeld	HL7-Feld	BDT-Datenfeld	HL7-Feld	BDT-Datenfeld	HL7-Feld
3000	PID.2	7803	ZST.7	7826	ZTO.10
3100	PID.5 XPN.5	7804	ZST.2	7827	ZTO.11
3101	PID.5 XPN.1	7805	ZST.3	7828	ZTO.6
3102	PID.5 XPN.2	7806	ZST.4	7841	ZKO.1
3103	PID.7	7910	ZST.5	7843	ZKO.3
3104	PID.5 XPN.6	7911	ZST.6	7853	ZTB.2
3105	PID.19	7920	ZAB.1	7854	ZTB.3
3106	PID.11 XAD.5	7922	PID.29, ZAB.2	7855	ZTB.4
3107	PID.2	7923	ZAB.11	7856	ZTB.5
3110	PID.8	7924	ZAB.6-10	7859	ZTB.6
3182	PID.23	7929	ZAB.13	7860	ZTB.9
3201	IN1.16 XPN.1	7712	ZLO.4	7862	ZTB.11
3202	IN1.16 XPN.2	7713	ZLO.5	7863	ZTB.12
3203	IN1.18	7720	ZHI.5	7866	ZTB.15
3204	IN1.19 XAD.5	7723	ZHI.6	7867	ZTB.16
3205	IN1.19 XAD.1	7726	ZHI.8	7869	ZTB.7
3210	IN1.43	7729	ZHI.9	7870	ZTB.13
3626	PID.13	7750	ZSM.4	7871	ZTB.14
3628	PID.15	7751	ZSM.5	7872	ZTB.17
4111	IN1.3	7752	ZSM.6	7873	ZTB.18
7700	ZST.1,ZLO.1, ZHI.1,ZSM.1, ZTO.1,ZTB.1, ZTI.1,ZME.1	7753	ZSM.7	7874	ZTB.19
7701	ZLO.2	7754	ZSM.8	7875	ZTB.20
7709	ZLO.3	7755	ZSM.9	7879	ZTB.21
7956	ZAI.3	7756	ZSM.10	7881	ZTI.5
7957	ZAI.9	7757	ZSM.11	7882	ZTI.6
7962	ZPZ.2	7758	ZSM.12	7883	ZTI.2
7810	ZPZ.4	7759	ZSM.13	3622	ZTI.10
7811	ZFE.2	7760	ZSM.14	3623	ZTI.11
7814	ZFE.3	7761	ZSM.15	7884	ZMV.1
3901	ZAN.6	7770	ZSM.16	7888	ZMV.3
3903	ZAN.7	7771	ZSM.17	7891	ZMV.5
3905	ZAN.8	7772	ZSM.18	7895	ZMV.6
3911	ZAN.13	7773	ZSM.19-30	7940	ZVM.1
3912	ZAN.14	7778	ZSM.32	7942	ZVM.2
3917	ZAN.15	7790	ZME.2	7943	ZVM.3
3940	ZAN.4	7794	ZME.3	7944	ZVM.4
3943	ZAN.5	7821	ZTO.3		
7704	ZAN.1	7830	ZTO.8		
7706	ZAN.3	7824	ZTO.7		
		7825	ZTO.9		

Abbildung 9.6: Komplette Zuordnung zwischen den BDT-Datenfeldern und den HL7-Feldern für das Beispiel Basistumordokumentation

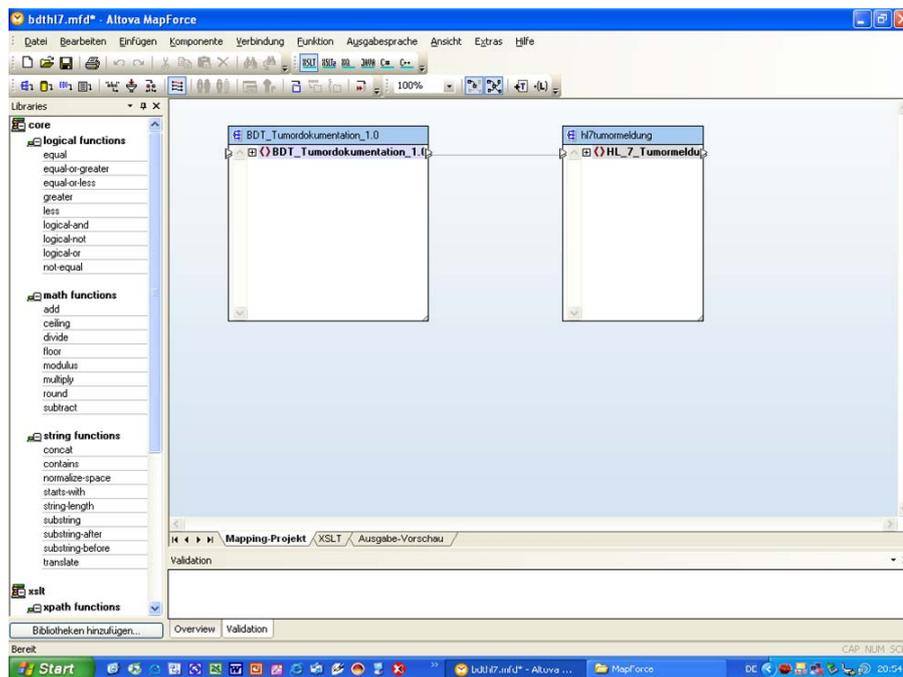


Abbildung 9.7: Importierte DTDs von BDT und HL7 in MapForce

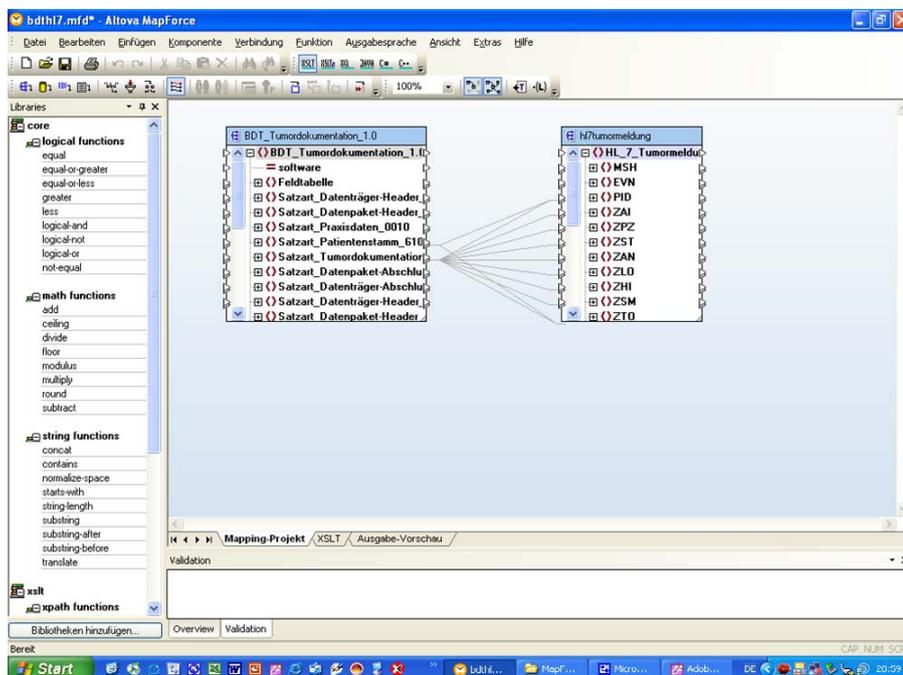


Abbildung 9.8: Abbildungsdefinition mittels Verbindungslinien zwischen BDT und HL7 in MapForce

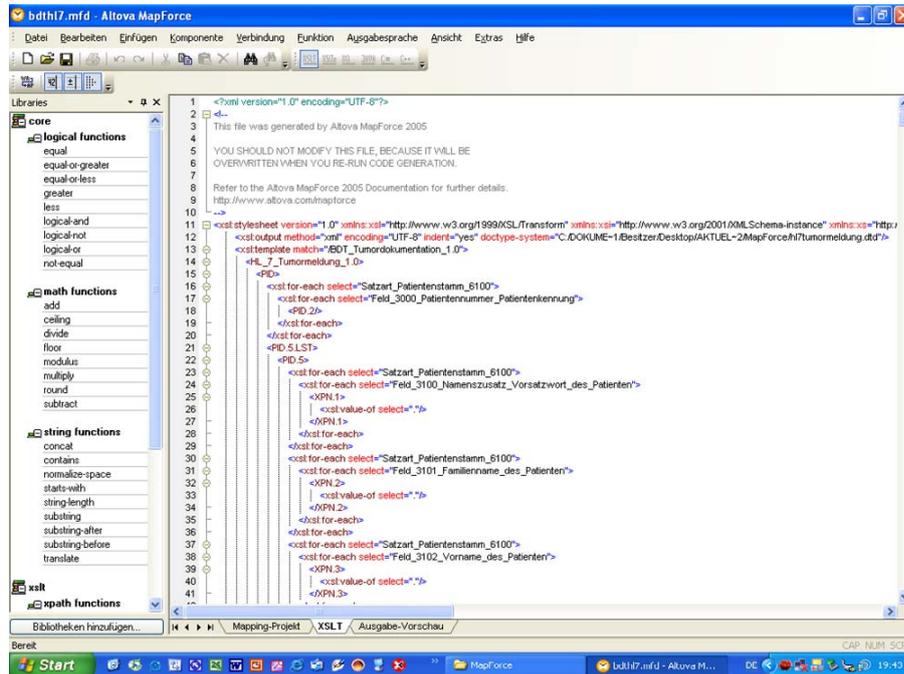


Abbildung 9.9: Generierte XSL-Transformationen von MapForce zwischen BDT und HL7

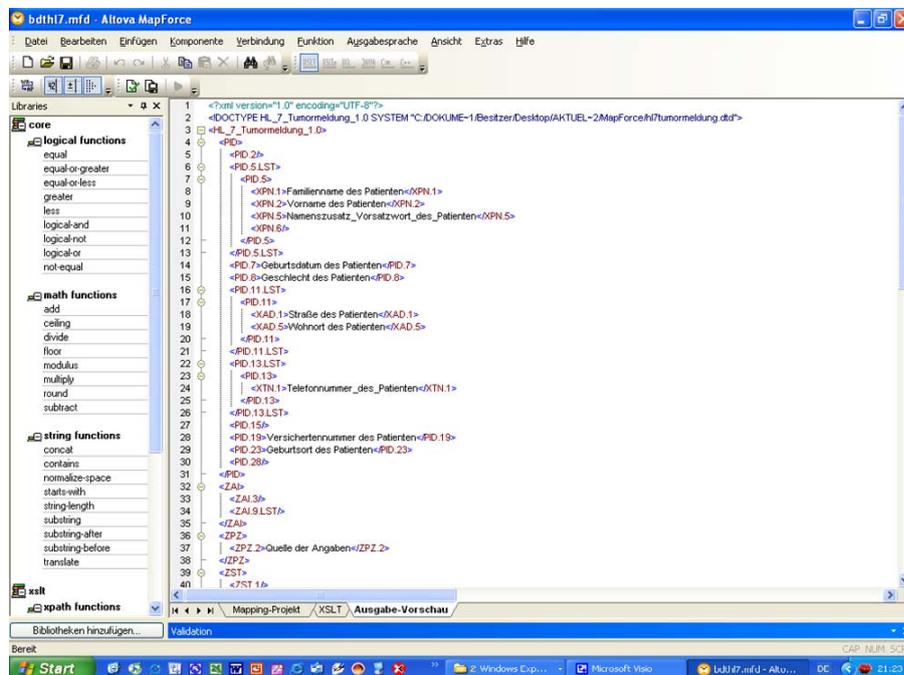


Abbildung 9.10: Transformation einer BDT-Beispielnachricht in HL7 in der MapForce-Vorschau

korrespondenzen herausgearbeitet werden können, nämlich diejenigen zwischen BDT-Datenfeldern und HL7-Feldern.

Die Qualität der Abbildung ist hier selbstverständlich von der Kompatibilität der beiden Standards bezüglich der Basisdokumentation für Tumorkranke abhängig (siehe auch Abschnitt 5.3 auf Seite 42). Diesbezüglich ist das Ergebnis nicht zufriedenstellend, da zum einen mit zahlreichen Z-Segmenten außerhalb des Standards gearbeitet werden muss, zum anderen, weil sich die Basisdokumentation von der 4. zur 5. Auflage spürbar verändert hat. Eine Weiterentwicklung der Standards in Richtung eines gemeinsamen Dokumentationsstandards unter Verwendung einheitlicher Begriffssysteme ist wünschenswert.

9.3 Abbildung Dokumentation

Im Folgenden soll aufgezeigt werden, wie zwischen den Dokumentationsstandards CDA und Basisdokumentation für Tumorkranke vermittelt werden kann. Als Domänenmodelle dienen für die CDA das OnkoDok-Modell von [Wol02], für die Basisdokumentation eine dafür erstellte DTD. Zunächst werden dazu wieder die Strukturen der DTD und XML-Schemata analysiert und anhand des vorgestellten Leitfadens die groben Korrespondenzen erstellt. Anschließend können auf einer Detailebene darunter feinere Korrespondenzen spezifiziert werden.

Zuordnungen, die dann noch offen bleiben, müssen manuell nach genauerer Prüfung über die gesamte Struktur gesucht werden. Einige Zuordnungen können überhaupt nicht getroffen werden.

9.3.1 Modellkorrespondenzen von Strukturierungselementen

Leider gehen hier aus dem Leitfaden keine ersten Abbildungen quasi automatisch hervor. Bei der Abbildung zwischen den Kommunikationsstandards HL7 und BDT waren bereits auf generischer Ebene wichtige Modellkorrespondenzen aufgefallen. Dies ist hier nicht möglich. Bei der Betrachtung der jeweiligen Strukturierungselemente, wie dies der Leitfaden vorsieht (siehe Abbildung 7.11 auf Seite 90), erscheinen aber die in Abbildung 9.11 skizzierten Korrespondenzen äußerst wahrscheinlich.

Innerhalb dieser ersten groben Zuordnung können anschließend weitere Strukturierungselemente feinerer Granularität verglichen und zugeordnet werden. Abbildung 9.12 zeigt beispielhaft die Abbildung zwischen dem Abschnitt D Allgemeine Identifikationsdaten und dem Header von OnkoDok.

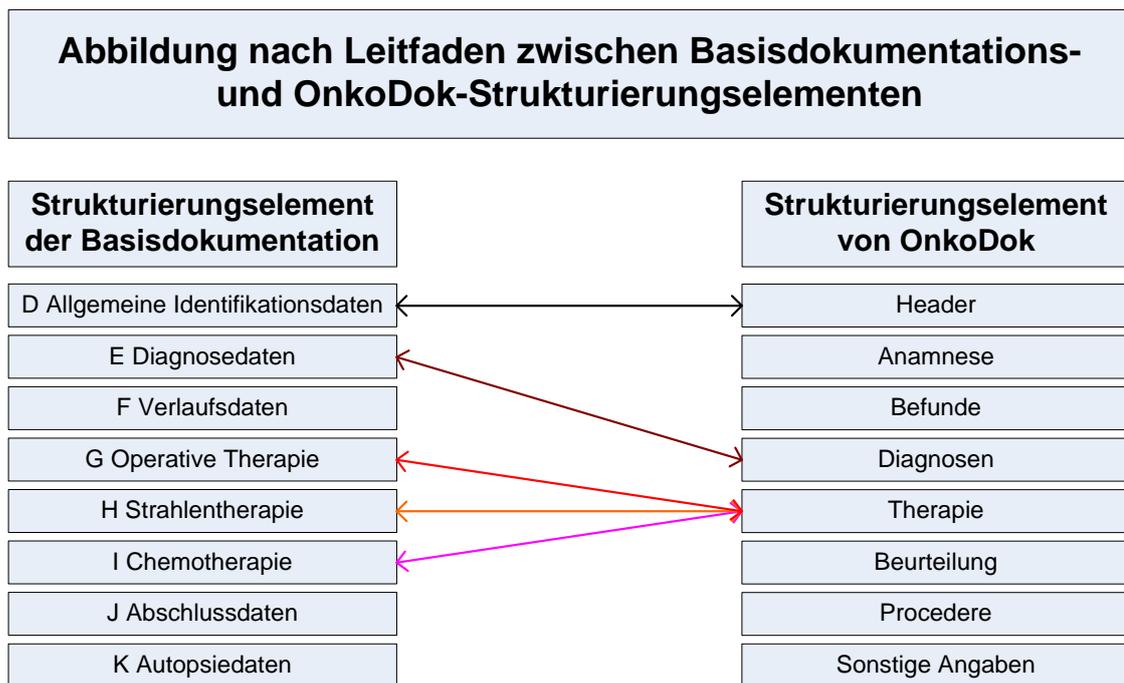


Abbildung 9.11: Erste grobe Abbildungen zwischen Strukturierungselementen der Basisdokumentation und OnkoDok nach dem vorgestellten Leitfaden

9.3.2 Übrige Abbildungen

Leider bleiben bei der Abbildung gemäß des Leitfadens viele Strukturierungselemente zunächst ohne Korrespondenz. Hier muss versucht werden im Detail eine Äquivalenz zu finden. Da die Basisdokumentation für Tumorkranke einen anderen Zweck erfüllt als die Dokumentation der Tumorpatientenfälle in Kliniken, an der sich OnkoDok orientiert, sind die Abbildungen zum Teil leider sehr lückenhaft. Abbildung 9.13 zeigt die Zuordnungen zwischen den Strukturierungselementen von Basisdokumentation und OnkoDok, die für beide Abbildungsrichtungen gelten.

9.3.3 XSLT mit MapForce

Wiederum in MapForce werden die erstellte Basisdokumentations-DTD der Basisdokumentation für Tumorkranke und die entsprechenden OnkoDok-Schemata importiert. Leider lassen sich nur die XML-Schemata Header, Anamnese, Beurteilung und Sonstige-Angaben importieren. Bei den OnkoDok-Schemata Befunde, Diagnosen, Therapie und Procedere verursachen jeweils einige Elemente, dass das Content Model nicht-deterministisch ist, was den Import sowohl in MapForce als auch bei e*Gate über das XML Toolkit verhindert. Aus diesem Grund wird im Folgenden beispielhaft eine Abbildung mit MapForce angegeben. Die komplett mögliche Abbildung in Form einer Tabelle wurde bereits in Abbildung 9.13 dargestellt.

Abbildung nach Leitfaden zwischen den Strukturierungselementen D der Basisdokumentation und Header von OnkoDok

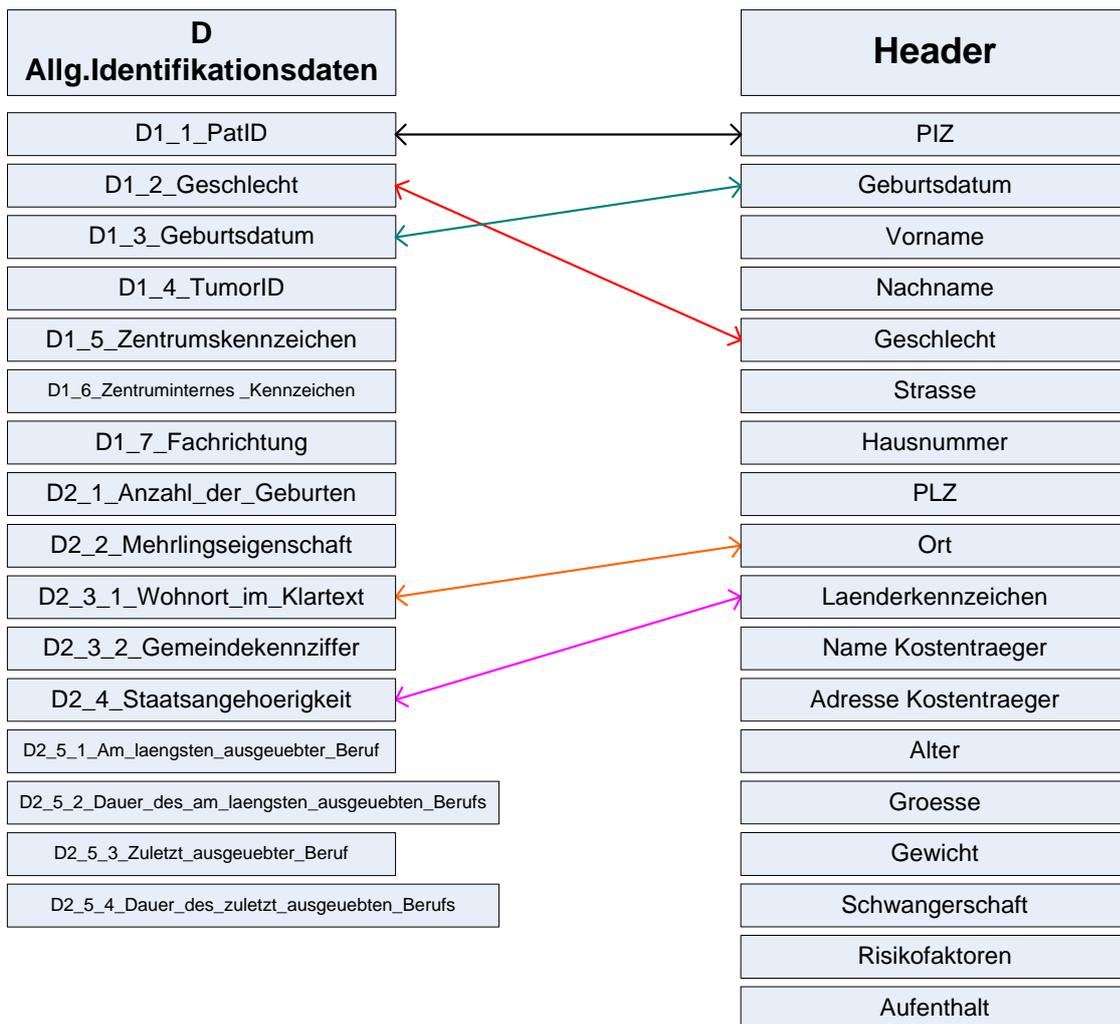


Abbildung 9.12: Abbildungen zwischen Strukturierungselementen von Basisdokumentation und OnkoDok, beispielhaft für Abschnitt D und OnkoDok-Header

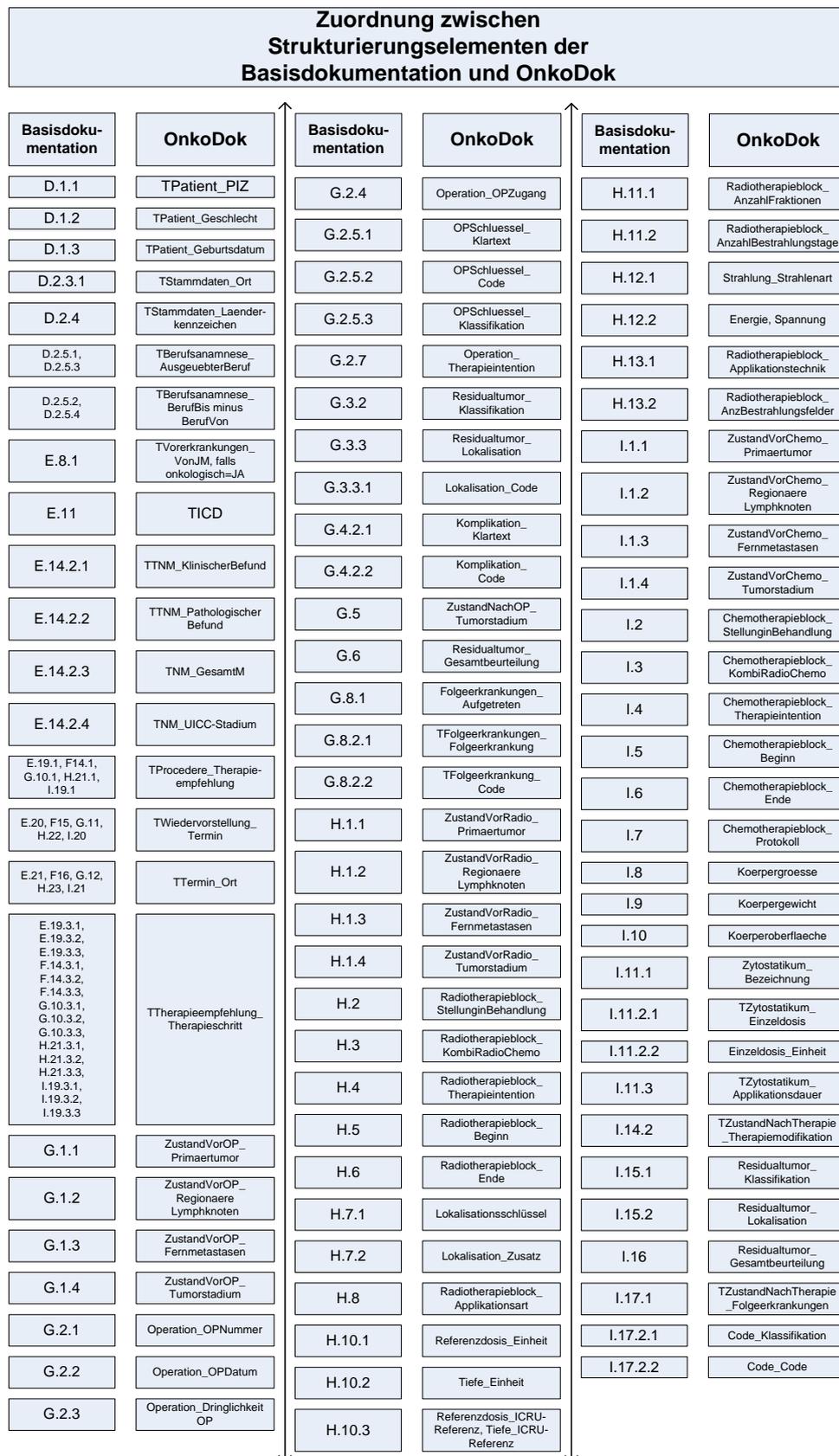


Abbildung 9.13: Komplette Zuordnung zwischen den Strukturierungselementen der Basisstamordokumentation und OnkoDok

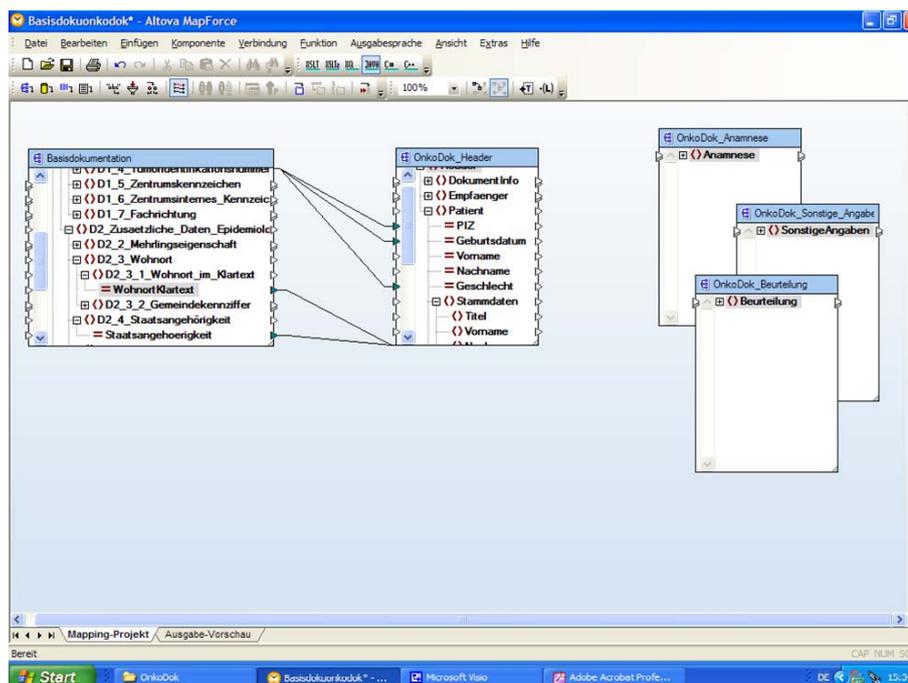


Abbildung 9.14: Abbildungsdefinition mittels Verbindungslinien zwischen Basisdokumentation D und OnkoDok-Header in MapForce

Die Verbindungslinien zwischen den beiden vorliegenden XML-Bäumen definieren die Abbildungen, aus denen MapForce automatisch die XSL-Transformationen generiert. Die Verbindungslinien zeigt Abbildung 9.14.

Um wiederum bidirektionale Abbildungen zu erreichen, müssen beide Richtungen getrennt beschrieben werden (siehe Abschnitt 5.2.4 auf Seite 40). Die Abbildungen wurden daher einmal zwischen der Basisdokumentation und OnkoDok und nochmals zwischen OnkoDok und der Basisdokumentation definiert. Es handelt sich insgesamt um bidirektionale Abbildungen zwischen der Basisdokumentation und OnkoDok.

Die mit MapForce erstellten XSL-Dateien für die XSL-Transformationen zwischen der Basisdokumentations-DTD und den OnkoDok-Schemata in der einen, wie in der anderen Richtung, befinden sich auf der Dissertations-CD (siehe Anhang E auf Seite 221).

9.3.4 Beurteilung

Mittels Leitfaden ergeben sich bei der Zuordnung von Strukturierungselementen der Basisdokumentation für Tumorkranke zu denen von OnkoDok nicht quasi automatisch Modellkorrespondenzen, wie dies bei der Abbildung zwischen HL7 und BDT der Fall ist. Es können jedoch sehr wahrscheinlich erscheinende, erste grobe Äquivalenzen entdeckt werden. Innerhalb dieser wiederum kleineren Arbeitspäckchen können in immer feineren

Granulationsgraden die jeweiligen Strukturierungselemente verglichen und zugeordnet werden.

Die Qualität der Abbildung ist wiederum von der Kompatibilität der beiden Standards abhängig (siehe auch Abschnitt 5.3 auf Seite 42). Diesbezüglich ist das Ergebnis nicht ganz zufriedenstellend. Die Basisdokumentation für Tumorkranke und die Dokumentation nach OnkoDok haben bezüglich ihres Zweckes Überschneidungen, aber keine komplette Kongruenz. Während bei der Basisdokumentation für Tumorkranke eine statistische Auswertung bezüglich unterschiedlicher Parameter im Vordergrund steht, dient das OnkoDok-Modell eher einer Dokumentation von Tumorpatientenfällen im klinischen Umfeld. Daraus ergibt sich, dass die Abschnitte E Diagnosedaten, F Verlaufsdaten nicht gut, die Abschnitte J Abschlussdaten und K Autopsiedaten gar nicht abgebildet werden können, die Abschnitte D Allgemeine Identifikationsdaten, G Operative Therapie, H Strahlentherapie und I Chemotherapie jedoch sehr gut. Positiv ist die Verwendung einheitlicher Begriffssysteme, wie z.B. ICD-O und TNM innerhalb der beiden Dokumentationsstandards. Damit können wesentliche Informationen ohne Probleme korrekt abgebildet werden.

9.4 Ergebnisse der Evaluation

Die Begriffssysteme können als Leitfaden zur Strukturanalyse den Weg zu den detaillierten Transformationen vereinfachen. Sie sind für Software-Ingenieure leicht lesbar, weil einheitlich im Standard UML spezifiziert. Dadurch kann die Einarbeitungszeit in den jeweiligen Standard verkürzt werden. Die Standards werden in ihrer Struktur durch die UML-Modelle leichter verständlich. Durch die Metasichten auf die Standards sowie die Begriffssysteme sind Äquivalenzen bzw. ähnliche Strukturen in verschiedenen Standards besser zu erkennen, z.B. HL7-Segment und BDT-Satz. Diese Äquivalenzen sind besonders dann hilfreich, wenn sie auf einer größeren Detailebene gefunden werden können. Die Abbildung lässt sich dann in kleineren Arbeitspaketen einfacher erstellen.

Das Begriffssystem für Kommunikationsstandards ist mächtiger, weil konkreter in seinen Aussagen. Daher ist es nicht verwunderlich, dass dieses Begriffssystem ein deutlich besserer „Wegweiser“ ist, als das Begriffssystem für Dokumentationsstandards. Implizit wurde bei den Kommunikationsstandards HL7 und BDT bereits ein Dokumentationsstandard, nämlich die Basisdokumentation für Tumorkranke, verfolgt. Ebenso kommen medizinische Begriffssysteme wie ICD und TNM zum Einsatz.

9.4.1 Bijektivität der Abbildungen

Bijektivität (bijektiv oder umkehrbar eindeutig) ist eine Eigenschaft einer mathematischen Funktion. Eine bijektive Funktion ist eine Funktion, die verschiedene Elemente ihres Definitionsbereichs auf verschiedene Elemente des Wertebereichs abbildet (sie ist injektiv),

und für die zusätzlich jedes Element des Wertebereichs als Bild auftritt (sie ist surjektiv) [Kom05]. Eine bijektive Funktion hat daher immer eine vollständig definierte Umkehrfunktion. Die „Anzahl der Elemente“ der Definitionsmenge, Bildmenge und der Wertemenge einer bijektiven Funktion sind stets gleich groß (mit Hilfe von Bijektionen wird der Begriff der Gleichmächtigkeit definiert), diese Eigenschaft reicht aber nicht zur Definition einer bijektiven Funktion aus.

Für die Abbildungen im Zusammenhang mit dieser Arbeit bedeutet Bijektivität, dass die Abbildung einer BDT-Tumormeldung auf eine HL7-Tumormeldung und eine zweite Abbildung dieser HL7-Tumormeldung in eine BDT-Tumormeldung zu einer identischen BDT-Tumormeldung führen muss. Dann wäre die gesamte Abbildungsdefinition (mit den beiden unidirektionalen Richtungen) bijektiv.

Bijektivität der Abbildung Kommunikation

Die gefundene Gesamtabbildung ist bijektiv. Dazu müssen jedoch ein paar Anmerkungen gemacht werden. Das BDT-Datenfeld 7700 Tumor-ID muss auf mehrere HL7-Felder abgebildet werden (siehe Abbildung 9.6 auf Seite 129), was umgekehrt nicht möglich ist. Wenn man allerdings davon ausgeht, dass die Tumor-ID in allen HL7-Feldern identisch ist, kann man eine einzige Abbildung, z.B. HL7-Feld ZST.1 auf BDT-Datenfeld 7700, definieren. Gleiches gilt für das BDT-Datenfeld 7922 Sterbedatum. Etwas aufwändiger wird die Angelegenheit bei den BDT-Datenfeldern 7924 Todesursache ICD-Schlüssel und 7773 Ann-Arbor Organ Schlüssel. Die zu dem ICD-Schlüssel der Todesursache korrespondierenden HL7-Felder sind ZAB.6 bis ZAB.10. Während das BDT-Datenfeld den kompletten Schlüssel in einem Feld aufnimmt, sind alle fünf HL7-Felder dafür zuständig. Bei beiden Abbildungsrichtungen müssen entsprechende Filter die Daten aufteilen bzw. entsprechend zusammenführen. Gleiches gilt für den Ann-Arbor Organ Schlüssel.

Bijektivität der Abbildung Dokumentation

Die hier gefundene Gesamtabbildung ist nicht bijektiv. Einzelne Abbildungen sind bijektiv nicht möglich, allerdings beschränkt sich dies auf einige wenige. Im Prinzip müssen auch hier diverse Filter bei der Transformation im Detail helfen. So müssen für die Dauer des zuletzt und am längsten ausgeübten Berufs jeweils die Angaben aus den Attributen `BerufVon` und `BerufBis` entsprechend voneinander subtrahiert werden. Dies ist beispielsweise genau das Problem: bei der Abbildung von der Basisdokumentation, die die Dauer angibt, auf `OnkoDok`, das eine Angabe verlangt, in welchem Jahr mit dem Beruf begonnen und in welchem Jahr aufgehört wurde. In diese Richtung kann eine Abbildung nicht erfolgen. Eine Abbildung von E.8.1 auf das Attribut `VonJM` von `TVorerkrankungen` kann nur erfolgen, wenn das Attribut `onkologisch` den Wert `JA` besitzt. Siehe dazu auch Abbildung 9.13 auf Seite 135.

Teil IV

Verwandte Ansätze

10 Verwandte Ansätze aus der Informatik

Verschiedene Forschungsprojekte und Organisationen bemühen sich mit diversen Lösungsansätzen Interoperabilität in Informatik und Medizin zu ermöglichen.

Strukturiert sind die vorgestellten Ansätze zunächst nach ihrem Schwerpunkt in die Kategorien Informatik und Medizininformatik, wobei diese Grenze zum Teil nur schwer gezogen werden kann, da die Übergänge fließend sind. Innerhalb dieser groben Kategorien werden zu einzelnen Themenbereichen jeweils exemplarisch relevante Ansätze aus der Vielzahl existierender verwandter Arbeiten beschrieben, mit dem eigenen Ansatz verglichen und bewertet.

Zu verwandten Ansätzen zählen Arbeiten auf dem Gebiet der Datenbank-Integration, der Ontologien, des Semantic Web, der semantischen Suche und des Matching. Für jedes dieser Teilgebiete werden exemplarisch ausgewählte Arbeiten vorgestellt.

10.1 Datenbank-Integration

10.1.1 Schemaintegration

Auf dem Gebiet der Föderierten Datenbanken werden verschiedene Schemaintegrationsverfahren angewandt, um die beteiligten Datenbanken zu integrieren. Auf Föderierte Datenbanken wird in Anhang A auf Seite 161 eingegangen. [Con02] stellt die grundsätzlichen Schemaintegrationstechniken vor. Neben zusicherungsbasierter Schemaintegration, gibt es die Integration von Klassenhierarchien mit Upward Inheritance und das Generic Integration Model (GIM) [SS05]. Ferner existieren formal festgelegte Sprachen zur Spezifikation der Abbildungen sowie Multidatenbanksprachen. Ziel der Schemaintegration ist es, ein gemeinsames, föderiertes Schema anzubieten, welches einen globalen Zugriff auf alle Datenquellen erlaubt. Integrationsrichtung ist bei der Datenbank-Integration traditionell Bottom-Up, was sich von der in dieser Arbeit verwendeten Top-Down-Methodologie grundlegend unterscheidet.

10.2 Ontologien und Semantic Web

10.2.1 Repräsentations- und Anfragesprachen für Ontologien

Es gibt eine Reihe von Repräsentations- und Anfragesprachen für Ontologien. [MM03] gibt eine Übersicht über den aktuellen Stand in diesem Bereich und stellt den Karlsruher Ansatz KAON zur Repräsentation und Abfrage von semantischen Modellen vor. Semantische Modelle können im Bereich der Datenbank- und Informationssysteme der Quellenbeschreibung, als auch als integriertes Modell dienen, um Heterogenität der Quellen zu überbrücken. UML wird als eine der bestehenden Repräsentationssprachen vorgestellt. Der Karlsruher Ansatz zur hybriden Ontologierepräsentation und -anfrage soll die gewünschten Eigenschaften existierender Ontologierepräsentationssprachen kombinieren, um die jeweiligen negativen Eigenschaften zu kompensieren. Es handelt sich um eine übergeordnete Grundlagenarbeit, die die Einsatzgebiete von Ontologien beleuchtet und versucht, einen besseren Ansatz durch Kombination von Stärken bestehender Ansätze anzubieten. Der OntoBroker nutzt formale Ontologien, um Metadaten im WWW zu extrahieren und zu generieren. Es wurden Tools entwickelt, um Anfragen zu formulieren, Ontologien zu definieren, Metadaten zu extrahieren und Metadaten im RDF-Format (Resource Description Framework=Metadatenrepräsentationssprache vom W3C) zu generieren.

Die vorliegende Arbeit ist den Weg der Begriffssysteme gegangen, weil in der Medizin das Arbeiten mit Begriffssysteme Gewohnheit ist und sie formal leichter handhabbar sind. Für eine semantische Abgrenzung der in den Standards verwendeten Begriffe sind sie ausreichend. Das Gebiet der Ontologien sollte nicht weiter vertieft werden.

10.2.2 A Semantic Web Service-based P2P Infrastructure for the Interoperability of Medical Information Systems (ARTEMIS)

In dem von der EU geförderten Forschungsprojekt ARTEMIS soll die Interoperabilität von medizinischen Informationssystemen (vor allem Krankenhausinformationssystemen) verbessert werden [DLK⁺05]. Es soll eine Infrastruktur entwickelt werden, die eine Kommunikation zwischen unterschiedlichen Informationssystemen mit Hilfe von Web-Services ermöglicht. Die verwendeten, unterschiedlichen Standards werden dazu in Ontologien beschrieben und Abbildungen zwischen diesen Standards definiert. Die ARTEMIS Web-Service-Architektur schlägt keine global abgestimmten Ontologien vor, vielmehr bringen die Institutionen ihre semantischen Unterschiede durch eine Mediator-Komponente in Einklang. Die Mediator-Komponente nutzt Ontologien basierend auf bekannten Standards in der Medizin als Referenzen, um eine semantische Vermittlung zwischen den involvierten Institutionen zu ermöglichen. Die Mediatoren haben eine P2P-Kommunikationsarchitektur, die Gesamtarchitektur ist angepasst an das Semantic Web.

10.2.3 Semantic Health Records

[Jun03] unternimmt den ersten Schritt Heterogenität bei Electronic Healthcare Records mittels Semantic Web-Technologien zu überwinden. Electronic Healthcare Records werden dabei als kleine Ausschnitte des WWW betrachtet. Es wird eine konsequente XML-Repräsentation der gesamten EHR-Datenstrukturen erreicht, die als Grundlage für eine Kombination von Semantic Web und Electronic Healthcare Records dienen kann. Die semantische Heterogenität ist allein mit der XML-Repräsentation jedoch noch nicht gelöst. Näheres zu Semantic Web ist im Anhang B.4 auf Seite 186 zu finden.

10.3 Semantische Suche

10.3.1 Semantische Komponentensuche auf Basis von Geschäftsprozessmodellen

In der komponentenbasierten Softwareentwicklung soll Software durch kundenindividuelle Komposition und Anpassung standardisierter Komponenten erstellt werden. Durch die Wiederverwendung von Komponenten sollen Entwicklungszeit und damit Kosten gespart werden. Eine semantische Suche nach Komponenten auf der Basis von Geschäftsprozessmodellen wird in [Tes03] vorgestellt. Angeboten werden Komponentenbeschreibungen, die Anforderungen werden mittels Geschäftsprozessmodellen ausgedrückt. Die Suchphase teilt sich in einen Protokoll- und einen Semantikvergleich. Dazu wurde eine Normsprache entwickelt, die gemeinsame fachsprachliche Grundlage für die Sprache der Komponentenhersteller (Komponentenbeschreibung) und der Komponentenverwender (Geschäftsprozessmodelle) darstellt. Während hier eine neuentwickelte Normsprache das gemeinsame Verständnis zwischen Komponentenherstellern und -verwendern herstellt, sind es in der vorgelegten Arbeit die verschiedenen Begriffssysteme, die für semantische Klarheit sorgen.

10.4 Matching

10.4.1 Metamodell-basiertes Matching

Integration von moderner Firmensoftware und angepassten Altsystemen durch eine kombinierten Top-Down- und Bottom up-Methodologie beschreibt [vdH02]. Die Altsysteme werden dazu mit Wrappern ausgestattet, damit Teile oder das Gesamtsystem sich wie Objekte verhalten und auch so aussehen. Anschließend wird auf einer konzeptuellen Ebene und einer Implementationsebene die Abbildung zwischen den Business-Objekten durchgeführt. So werden in einer neuen Firmensoftware Funktionalitäten der Altsysteme ebenso verwendet wie neue, die sich aus den Anforderungen ergeben haben.

10.5 Tabellarische Zusammenfassung

	Conrad	OntoBroker	ARTEMIS	Jung	Teschke	Heuvel	Eigener Ansatz
Schwerpunkt Anwendungsarchitektur		●	●	●			●
Schwerpunkt Technologiearchitektur	●						
Schwerpunkt Geschäftsarchitektur					●	●	
Einsatz von Domänenstandards			●	●	●	●	●
Flexible Kombination von Standards			●				●
Schwerpunkt Integration	●	●	●	●	●	●	●
TOP-DOWN-Integration			●				●
JO-JO-Integration						●	
BOTTOM-UP-Integration	●						
XML als gemeinsame Syntax		●	●	●			●
Verwendung von Mediatoren			●				●
Schwerpunkt Modellierung							
Beschränkt auf Teilgebiet							

Abbildung 10.1: Eine tabellarische Übersicht der verwandten Ansätze aus der Informatik einschließlich des eigenen Ansatzes

11 Verwandte Ansätze aus der Medizininformatik

Mehrere Integrationsbemühungen gibt es auf dem Gebiet der Standards CDA, HL7, DICOM und beispielsweise der Tumordokumentation. Ferner gibt es Ansätze für Integrations-Architekturen. Für jedes dieser Teilgebiete werden wiederum exemplarisch ausgewählte Arbeiten vorgestellt.

11.1 Clinical Document Architecture

Der Dokumentationsstandard der HL7 Gruppe befindet sich im Entwicklungsprozess. So muss der Standard z.B. auf nationale Bedürfnisse angepasst werden.

11.1.1 Standardisation of Communication between Information Systems in Physician Offices and Hospitals using XML (SCIPHOX)

Angesiedelt ist das Projekt bei der HL7 Benutzergruppe D, Technisches Komitee: XML-Anwendungen im Gesundheitswesen. Aus der Projektgruppe: Harmonisierung der Datenkommunikation BDT und HL7 mittels XML entstand Anfang 2000 das Projekt SCIPHOX [SCI02]. Phase I des Projektes SCIPHOX beschäftigte sich mit dem Kommunikationsumfang und -inhalt. So ist eine Spezifikation zu standardisierten elektronischen Kurzberichten (Entlassungsbrief, Überweisung, Einweisung) entstanden. XML steht als Transfersyntax im Vordergrund. SCIPHOX liefert die Spezifikation eines Kommunikationsstandards, der Erfahrungen der Domänen xDT und HL7 berücksichtigt. Die CDA (Clinical Document Architecture) ist die zentrale Dokumentenspezifikation. Die CDA wurde im Detail mit SSUs (Small Semantic Units) verfeinert. Es wurde zwar von der CDA als gemeinsamer Dokumentenstruktur ausgegangen, aber trotzdem wurden die BDT- und HL7-Kurzberichte in XML bottom up abgebildet und die CDA um fehlende Konstrukte erweitert, also eine JO-JO-Integration.

Dieses Projekt bildet also zwei Kommunikationsstandards für einen elektronischen Kurzbericht ab auf einen gemeinsamen Dokumentationsstandard CDA. Damit die Abbildungen auf den gemeinsamen Standard komplett sind, wurde sowohl bottom-up als auch top-down integriert. Eine flexible Kombination von Standards über Mediatoren, wie in dem hier vorgestellten Ansatz, findet nicht statt.

11.2 Health Level Seven (HL7)

11.2.1 MIRApplus (Universität Freiburg)

Der Einsatz des Medizinischen Informations-, Retrieval- und Archivsystems MIRApplus in Verbindung mit dem Patientenmanagementsystem PDV-FR hat das Ziel, alle klinisch relevanten Informationen zum Patienten fachabteilungsübergreifend zusammenzuführen und zu speichern [MIR04]. Diese Informationen werden dem Anwender entsprechend seiner spezifischen Anforderungen zentral zur Verfügung gestellt. Informationen von Stationen, Ambulanzen, Funktionsbereichen usw. werden an MIRApplus zur integrierten und strukturierten Verwaltung und Speicherung übermittelt. Hier sind die Informationen stets aktuell von allen klinischen Arbeitsplätzen aus abrufbar. Damit bildet MIRApplus als zentraler HL7-Informationsserver die zentrale Datenbank- und Kommunikationsplattform für die Elektronische Patientenakte (EPA). So werden redundante verteilte Datenbestände vermieden und auf einen Kommunikationsserver kann verzichtet werden. Die Integration mit anderen Anwendungssystemen erfolgt über offene, erweiterbare Schnittstellen bei durchgängigem Einsatz des Kommunikationsstandards HL7.

Der in der vorliegenden Arbeit vorgestellte Leitfaden zur Strukturanalyse, der bei der Erstellung von Abbildungsspezifikationen behilflich ist, kann auch sinnvoll in diesem Szenario eingesetzt werden. Nämlich dann, wenn es darum geht, Applikationen, die andere Standards nutzen, mit entsprechenden Abbildungsspezifikationen durch Transformationen zu integrieren.

11.3 Digital Imaging and Communications in Medicine Structured Reporting (DICOM SR)

11.3.1 Ein Verfahren zur Bewertung der Interoperabilität medizinischer Bildkommunikationssysteme

Interoperabilität von DICOM-kompatiblen Systemen ist nicht selbstverständlich. Ursache sind häufig unvollständige bzw. fehlerhafte Erzeugung oder Interpretation der DICOM-Datenstrukturen. Entwickelt wurde in [Eic01] ein Ansatz zur Konformitätsprüfung von

DICOM-Strukturen, der von einem Prüfwerkzeug umgesetzt wird. Aus dem Ergebnis der Konformitätsprüfung leitet sich danach eine automatisierte Interoperabilitätsbewertung ab. Problem für die semantische Interoperabilität sind demnach nicht nur unterschiedliche Standards, sondern auch unterschiedliche Umsetzungen von Standards.

11.3.2 Ein generisches Verfahren zur adaptiven Visualisierung von strukturierten medizinischen Befundberichten

Mit dem noch relativ neuen DICOM Structured Reporting gibt es ein standardisiertes Format für strukturierte medizinische Befundberichte, die neben textuellen Informationen auch Bilder und Signale aufnehmen können. Die äußerst flexible Struktur, in der Inhalte abgelegt werden, erleichtert jedoch nicht unbedingt die Einführung in die Praxis. Entgegen den in der Praxis bislang favorisierten, spezifischen Ansätzen mit ihren Einschränkungen bzw. generalisierten Lösungen, ist es Ziel von [REvGJ04], ein Verfahren zu entwickeln, das es erlaubt, dem Anwender beliebige Structured Reporting-Dokumente in angemessener Weise zu präsentieren. Dabei geht dieser Ansatz davon aus, dass in einem strukturierten, und damit selbstbeschreibenden Dokument, bereits genügend Informationen enthalten sind, die für eine geeignete Visualisierung benötigt werden. Dieser dokumenten-gesteuerte Ansatz scheint damit ein probates Mittel zu sein, sowohl die Hersteller bei der aufwendigen Erstellung von angepassten Anwendungen zu unterstützen, als auch dem Bedarf nach einer allgemeinen Arbeitsstation für Mediziner nachzukommen.

Der konforme Einsatz von Standards kann durch entsprechend geeignete Visualisierung der Strukturierungselemente unterstützt werden. Die Standards sind in der Regel sehr komplex aufgebaut und können durch Visualisierung dem Nutzer besser zugänglich gemacht werden. Durch die einheitliche Modellierung der Standards in UML erleichtert der Leitfaden ebenfalls den Zugang zu den Standards.

11.4 Tumordokumentation

Für diverse Gebiete in der Medizin, beispielsweise die Tumordokumentation, müssen Standards entsprechend angepasst und erweitert werden, damit sie überhaupt eingesetzt werden können.

11.4.1 Schnittstellenkonzepte in Tumordokumentationssystemen

Ziel von [Hae99] war die Erstellung einer konfigurierbaren Datenschnittstelle für das Giessener Tumordokumentationssystem (auch Kommunikationsmodul genannt) auf der

Basis der domänenspezifischen Standards BDT, HL7 und Basisdokumentation für Tumorkranke. Der Datenaustausch wird ermöglicht durch Abbildung der Basisdokumentation auf HL7 (mit Z-Segmenten) und BDT sowie entsprechend programmierter Schnittstellenmodule. Schwerpunkt dieser Arbeit liegt auf der Ebene der Anwendungsarchitektur, außerdem wurden Domänenstandards eingesetzt. HL7 wurde allerdings um Z-Segmente erweitert, um dem vorhandenen Datenschema ganz gerecht zu werden. Dies entspricht einer Jo-Jo-Integration, der HL7-Standard wurde lokal verändert.

Bei dieser Arbeit gilt das bei SCIPHOX genannte analog. Der gemeinsame Dokumentationsstandard ist hier die Basisdokumentation für Tumorkranke.

11.4.2 Patientenzentrierte Dokumentation onkologischer Erkrankungen

Ziel des Projektes ePA-Tumor an der Universität Heidelberg ist es, die Möglichkeiten für eine verteilte elektronische Patientenakte für Tumorkranke der Einrichtungen Thoraxklinik Heidelberg und Radiologische Universitätsklinik in Zusammenarbeit mit dem Tumorzentrum Heidelberg / Mannheim auszutesten. Das Projekt gliedert sich in drei Teilprojekte: Dokumentationskonzept, Architekturkonzept und Wissensintegration.

1. Dokumentationskonzept: Analyse der Dokumentationen beider Einrichtungen, um Überlappungen festzustellen. Syntax für die Strukturen ist XML. Nach einer Analyse der Patientenakten erfolgte eine Analyse der Dokumententypen. Anschließend die Identifikation von Informationseinheiten und -elementen. Danach ein Vergleich mit internationalen Standards. Die Auswahl ist auf die CDA gefallen. Im Rahmen von [Wol02] wurde ein generisches Informationsmodell in XML, das Onko-Dok-Modell, in Anlehnung an die CDA erstellt. Auch hierbei gab es Vollständigkeitsprobleme bei der Abbildung.
2. Architekturkonzept: Technische Realisierbarkeit soll untersucht werden unter Beachtung der Aspekte des Datenschutzes und der digitalen Signatur.
3. Wissensintegration: Integration von Leitlinien, Literatur u.ä. in die elektronische Patientenakte. Es wurden für zwei Leitlinien des Tumorzentrums Heidelberg / Mannheim und der Deutschen Krebsgesellschaft zu Diagnostik und Therapie des Bronchialkarzinoms eine XML-basierte Repräsentation entwickelt und die gemeinsamen Strukturelemente in einer DTD abgebildet.

[Wol02] wendet die CDA als Dokumentationsstandard für ein konkretes Szenario an. Es wurde eine Abbildung der Tumordokumentation der Kliniken und des Tumorzentrums in Heidelberg auf einen gemeinsamen Standard CDA durchgeführt, mit den Problemen der Vollständigkeit. Lösung wäre auch hier die Ergänzung bzw. Gestaltung des Standards.

11.4.3 eMamma-Akte (Fachhochschule Dortmund)

Bereits 2000 gab es erste Ansätze bei der Kassenärztlichen Vereinigung Nordrhein (KVNo) die Behandlung von Brustkrebs-Patientinnen mit D2D (PaDok) zu verbessern [Moh03]. D2D wird in Abschnitt 6.5.2 auf Seite 73 genauer beschrieben. Gemeinsame Dokumentationsregeln wurden 2002 auf der Medica vorgestellt. Die Daten in der eMamma-Akte stellen die Schnittmenge der in der ambulanten und stationären Versorgung erfassten Daten einer Brustkrebspatientin dar. Als Datenformat hatten die Beteiligten XML gewählt. Im Frühjahr 2003 wurde die eMamma-Akte in SCIPHOX-XML transformiert. Die Initiative SCIPHOX wird in Abschnitt 11.1.1 auf Seite 145 vorgestellt. Da die Marktführer der Arztpraxissoftwarehersteller mit VCS (siehe Abschnitt 6.5.2 auf Seite 72) einen eigenen Standard entwickelt hatten, gelang erst nach Verhandlungen eine Zusammenarbeit der konkurrierenden Systeme D2D/PaDok und VCS unter Leitung der Fachhochschule Dortmund. Für die Kommunikation mit den Krankenhäusern wurde, wegen anhaltender Sicherheitsbedenken, eine unabhängige Kommunikationsplattform (Health Telematic Broker (HTB)) entwickelt [eMa05]. Hier wurde am Fachbereich Informatik die eMamma-Akte beispielhaft evaluiert. Im Moment wird die eMamma-Akte in der Pilotregion Essen eingesetzt. Fünf niedergelassene Gynäkologen, eine radiologische Gemeinschaftspraxis und drei Krankenhäuser sind derzeit beteiligt. An einer räumlichen Ausdehnung wird gearbeitet, eine Ausdehnung dieser Art elektronischer Krankenakte auf andere Krebsarten (Kolonrektalkarzinom, Prostatakarzinom und Lymphom) ist geplant.

Bei der eMamma-Akte wurde, mit der Einigung auf eine einzige Dokumentationsstruktur, ein gemeinsamer Dokumentationsstandard für eine Brustkrebsakte geschaffen. Mit der Transformation in SCIPHOX-XML wurde auch die Darstellung standardisiert. Die eMamma-Akte ist ein Beispiel, wie mit der Verständigung auf Standards auf verschiedenen Ebenen Interoperabilität gelingen kann.

11.5 Integrations-Architekturen

11.5.1 Modellierung von Integration in Krankenhausinformationssystemen

Interoperabilität durch eine gemeinsame Darstellung der Krankenhausinformationssystem-Architekturen über eine UML-basierte Ontologie [WBW01b, WBW01a]. Erweiterung des Metamodells um weitere Domänen, z.B. Master-Slave-Domänen wie Datendomäne, Semantikdomäne, Kontextdomäne und Funktionsdomäne und Middleware-Domänen. Eine Bewertung der Integration erfolgt über den Grad der Überlappung dieser Domänen. [Wen05] beschäftigt sich mit allen drei Architekturebenen (Anwendungs-, Technologie- und Geschäftsarchitektur) und hat den Schwerpunkt des Modellierens der Integration.

11.6 Tabellarische Zusammenfassung

	SCIPHOX	MIRA plus	Eichelberg	Riesmeier	Häberlin	Wolff	eMamma Akte	Wendt	Eigener Ansatz
Schwerpunkt Anwendungsarchitektur	●	●		●	●	●	●	●	●
Schwerpunkt Technologiearchitektur		●	●			●		●	
Schwerpunkt Geschäftsarchitektur								●	
Einsatz von Domänenstandards	●	●	●	●	●	●	●		●
Flexible Kombination von Standards							●		●
Schwerpunkt Integration	●	●				●	●	●	●
TOP-DOWN-Integration									●
JO-JO-Integration	●				●	●	●		
BOTTOM-UP-Integration									
XML als gemeinsame Syntax	●					●	●		●
Verwendung von Mediatoren									●
Schwerpunkt Modellierung								●	
Beschränkt auf Teilgebiet						●			

Abbildung 11.1: Eine tabellarische Übersicht der verwandten Ansätze aus der Medizininformatik einschließlich des eigenen Ansatzes

Teil V

Zusammenfassung und Ausblick

12 Zusammenfassung

Domänenspezifische Standards für das Gesundheitswesen wurden analysiert, gleichförmig strukturiert und zueinander in Beziehung gesetzt [PH03]. Dabei wurden in Begriffssystemen für Kommunikation und Dokumentation Metadaten zusammengestellt, die die relevanten Standards und ihre Beziehungen zueinander beschreiben. Diese Begriffssysteme erfassen jeweils die Gemeinsamkeiten der Standards. Da dies nur auf einer übergeordneten Ebene stattfinden kann und die Standards zum Teil stark differieren, können diese Zusammenhänge nicht bei den eigentlichen Transformationen helfen. Es existiert kein gemeinsames, mehrere Standards umfassendes Datenschema.

Die in dieser Arbeit vorgestellte mediator-basierte Architektur ist durch ihren modularen, schichtenorientierten Aufbau und dem Konzept der Mediatoren eine flexible und erweiterbare Software-Architektur. Ferner erhöht dieser Ansatz die Skalierbarkeit, indem eine Trennung der Verwaltung der globalen Modelle und der Integration von Komponentenmodellen in domänenspezifische Modelle erreicht wird. Durch eine einheitliche Spezifikation der relevanten Kommunikations- und Dokumentenstandards, quasi als Instanz des MOF-Modells, sowie eine geeignete Kombination beider, stehen geeignete Metamodelle als Entwurfshilfe für Transformationen zwischen den verschiedenen Standards zur Verfügung. Sie stellen einen geeigneten Leitfaden zur Strukturanalyse der Standards und deren Abbildungen dar und unterstützen den Entwickler bei der Erstellung der Abbildungen dadurch, dass er weiß, auf welche Kriterien er bei der Erfassung der Korrespondenzen zu achten hat. Mit Hilfe von domänenspezifischen Standards für Kommunikation und Dokumentation sowie entsprechenden Mediatoren, die zwischen den einzelnen Standards mit Hilfe von Abbildungsspezifikationen vermitteln, kann eine institutionsübergreifende Interoperabilität gelingen. Die Architektur bietet die in Abschnitt 7.1 geforderte Skalierbarkeit, Evolution und Flexibilität.

Die Architektur und die in Abschnitt 7.3 vorgestellten Metamodelle dienen als Grundlage für eine prototypische Evaluierung am Beispiel des Epidemiologischen Krebsregisters Niedersachsen (Teil III).

Bei der Evaluation der Architektur gestaltete sich das Erstellen des HL7-Domänenmodells, ausgehend vom generischen HL7-RIM, als aufwändigster Schritt. Die dann eingesetzte HL7-DTD der Version 2.4 war gut zu erstellen, allerdings nur mit Hilfe der Z-Segmente aus [Hae99]. Die Abbildung CDA und Basisdokumentation für Tumorkranke war durch das OnkoDok-Modell gut zu erstellen. Auch die Abbildung zwischen dem HL7- und dem BDT-Domänenmodell war kein Problem. Allen Abbildungen gemeinsam ist jedoch, dass

sie nicht vollständig erstellt werden konnten. Die Standards müssen zum einen weiterentwickelt werden, um auch den Anforderungen in einem etwas abseits des ärztlichen Alltags liegenden Bereich gerecht zu werden, zum anderen ist deren ständige Evolution bei Abbildungen ein Problem. So hätte beispielsweise die Abbildung zwischen der eingesetzten HL7-DTD und der vom EKN zur Verfügung gestellten BDT-DTD nahezu komplett gelingen können, wenn nicht zwei unterschiedliche Versionen der Basisdokumentation für Tumorkranke zugrunde gelegen hätten.

Mit der Software XMLSPY konnten die fehlenden DTDs von HL7 sowie der Basisdokumentation für Tumorkranke bequem erstellt werden. Beim Import der jeweiligen DTDs in MapForce, war es möglich, Beispieldateien anzugeben, um die spezifizierte XSL-Transformation sofort anzusehen. Ferner sind die importierten DTDs über XMLSPY veränderbar, was auch sofort in MapForce aktualisiert wird. Die Applikationen MapForce und XMLSPY arbeiten damit komfortabel zusammen. Die Abbildungen konnten durch graphische Verknüpfungen einfach realisiert werden. Auch das Einsetzen von Filtern kann graphisch per Maus vorgenommen werden. Dabei war ein Vorgehen nach dem Leitfaden möglich. Die Korrespondenzen, die sich aus dem Metamodell ableiten ließen, waren in der Regel direkt umsetzbar, da die Baumstruktur der DTDs in MapForce gut dargestellt wird. Wenn die größeren Arbeitspäckchen verknüpft waren, konnte eine detaillierte Zuordnung erfolgen, indem in beiden DTD-Bäumen die entsprechende Struktur ausgeklappt wurde. Nach anfänglichen Installationsschwierigkeiten von e*Gate wegen des Microsoft Windows XP Service Packs 2, war es anschließend kein Problem die erstellten XSLT-Dateien e*Gate zu übergeben.

Der vorgestellte Leitfaden zur Strukturanalyse domänenspezifischer Standards ermöglicht auch in anderen Domänen einen leichteren Zugang für Softwareentwickler zu den Standards der Domäne. Durch eine einheitliche Modellierung in UML werden die Strukturen der einzelnen Standards klar visualisiert. Die Begriffe der Domänenstandards können mit einem Metamodell bzw. Begriffssystem zu den einheitlich spezifizierten Modellen semantisch eindeutig abgegrenzt werden. Zusätzlich können mit Hilfe des Metamodells Beziehungen zwischen den verschiedenen Standards, die Grundlage für die Transformationen zwischen ihnen sind, schneller erkannt werden. Konkrete Transformationsregeln können, wie am Evaluationsbeispiel gezeigt, mit Hilfe des vorliegenden Metamodells, der Modelle und den aufgedeckten Beziehungen zügiger erstellt werden. Die in dieser Arbeit vorgestellte mediator-basierte Architektur ist grundsätzlich in allen Domänen einsetzbar, bei denen zwischen diversen Standards vermittelt werden muss, um Interoperabilität zu erreichen. Nachdem in Teil I die domänenunabhängigen Grundlagen vorgestellt worden sind, befasst sich in Teil II das Kapitel 6 mit der Analyse der domänenspezifischen Standards am Beispiel Medizin. Die in Abschnitt 6.1 ausgeführte Bedeutung von Standards bei der Interoperabilität gilt allerdings domänenunabhängig. Kapitel 7 beschreibt zunächst domänenunabhängig in Abschnitt 7.1 die erweiterte Mediator-basierte Architektur, bevor diese in Abschnitt 7.2 auf die Domäne Gesundheitswesen beispielhaft angewendet wird.

Wird um einen weiteren Domänenstandard erweitert, so muss dieser zunächst ebenfalls in UML modelliert werden. Ausgehend vom zugehörigen Metamodell kann nach Beziehungen zwischen den Standardstrukturen gefahndet werden. Diese Beziehungen erlauben

dann einen schnelleren Zugang zu den konkreten Transformationsregeln in einem konkreten Kontext.

Aus dieser Arbeit lassen sich indirekt Empfehlungen für die Weiterentwicklung von Standards insofern ableiten, dass zunehmende Gemeinsamkeiten von Standards das zugehörige Metamodell bzw. Begriffssystem konkreter und detaillierter werden lassen würde. So könnten vor allem genauere Spezifikationen im Bereich der Dokumentation zu einem geeigneteren Metamodell bzw. Begriffssystem für die Dokumentationsstandards führen. Damit würden die Möglichkeiten für umfassendere Transformationsregeln entstehen. Wären diese Transformationsregeln schon anhand der allgemeinen Struktur zahlreich, wäre dies ein Grad für eine Konvergenz der Standards.

Nachdem sich die Kassenärztliche Bundesvereinigung (KBV) aktiv an der Arbeitsgemeinschaft SCIPHON beteiligt und bereits erste Spezifikationen, die auf der CDA beruhen, veröffentlicht hat [KBV03b, SCI03], könnte der BDT-Standard mittelfristig verschwinden. Jedoch muss auch der „Weg“ bis zu einer möglichen Konzentration auf HL7 und CDA bewältigt werden. Zudem werden neue Standards bzw. neue Versionen bestehender Standards entwickelt werden. Probleme kann es bei einer unterschiedlichen, manchmal mangelhaften Umsetzung der Standards geben, wie sie die medizinischen Bildkommunikationssysteme beim Standard DICOM (Digital Imaging and Communications in Medicine) aufweisen [Eic01]. Die vorgestellte Architektur wird daher nicht wesentlich tangiert, wenngleich natürlich die Abbildungsspezifikationen und -regeln jeweils angepasst werden müssen. Durch die „Umwege“ über entsprechende Standards entsteht zwar zunächst ein gewisser Overhead, sie erleichtern aber auch den Übergang zu modernen Systemen, die die Standards unterstützen [Has00b]. Bei standardkonformen Systemen verschwindet dieser Overhead.

13 Ausblick

In Abschnitt 5.2 auf Seite 39 sind in einer Übersicht verschiedene Modelltransformationsansätze vorgestellt worden. Da die Ansätze unterschiedliche Schwächen aufweisen, wird vermehrt versucht, hybride Ansätze zu entwickeln und zu evaluieren, um diese Schwäche auszugleichen. Solch weiterentwickelte Ansätze von Modelltransformation würden sicherlich nutzbringend einsetzbar sein, wenn es dadurch gelänge, die Beziehungen zwischen diversen Standards, die sich aus den UML-Modellen und dem zugehörigen Metamodell ergeben, technisch noch verwertbarer zu modellieren.

Ziel könnte ferner sein, auf Basis der vorhandenen UML-Metamodelle, ein übergeordnetes (UML-) Sprachprofil zu erstellen und dieses in der Softwareumgebung der Nutzer (bidirektional) zu reflektieren. Die Metamodelle für die Kommunikations- und Dokumentationsstandards könnten dabei als grobgranulare, strukturelle Basis dienen und durch ein zu implementierendes Werkzeug software-technisch, interaktiv verfeinert werden.

Frage ist, ob nicht schon die jeweiligen Domänenmodelle der verschiedenen Standards entsprechende standardisierte Metadaten enthalten könnten, die für die Mediatoren die notwendigen Informationen für die jeweiligen Transformationsregeln mitbringen.

Unterschiedliche Matching-Techniken, die durch Abgleich die Korrespondenzen identifizieren sollen, sind von [vdHR05] in einer Taxonomie erfasst worden, die Abbildung 13.1 zeigt. Danach werden zunächst Einzel- und kombinierte (hybride) Matching-Ansätze unterschieden. Bei den hybriden Ansätzen können ein oder mehrere der Einzel-Matching-Ansätze sequentiell oder parallel kombiniert werden. Die sequentielle Kombination kann als Filterprozess in Stufen betrachtet werden, der mit jedem Schritt das Matching verfeinert. So könnten z.B. zunächst semantische Korrespondenzen identifiziert werden und bei großer semantischer Ähnlichkeit nach strukturellen Aspekten abgeglichen werden. Die parallele Kombination ermöglicht dagegen eine Analyse mehrerer Kriterien gleichzeitig durch das gleichzeitige Ausführen unterschiedlicher Matching-Techniken. Einzel-Matching-Ansätze sind semantisches und strukturelles Matching, Metamodell-Matching und AI-Matching. Semantisches Matching führt einen Abgleich nach semantischen Kriterien durch, um semantische Korrespondenzen zu ermitteln. Strukturelles Matching orientiert sich beim Matching an der Struktur. Werden Informationen über Korrespondenzen von einem Metamodell geliefert, liegt das Metamodell-basierte Matching vor. Bei AI-Matching sollen neuronale Netze bei der Erkennung ähnlicher Muster helfen.

Diese Arbeit liefert einerseits Metamodelle für ein Metamodell-basiertes Matching, andererseits ist im Detail nur ein manuelles semantisches Matching durchführbar.

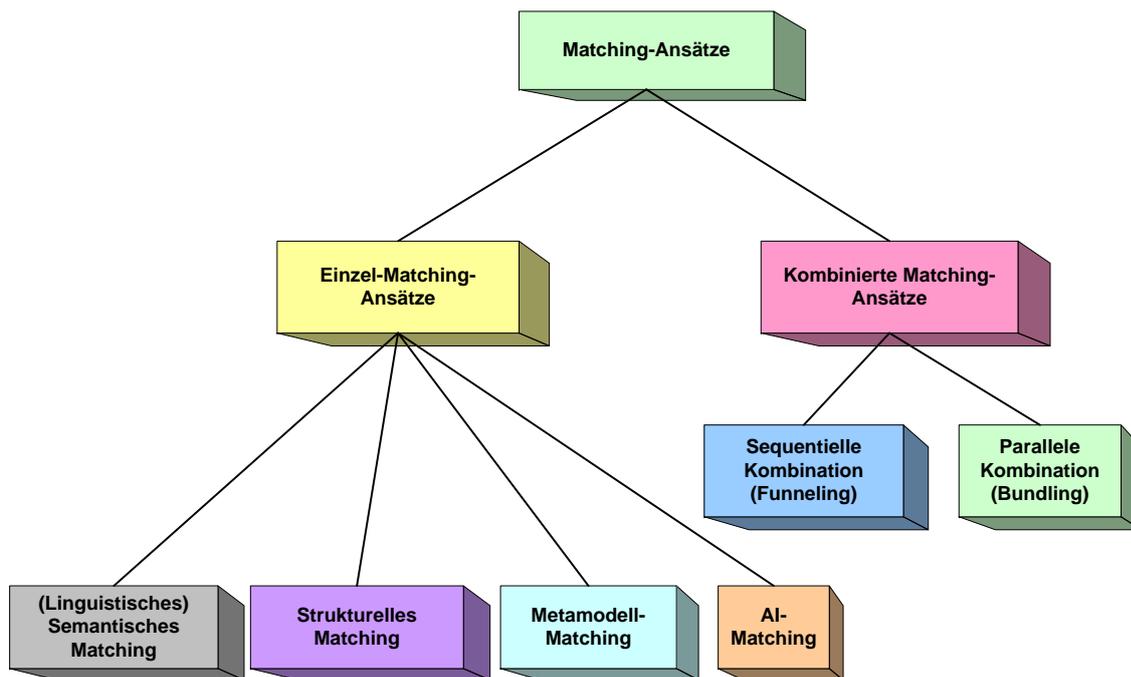


Abbildung 13.1: Eine Matching-Taxonomie aus [vdHR05]

Der vorgestellte Ansatz könnte durch eine formale Sprache automatisierbare Korrespondenzen aus den Metamodellen nutzen, wie dies mit den Techniken des Semantic Web bereits kurz angedeutet worden ist. Dies würde in der Konsequenz auch eine andere Matching-Technik (semantisches Matching auf der Basis einer Ontologie) ermöglichen. So könnte untersucht werden, ob der vorgestellte Ansatz nicht auch mit anderen Matching-Techniken angewandt werden kann.

Teil VI

Anhänge

A Föderierte Informationssysteme

Es folgt eine Zusammenstellung von Terminologie und Architekturen aus [BKLW99], um die wichtigsten Grundlagen auf diesem Gebiet für diese Arbeit vorzustellen.

A.1 Einleitung

Aktuell in der Diskussion ist eine neue Generation von Softwaresystemen, die so genannten „föderierten Informationssysteme“. Ihre Hauptcharakteristik ist, dass sie als integrierende Schicht über existierende Alt-Anwendungen und -Datenbanken konstruiert werden. Allgemein können sie nach drei Dimensionen klassifiziert werden:

- Grad der Autonomie der integrierten Komponenten
- Grad der Heterogenität zwischen den Komponenten
- Verteilung

Zunächst soll eine Klassifikation der Informationssysteme bezüglich der Dimensionen Autonomie, Heterogenität, Verteilung und Flexibilität, in Bezug auf die Weiterentwicklung, vorgenommen werden. Dies führt dann zu einer ersten Definition von „Föderierten Informationssystemen“, bei denen insbesondere die Heterogenität eine wichtige Rolle spielt. Sie kann mit dem Einsatz von Metadaten-Konzepten gelöst werden.

A.2 Dimensionen für die Klassifikation von Informationssystemen

A.2.1 Autonomie

Wenn ein Informationssystem auf mehreren Komponenten basiert, kann die Autonomie, die integrierte Komponenten behalten können, ein kritischer Punkt sein. [BKLW99] unterscheidet dabei Entwurfs-, Kommunikations- und Ausführungs-Autonomie. Entwurfs-Autonomie meint, dass eine Komponente von anderen in ihrem Entwurf unabhängig ist

bzgl. Datenmodell etc. Kommunikations-Autonomie ist gegeben, wenn eine Komponente unabhängig entscheiden kann, mit welchen anderen Systemen sie kommuniziert. Innerhalb von Föderationen meint Kommunikationsautonomie, dass jede Komponente die Föderation jederzeit verlassen bzw. in sie eintreten kann. Ausführungs-Autonomie bedeutet die Unabhängigkeit der Komponente bei der Ausführung und dem Scheduling hereinkommender Anfragen. Es ist unmöglich Ausführungs-Autonomie zu erhalten, wenn eine globale Transaktionsverwaltung involviert ist.

A.2.2 Heterogenität

Heterogenität ist natürlich auf die Tatsache zurückzuführen, dass eine autonome Entwicklung von Systemen immer unterschiedliche Lösungen ergibt, z.B. aufgrund eines anderen Verständnisses und Modellierung der Objekte der realen Welt. [BKLW99] unterscheidet syntaktische, datenmodellbasierte und logische Heterogenität mit verschiedenen Untertypen:

A.2.3 Syntaktische Heterogenität

- Technische Heterogenität. Sie bezieht sich auf Unterschiede in technischen Aspekten, wie Hardware-Plattformen, Betriebssysteme u.ä.
- Schnittstellen-Heterogenität existiert, wenn auf verschiedene Komponenten über verschiedene Zugriffssprachen zugegriffen werden kann. Dabei ist nicht der technische Aspekt gemeint, sondern die Restriktionen der möglichen Zugriffsmethoden.

A.2.4 Datenmodellbasierte Heterogenität

Diese Art von Heterogenität betrifft die Tatsache, dass unterschiedliche Datenmodelle eine unterschiedliche Semantik für ihre Konzepte haben. So kennt z.B. das relationale Modell keine Vererbung im Gegensatz zum objekt-orientierten Modell. Obwohl diese Heterogenität ein semantisches Problem ist, wird sie getrennt davon behandelt. Bei so genannten mediator-basierten Systemen wird die Datenmodelldiversität von so genannten Wrappern gekapselt, nicht aber die semantische Heterogenität. Die klassische Fünf-Schichten-Architektur für föderierte Datenbanksysteme hat eine separate Schicht, um Datenmodelle zu transformieren (Komponenten-Schema).

A.2.5 Logische Heterogenität

- Semantische Heterogenität betrifft die Semantik von Daten und Schemata. Obwohl Schemata im gleichen Modell formuliert sind, können sie unterschiedliche Semantik haben. Bei einem Schema mit Relationen und Attributen z.B. haben diese Relationen und Attribute Namen und tragen eine implizite Semantik, welche der Begriff ist, für den sie stehen. Das Schema enthält nun aber nur den Namen, jedoch nicht den Begriff. Die Interpretation von Namen durch verschiedene Personen muss sich nicht notwendigerweise decken. Daher können unterschiedliche semantische Konflikte auftreten: gleiche Bezeichnungen können verschiedene Begriffe bedeuten (Homonyme), unterschiedliche Bezeichnungen den gleichen Begriff (Synonyme), unterschiedliches Verständnis von einem Begriff selbst, usw.. Attribute können auch dieselbe Semantik haben, aber unterschiedliche Einheiten (z.B. EURO oder USD für Preise). Die Semantik von Datenwerten ist definiert durch das Schemaelement, unter dem sie repräsentiert ist. Das Schema ist der Hauptort die Semantik einer Datenquelle zu verschlüsseln.
- Schemabasierte Heterogenität ist das Verschlüsseln von Begriffen mit verschiedenen Elementen eines Datenmodells. Im relationalen Modell existieren drei Typen solcher Konflikte: Relation \leftrightarrow Attributname, Attributname \leftrightarrow Attributwert und Relation \leftrightarrow Attributwert.
- Strukturelle Heterogenität existiert, wenn Elemente die gleiche Bedeutung haben, im selben Datenmodell modelliert sind, schematisch homogen sind, aber unterschiedlich strukturiert sind.

A.2.6 Verteilung

Ein drittes Problem ist die physische Verteilung von Datenquellen. Seit Netzwerke und insbesondere das Internet eine Kommunikation zwischen Computern ermöglichen, ist es naheliegend, an das Kombinieren von Anwendungen und Datenquellen zu denken, die physisch getrennt voneinander liegen. Verteilung wird von einer Reihe von Techniken beherrscht, wie z.B. http oder CORBA. Zum Beispiel können bei Einsatz von CORBA Anwendungen entwickelt werden, die die physische Lokalisation von Komponenten zu einem hohen Grade ignorieren können. Auf CORBA ist in dieser Arbeit nicht näher eingegangen worden.

A.3 Klassifikation von Informationssystemen

Basierend auf den Dimensionen Verteilung und Heterogenität kann man drei allgemeine Klassen von Informationssystemen trennen, wie dies in Abbildung A.1 mittels UML dargestellt ist:

1. Single Information System (SIS)
2. Distributed Information System (DIS)
3. Heterogeneous Information System (HIS)

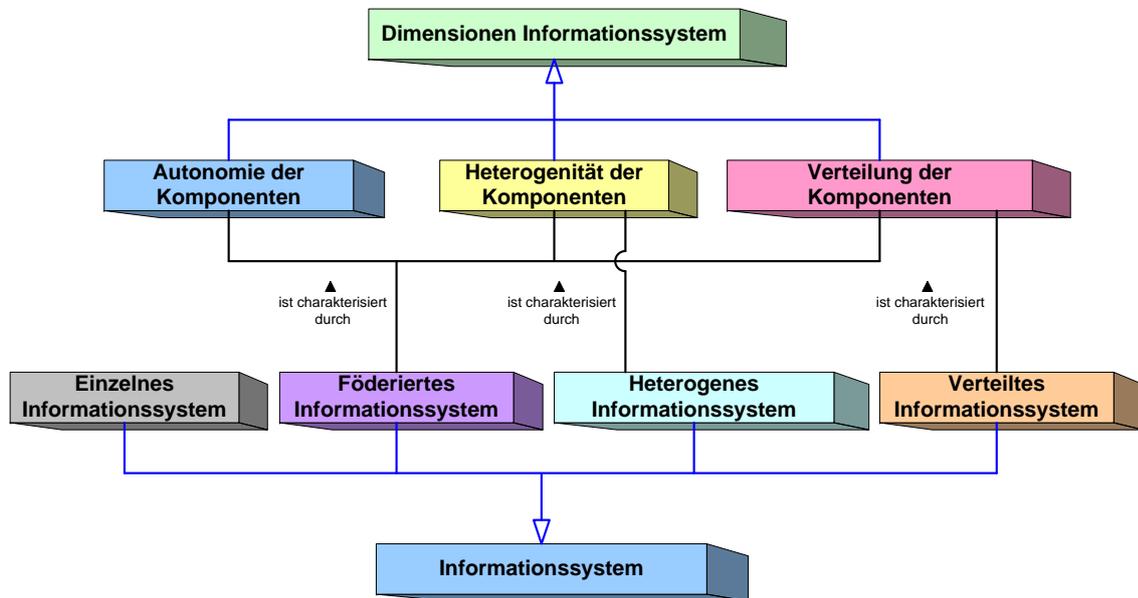


Abbildung A.1: UML-Diagramm einer Klassifikation für Informationssysteme

Ein SIS läuft als eine monolithische Anwendung auf einem Computer. Bei einem DIS sind die Daten physisch über mehrere Orte verteilt, die durch irgendein Kommunikationsnetzwerk miteinander verbunden sind. Ein HIS ist eine Sammlung von Informationssystemen, die sich in syntaktischen oder logischen Aspekten wie Hardware-Plattform, Datenmodell oder Semantik unterscheiden.

Wenn man den Komponenten zusätzlich noch Autonomie gibt, kommt man zu folgender Definition föderierter Informationssysteme (Federated Information System = FIS): Ein FIS enthält eine Menge von heterogenen und autonomen Informationssystem-Komponenten, die Teilnehmer dieser Föderation.

Anwendung und Nutzer greifen auf eine Menge von heterogenen Datenquellen über eine Föderationsschicht bzw. Integrationsschicht zu, die einen einheitlichen Zugang zu den gespeicherten Daten in den Datenquellen bietet (siehe Abbildung 3.1). Zum Beispiel kann diese Schicht ein föderiertes bzw. integriertes Schema anbieten, eine einheitliche Anfragesprache oder eine einheitliche Menge von Quell- und Inhaltsbeschreibungen als Metadaten.

Zudem müssen zukünftig entwickelte Systeme flexibel dem stetigen Wechsel und der Weiterentwicklung der Komponenten gerecht werden. Ein wichtiges Konzept, dies zu erreichen sowie Heterogenität und Verteilung zu begegnen, ist das Konzept der Metadaten.

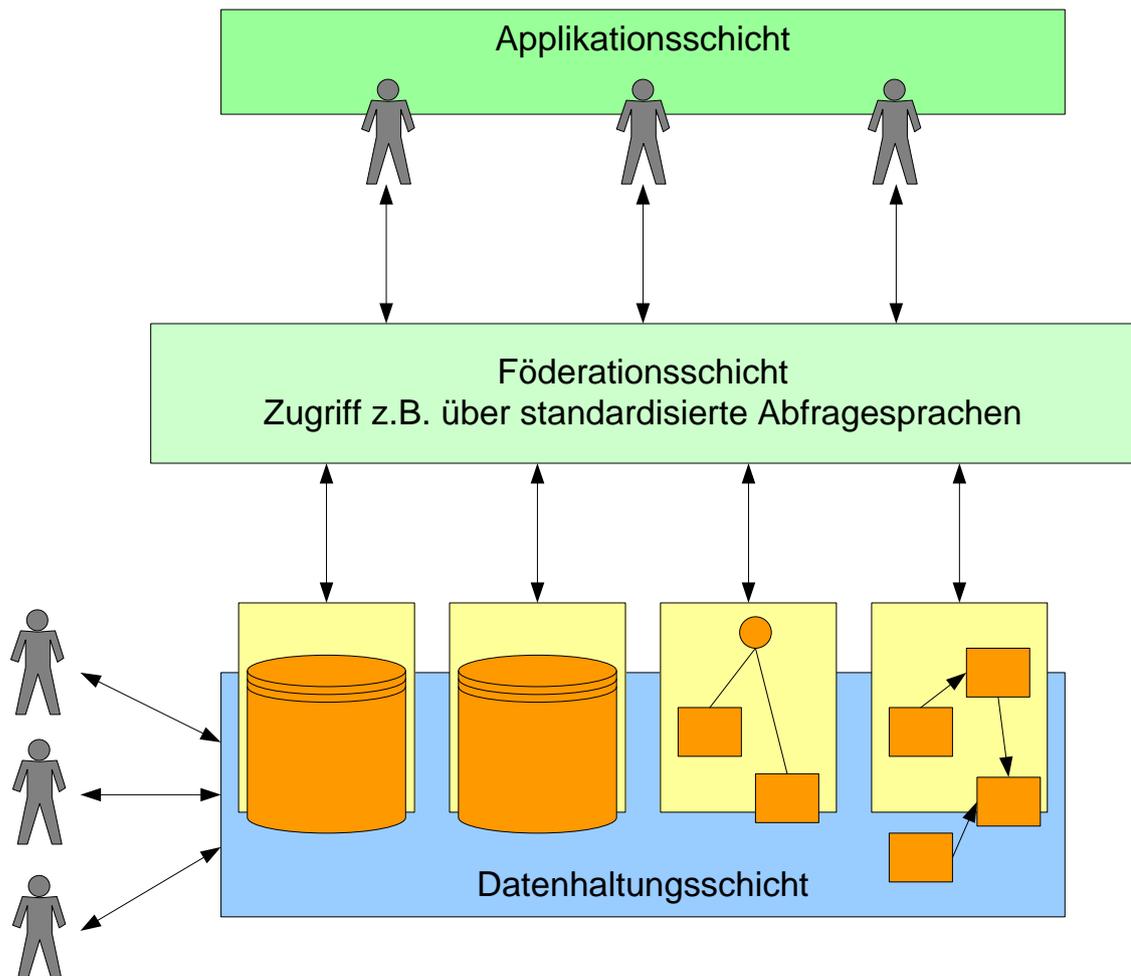


Abbildung A.2: Schichten-Architektur eines Föderierten Informationssystems aus [Wil02]

A.3.1 Metadaten für föderierte Informationssysteme

[BKLW99] unterscheidet folgende Arten von Metadaten bezüglich föderierter Informationssysteme, speziell der Föderationsschicht:

1. **Technische Metadaten** beschreiben Informationen bezüglich der technischen Zugriffsmechanismen von Komponenten, wie z.B. das Protokoll etc. Sie werden eingesetzt um technische und Schnittstellen-Heterogenitäten zu überbrücken.
2. **Logische Metadaten** beziehen sich auf Schemata und deren logische Beziehungen. Logische Metadaten sind z.B. erreichbar durch die Data Dictionaries in RDBMS oder als Klassendiagramme in OODBMS. Insbesondere sind die Beziehungen und Abhängigkeiten zwischen verschiedenen Schemata eines Datenmodells eine wichtige Art von logischen Metadaten in FIS.
3. **Metamodelle** als Metadaten unterstützen die Interoperabilität von Schemata in verschiedenen Datenmodellen. Sie beziehen sich hauptsächlich auf Datenmodellheterogenität.
4. **Semantische Metadaten** sind Informationen, die helfen, die Semantik von Begriffen zu beschreiben. Insbesondere Ontologien und Thesauren werden für diesen Zweck eingesetzt. Alle domänenspezifischen Beschreibungen gehören in diese Klasse.
5. **Qualitätsbezogene Metadaten** beschreiben quellenspezifische Eigenschaften von Informationssystemen bezüglich deren Qualität, wie z.B. Zuverlässigkeit, Aktualität usw..
6. **Infrastruktur-Metadaten** helfen Nutzern relevante Daten zu finden, wie z.B. Navigationshilfe, kommentierte Lesezeichen usw..
7. **Nutzerbezogene Metadaten** beschreiben Verantwortlichkeiten und Präferenzen von Nutzern, z.B. Nutzer-Profile.

Ein FIS nutzt typischerweise Metadaten, die üblicherweise als Daten in einem Repository abgespeichert werden. Nicht näher vertieft werden in dieser Arbeit die technischen, qualitätsbezogenen, infrastruktur- und nutzerbezogene Metadaten.

A.4 Klassifikationskriterien für FIS

[BKLW99] unterscheidet FIS z.B. nach der Art der Komponenten, die integriert werden können, nach dem Grad von Transparenz, die für einen Nutzer erreicht wird, nach dem Grad der semantischen Integration sowie nach der allgemeinen Entwicklungsmethodologie. Es handelt sich dabei um keine orthogonalen Kriterien. Abbildung A.3 zeigt diese Zusammenhänge in einem UML-Diagramm.

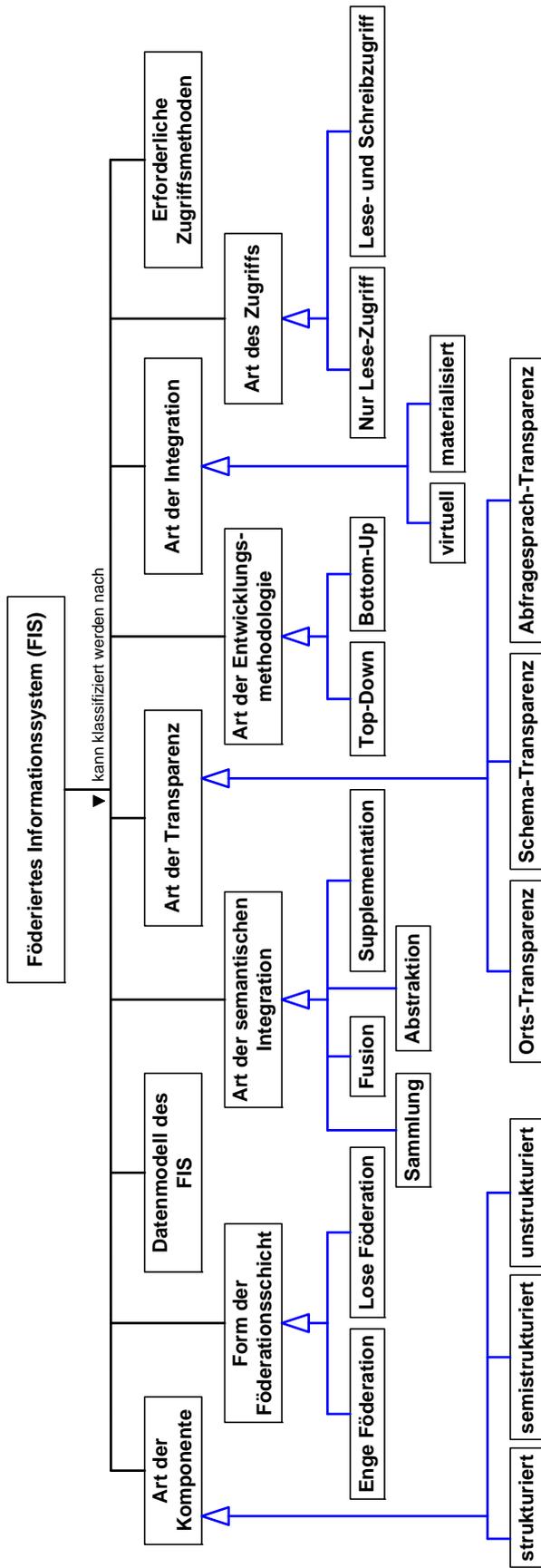


Abbildung A.3: UML-Diagramm von Klassifikationskriterien für Föderierte Informationssysteme

A.4.1 Arten von Komponenten: strukturiert, semi-strukturiert, unstrukturiert

FIS unterscheiden sich in den Typen von Komponenten, die sie integrieren können. Ein FIS kann die Integration von strukturierten, semi-strukturierten und unstrukturierten Komponenten erlauben oder nicht. Strukturierte Quellen haben nach dem Verständnis von [BKLW99] ein vordefiniertes Schema. Alle Datenitems sind im Vorwege definiert durch das Schemaelement, zu dem es Bezug hat. Des Weiteren diktiert das Schema das Format von großen Datenitems; Items, die nicht in das Schema passen, können nicht in die Datenmenge integriert werden.

Eine semi-strukturierte Datenquelle hat eine Struktur, aber diese Struktur ist nicht vordefiniert in Form eines strikten Schemas. Daher muss jedes einzelne Datenitem seine eigene Semantik-Definition tragen, üblicherweise in Form eines Labels. Zu einem gegebenen Zeitpunkt kann die Summe aller Labels einer semi-strukturierten Datenquelle als sein Schema angesehen werden. Wie auch immer, dieses Schema kann sich potentiell jederzeit durch Hinzufügen neuer Daten ändern, während in strukturierten Quellen eine Schemaänderung nicht so häufig auftritt.

Unstrukturierte Datenquellen haben keinerlei Struktur, wie z.B. Textdokumente.

A.4.2 Enge versus lose Föderation

Bei der Form der Föderationsschicht kann man hauptsächlich zwei Klassen unterscheiden: enge und lose Föderationen. Eine enge Föderation hat ein einheitliches, globales Schema, welches das Zugriffsschema für jeden Nutzer darstellt. Eine lose Föderation hat kein solches Schema, aber bietet eine einheitliche Sprache, um im Dateninhalt der Komponenten zu suchen. Daher bieten enge Föderationen eine Schema-, eine Sprach- und Schnittstellen-Transparenz, während lose Föderationen nur letzteres bieten.

A.4.3 Enge Föderation

Eine enge Föderation bietet ein einheitliches Schema (integriertes oder föderiertes Schema) als Zugriffsschnittstelle auf die Föderation (Abbildung A.4). In jedem Fall muss die semantische Essenz des föderierten Schemas eine Untermenge der Vereinigung der semantischen Essenzen der Komponenten-Schemata sein. Mit semantischer Essenz meinen wir eine Menge von Begriffen aus der realen Welt, die durch ein Schema beschrieben werden. Die Hauptaufgabe beim Einsatz eines föderierten Schemas ist die Auflösung und

Handhabung der logischen Heterogenität in den Quellschemata. Es muss beides betrachtet werden während der Schemaintegration und während der Suchanfragen. Um semantische Äquivalenz zu garantieren, muss das FIS über Korrespondenzen (logische Metadaten) zwischen Anfragen, föderierten und Komponenten-Schemata Bescheid wissen. Diese Korrespondenzen können z.B. ausgedrückt werden durch ontologische Beschreibungen oder Regeln. Sie können definiert werden durch Menschen, durch eine Sprache (MoCa oder BRITY) oder automatisch eingebracht werden. Enge Föderationen sind komfortabel für Nutzer des FIS, da sie nicht über die Schemata aller Komponenten Bescheid wissen müssen, sondern nur über das föderierte Schema. Auf der anderen Seite müssen sich Nutzer einer engen Föderation auf den Übersetzungsmechanismus verlassen und daher auf die Korrespondenzen. Daher sollten die Korrespondenzen durch einen Domänenexperten definiert werden. Der Einsatz und das Anbieten globaler Schemata ist dann essentiell, wenn sehr viele Quellen vorhanden sind, wo es unmöglich ist, von jedem Nutzer zu erwarten, dass er alle im Detail kennt, oder wenn Quellen häufig ihre Schemata weiterentwickeln und Nutzer nicht allen Änderungen folgen können. Weiterhin erfordern bestimmte Situationen unvermeidlich den Einsatz eines globalen Schemas, nämlich dann, wenn Standard-Schemata genutzt werden (z.B. STEP Schemata, OMGs Domänenstandards oder im Falle epidemiologischer Krebsregister die Basisdokumentation für Tumorkranke). In diesen Fällen müssen existierende Komponenten in dieses Standard-Schema eingepasst werden.

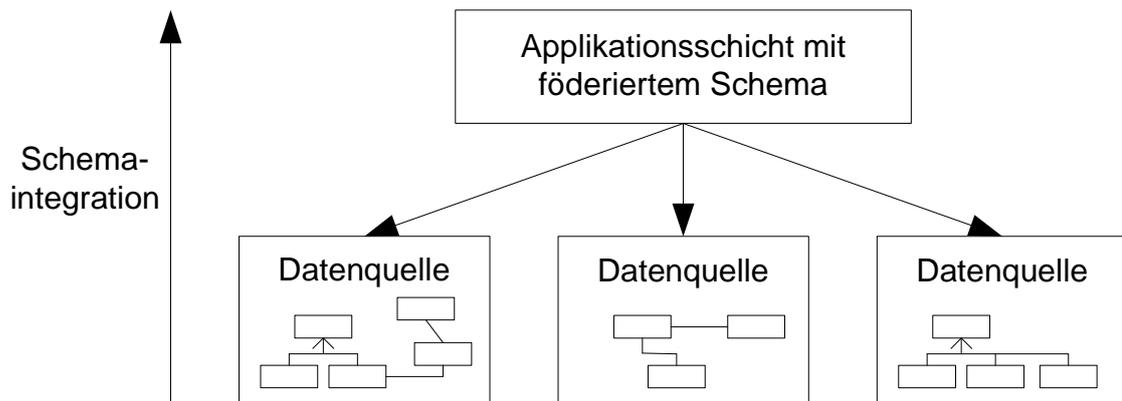


Abbildung A.4: Schemata in eng integrierten FIS

A.4.4 Lose Föderationen

Lose Föderationen bieten kein einheitliches Schema für Suchanfragen gegenüber der Föderation (Abbildung A.5). Aber sie bieten eine einheitliche Anfragesprache (Multidatabase Query Language, MDBQL), die von der Anfragesprache der Komponenten abstrahiert und technische und sprachbasierte Heterogenität versteckt. Daher ist jeder Nutzer selbst verantwortlich für die Handhabung der logischen Heterogenität in den Komponenten. Um

die schemabasierte Heterogenität abzudecken, braucht die MDBQL die Möglichkeit über die Schemaelemente zu rangieren, als wären sie Daten, was zum Beispiel mit SQL nicht möglich ist. Lose Föderationen können nur erstellt werden, wenn die Datenquellen selbst einen Zugriff über Anfragesprache anbieten. Um zu vermeiden, dass die Last des Integrationsprozesses komplett auf den Nutzer geschoben wird, nutzen Systeme, basierend auf MDBQL, oft integrierende Ansichten. Dies bedeutet, dass Nutzer Views auf den Komponenten definieren und sie zugänglich für andere Nutzer machen können, die sie dann nutzen, als wären es globale Relationen.

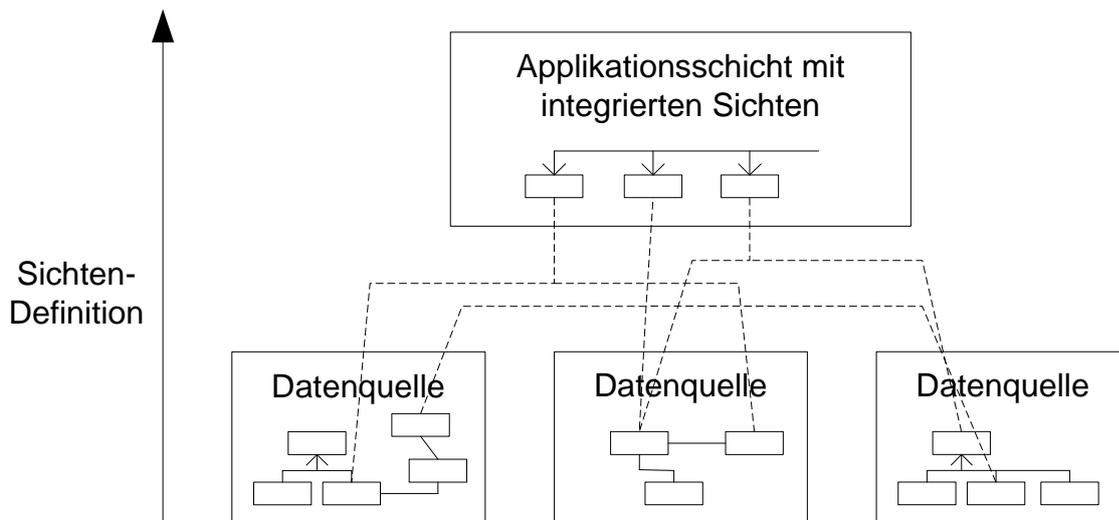


Abbildung A.5: Lose integrierte föderierte Informationssysteme

Lose integrierte Systeme, die umfangreich integrierte Views nutzen, die sie vielleicht durch Domänenexperten haben definieren lassen, können als eng integrierte FIS angesehen werden. Das Hauptkriterium, enge von loser Föderation zu unterscheiden, ist daher, ob oder ob nicht die Komponenten-Schemata vor den Nutzern versteckt sind. In einer engen Integration sind die Quell-Schemata nicht länger sichtbar; in einer losen Integration sind Quell-Schemata nach wie vor sichtbar, obwohl zusätzliche Sichten eingesetzt werden können, um auf sie in einer vorgefertigten Weise zuzugreifen.

A.4.5 Modellierungssprache des FIS

Die Föderationsschicht eines FIS muss auf einem spezifischen Datenmodell basieren, genannt kanonisches (von allen anerkanntes) Datenmodell oder gemeinsames Datenmodell. Die Schemata von engen Föderationen sind Schemata in diesem Modell. In losen Föderationen basiert nur die Anfragesprache, die für den Zugriff auf die Datenquellen genutzt wird, auf diesem Datenmodell. Das Datenmodell beschränkt inhärent die Art von Komponenten, die in das FIS integriert werden können, was auf fehlende Übersetzungsmöglichkeiten zwischen einigen Modellen zurückzuführen ist. Inkompatibilitäten entstehen, wenn

semi-strukturierte Komponenten in strukturierte Daten integriert werden sollen. Probleme tauchen ebenfalls auf, wenn semantisch reichere Modelle in ärmere Modelle integriert werden sollen. Zum Beispiel ist die Integration objekt-orientierter Schemata in ein relationales Schema nur möglich mit einem Verlust an semantischem Wissen.

A.4.6 Arten von semantischer Integration

Die Integrationsschicht kapselt die Datenquellen für die Nutzer. Im Gegensatz zu lose gekoppelten FIS, wo die Daten im Prinzip unverändert gegenüber der Datenquelle gesammelt sind, können Föderationsdienste eines eng gekoppelten FIS eine bessere semantische Integration unterstützen. Wir unterscheiden unterschiedliche Arten von semantischer Integration auf der Datenebene:

- **Sammlung:** Daten von Komponenten werden unverändert gesammelt ohne Übereinstimmungen von äquivalenten Datenobjekten verschiedener Quellen zu beachten.
- **Fusion:** Die Integration von Komponentendaten wird durch eine einfache Extraktion vorgenommen (ausdrückbar mit einer Anfrage an das Komponenten-Schema); keine weiteren abstrahierenden Berechnungen werden gemacht. Aber im Gegensatz zum Sammlungs-Ansatz, funktioniert die Objektfusion, um semantisch äquivalente Entities von verschiedenen Quellen zu identifizieren. Weiterhin versucht das FIS eine konsistente Repräsentation festzulegen, z.B. wenn Quellen widersprüchliche Werte für das gleiche Datenitem berichten, wie z.B. unterschiedliche Sozialversicherungsnummern für die gleiche Person, nutzt das FIS Regeln, um diesen Konflikt zu beseitigen. Bedenke, dass Datenfusion sehr schwierig ist; häufig ist es unmöglich, Objekte zu identifizieren oder zu entscheiden, welcher Datenwert der richtige ist.
- **Abstraktion:** Hierbei basieren die föderierten Daten auf extrahierten Daten der Komponenten, aber weitere Funktionen können angewendet werden, um die Quelldaten auf einen höheren oder niedrigeren Abstraktionslevel des föderierten Schemas zu heben. Die Notwendigkeit zu abstrahieren wird allgemein durch semantische Konflikte verursacht. Es umfasst Funktionen zum Aggregieren von Daten, Reklassifizieren von Entities oder eben komplexere, logische Denk-Prozesse. Die Anwendung einer Abstraktion während der Integration impliziert, dass keine Schreiboperationen möglich sind, während es üblicherweise unmöglich ist Rückwärts-Operationen zu spezifizieren.
- **Supplementation:** Daten werden nicht abgeleitet von den Komponentendaten, sondern es werden einige andere Daten hinzugefügt, die den Inhalt oder den semantischen Zusammenhang der Daten (semantische Metadaten) beschreiben. Solch eine Integration wird eingesetzt um implizit die Semantik der Komponenten zu handhaben. Dies ist zum Beispiel notwendig, wenn Datenquellen kein Schema anbieten, aber die Föderationsschicht auf einem Metadaten-Schema basiert.

A.4.7 **Transparenz**

Wir betrachten Transparenz für den Endnutzer als das ultimative Ziel der Integration. Ein perfekt integriertes Informationssystem vermittelt die Illusion, das die Nutzer nur mit einem einzigen, zentralen, lokal laufenden, homogenen und konsistenten Informationssystem interagieren. Wir unterscheiden die folgenden Typen von Transparenz:

- **Orts-Transparenz:** Nutzer brauchen die physische Lokalisation der Information nicht zu kennen.
- **Schema-Transparenz:** Nutzer brauchen die unterschiedlichen Benennungen, die Entities oder Attribute in unterschiedlichen Datenquellen haben, nicht zu kennen. Bei einem relationalen Beispielszenario brauchen sie z.B. die verschiedenen Relationen oder Attributnamen nicht zu kennen. Mit anderen Worten gesagt, sind alle logischen Konflikte maskiert. Natürlich kann Schematransparenz nur erreicht werden, wenn ein föderiertes Schema existiert.
- **Sprach-Transparenz:** Nutzer müssen sich nicht mit verschiedenen Anfragemechanismen und Sprachen auskennen. Dies beinhaltet die Anfragesprache und daher implizit das Datenmodell und der Zugriffsmechanismus, z.B. ob Anfragen am Ende durch die Anweisungen einer deklarativen Anfragesprache, wie z.B. SQL, ausgeführt werden oder durch einige Anwendungsmethoden über RPC gestartet werden.

Es gibt eine klare Beziehung zwischen der Behandlung von Heterogenität und dem Grad der Transparenz, die ein FIS anbietet. Schema-Transparenz ist da, um logische Heterogenität zu verstecken, während Sprach-Transparenz durch das Kapseln von Schnittstellen-Heterogenität erreicht wird. Orts-Transparenz steht in Beziehung zur technischen Heterogenität. Das Erreichen kompletter Transparenz ist sehr schwierig oder gar unmöglich.

A.4.8 **Anfrage-Paradigmen**

Informationssysteme werden oft durch die Typen von Anfragen klassifiziert, die sie erlauben: strukturierte Anfragen oder Informationsretrieval (IR)-Anfragen. Während ersteres einige Struktur in der Information voraussetzt, die genutzt wird um Datenitems in einer Anfrage zu spezifizieren, verwendet letzteres Ähnlichkeitssuchen in Dokumenten. Eine dritte Klasse von Anfrage-Paradigmen ist der Einsatz von spezifischen Metadaten, die die Datenobjekte beschreiben, obwohl es kein Teil der Daten selbst sein kann. Zum Beispiel kann Dokumentensuche Suchkriterien zulassen, wie z.B. Dokumentengröße oder Erstellungsdatum, die nicht explizit mit dem Dokument gespeichert sind.

A.4.9 Bottom-up versus Top-down

Wir unterscheiden zwischen zwei fundamentalen Wegen ein eng gekoppeltes FIS zu entwickeln: Top-Down, z.B. angefangen bei einem globalen Informationsbedürfnis und späteres Einklinken in Quellen, die zu diesem Bedürfnis beitragen, oder Bottom-Up, welches bei den Integrationsanforderungen bei einer Menge von Informationsquellen startet.

Top-Down-Strategie

Top-Down-Ansätze sind auf ein globales Informationsbedürfnis ausgerichtet. Zum Beispiel möchte eine Firma einen Service anbieten, um die niedrigsten Buchpreise in verschiedenen Internet-Läden zu finden; oder ein Entscheidungsunterstützungssystem möchte bestimmte Kundeninformation integrieren, die über mehrere Abteilungsdatenbanken verteilt ist. In diesen Fällen spielt das aktuelle Schema der Komponenten keine Rolle für den Entwurf des föderierten Schemas. Von den vier klassischen Anforderungen für die Schemaintegration, z.B. Vollständigkeit, Richtigkeit, Verständlichkeit und Minimalität, werden zwei nicht angewendet. Erstens gibt es keine Notwendigkeit Schemata vollständig einzubinden, wenn nur Kundendaten gewünscht sind (Vollständigkeit). Zweitens gibt es keine Notwendigkeit die Daten auf der globalen Ebene exakt so zu repräsentieren wie in den Komponenten, wenn nur abgeleitete oder abstrahierte Daten gewünscht werden (Richtigkeit; die Abbildung wird unidirektional). Einer mag z.B. entscheiden die Kunden in Gehaltsgruppen zu gliedern und nicht das genaue Einkommen zu speichern. Das globale Schema in Top-Down-Ansätzen kann entweder ad-hoc generiert werden oder das Ergebnis von formalen Analyseprozessen sein, angefangen bei Anwendungsfallbeschreibungen und endend bei View-Integrationstechniken. Das globale Schema kann ebenso durch einen Standard vordefiniert sein. In jedem Fall werden Komponenten-Schemata nur in einem zweiten Schritt betrachtet, wenn Übereinstimmungen zwischen den globalen Schemata und den Quell-Schemata etabliert sind, um die Umsetzung von Anfragen zu erlauben. Top-Down-Ansätze haben viele Vorteile in Szenarien, wo die Quellen sich schnell weiterentwickeln, wenn es häufig vorkommt, dass Quellen entfernt oder neue hinzugefügt werden, wenn Schema-Integration unmöglich oder zu teuer ist oder wenn globale Anforderungen selbst sich verändern. Dies liegt in der Hauptsache daran, dass schemaintegrationsbasierte Ansätze extrem empfindlich gegenüber jeder Veränderung sind. Wie auch immer, Top-Down-Ansätze führen typischerweise zu einer weniger engen Integration als Bottom-Up Ansätze.

Bottom-Up-Strategie

Ein eng integriertes FIS zu bauen bedeutet, dass die Eingangsanforderung die Notwendigkeit ist, einen integrierten Zugriff auf eine gegebene Menge von Datenquellen zu besitzen. Ein typisches Szenario ist die Notwendigkeit, einen detaillierten und einheitlichen Zugriff

auf alle Datenbanken einer Firma zu haben, um globale Anwendungen zu erstellen, vielleicht vor einer Migration. Der Manager hier würde sagen: „Gib mir einen integrierten Zugriff auf (exakt und komplett) diese Datenbanken“, während der typische Satz für eine Top-Down-Entwicklung nahezu ist: „Gib mit einen integrierten Zugriff auf alle Patientendaten (es spielt keine Rolle von wo)“. In solch einem Zusammenhang gibt es ein größeres Bedürfnis, Vollständigkeit und Richtigkeit des integrierten Schemas zu garantieren, und daher sind (semi-) formale Integrationstechniken passender. Bottom-Up-Integration führt zu semantisch gut integrierten Systemen, weil vorausgesetzt wird, dass die Komponenten vor dem Integrationsprozess komplett bekannt sind. Eine Änderung in der Konfiguration führt zu einem neuen Integrationsprozess.

Ein besonderes Problem bei der Konstruktion eines FIS ist die Notwendigkeit, Daten in Datenquellen über das globale Schema zu aktualisieren. Natürlich sind Updates nur möglich, wenn die Verbindung zwischen dem globalen und den Komponenten-Schemata sehr eng ist, z.B. wenn Updates einmalig bekannt gegeben werden können. Top-Down-Ansätze bieten selten diese Möglichkeit, und auch viele Bottom-Up-Ansätze bewahren nicht einmalige Korrespondenzen.

A.4.10 Virtuelle versus materialisierte Integration

Föderationen können danach unterschieden werden, ob oder ob nicht die Daten der Komponenten persistent in der Integrationsschicht gespeichert werden. Virtuelle Integrationsarchitekturen materialisieren nur temporär das Ergebnis der Anfragen zu der Zeit, wo die Anfrage gestellt ist. Dies erfordert einen Mechanismus, Anfragen an das föderierte Schema zu übersetzen in eine oder mehr semantisch bedeutungsvolle und ausführbare Anfragen an die Komponenten, die dynamisch an die Quellen verbreitet werden.

Das andere Extrem sind Architekturen, die die Quellen komplett oder teilweise auf der Föderationsebene materialisieren. Sie sind gut bekannt im Zusammenhang von Data Warehousing.

Die Materialisierung hat Vorteile:

- Hohe Geschwindigkeit bei Anfragen an die materialisierte Datenmenge
- Kontrolle über materialisierte Daten, insbesondere Pflege, ist möglich

und Nachteile:

- Die Daten up-to-date zu halten erfordert möglicherweise komplexe Update-Prozeduren, insbesondere wenn keine differenzierten Snapshots angeboten werden
- Die Föderationsebene muss einen möglicherweise großen Speicherplatz anbieten

Es gibt auch hybride Architekturen, die nur ausgewählte Datentypen materialisieren, z.B. sind Objektnamen Verweise zu den Datenquellen.

A.4.11 Nur Lese- oder Lese- und Schreib-Zugriff

Wir unterscheiden zwischen FIS, die das Einfügen (oder Aktualisieren) von Daten in Komponentensysteme hinein über die Föderationsebene erlauben und jene, die dies nicht tun. Schreibzugriff wird oft in Integrationsprojekten missachtet aufgrund der folgenden Punkte:

- Viele Schnittstellen, z.B. die WWW-Schnittstelle erlaubt kein Durchschreiben;
- Schreiben durch integrierte Views bringt alle Probleme der Datenaktualisierung über Views zum Vorschein;
- Globale Transaktionen erfordern komplexe Protokolle.

Im allgemeinen verringert der Durchschreibe-Zugriff die Quell-Autonomie zu einem hohen Grad. Es wird daher nicht oft in FIS-Projekten betrachtet.

A.4.12 Erforderliche Zugriffsmethoden

Insbesondere wenn Daten dynamisch von Datenquellen beschafft werden (virtuelle Integration), müssen die unterschiedlichen Zugriffsmethoden von Komponenten-Systemen (oder ihren Wrappern) betrachtet werden. FIS können danach unterschieden werden, inwieweit sie potentielle Einschränkungen auf dieser Ebene abdecken. Ein Client hat typischerweise eine oder mehrere der folgenden Möglichkeiten auf die Daten zuzugreifen:

1. Zugriff über eine **Anfragesprache**, z.B. SQL oder OQL. Dieser Zugriff kann über eine ursprüngliche Schnittstelle gewährt werden, über ODBC/JDBS, über eine spezielle Methode in einer CORBA API oder über ein spezielles Formular in einer WWW-Schnittstelle.
2. Zugriff über **parametrisierte, gekapselte Anfragen (PCQ)**. Parametrisierte, gekapselte Anfragen sind vordefinierte Anfragen mit einigen variablen Positionen. Gesprochen in SQL, hat eine PCQ typischerweise einen festen Select- und From-Abschnitt, feste Join-Bedingungen im Where-Abschnitt und festgelegte Typen anderer Bedingungen im Where-Abschnitt, erlaubt aber eine freie Spezifikation von Werten, mit welchen die Attributwerte in diesen Bedingungen verglichen werden. Gekapselte Anfragen werden typischerweise in IDL-Methoden verwendet (unter Einsatz von Funktionsparametern für die variablen Positionen), WWW-Formblätter (unter Einsatz von Eingabeelementen für die variablen Positionen) oder beliebige API-Prozeduren.

- Zugriff über **Browsing**. Insbesondere in einem WWW-Umfeld können Daten oft nur gebrowst werden als gesucht. In solchen Fällen ist eine Suche typischerweise auf ein navigierendes Durchqueren der Daten beschränkt.

Unterschiede in den Zugriffsmethoden betreffen zwei Typen von Heterogenität: technische Heterogenität und Zugriffssprachen-Heterogenität. Vom logischen Standpunkt aus spielt nur die Zugriffssprachen-Heterogenität eine Rolle. Es ist insbesondere sehr schwierig z.B. Beschränkungen in den möglichen Anfragen in einer Multidatenbank-Anfragesprache zu betrachten, wo Clients im Prinzip beliebige Anfragen formulieren können.

A.5 Typen von Föderierten Informationssystemen

Basierend auf den Kriterien, die zuvor vorgestellt worden sind, gibt [BKLW99] eine Definition von drei Typen von föderierten Informationssystemen: lose gekoppelte Informationssysteme, föderierte Datenbanksysteme und mediator-basierte Informationssysteme. Tabelle A.6 beschreibt knapp ihre Charakteristiken.

	Lose gekoppelte Informationssysteme	Föderierte Datenbanken	Mediator-basierte Informationssysteme
Typen von Heterogenität	Technische und linguistische Heterogenität	Jede, außer Heterogenität Anfragebeschränkung; Schemaintegration schwierig bei Schemaheterogenität	jede
Verlust von Autonomie	Ausführungsautonomie	Ausführungsautonomie; Kenntnisnahme von Schemaänderungen	Ausführungsautonomie
Transparenz	Sprachtransparenz	Orts-, Schema- und z.T. Sprachtransparenz	Orts-, Schema- und Sprachtransparenz
Art von Komponenten	strukturiert	strukturiert	jede Art
Zugriffsmethoden	Anfragesprache	Anfragesprache	jede
Zugriffsbeschränkungen	nein	nein	ja
Schreibzugriff	ja	ja	nein
Enge oder lose Kopplung	lose	eng	eng
Arten der semantischen Integration	Sammlung	Sammlung und Fusionen	Sammlung, Fusion, manchmal Abstraktion
Notwendige Metadaten	technische, über Infrastruktur	logische, technische und semantische	logische, technische und semantische
Bottom up oder Top-Down	-	Bottom-up	Top-Down
Virtuell oder materialisiert	virtuell	virtuell	virtuell
Weiterentwickelbarkeit	gut	schlecht	gut

Abbildung A.6: Arten von föderierten Informationssystemen

Aus der Tabelle wird klar, dass es schwierig ist eine baumartige Klassifikation wie in [SL90] zu geben. Ein Entwurf nach [BKLW99] für eine Klassifikation zeigt Abbildung A.3, wobei für die Mediator-basierten Informationssysteme von der Quelle abweichend nicht nur ein Lesezugriff gesehen wird:

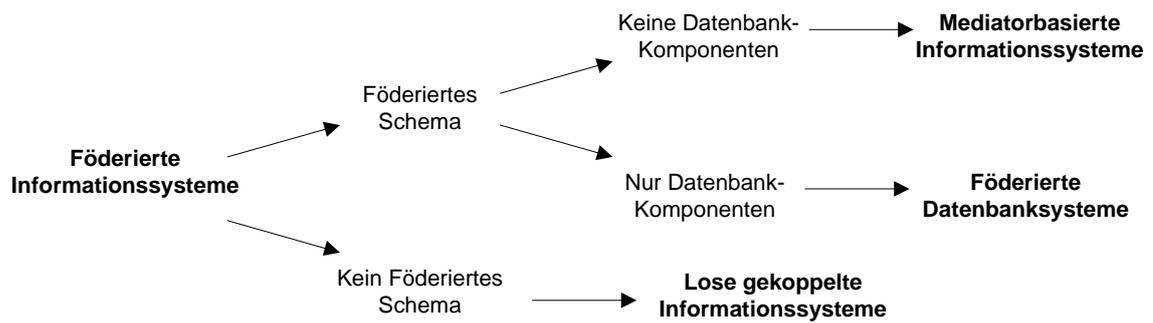


Abbildung A.7: Klassifikation föderierter Informationssysteme

A.5.1 Lose gekoppelte Informationssysteme

Lose gekoppelte Informationssysteme bieten kein föderiertes Schema, sondern nur eine Multidatenbank-Anfragesprache, um auf die Komponenten zuzugreifen. Dies hat den Vorteil, dass Komponenten nicht ihre Autonomie aufgeben, um an der Föderation teilzunehmen. Aber auf der anderen Seite wird keine Orts- oder Schema-Transparenz angeboten: der Nutzer muss in seinen Anfragen die entsprechende Komponente und das entsprechende Element im Komponenten-Schema adressieren. Wird eine einheitliche Anfragesprache angeboten, wird technische und sprachbasierte Heterogenität durch das FIS überbrückt. Alle logischen Konflikte müssen durch den Nutzer oder Dienste der Präsentationsschicht gelöst werden. Sie sind verantwortlich für die Datenintegration mit allen seinen Aspekten von Sammlung, Fusion und Abstraktion. Die Föderationsebene ist unabhängig vom logischen Entwurf der Komponenten. Da kein globales Schema existiert, haben Änderungen der Komponenten-Schemata keinen Einfluss auf das System. Aber die fehlende logische Integration führt zu zahlreichen Abhängigkeiten zwischen Anwendungen und Komponenten-Systemen mit allen negativen Effekten von Weiterentwicklungen, die von Zwei-Schichten-Systemen bekannt sind.

In der Literatur werden lose gekoppelte Informationssysteme manchmal Multidatenbanksysteme genannt. Andere Autoren betrachten sie als eine spezielle Art von föderierten Datenbanksystemen. Wegen der Unterschiede zwischen lose und eng gekoppelten FIS mit ihren Konsequenzen für Autonomie und Weiterentwicklung klassifizieren wir lose gekoppelte Informationssysteme separat.

A.5.2 Föderierte Datenbanksysteme (FDBS)

Ein Datenbanksystem besteht aus einem Datenbank-Managementsystem und einer oder mehreren Datenbanken, das es managt []. In vielen Anwendungsbereichen sind Daten über eine Vielzahl von heterogenen, autonomen Datenbanksystemen verteilt. Diese Systeme sind häufig isoliert und ein Austausch von Daten zwischen ihnen ist nicht einfach.

Auf der anderen Seite ist die Unterstützung für dynamischen Datenaustausch erforderlich, um Geschäftsprozesse zu fördern. Globaler Zugriff auf unterschiedliche lokale Systeme und konsistentes Replizieren / Bewegen von Daten zwischen Systemen sind typische Anforderungen. Ein föderiertes Datenbanksystem ist eine Integration solcher autonomer Datenbanksysteme, das gleichermaßen lokale und globale Applikationen bei Zugriffen auf multiple Datenbanksysteme unterstützt.

Für föderierte Datenbanksysteme muss die traditionelle Drei-Schichten-Schema-Architektur erweitert werden, um die Dimensionen Verteilung, Heterogenität und Autonomie zu unterstützen. Die allgemein akzeptierte Referenz-Architektur für Schemata in föderierten Datenbanksystemen wird in Abbildung A.8 gezeigt.

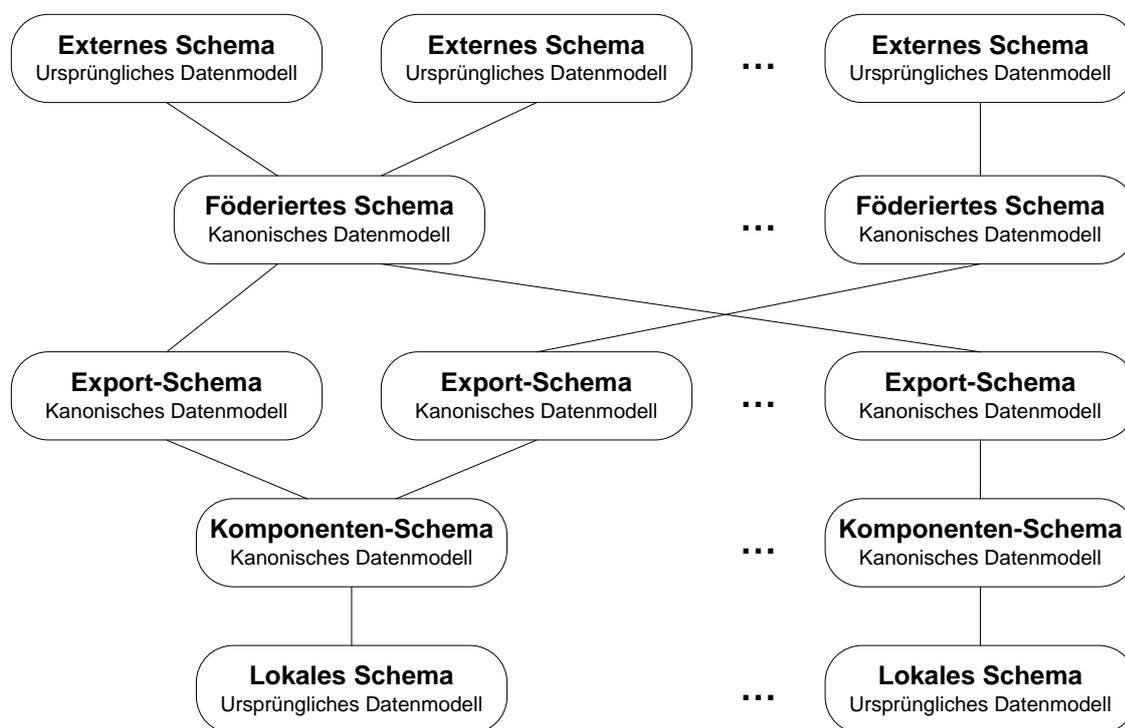


Abbildung A.8: Fünf-Schichten-Schema-Architektur von föderierten Datenbanksystemen (FDBS) nach [SL90]

Die unterschiedlichen Schematypen sind:

- **Lokales Schema:** Ein lokales Schema ist das konzeptuelle Schema eines Komponenten-Datenbanksystems, welches im ursprünglichen Datenmodell dieses Komponenten-Datenbanksystems ausgedrückt ist.
- **Komponenten-Schema:** Ein Komponenten-Schema ist ein lokales Schema transformiert in das (kanonische) Datenmodell der Föderationsschicht.

- **Export-Schema:** Ein Export-Schema wird vom Komponenten-Schema abgeleitet und definiert eine Schnittstelle zu den lokalen Daten, die der Föderation zugänglich gemacht wurden. Daher kann nur auf Elemente der Export-Schemata und deren Daten auf der integrierten Ebene zugegriffen werden.
- **Föderiertes Schema:** Wenn Export-Schemata semantisch heterogen sind, ist es notwendig, sie auf einer anderen Ebene zu integrieren. Ein föderiertes Schema auf dieser höheren Ebene ist das Ergebnis einer Integration mehrerer Export-Schemata; daher bieten sie eine integrierte Sicht.
- **Externes Schema:** Ein externes Schema ist eine spezielle Sicht auf ein föderiertes Schema oder auf ein lokales Schema. Externe Schemata können auf einem spezifischen Datenmodell basieren, welches unterschiedlich ist zum kanonischen Datenmodell. Grundsätzlich werden externe Schemata als spezielle Schnittstellen für Anwendungen (lokal oder global) angeboten.

Diese Schema-Architektur (Abbildung A.8), die durch die Föderationsschicht gemangt wird, spezifiziert die Abhängigkeiten / Korrespondenzen zwischen den individuellen Schemata.

A.5.3 Mediator-basierte Informationssysteme

Der Begriff Mediator wurde durch Wiederhold eingeführt und wird seither in vielen Publikationen über Datenintegrationsprojekte und -techniken genutzt. Bei der ersten Verwendung gab es noch keine exakte Definition von einem Mediator oder seine Beziehung zum FDBS. Im allgemeinen sollte ein Mediator eine Softwarekomponente sein, die zwischen Nutzer und physischen Datenquellen vermittelt. Sie sollte leichtgewichtig sein (verwaltbar von einer Gruppe, aber kein Komitee erfordern), flexibel und wiederbenutzbar. Insbesondere werden Mediatoren entworfen, um andere Mediatoren als Komponenten zu nutzen; eine Eigenschaft, welche auch als Möglichkeit in [SL90] erwähnt wird, aber was irgendwie in der FDBS-Gemeinschaft ignoriert wurde.

Die Definition von mediator-basierten Informationssystemen (MBIS) berücksichtigt sowohl den Stand der Dinge der Systeme, die sich selbst mediator-basierend nennen, als auch jene, die sich auf die originalen Anregungen von Wiederhold beziehen. Ein offensichtlicher Unterschied zwischen MBIS und FDBS ist der nur Lesezugriff auf die Datenquellen. MBIS sind eng gekoppelte Informationssysteme, weil ein föderiertes Schema gebraucht wird, um integrierten Zugriff auf Daten unterschiedlicher Komponenten (semantische Heterogenität) anzubieten. Im Gegensatz zu FDBS wird ein föderiertes Schema üblicherweise Top-Down erstellt, mit Bezug zu den Informationsbedürfnissen. Darauf bezogen ist das Verständnis von Mediatoren als Dienste, die für Kunden konstruiert und angeboten werden. Dies impliziert die wichtige Anforderung von Flexibilität bezüglich der Weiterentwicklung des Systems. Schließlich muss es möglich sein auf einfache Weise

Komponenten ein- bzw. auszustöpseln, da Quellen in einem MBIS typischerweise komplette Kommunikationsautonomie behalten. Eine weitere Klassifikation von MBIS enthält einige Unterschiede. In der Literatur werden sowohl strukturierte, semi-strukturierte und unstrukturierte Komponenten betrachtet. Die Heterogenität von Zugriffsmethoden ist ein wichtiger Forschungspunkt, z.B. bezüglich Binding-Patterns (überall gegenwärtig bei der Integration von WWW-Quellen) und beschränkten Suchmöglichkeiten (wichtig in vielen Wissenschaftsdomänen, wie z.B. räumliche Datenbanken). MBIS können verschiedene Integrationsmechanismen wie Abstraktion, Aggregation oder Metainformationsansätze integrieren. Typischerweise berücksichtigt der Mediator nicht alle diese Aspekte, aber er sollte letztlich einen von ihnen handhaben.

Abbildung A.9 zeigt die klassische Mediator-Wrapper-Architektur entlehnt von der allgemeinen Architektur für FIS (siehe Abbildung A.2). Die Föderationsschicht enthält verschiedene Mediatoren, die Vermittlungsdienste anbieten; deshalb wird diese Schicht nun Vermittlungsschicht genannt. Jeder Mediator hat sein eigenes föderiertes Schema und Mediatoren können andere Mediatoren als Datenquellen benutzen (Mediatoren-Netzwerk). Wrapper verstecken technische und datenmodellbasierte Heterogenität; wie sie auf ihre Datenquellen zugreifen, ist für einen Mediator transparent. Anfragen an föderierte Schemata (Nutzeranfragen) sind gestrichelt dargestellt und Anfragen an den Wrapper (Quellanfragen) sind liniert gezeichnet.

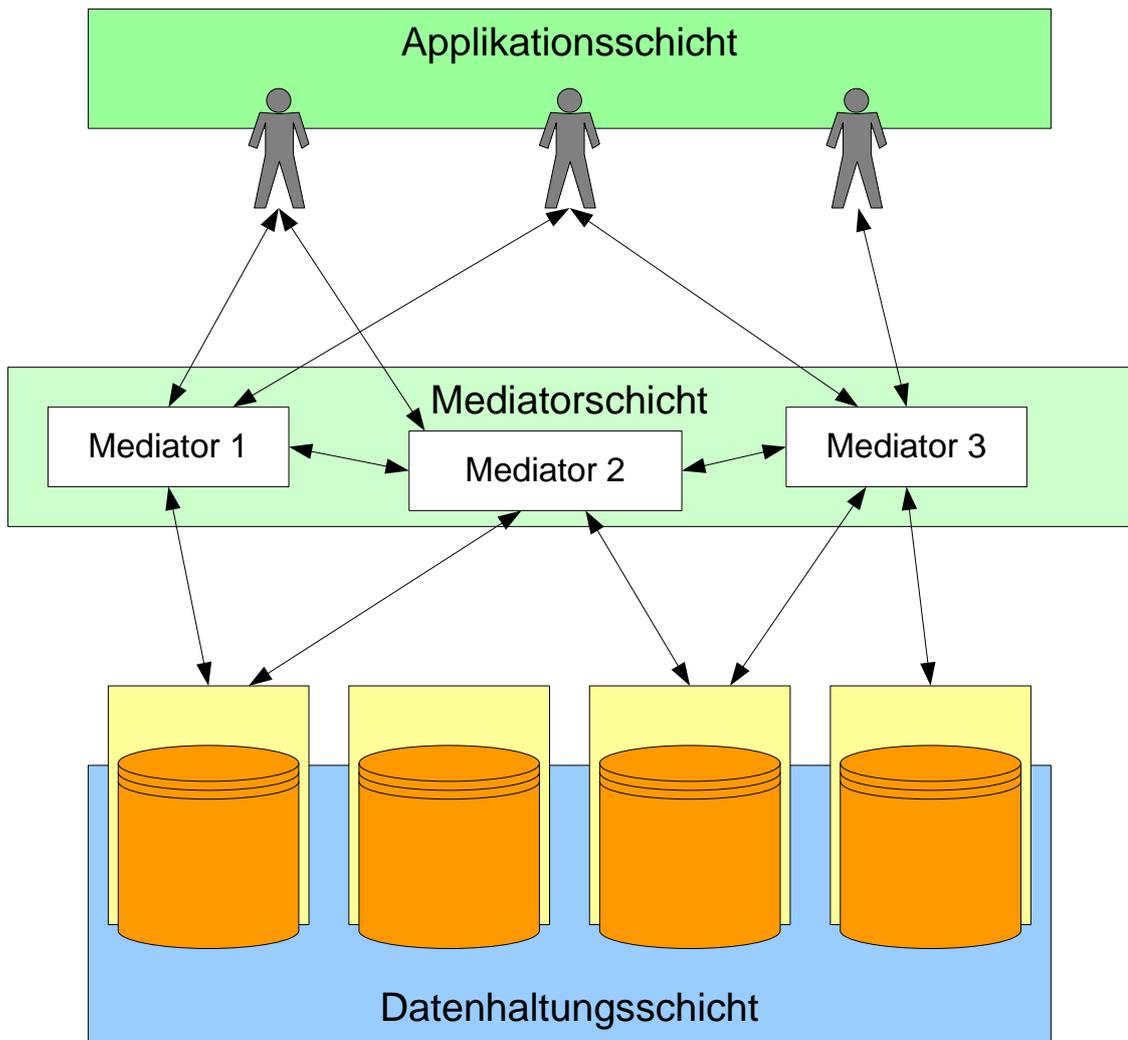


Abbildung A.9: Schichten-Architektur von mediatorbasierten Informationssystemen (MBIS) nach [Wil02]

B Resource Description Framework und Semantic Web

Die Arbeit des W3C konzentriert sich auf Standardwege Metadaten über den eigenen Standard RDF zu modellieren und zu verschlüsseln [Kor01]. RDF wurde konzipiert, um Interoperabilität zwischen Anwendungen zu ermöglichen, die maschinen-lesbare Information im Web austauschen. Innerhalb des RDF-Frameworks können unabhängig voneinander Vokabularien entwickelt werden, die auf spezielle Bedürfnisse abgestimmt sind und die mit anderen geteilt werden können. Die Beschreibungen dieser Vokabularien werden RDF-Schemata genannt. Eines der besser bekannten Schemata ist das Dublin Core, eingeführt für Bibliotheken (<http://www.dublincore.org>).

B.1 Grundlagen RDF

RDF identifiziert Ressourcen mit Uniform Resource Identifiers (URI) [URI04]. RDF betrachtet dabei jede URI selbst als komplette Ressource. Die fundamentale Syntax beschreibt das RDF-Modell. Dessen Basiselement ist ein Tripel (Subjekt,Prädikat,Objekt). Für einen RDF-Graphen bedeutet dies, dass eine Ressource (Subjekt) verknüpft ist mit einer weiteren Ressource (Objekt) durch eine Kante, die beschriftet ist mit einer dritten Ressource (Prädikat). Abbildung B.1 veranschaulicht diesen Zusammenhang graphisch.

Eine Aussage (*statement*) beschreibt in RDF eine Eigenschaft einer Ressource, beispielsweise: Dokumentenstandard vereinheitlicht Dokumentation medizinischer Daten. Die Komponenten eines *statements* sind:

- Resource (Subjekt): Dokumentenstandard

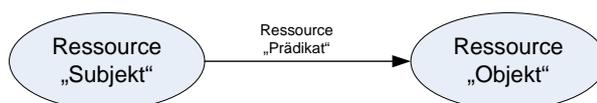


Abbildung B.1: Ein RDF-Tripel als Graph

- Property (Prädikat): vereinheitlicht
- Value (Objekt): Dokumentation medizinischer Daten

Diese Aussagen sind stark kontext-abhängig. Mit einem globalen Identifizierungsmechanismus, wie dem URI, gibt es ein Konzept zur globalen Identifikation beliebiger Objekte, welches den Kontext weitgehend festlegt. Identifiziert werden können mit URIs:

1. Sämtliche Web-Ressourcen mittels Uniform Resource Locator (URL)
2. Objekte ohne Lokalisierung im Web
3. Gegenständliches und Abstraktes

Alle Ressourcen werden mit URIs benannt, wobei es eine Ausnahme gibt: das Objekt. Nur dieses darf ein *Literal* sein, also eine Zeichenkette oder eine Ganzzahl. URIs werden in XML als XML-Namespace-Deklarationen spezifiziert.

Ein RDF-Statement repräsentiert das folgende Denkmodell:

- Ein Objekt (Ressource) erhält zu einer Eigenschaft (Property) einen Wert (Value) zugeordnet.
- Eine Ressource wird immer durch einen URI identifiziert (Ausnahme Objekt mit Literal).
- Property meint Eigenschaft oder Beziehung zwischen Ressourcen.
- Wert oder Ausprägung der Eigenschaft ist der Value.

B.2 Notation 3

Notation 3, oder auch kurz N3, ist eine einfache Lehrsprache, die gleichwertig zur XML-Syntax von RDF ist, nur einfacher zu schreiben und zu lesen [Not00]. Ein RDF-Tripel lautet in N3 beispielsweise `<#alex> <#kennt> <#jojo>`. Es gibt zwei Abkürzungen für mehrere Sätze mit einem Subjekt. Das Semikolon „;“ leitet eine neue Eigenschaft desselben Subjekts, das Komma „;“ ein weiteres Objekt zum selben Subjekt und selben Prädikat ein.

```
<#alex> <#kind> <#max>, <#lilli>, <#moni>;  
<#alter> "24";  
<#augenfarbe> "blue".
```

Wenn kein Bezeichner angegeben werden soll, sondern nur die Eigenschaften bezeichnet werden sollen, werden in N3 die Eigenschaften innerhalb von eckigen Klammern notiert.

```
<#alex> <#kind> [<#alter> "4"], [<#alter> "3"].
```

Im Dublin Core [Dub04] sind Bezeichner für eine zusammengestellte Liste von Eigenschaften standardisiert.

```
<> <#titel> "Ein einfaches N3-Beispiel". würde mit dem entsprechenden Dublin Core-Bezeichner lauten: <> <http://purl.org/dc/elements/1.1/title> "Ein einfaches N3-Beispiel".
```

N3 erlaubt einen Präfix als Kürzel für den Namensraum, beispielsweise:

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.
```

Damit lautet das RDF-Tripel: <> dc:title "Ein einfaches N3-Beispiel".

Es sind bereits einige RDF-Vokabularien entstanden, die auf der RDF-Homepage [RDF04a] zu finden sind. Folgende Standardpräfixe sind definiert:

- @prefix rdf: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210#>.
- @prefix rdfs: <http://www.w3.org/TR/2004/REC-rdf-schema-20040210#>.
- @prefix owl: <http://www.w3.org/2002/07/owl#>.
- @prefix <#>., so steht das leere Präfix für das Dokument, in dem es sich befindet.

Damit vereinfacht sich die Notation des Beispiels von weiter oben:

```
:alex :kind [:alter "4"], [:alter "3"].
```

B.3 Erstellung eines Vokabulars

Bei `dc:title` handelt es sich um eine RDF-Eigenschaft. Um ein neues Vokabular zu definieren, werden Definitionen von neuen Klassen und neuen Eigenschaften benötigt. Wenn ausgedrückt werden soll, von welchem Typ etwas ist, wird die Klasse angegeben, zu der es gehört. Diejenige Eigenschaft, die aussagt, von welchem Typ etwas ist, heißt `rdf:type`. In N3 abgekürzt mit `a` (englisch für ein). So definiert man die Klasse `Person` folgendermaßen:

```
:Person a rdfs:Class.
```

Im selben Dokument kann nun mit `:alex a :Person.` eine konkrete Person eingeführt werden. Klassen sagen etwas über die Objekte aus, die in ihnen enthalten sind. Ein Objekt kann sich in vielen Klassen befinden, dabei müssen diese keine hierarchischen Beziehungen zueinander haben. Beziehungen zwischen zwei Klassen werden als *statement* formuliert, beispielsweise:

```
:Frau a rdfs:Class; rdfs:subClassOf :Person.
```

Eine Eigenschaft beschreibt eine Beziehung zwischen zwei Dingen, z.B.:

```
:schwester a rdf:Property.
```

Wenn in einer Aussage das Subjekt Instanz einer Klasse ist, dann wird diese Klasse als *Domäne* (englisch: *domain*) des Prädikats bezeichnet. Wenn dagegen das Objekt Instanz einer Klasse ist, dann wird diese Klasse als *Zielbereich* (englisch: *range*) des Prädikats bezeichnet. Eine Eigenschaft kann viele Domänen und Zielbereiche haben, aber typischerweise genügt jeweils eine, um das Prädikat zu spezifizieren. Mathematisch betrachtet, handelt es sich um eine Funktion, die Werte aus einer Domäne auf einen Zielbereich abbildet.

```
:schwester rdfs:domain :Person; rdfs:range :Frau.
```

B.4 Semantic Web und Ontology Web Language (OWL)

Nach der Vorstellung des Web-Erfinders Tim Berners-Lee soll das bestehende Web um eine semantische Ebene mit Metadaten erweitert werden, um eine automatisierte und wissensbasierte Verarbeitung von Web-Ressourcen zu ermöglichen [SL03]. Die Techniken, die in diesem Rahmen entwickelt wurden, umfassen Metamodell- und Ontologiesprache, ontologische Anfragesprachen oder auch Mapping-Techniken. Die verteilten Metadatenschichten sowie Ontologien sind auch im Bereich der Systemintegration hilfreich, da die Informationen präzise und navigierbar modelliert werden können [Oh103]. Um

die Web-Ressourcen mit formaler Semantik beschreiben zu können, was die Voraussetzung ist, Information maschinenverständlich zu machen, hat das World Wide Web Consortium zwei Sprachen zur Metadatenrepräsentation entwickelt und als Standard propagiert [MM03]. Diese Sprachen sind das RDF und das darauf aufgebaute RDF-Schema. RDF-Schema ist eine sehr einfache Ontologierepräsentationssprache. Für die Beschreibung von Web Services steht OWL-S zur Verfügung, wobei es sich dabei um eine OWL-basierte Webservice-Ontologie handelt [OWL03]. Vorherige Versionen von OWL waren unter DAML-based Web Service Ontology (DAML-S) bekannt, welches seinerseits auf DAML+OIL aufbaut. Mit den angebotenen Sprachkonstrukten können Eigenschaften und Möglichkeiten der Web Services in einer eindeutigen, vom Computer interpretierbaren Form, beschrieben werden. Das Semantic Web baut auf den Möglichkeiten von XML auf, angepasste Auszeichnungsschemata zu definieren und auf den flexiblen Ansatz der Datenrepräsentation von RDF. Die erste Schicht oberhalb von RDF, die für das Semantic Web notwendig ist, ist eine Ontologiesprache, die formal die Bedeutung der in Web-Dokumenten verwendeten Terminologie beschreibt. Wenn man von Maschinen erwarten will, dass sie nützliche Folgerungen aus diesen Dokumenten ziehen, muss die Sprache über die grundlegende Semantik von RDF-Schema (RDFS) hinausgehen. OWL wurde entwickelt, um das Bedürfnis nach einer Ontologiesprache für das Netz zu befriedigen. OWL ist Teil des wachsenden Bestandes von W3C Empfehlungen für das Semantic Web.

- XML stellt eine Syntax für strukturierte Dokumente zur Verfügung, nutzt aber keine semantische Eingrenzung der Bedeutung dieser Dokumente.
- XML-Schema ist die Sprache um die Struktur von XML Dokumenten festzulegen und erweitert XML auch um die Datentypen.
- RDF ist ein Datenmodell für Ressourcen und Relationen zwischen diesen. Es stellt eine einfache Semantik für das Datenmodell zur Verfügung. Diese Datenmodelle können in XML repräsentiert werden.
- RDF-Schema ist ein Vokabular zur Beschreibung der Eigenschaften und Klassen von RDF-Ressourcen, mit einer Semantik für verallgemeinernde Hierarchien solcher Eigenschaften und Klassen.
- OWL fügt mehr Vokabeln zur Beschreibung von Eigenschaften und Klassen hinzu, unter anderem: Relationen zwischen Klassen (z.B. Disjunktheit), Kardinalität (z.B. „genau einer“), Äquivalenz, mehr Propertytypen, Charakteristika von Properties (z.B. Symmetrie), und Klassen, die durch die Aufzählung ihrer Mitglieder definiert werden.

OWL bietet drei Untersprachen mit zunehmender Ausdrucksstärke. Diese wurden für die Verwendung durch spezielle Gruppen von Entwicklern und Anwendern entwickelt:

1. OWL Lite unterstützt primär den Benutzer, der eine Klassifikationshierarchie und einfache Restriktionen benötigt. Zum Beispiel erlaubt es zwar Kardinalitätsrestriktionen, aber nur die Kardinalitätswerte 0 und 1. Es sollte einfacher sein, Tools für

OWL Lite zu entwickeln, als für seine ausdrucksstärkeren Verwandten. Außerdem bietet OWL Lite eine schnelle Möglichkeit zur Migration von Thesauri und anderen Taxonomien. OWL hat auch eine geringere formale Komplexität als OWL DL, siehe Abschnitte zu OWL Lite in der OWL Referenz für weitere Details.

2. OWL DL unterstützt die Benutzer, die das Maximum an Ausdrucksstärke haben möchten und behält dabei die vollständige Verarbeitbarkeit (alle Folgerungen können garantiert gezogen werden) und Entscheidbarkeit (jede Verarbeitung wird in endlicher Zeit durchgeführt). OWL DL schließt alle OWL Sprachkonstrukte ein, diese können aber nur unter bestimmten Bedingungen verwendet werden (z.B. kann eine Klasse Unterklasse vieler Klassen sein, eine Klasse kann aber keine Instanz einer anderen Klasse sein). OWL DL wird so genannt, weil es den Anforderungen der „Description Logics“ genügt, einem Forschungsfeld, das die Logik behandelt, die die formelle Basis von OWL begründet. Description Logic (DL) repräsentiert eine entscheidbare terminologische Logik, also eine Teilmenge der Prädikatenlogik erster Stufe.
3. OWL Full ist für Benutzer gedacht, die maximale Ausdrucksstärke und die syntaktische Freiheit von RDF möchten, allerdings ohne eine Garantie der Verarbeitbarkeit. Zum Beispiel kann in OWL Full eine Klasse gleichzeitig als Collection von Individuals behandelt werden und als eigenes Individual. OWL Full erlaubt es, einer Ontologie die Bedeutung von vordefiniertem (RDF oder OWL) Vokabular zu erweitern. Es ist unwahrscheinlich, dass irgendeine Schlussfolgerungssoftware das komplette Schlussfolgern für jedes Feature von OWL Full unterstützen können wird.

C Medizinische Dokumentation

C.1 Einleitung

Was bedeutet medizinische Dokumentation? Als Dokumentation bezeichnet man die Tätigkeiten des Sammelns, Erschließens, Ordnen und Aufbewahrens von Information oder von Wissen, um beides zu einem späteren Zeitpunkt und für eine gegebenes Ziel nutzbar zu machen. Erst durch die inhaltliche Erschließung werden die in den Dokumenten enthaltenen Informationen zugänglich [LGHKG99]. Daraus geht hervor, dass mit einer konkreten Nutzungsabsicht dokumentiert wird und die Dokumentation keinen Selbstzweck darstellt. Allgemeines Ziel ist es, berechtigten Personen gezielt Informationen oder Wissen zum richtigen Zeitpunkt, in der richtigen Form und am richtigen Ort zur Verfügung zu stellen. Die Dokumentation ist also Mittel zur Informations- und Wissenslogistik. Bei der medizinischen Dokumentation können ganz unterschiedliche Informationen gemeint sein, z.B. Befunde und Therapien einzelner Patienten, Diagnose- und Therapiewissen über spezielle Erkrankungen, ein Verzeichnis medizinischer Veröffentlichungen sowie Resultate aus Arzneimittelstudien [LGHKG99]. Bei der medizinischen Dokumentation werden zur Unterstützung diverse medizinische Begriffssysteme eingesetzt.

C.2 Ziele der medizinischen Dokumentation

Das allgemeine Ziel der medizinischen Dokumentation ist es, wie bereits erwähnt, die richtige Information bzw. das richtige Wissen zum richtigen Zeitpunkt am richtigen Ort den richtigen und berechtigten Personen in der richtigen Form anzubieten.

Weitere inhaltliche Ziele nach [LGHKG99] sind:

- Unterstützen der Patientenversorgung (Erinnerungs-, Kommunikations- und Organisationshilfe);
- Erfüllen rechtlicher Erfordernisse (gesetzliche Dokumentations- und Meldepflichten, nachträgliche Rechtfertigung des Vorgehens);

- Unterstützen der Administration (patientenbezogene Darstellung der erbrachten Leistungen);
- Unterstützen des Qualitätsmanagements (nachträgliche Beurteilung des Vorgehens, Qualitätsmonitoring);
- Unterstützen der klinisch-wissenschaftlichen Forschung (Patientenauswahl, statistische Auswertungen);
- Unterstützen der klinischen Aus- und Fortbildung (nachträgliche Beurteilung des Vorgehens, Fallbeispiele).

C.3 Grundlagen der Terminologielehre

Ein Hauptziel der medizinischen Dokumentation besteht in der Einordnung von Informationen und deren Rückführung auf ihren begrifflichen Inhalt. Dies liegt vor allem an der Tatsache, dass die patientenorientierte Information überwiegend in sprachlicher Form dargestellt wird. Aus diesem Grund sind die Bereitstellung und Anwendung von Begriffssystemen (Klassifikationen, Nomenklaturen, Thesauren) sehr wichtig. Die allgemeine Terminologielehre liefert die Grundlagen für den Aufbau von Begriffssystemen.

C.3.1 Semiotische Triade

Die Terminologielehre ist definiert als die Wissenschaft von den Begriffen und ihren Benennungen im Bereich der Fachsprachen [DIN 2342 1992]. Die allgemeine Terminologielehre ist fach- und sprachübergreifend. Die spezielle Terminologielehre, wie z.B. die medizinische Terminologielehre, hat eine fachbezogene Ausrichtung [Ber00]. Die so genannte semiotische Triade stellt die Grundelemente der allgemeinen Terminologielehre, nämlich Begriff, Gegenstand und Bezeichnung, entsprechend ihrer Verhältnisse zueinander graphisch dar:

Aus der semiotischen Triade in Abbildung C.1 gehen also folgende drei Beziehungen hervor:

- Eine Bezeichnung bezeichnet einen Begriff
- Eine Bezeichnung bedeutet einen Gegenstand
- Ein Begriff vertritt oder bezieht sich auf einen Gegenstand

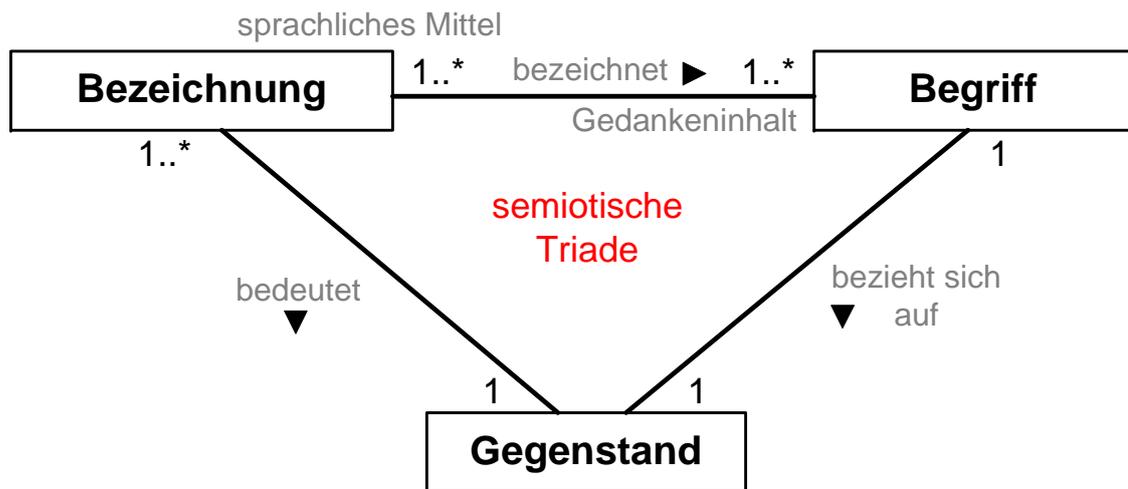


Abbildung C.1: Semiotische Triade

In der Deutschen Industrienorm (DIN) 2342 finden sich für die drei wichtigsten Grundbegriffe entsprechende Definitionen:

Gegenstand (auch Objekt):

Beliebiger Ausschnitt aus der wahrnehmbaren oder vorstellbaren Welt.

Begriff:

Denkeinheit, die aus einer Menge von Gegenständen, unter Ermittlung der diesen Gegenständen gemeinsamen Eigenschaften mittels Abstraktion, gebildet wird.

Bezeichnung:

Repräsentation eines Begriffs mit sprachlichen oder anderen Mitteln.

Werden bei einer Menge von Gegenständen die gemeinsamen Eigenschaften abstrahiert, entsteht dadurch eine gedankliche Einheit, der Begriff. In Bezug auf die Terminologielehre ist ein Gegenstand ein beliebiger Ausschnitt aus der Welt, die wir wahrnehmen oder uns vorstellen können. Somit können Gegenstände konkret, materiell oder abstrakt, immateriell sein. Sie können sich dabei sogar auf komplexere Tatsachen oder Prozesse beziehen. Wahrnehmbare Gegenstände gelten als konkret, vorstellbare Gegenstände als abstrakt. Gegenstände werden auch in individuelle und allgemeine Gegenstände unterschieden. Sind Gegenstände einmalig in Raum und Zeit, so sind sie individuell, haben sie dagegen keinen Bezug zu Raum und Zeit, sind sie allgemein. Entsprechend vertreten Individualbegriffe wie z.B. „Ulmer Münster“ individuelle Gegenstände, und Allgemeinbegriffe wie z.B. „Krankenhaus“ allgemeine Begriffe. Begriffe dienen dem Erkennen von Gegenständen, der Verständigung über Gegenstände sowie dem gedanklichen Ordnen von Gegenständen. Eine Bezeichnung dient dazu einen Begriff mit natürlichsprachlichen oder anderen Mitteln mitzuteilen. Die natürlichsprachliche Bezeichnung heißt auch Benennung [Ber00].

C.3.2 Merkmale

Ein Merkmal beschreibt diejenige Eigenschaft von Gegenständen, die zur Begriffsbildung und -abgrenzung dient. Merkmale sind ebenfalls Begriffe. Das Merkmal einer Fraktur könnte beispielsweise die Lokalisation am Femur sein. Die Eigenschaften von Gegenständen, die Merkmale beschreiben können, sind beobachtete, gemessene oder allgemein akzeptierte Eigenschaften. Merkmale spielen eine wichtige Rolle bei der Begriffsbestimmung und Bestimmung von Begriffsbeziehungen [Ber00]. Nach [Ber00] liegt ihre Bedeutung in folgenden Aspekten:

- Merkmale dienen zur Festlegung des Begriffsinhalts als der Gesamtheit der Merkmale eines Begriffs.
- Merkmale haben einen entscheidenden Einfluss auf die Strukturierung von Begriffssystemen, in dem sie die Über- bzw. Unterordnung und Nebenordnung von Begriffen bestimmen.
- Merkmale sind die Grundlage der Benennungsbildung, wenn „motivierte“ Benennungen geschaffen werden sollen. Eine motivierte Benennung drückt durch die Wahl der verwendeten Worte die Merkmale des bezeichneten Begriffs aus. Die Benennung „Innenmeniskusverletzung“ ist eine motivierte Benennung für eine Verletzung des inneren Meniskus.
- an gleichen Merkmalen kann die Identität von Begriffen festgestellt werden.

C.3.3 Merkmalart

Eine Merkmalart ist eine Zusammenfassung von Merkmalen, die zur Gliederung eines Begriffssystems herangezogen wird [DIN 2342 1992]. Merkmalarten haben die Funktion von Kriterien, die zur Abgrenzung und Ordnung von Begriffen verwendet werden können. Merkmalarten von Krankheiten sind z.B. die Genese (angeboren oder erworben), das betroffene Organsystem, Krankheitsmechanismus. DIN 2330 unterteilt die Merkmale von Begriffen, die konkrete Gegenstände vertreten, in Beschaffenheitsmerkmale, Beziehungsmerkmale und Funktionsmerkmale. Ein weiteres Kriterium zur Unterteilung von Merkmalen betrifft ihre Bedeutung in Bezug auf ihren Begriff. DIN 2330 unterscheidet zwischen Ordnungsmerkmalen (oder wesentlichen (essentiellen) Merkmalen) und akzidentiellen (oder unwesentlichen) Merkmalen. Essentielle Merkmale sind wesentliche oder notwendige Merkmale eines Begriffs und werden zur Ordnung von Begriffen herangezogen. Akzidentielle Merkmale sind mögliche oder eher zufällige Merkmale eines Begriffs. In der Dokumentation sind die wesentlichen Merkmale von Begriffen wichtig für die Erstellung von Ordnungssystemen (Klassifikationen, Thesauren). Die akzidentiellen Merkmale sind wichtig als Merkmale von Dokumentationseinheiten. In Begriffssystemen

mit mehreren Ebenen von Abstraktionsbeziehungen spielt die Abhängigkeit von Merkmalen eine Rolle. Merkmale sind voneinander abhängig, wenn das übergeordnete Merkmal gegeben sein muss, ehe das spezielle hinzutritt. Die Abhängigkeit bzw. Unabhängigkeit von Merkmalen bestimmt häufig die Struktur von Begriffssystemen. Abhängige Merkmale erzwingen eine bestimmte hierarchische Anordnung von Begriffen. Unabhängige Merkmale können in verschiedener Weise aufeinander folgen und beliebig miteinander kombiniert werden. Der Begriffsinhalt ist nach DIN 2342 1992 die Gesamtheit der Merkmale eines Begriffs. Im Unterschied dazu ist der Begriffsumfang definiert als Gesamtheit der einem Begriff auf derselben Hierarchiestufe untergeordneten Begriffe. Klasse im Sinne der Terminologiearbeit wird definiert als Gesamtheit der Gegenstände, die unter einen Begriff fallen. Nach DIN 2342 ist eine Definition eine Begriffsbestimmung mit sprachlichen Mitteln. Definitionen dienen dazu, einen möglichst eindeutigen Zusammenhang zwischen Begriffen und Benennungen herzustellen [Ber00].

C.3.4 Zusammenhänge Begriff und Benennung

Existieren für denselben Begriff verschiedene Benennungen, so nennt man diese Synonyme, bzw. man sagt, diese Benennungen sind untereinander synonym. Bezeichnet eine Benennung verschiedene Begriffe, so ist diese Bezeichnung ein Homonym.

Synonymie

Liegen synonyme Benennungen vor, ist es vorteilhaft, eine als Vorzugsbenennung auszuzeichnen. In Ordnungssystemen kann man auf diese Weise Redundanz vermeiden: Information, die zu dem Begriff gehört, auf den sich die synonymen Benennungen beziehen, muss nur einmal vorgehalten werden.

Homonymie

Umgangssprachlich bereiten Homonyme kaum Schwierigkeiten, da sich die jeweilige Bedeutung aus dem Kontext ergibt. In der Dokumentation werden Deskriptoren jedoch ohne Textzusammenhang verwendet, daher Gefahr von Missverständnissen. Eine Recherche liefert zu viele und nicht-relevante Dokumentationseinheiten.

C.4 Grundlagen medizinischer Begriffssysteme

C.4.1 Allgemeines

Überall in der Medizin und ihren Teilgebieten begegnet man inzwischen medizinischen Begriffssystemen. Unter den Aspekten der Qualitätssicherung, Gesundheitsökonomie, Kommunikation und Informationsverarbeitung haben sie eine wachsende Wichtigkeit bei

der Integration, da die Informationstechnologie sich zunehmend in allen Bereichen etabliert. Das Spektrum reicht von eher impliziten Begriffsordnungen, wie sie z.B. der Gliederung von Lehrbüchern zugrunde liegen und der Strukturierung von Fachwissen dienen, bis zu expliziten Begriffssystemen, bei denen die systematische Anordnung einer beschränkten Menge von Begriffen im Vordergrund steht. Die medizinischen Begriffssysteme sind meist Klassifikationssysteme, aber auch Nomenklaturen oder Thesauern können zu den Begriffssystemen gezählt werden [Ber95].

Begriffssysteme können meist nach folgenden Aspekten charakterisiert werden [Ber95]:

- Art der Begriffsbeziehungen
- Anzahl der Achsen
- Art der hierarchischen Struktur
- Codierprinzipien

C.4.2 Art der Begriffsbeziehungen

Bei den Begriffsbezeichnungen unterscheidet man hierarchische und nicht-hierarchische Begriffsbeziehungen.

Hierarchische Begriffsbeziehungen

Bei den hierarchischen Begriffsbeziehungen unterscheidet man zwischen Abstraktionsbeziehungen (generische Beziehungen) und Bestandsbeziehungen (partitive Beziehungen).

Abstraktionsbeziehung bzw. generische Beziehung

Zwischen zwei Begriffen besteht eine Abstraktionsbeziehung, wenn bei gleichem Merkmalbesitz ein Begriff mindestens ein zusätzliches, differenzierendes Merkmal aufweist. D.h. je tiefer ein Begriff in der Abstraktionshierarchie steht, desto mehr Merkmale weist er auf. Eine Abstraktionshierarchie wird auch als Taxonomie bezeichnet.

Es ist oft zweckmäßig, zu mehreren gleichgeordneten Begriffen einen Begriff mit der Bedeutung „sonstiges zum Oberbegriff“ aufzunehmen, um ein Begriffssystem vollständig zu machen.

Bestandsbeziehung bzw. partitive Beziehung

Eine Bestandsbeziehung besteht zwischen zwei Begriffen, wenn ein Begriff sich auf ein Ganzes und der andere auf einen Teil des Ganzen bezieht. Der erste Begriff heißt Verbandsbegriff, der zweite Teilbegriff. Die Bestandsbegriffe einer Stufe bilden eine Bestandsreihe (analog zur Abstraktionsreihe), die Bestandsbegriffe mehrerer Stufen bilden eine Bestandsleiter (analog zur Abstraktionsleiter). Eine Bestandshierarchie wird auch als Partonomie bezeichnet [Ber00].

Nicht-hierarchische Begriffsbeziehungen

Bei den nicht-hierarchischen Begriffsbeziehungen unterscheidet man zwischen sequentiellen Beziehungen, pragmatischen Beziehungen und Komplementärbeziehungen.

Sequentielle Begriffsbeziehung

Eine sequentielle Begriffsbeziehung ist eine Begriffsbeziehung, die auf einer direkten Abhängigkeit von Begriffen im Sinne der Vor- und Nachordnung beruht [DIN 2342]. In medizinischen Begriffsordnungen spielen sequentielle Beziehungen z.B. eine Rolle:

- bei anatomischen Begriffen (räumlich sequentielle Anordnung)
- bei Krankheitsbegriffen (Ausbreitungs-, Schweregrad)
- bei Behandlungsprozessen (sequentielle Teilschritte)

Pragmatische Begriffsbeziehung bzw. Assoziationsbeziehung

Eine pragmatische Begriffsbeziehung ist eine Begriffsbeziehung, die auf thematischen Zusammenhängen zwischen Begriffen beruht, die jedoch weder der hierarchischen noch der sequentiellen Begriffsbeziehung zugeordnet werden kann [DIN 2342]. Beispiel: Hüftgelenk - Hüftgelenksprothese

Komplementärbeziehung bzw. Oppositionsbeziehung

Die Komplementärbeziehung besteht zwischen zwei Begriffen, wenn bei sonst gleichbleibenden Merkmalen je ein Merkmalpaar sich komplementär verhält. Beispiel: endogen - exogen; oben - mitte - unten

C.4.3 Anzahl der Achsen

Komplexe Begriffe wie z.B. der einer akuten, serösen und fokalen Entzündung oder der einer akuten Nephritis lassen sich bezüglich mehrerer Bedeutungsdimensionen charakterisieren. Nephritis beispielsweise enthält einen topographischen (Lokalisation ist die Niere) und einen morphologischen Aspekt (Art der pathologischen Veränderung ist Entzündung). Ein Begriff wie Nephritis lässt sich daher aspektbezogen in mehrere Begriffshierarchien einordnen: bezogen auf die Topographie ist Pyelonephritis ein nachgeordneter Begriff (in einer partitiven Hierarchie, da das Nierenbecken Teil der Niere ist); bezogen auf die Morphologie ist akute Nephritis ein nachgeordneter Begriff (in einer generischen Hierarchie, da eine akute Entzündung eine Spezialisierung des Begriffs Entzündung ist). Wird eine bestimmte Bedeutungsdimension festgehalten (z.B. die topographische, so dass die Begriffe hinsichtlich der Ausprägung des Merkmals „Lokalisation“ angeordnet werden), ergibt sich eine Begriffshierarchie, die genau auf diese Bedeutungsdimension oder

Achse bezogen ist. Statt einen Vorrat an Begriffen lediglich bezogen auf eine einzige Achse zu ordnen, können dieselben Begriffe gleichzeitig bezogen auf mehrere Achsen geordnet werden. Die Achsen sind idealerweise voneinander unabhängig: auf der einen Achse können nachgeordnete Begriffe aufgesucht werden, während hinsichtlich der Bedeutungsdimension einer anderen Achse keine Veränderung auftritt (Nephritis > akute Nephritis bzw. Nephritis > Pyelonephritis). Genauso ist es aber auch möglich, hinsichtlich mehrerer Achsen, nachgeordnete Begriffe aufzusuchen (Nephritis > akute Pyelonephritis). Anschaulich bewegt man sich im ersten Fall nur in Richtung einer einzigen Achse durch die Begriffsordnung, während die Bewegung im zweiten Falle auch in Richtung der Raumdiagonalen erfolgt. Begriffsordnungen lassen sich durch die Anzahl der systematischen Achsen charakterisieren, gemäß derer die Begriffe angeordnet werden. Werden mehrere Achsen verwendet, so spricht man pauschal von multiaxialer Ordnung, sonst von uniaxialer. Typisch für bestehende uniaxiale Systeme mit einer generischen Begriffshierarchie ist, dass zwar alle Begriffe einheitlich geordnet sind, jedoch der Aspekt, bezogen auf den die Spezialisierungen vorgenommen werden, wechseln kann. Multiaxiale Begriffsordnungen sind aus mindestens zweierlei Gründen vorteilhaft: Erstens erlauben sie eine sehr differenzierte Einordnung von Begriffen hinsichtlich der durch die Achsen genau spezifizierten Ordnungskriterien. Sind zunächst nur die Hierarchien der achsenspezifischen Eigenschaften gegeben, so lässt sich ein komplexer, sehr differenzierter Begriff (genauer seine Definition) durch die Kombination dieser Eigenschaften gewissermaßen konstruieren. In diesem Sinne ist das Ausdrucksvermögen multiaxialer Ordnung höher als das uniaxialer Systeme. Zweitens erlaubt es eine multiaxiale Begriffsordnung Begriffe hinsichtlich bestimmter Kriterien selektiv herauszugreifen. Ein typisches Kriterium könnte es sein, eine bestimmte Lokalisation (z.B. Nierenbecken) festzuhalten. Das Ergebnis wären alle eingeordneten Begriffe (z.B. Erkrankungen), deren Lokalisation das Nierenbecken ist. In der Dokumentation beschleunigen Achsen das Auffinden relevanter Dokumente; hinsichtlich statistischer Fragestellungen lassen sich Klassierungen vornehmen, die der Fragestellung genau angepasst sind [RWT01].

C.4.4 Art der hierarchischen Struktur

Monohierarchie

Eine Monohierarchie liegt vor, wenn in einer hierarchischen Anordnung von Begriffen jeder Begriff nur einen einzigen übergeordneten Begriff hat.

Polyhierarchie

Eine Polyhierarchie liegt vor, wenn es Begriffe in einer hierarchischen Anordnung gibt, die mehrere übergeordnete Begriffe haben.

C.4.5 Codierprinzipien

Code-Typen

In Ordnungssystemen (Klassifikationssystemen und Thesauren) werden Begriffe häufig auf Codes abgebildet, um sie sprachunabhängig auf eine eindeutige Weise identifizieren zu können. Oft haben diese Codes eine eigene Struktur und tragen in dieser eine gewisse Bedeutung, d.h. aus dem Code kann in gewisser Hinsicht auf das, was er codiert, geschlossen werden. Die folgenden Definitionen betreffen Prinzipien der Codierung (vgl.: ISO/IEC TR 9789: 1989: Guidelines for the organization and representation of data element types for data interchange - Coding methods and principles)

Code (Schlüssel) Ein Code ist eine Zeichenkette oder Symbolkette, die dazu dient, einen Gegenstand oder einen Begriff zu bezeichnen.

Code (Codier-Schema) Menge von Regeln zur Abbildung von Elementen einer Menge auf die Elemente einer anderen Menge.

Code-Typen

- Nichtbedeutungstragende Codes
 - Sequentieller Code
 - * Inkrementeller Code
 - * Gruppencode
 - * Angeordneter Code
 - Zufallscode
- Bedeutungstragende Codes
 - Mnemonischer Code
 - Hierarchischer Code
 - Juxtapositionscode
 - Additionscode

Nicht-bedeutungstragende Codes

Bei einem nichtbedeutungstragenden Code stehen die Code-Werte nicht oder nur teilweise in Bezug zu den Datenelementen.

1. Sequentieller Code: Bei einem sequentiellen Code werden Ziffern oder Zeichen, die einer geordneten Menge entstammen, sequentiell den Datenelementen zugeordnet.
 - a) Inkrementeller Code: Ein inkrementeller Code ist ein sequentieller Code, bei dem die Code-Werte inkrementell (um gleiche Schritte gegenüber dem Vorgänger erhöht) den Datenelementen zugewiesen werden.
 - b) Gruppensequentieller Code: Bei einem gruppensequentiellen Code werden Datenkategorien, mit gemeinsamen Merkmalen, sequentielle Code-Bereiche zugeordnet.
 - c) Angeordneter sequentieller Code: Ein angeordnet sequentieller Code ist ein sequentieller Code, bei dem die Code-Werte einer bereits geordneten Menge von Begriffen zugeordnet werden. Ein angeordneter sequentieller Code setzt die Kenntnis und Vollständigkeit der Datenelemente voraus.
2. Zufallscode: Bei einem Zufallscode werden den Datenelementen zufällig Code-Werte aus einer möglichen Menge von ungeordneten Code-Werten zugewiesen.

Bedeutungstragende Codes

Bei einem bedeutungstragenden Code stehen die Code-Werte in Bezug zu den Datenelementen.

1. Mnemonischer Code: Bei einem mnemonischen Code enthält der Code-Wert ein oder mehrere Zeichen der Benennung der Datenelemente. Zweck ist, dem Benutzer die Erinnerung der Code-Werte zu erleichtern.
2. Hierarchischer Code: Bei einem hierarchischen Code entsprechen die Code-Werte der hierarchischen Anordnung der Datenelemente.
3. Juxtapositionscode: Ein Juxtapositionscode ist ein zusammengesetzter Code, der aus einzelnen Segmenten besteht, die voneinander unabhängigen Merkmalen der Datenelemente entsprechen. Die in dieser Weise kombinierten Codes können von beliebigem Typ sein (sequentiell, mnemonisch, hierarchisch).
4. Kombinationscode: Ein Kombinationscode ist ein zusammengesetzter Code, der aus einzelnen Segmenten besteht, die voneinander abhängigen Merkmalen der Datenelemente entsprechen. Die in dieser Weise kombinierten Codes können von beliebigem Typ sein.

5. Additionscode: Bei einem Additionscode dient ein Code-Wert zur eindeutigen Darstellung einer Menge von Datenelementen. Dabei wird jedes Datenelement durch eine Zweier-Potenz dargestellt, und die Menge durch die Summe dieser Zahlen.

C.4.6 Klassifikation, Thesaurus und Nomenklatur

Klassifikation

Bisweilen werden im Terminus Klassifikation Prozess und Produkt miteinander vermennt. Grundsätzlich ist zu unterscheiden zwischen:

- Klassifikation als dem Prozess der Klassifikationserstellung, d.h. der Klassenbildung
- Klassifikation als dem Produkt des Klassenbildungsprozesses vorliegt, d.h. dem Klassifikationssystem
- Klassifikation als dem Prozess des Klassifizierens, d.h. dem Zuordnen von Elementen zu Klassen eines Klassifikationssystems

Der Wunsch, Ordnung zu erfassen oder Ordnung zu schaffen, scheint ein Grundbedürfnis des Menschen zu sein. Bei der Bildung von Klassifikationen ist unterschiedlich motiviert:

- erkenntnis- oder wissensorientiert: Suche nach der wahren Struktur von Objekten und Phänomenen (z.B. Klassifikation von Krankheiten)
- pragmatisch oder dokumentarisch orientiert: Aufbewahren und Wiederfinden von Dokumenten und Gegenständen, statistische Analyse von Phänomenen (Dokumentationssprachen, Thesauren)

Ein Klassifikationssystem ist die strukturierte Darstellung von Klassen und der zwischen ihnen bestehenden Begriffsbeziehungen. Klassifikationssysteme sind Hilfsmittel zur Ordnung von Gegenständen oder Wissen über Gegenstände. Die Elemente von Klassifikationssystemen sind Klassen und einzelne Begriffe und deren Bezeichnungen. Klassifikationen sind deshalb spezielle Begriffssysteme. Ein Klassifikationssystem entsteht durch den mehrfach nacheinander durchgeführten Prozess der Klassenbildung oder Untergliederung. Bei der Bildung von Klassen wird Gleiches bzw. Ähnliches zusammengefasst [Ber00].

Thesaurus

Ein Thesaurus, im Bereich der Information und Dokumentation, ist eine geordnete Zusammenstellung von Begriffen und Benennungen, die in einem Dokumentationsgebiet zum Indexieren, Speichern und Wiederauffinden dient. Ein Thesaurus ist eine pragmatisch oder dokumentarisch orientierte Klassifikation. Er ist durch folgende Merkmale gekennzeichnet:

- Begriffe und Bezeichnungen werden eindeutig aufeinander bezogen (*terminologische Kontrolle*), indem
 - Synonyme möglichst vollständig erfasst werden,
 - für jeden Begriff eine Bezeichnung (Vorzugsbezeichnung, Begriffsnummer oder Notation) festgelegt wird, die den Begriff eindeutig vertritt.

Beziehungen zwischen Begriffen (repräsentiert durch ihre Bezeichnungen) werden dargestellt (DIN 1463 Teil 1) [Ber00].

Nomenklatur

Nach Wingert ist eine Nomenklatur eine systematische Ordnung von Namen [Win79]. Bernauer spricht von einem nach vorab festgelegten Regeln erarbeiteten System von Termini in einem Fachgebiet [Ber00]. In der Regel basiert diese Systematik auf einer Begriffsordnung. Meist ist dies eine bedeutungsbezogene Ordnung. Eine Nomenklatur kann mittels Definitionen, Angaben zu Bezeichnungsproblematik (Synonymie, Homonymie, etc.), sprachliche Varianten und Wortformen erweitert werden zu einem Thesaurus [RWT01].

C.5 Medizinische Begriffssysteme

C.5.1 Übersicht

Die UML-Diagramme C.2 und C.3 geben einen Überblick über die wichtigsten Zusammenhänge in der Welt der medizinischen Begriffssysteme und bilden ein systematisches Begriffssystem für die Begriffssysteme.

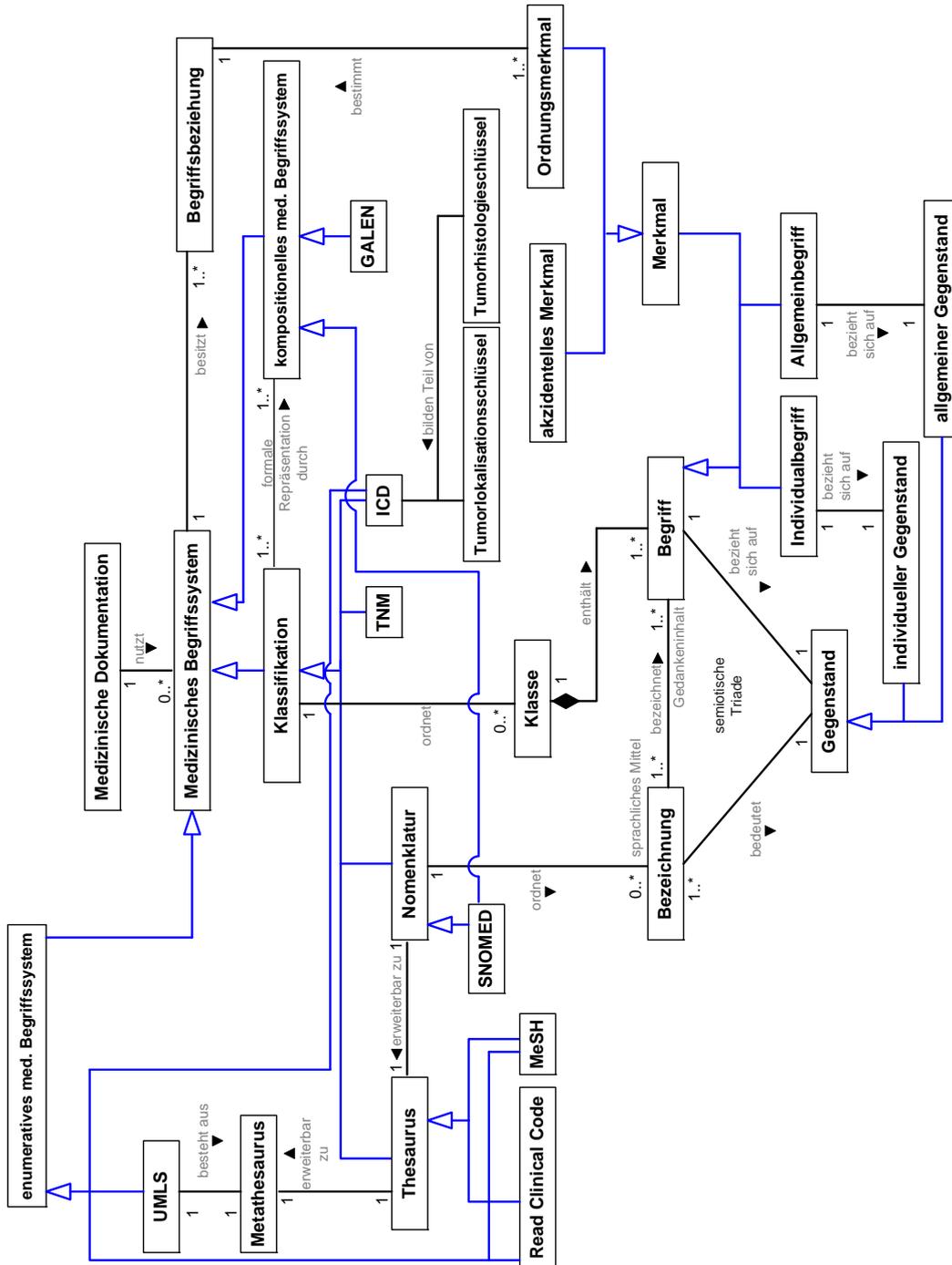


Abbildung C.2: Ausschnitt medizinische Begriffssysteme als UML-Diagramm

C.5.2 TNM-Klassifikation und Tumorlokalisationsschlüssel

Das TNM-System dient der Beschreibung der Stadien bösartiger Tumoren und wurde durch die UICC (Union Internationale contre le Cancer) festgelegt. Es wird unterschieden zwischen der prätherapeutischen klinischen Klassifikation ((c)TNM), die vor der Behandlung erfolgt und der postoperativen pathohistologischen Klassifikation (pTNM). Zur Beschreibung der anatomischen Ausbreitung und der Malignität der Tumoren werden folgende Merkmale benutzt:

- T = Tumor: Ausdehnung des Primärtumors, wobei T von T0 für keinen Primärtumor über T1 für kleine Tumoren bis T3 oder T4 für große Tumoren reicht.
- N = Node: Zustand der regionären Lymphknoten, wobei N0 für keine regionären Lymphknotenmetastasen steht, N1 für den Befall von Lymphknoten in der nächsten Umgebung des Tumors bis 3 cm Größe und N3 für den Befall der regionären Lymphknoten mit mehr als 3 cm in größter Ausdehnung.
- M = Metastasis: Fehlen bzw. Vorhandensein von Fernmetastasen, M0 für keine Fernmetastasen und M1 für Fernmetastasen.
- C = Certainty: Grad der Sicherung des Befundes; C1 für die Diagnosesicherung aufgrund von diagnostischen Standardmethoden, C2 bei Aussage aufgrund spezieller diagnostischer Verfahren und C3 bei chirurgischen Explorationen.
- Präfix y, Präfix r: Der Präfix y wird verwendet, wenn die Klassifikation während oder nach initialer multimodaler Therapie erfolgt. Der Präfix r dient dem Kennzeichnen vom Vorhandensein von Tumorrezidiven.

Somit beschreibt die TNM-Klassifikation einheitlich die anatomische Ausdehnung maligner Tumorerkrankungen und soll die topographischen und histologischen Beschreibungen der Tumoren, z.B. durch den ICD-O, ergänzen [LGHKG99].

C.5.3 ICD

Die International Classification of Diseases ist eine uniaxiale, multihierarchische Klassifikation mit generischen und pragmatischen Begriffsbeziehungen. Seine Codierprinzipien sind der gruppensequentielle und der hierarchische Code.

C.5.4 SNOMED

SNOMED ist die Abkürzung für Systematized Nomenclature of Human and Veterinary Medicine, gehört also zu den Nomenklaturen. SNOMED ist multiaxial, monohierarchisch mit generischen und partitiven Begriffsbeziehungen. Der mnemonische, hierarchische, gruppensequentielle, inkrementelle und Kombinationscode bilden die Codierprinzipien für SNOMED.

C.5.5 RCC, MeSH, UMLS und GALEN

Eine Nomenklatur kann erweitert werden zu einem Thesaurus. Hierzu gehören der Read Clinical Code (RCC) und die Medical Subject Headings (MeSH). Wird wiederum der Thesaurus erweitert um semantische und linguistische Informationen, so entsteht ein Metathesaurus, wie ihn der Kern von UMLS (Unified Medical Language System) darstellt. UMLS gehört zu den so genannten enumerativen Begriffssystemen und versucht, die wichtigsten medizinischen Begriffssysteme in einem einzigen zu integrieren und dabei möglichst viele Begriffsbeziehungen explizit darzustellen. UMLS ist ein Langzeit-Forschungs- und Entwicklungsprojekt der NLM (U.S. National Library of Medicine) mit dem Zweck, das Zurückgewinnen (Information Retrieval) und die Integration von Informationen unterschiedlicher Quellen zu unterstützen. Dazu abgegrenzt werden kompositionelle Begriffssysteme, wie z.B. GALEN (Generalized Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine), welches versucht, methodische Grundlagen für die formale Repräsentation medizinischer Begriffssysteme zu entwickeln. GALEN erlaubt die formale Konstruktion von allen anderen fachrelevanten Begriffen durch die Komposition einer Anzahl primitiver Begriffe [Ber95].

D Epidemiologisches Krebsregister Niedersachsen

D.1 Einleitung

Krebserkrankungen in Deutschland nehmen immer mehr an Bedeutung zu. Bereits ca. jeder Vierte bis Fünfte stirbt an einem bösartigen Tumor, die Tendenz ist steigend [Mut96]. Dies liegt zum Teil auch an der gestiegenen Lebenserwartung der Menschen, da Krebs zu meist eine Erkrankung des höheren Alters ist. So gehört Krebs neben den Herz-Kreislauf-erkrankungen zu den häufigsten Todesursachen. 80% der Krebstodesfälle sind auf Karzino me zurückzuführen. In den westlichen Industriestaaten bedeutet dies jährlich mindes tens 1 Million Opfer, in der Bundesrepublik Deutschland allein über 200.000 Menschen [Abe95].

Es besteht daher zweifelsfrei ein öffentliches Interesse, räumliche und zeitliche Vertei lungen von Krebserkrankungen möglichst vollständig und genau zu erfassen und wissen schaftlich zu analysieren. Mit Hilfe der Ergebnisse wird eine Erforschung der Ursachen und Risiken für maligne Tumoren betrieben werden, um dann bessere und gezieltere Prä ven tion und Therapie anbieten zu können [ABK97]. Um alle benötigten Daten für ein epidemiologisches Krebsregister sammeln zu können, müssen verschiedene Institutionen einbezogen werden, die Daten über Krebserkrankungen erheben. Dies sind in Nieder sachsen Nachsorgeleitstellen, klinische Register, Ärzte und Zahnärzte, Pathologen und Gesundheitsämter. Meldende Einrichtungen in einem anderen Sinne sind das Landesver messungsamt, das Landesamt für Statistik und die Einwohnermeldeämter [RW98].

Es bleibt zu wünschen, dass die flächendeckende Krebsregistrierung den erhofften Beitrag im Kampf gegen den Krebs leistet.

D.2 Epidemiologische Krebsregister

Epidemiologische, das heißt bevölkerungsbezogene Krebsregister messen z.B. die Krebs in zidenz, analysieren Häufungen von Krebserkrankungen, errechnen Überlebensraten und mittleres Erkrankungsalter, liefern Aussagen über die Krebsprävalenz und sind Basis

für Ursachen- und Risikostudien. Nur eine flächendeckende Registrierung mit über 90% erfasster Krebsfälle führt zu statistisch-epidemiologischen Auswertungen mit ausreichender Qualität.

D.2.1 Allgemeines

Am 1.1.1995 trat das vom Bundestag und Bundesrat verabschiedete Bundeskrebsregistergesetz (KRG) in Kraft, das alle Bundesländer verpflichtet, ein epidemiologisches Krebsregister bis zum 1. Januar 1999 aufzubauen. Durch jeweilige Landesgesetze konnten abweichende Regelungen für die Einrichtung und Führung der Krebsregister getroffen werden. Am 16. November 1999 wurde das Gesetz über das Epidemiologische Krebsregister Niedersachsen (GEKN) vom Niedersächsischen Landtag beschlossen und ist am 1. Januar 2000 in Kraft getreten.

Laut § 1 (3) des Bundeskrebsregistergesetzes (KRG) müssen die Krebsregister aus selbstständigen, räumlich, organisatorisch und personell voneinander getrennten Vertrauensstellen und Registerstellen bestehen. Damit wurde das von Prof. Dr. Michaelis vorgeschlagene Meldemodell für alle epidemiologischen Krebsregister zur Grundlage. Die Vertrauensstelle (VST) sammelt patientenbezogene Krebsmeldungen von Ärzten (einschließlich Pathologen und Zahnärzten), teilweise über Nachsorgeleitstellen und klinische Register. Sie chiffriert die personenidentifizierenden Daten und übermittelt die chiffrierten personenidentifizierenden Daten, zusammen mit den im Klartext vorliegenden epidemiologischen Daten, an die Registerstelle. Die Registerstelle (RST) speichert die chiffrierten personenidentifizierenden und die epidemiologischen Daten. Hier werden mehrere zu einem Patienten gehörende Meldungen zusammengeführt [AMST96].

Die Vertrauensstelle muss nach dem KRG (§ 4) folgende Aufgaben erfüllen:

1. Prüfung der Meldungen auf Schlüssigkeit und Vollständigkeit, Klärung von Unklarheiten durch Rückfragen beim Melder
2. Bearbeitung der Totenscheindaten wie eine Meldung und Übernahme der Daten in den entsprechenden Datensatz
3. Chiffrierung der personenidentifizierenden Daten sowie Generierung von Kontrollnummern
4. Übermittlung der chiffrierten personenidentifizierenden Daten und der epidemiologischen Daten im Klartext an die Registerstelle sowie Vernichtung aller gemeldeten Daten nach der Übermittlung, spätestens jedoch nach drei Monaten
5. Abgleichung, Entschlüsselung und Übermittlung von Identitätsdaten bei Forschungsaufgaben des öffentlichen Interesses sowie Erhebung zusätzlicher Daten beim Patienten und Einholung der Einwilligung des Patienten, falls dies erforderlich ist

6. Hat ein Patient einen Arzt oder Zahnarzt benannt, um über diesen Auskunft über seine Eintragungen ins Krebsregister zu erhalten, Erteilung dieser Auskunft. Unter Umständen mit vorheriger Anforderung der Daten bei der Registerstelle
7. Veranlassung einer Löschung, Zählung der Löschung und schriftliche Benachrichtigung des meldenden Arztes oder Zahnarztes, falls ein Patient der Meldung widersprochen hat.
8. Installation entsprechender technischer und organisatorischer Maßnahmen nach § 9 Bundesdatenschutzgesetz und vor allem, damit personenidentifizierende Daten nicht von Unbefugten eingesehen oder genutzt werden können.

Die Registerstelle muss nach dem KRG (§ 5) folgende Aufgaben erfüllen:

1. Speicherung der von der Vertrauensstelle übermittelten Daten
2. Abgleich der neuen Daten mit den vorhandenen Datensätzen im Register mittels der zuvor von der Vertrauensstelle generierten Kontrollnummern
3. Prüfung auf Schlüssigkeit; Berichtigung und Ergänzung gegebenenfalls nach Rücksprache mit der Vertrauensstelle
4. Abschlussmeldung an die Vertrauensstelle nach Verarbeitung der Daten im Register
5. Regelmäßiger Abgleich der Kontrollnummern mit anderen epidemiologischen Registern, um die Daten zu berichtigen und zu ergänzen
6. Auswertungen und Analysen der anonymisierten Daten gemäß den Aufgaben eines epidemiologischen Registers
7. Jährliche anonymisierte Übermittlung an das Robert-Koch-Institut in Berlin für ein bundesweites Register
8. Übermittlung der angeforderten Daten an die Vertrauensstelle für Forschungsvorhaben mit öffentlichem Interesse und bei Auskünften für Patienten
9. Löschung der entsprechenden Daten bei Widerspruch eines Patienten nach Mitteilung von der Vertrauensstelle

D.3 Epidemiologisches Krebsregister Niedersachsen (EKN)

Die ärztlich geleitete Vertrauensstelle ist in Niedersachsen beim Niedersächsischen Landesgesundheitsamt unter der Leitung von Prof. Dr. Windorfer angesiedelt. Seit dem 1. Januar 2003 nimmt die OFFIS CARE GmbH, ein zwei Jahre zuvor aus dem Informatikinstitut OFFIS heraus gegründetes Unternehmen, die Aufgaben der Registerstelle des Epidemiologischen Krebsregisters Niedersachsen (EKN) wahr und sammelt Daten zur Krebsforschung und -Prävention. Hierzu wurde die OFFIS CARE durch das Land Niedersachsen mit der Wahrung der Aufgaben der Registerstelle des EKN beliehen, Geschäftsführer der OFFIS CARE GmbH ist Dr. Wilfried Thoben. Die Softwarewerkzeuge, die der Vertrauensstelle für die Erfüllung ihrer Aufgaben zur Zeit zur Verfügung stehen, sind CARAMEL für die Anbindung der Melder und Plausibilitätskontrollen sowie CARTRUST. CARTRUST unterstützt die Bearbeitung von Meldungen einschließlich Prüfung, Chiffrierung der personenidentifizierenden Daten, die Kontrollnummerngenerierung (UNICON), eine einheitliche Geokodierung von Adressen sowie die datentechnische Kommunikation mit der Registerstelle. Die Registerstelle wird zur Zeit von den Software-Werkzeugen CARELIS und CARESS unterstützt. CARELIS dient der vollständigen Bearbeitung der Meldungen in der Registerstelle, die wichtigste Funktion ist der automatische stochastische Abgleich der Datensätze auf Basis der UNICON-Kontrollnummern. Danach können Mitarbeiter manuell weitere Zusammenführungen von Daten (z.B. Best-Of-Generierung) vornehmen. CARESS ist das Werkzeug für die umfassenden epidemiologischen Auswertungen und Aufbereitungen wie:

- Qualitätssicherung im Register durch ein Angebot von Indikatoren wie M/I, DCO bzw. DCN oder HV
- Inzidenzmonitoring, Durchführung von Überlebenszeit- und Clusteranalysen auf der Basis eines einheitlichen Raumbezugs
- Hypothesenbildung über Risikofaktoren von Krebs durch die vergleichende Einbeziehung raumbezogener und soziodemographischer Hintergrunddaten sowie
- Berichterstellung und Export von Daten an die Forschung

Für den Raumbezug wird das Amtliche Topographisch - Kartographische Informationssystem (ATKIS) genutzt. [www.offis.uni-oldenburg.de/projekte/...].

D.4 Meldungsarten an das EKN

Grob kann man zur Zeit vier Meldungsarten beim EKN unterscheiden, die auch so im Datenbankschema des EKN definiert sind [ABH99]:

- Standardmeldungen
- Pathologenmeldungen
- Sterbemeldungen
- Anfragemeldungen

Standardmeldungen

Diese umfassen Direktmeldungen von Ärzten und Zahnärzten sowie Meldungen über Nachsorgeleitstellen oder Klinische Register. In der Regel liegen die Angaben aus dem Patientenfragebogen und die Einwilligung des Patienten vor.

Pathologenmeldungen

Diese Meldergruppe hat wichtige Funktionen in der Diagnostik oder Begutachtung, hat aber in der Regel keinen direkten Kontakt zum Betroffenen. Da in diesen Fällen die Einholung der Patienteneinwilligung äußerst umständlich wäre, gibt es den Weg der so genannten „peripheren Teilanonymisierung mit zentraler Verschlüsselung“, der den Anforderungen des Datenschutzes gerecht wird.

Sterbemeldungen

Gesundheitsämter melden über Todesbescheinigungen einen Todesfall, aber auch Standardmelder berichten über den Tod eines Tumorpatienten (klinischer Abschluss). In ca. 30% In diese Meldeart sollen die Einwohnermeldeämter mit Meldungen über Adressänderungen eingebunden werden [RW98].

Anfragemeldungen

Hiermit sind Anfragen von Betroffenen auf Auskunft oder Anträge von Betroffenen auf Löschung ihrer Daten gemeint.

D.5 Problemstellung und Zielsetzung der Evaluation

Alle meldenden Institutionen an ein epidemiologisches Krebsregister können einen ganz unterschiedlichen technischen Stand ihrer EDV aufweisen sowie entsprechend ihrer jeweiligen Aufgabengebiete völlig unterschiedlich aufgebaute Softwaresysteme haben. Die gesetzlichen Grundlagen (Bundes- und Landeskrebsregistergesetz) schreiben ganz klare Vorgehensweisen vor. So soll in punkto Datenschutz und Datensicherheit eine hohe Qualität erfüllt werden. Andererseits müssen mindestens 90% der gewünschten Daten erfasst werden können, damit ein epidemiologisches Krebsregister seine Aufgaben erfüllen kann. Diverse Meldergruppen sind nach den jeweiligen Landeskrebsregistergesetzen zu lückenlosen Meldungen verpflichtet worden, die Gesundheitsämter bereits durch das Bundeskrebsregistergesetz. Nun kann man von den beiden Standpunkten aus, nämlich Melder und Krebsregister, folgende Situation festhalten:

Sicht der Melder:

Für die meldenden Institutionen bedeutet die Meldung an ein Krebsregister zusätzlichen Aufwand, der bei meist ohnehin knapper Personalkapazität, oftmals als Last empfunden wird. Motivation stellt natürlich die gesetzliche Meldepflicht für einzelne Gruppen sowie die Vergütung pro Meldung dar. Für den Melder ist also eine einfache Erfassung der zu meldenden Daten für das Register von Bedeutung. Einfach meint hier, möglichst geringe Investitionen, Lernaufwand, Zeitaufwand und Wartung für den Melder. Wichtig sind weiter, eine schnelle, einfache, sichere und kostengünstige Übertragung der Daten an das Krebsregister sowie gut funktionierende Vergütungsmechanismen, die transparent sind und kaum geprüft werden müssen.

Sicht des Krebsregisters:

Das Krebsregister braucht, um überhaupt qualitativ gut arbeiten zu können, möglichst komplette und korrekte Datensätze. Dies bedeutet auf der einen Seite eine möglichst komplette und korrekte Erhebung aller Krebsfälle und auf der anderen Seite möglichst alle Daten zu einem einzelnen Krebspatienten. So ist eine gute Qualität der gemeldeten Daten im eben genannten Sinne eine „conditio sine qua non“ für ein Krebsregister. Die Daten sollen sicher übertragen werden, unverfälscht sein und einen nachvollziehbaren Urheber haben. Das Krebsregister legt außerdem gesteigerten Wert auf einen regelmäßigen, möglichst aktuellen Eingang der Daten in einem Format, welches sofort weiterverarbeitet werden kann. Schließlich muss für das Krebsregister die Vergütungshöhe für den einzelnen Melder einfach zu ermitteln sein und die nachfolgende Abwicklung muss ebenfalls entsprechend gut funktionieren.

Derzeit macht die Anbindung der Melder an die epidemiologischen Krebsregister aufgrund ihrer Heterogenität Probleme. Mühevoll muss analysiert werden, welche Gegebenheiten beim Melder vorliegen und viele individuelle Lösungen wurden geschaffen. Hinzu kommt, dass noch vielfach zum papiernen Meldebogen gegriffen werden muss, was die mehrmalige Erfassung der Daten zur Folge hat. Kleine eigenständige Software, die nur zur elektronischen Datenerfassung im PC dient und allen Meldergruppen in gleicher Form

angeboten wird, lässt den Melder die Daten zweimalig erfassen. Die Übertragung der Daten zu den Registern erfolgt oft über den Postweg. Folge ist, dass viele Krebserkrankungen nicht oder nicht vollständig gemeldet werden, weil die Meldetätigkeit zu unkomfortabel und mühevoll ist.

D.6 Grundlagen zum Epidemiologischen Krebsregister

In diesem Kapitel werden die für diese Arbeit notwendigen Grundlagen vorgestellt. Zunächst wird das Szenario für die Meldewege erläutert, wie sie zum Beispiel beim epidemiologischen Krebsregister in Niedersachsen vorkommen. Danach folgt eine Betrachtung der Tumordokumentationen. Diese Betrachtung und auch die Bewertung der Möglichkeiten erfolgen vor allem unter Berücksichtigung der hier angestrebten Zielsetzung und erheben keinesfalls den Anspruch auf Vollständigkeit.

D.6.1 Anwendungsszenario

Im vorangegangenen Kapitel wurde bereits verdeutlicht, warum von mehreren Institutionen die Daten über Krebsfälle für epidemiologische Krebsregister möglichst vollständig gesammelt werden müssen.

Meldende Institutionen

Meldende Institutionen sind in Niedersachsen im wesentlichen Nachsorgeleitstellen, Ärzte, Zahnärzte, Klinische Register, Pathologen, Gesundheitsämter und Einwohnermeldeämter (siehe Abbildung D.1). Hinzu kommen das Landesvermessungsamt und das Landesamt für Statistik, die hier nicht näher betrachtet werden [RW98].

Nachsorgeleitstellen In Niedersachsen gibt es insgesamt sieben Nachsorgeleitstellen, welche in Braunschweig, Bremen, Göttingen, Hannover, Oldenburg, Osnabrück und in Stade angesiedelt sind [RW98]. Nachsorgeleitstellen dienen hauptsächlich der flächendeckenden und wohnortnahen Versorgung der Tumorpatienten. Sie sollen Sorge dafür tragen, dass Tumorpatienten gemäß ihrer Tumordiagnose optimal nachgesorgt werden, dass die Kommunikation zwischen den behandelnden Ärzten verbessert wird und dass die verschiedenen Befund- und Behandlungsdaten zusammengeführt werden [Tum98].

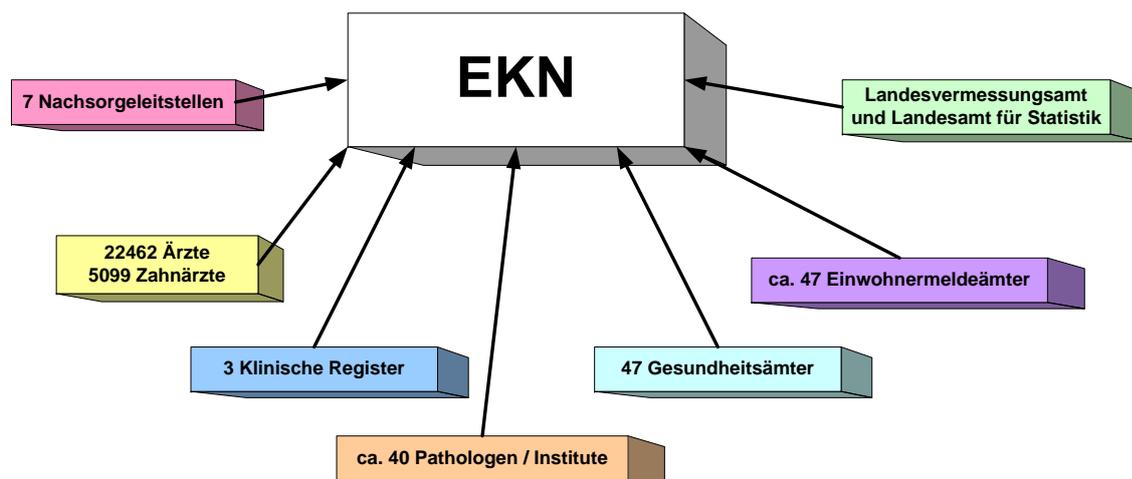


Abbildung D.1: Meldende Institutionen an das EKN [RW98]

Ärzte / Zahnärzte In Niedersachsen sind ca. 23.000 Ärzte und ca. 5.100 Zahnärzte tätig [RW98]. Nicht alle Fachrichtungen sind gleichermaßen mit Tumorpatienten befasst. So werden die meisten Fachärzte für Allgemeinmedizin als Lotsen und erste Anlaufstelle für den Patienten zwar Tumorverdachte haben, diese jedoch bei entsprechenden Kollegen anderer Fachrichtungen abklären lassen. Erst wenn beim eingeschalteten Fachkollegen die Diagnose feststeht, kann gemeldet werden. Daher sind besonders die Ärzte der Fachrichtungen Dermatologie, Gynäkologie, Urologie, Innere Medizin und Chirurgie potentielle Melder. Die Zahnärzte werden in der Regel bei Verdacht auf Tumore zu Kieferchirurgen, Hals-Nasen-Ohren-Ärzten und Dermatologen überweisen.

Klinische Register Sie sind, wie die Nachsorgeleitstellen, so genannte versorgungsorientierte Krebsregister. Erfasst werden Daten zur Diagnose, Behandlung und Verlauf der Tumorerkrankung möglichst aller in der entsprechenden Klinik betreuten Patienten, egal woher diese stammen. Risikofaktoren wie Tätigkeiten, Wohnortanamnese oder Raucherstatus werden in der Regel nicht aufgenommen. Diese Daten fehlen für die Auswertung im epidemiologischen Krebsregister.

Pathologen bzw. Pathologische Institute Erst der Pathologe macht eine Tumordiagnose komplett. Nur er kann durch die histologische Untersuchung die genaue Tumordiagnose liefern, kann Lymphknotenbefall oder Veränderungen in anderen Geweben im Frühstadium erkennen und wird Differenzierungsgrad und Dignität einstufen. Der Pathologe hat in der Regel keinen direkten Patientenkontakt und erfasst meist nur die nötigsten Patientendaten, wie Personalien und Abrechnungsdaten. Aufgrund der Genauigkeit seiner Tumordiagnose ist die Meldung des Pathologen immens wichtig für das Krebsregister. In Niedersachsen ist der Pathologe nach dem Landeskrebsregistergesetz meldepflichtig.

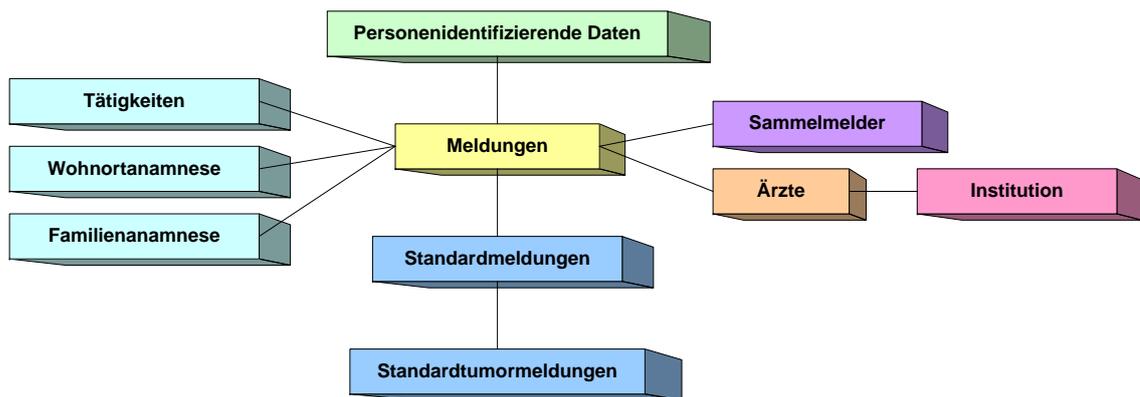


Abbildung D.2: Datenschema Standardmeldeweg an das EKN [RW99]

Gesundheitsämter Sie verwalten die Todesbescheinigungen aller Sterbefälle und somit auch die der Tumorpatienten. Alle Sterbemeldungen werden ungeachtet der Todesursache an das epidemiologische Krebsregister geliefert, da erst im Datenabgleich im epidemiologischen Krebsregister ermittelt werden kann, ob der Tote als Tumorpatient gemeldet war. Sterbemeldungen werden auch von Nachsorgeleitstellen, Ärzten, Zahnärzten und Klinischen Registern unter dem Stichwort *klinischer Abschluss* gemacht.

Einwohnermeldeämter Sie sollen die Datenqualität der Patientenadressen im epidemiologischen Krebsregister erhöhen. Von den Ämtern werden Adressänderungsmeldungen erwartet, Meldeweg ist der der Gesundheitsämter.

Landesamt für Statistik Zu vereinbarten Zeiten werden benötigte Bevölkerungszahlen an das EKN geliefert. Außerdem ist es in den Meldeweg der Gesundheitsämter integriert [RW98].

Landesvermessungsamt Es liefert zu vereinbarten Zeiten die benötigten Geodaten an das EKN [RW98].

Meldungsarten

Standardmeldeweg Dieser Weg wird von den Nachsorgeleitstellen, Ärzten und Zahnärzten sowie den klinischen Registern besritten. Im Entitäten-Relationen-Schema der Vertrauensstelle ist er, wie Abbildung D.2 zeigt, abgebildet [RW99]:

Die Entitäten erfassen im wesentlichen jeweils folgende Daten:

- Personenidentifizierende Daten: Name, Vorname, Geburtsname, früherer Name, Geschlecht, Geburtsdatum, Adresse, Sterbedatum, Diagnosedatum
- Sammelmelder: Institutionsname, Abteilung, Adresse, Bankdaten, Ansprechpartner
- Ärzte: Name, Fachrichtung, Bankdaten, KV-Nummer
- Institution: Name, Typ, Adresse, Abteilung, Chefarzt
- Tätigkeiten: Beruf, Dauer der Tätigkeit, letzte und längste Tätigkeit
- Wohnortanamnese: seit wann aktueller Wohnort, Geburtsort, Geburtsland, wo aufgewachsen, längster Aufenthalt
- Familienanamnese: Krebserkrankungen in der Verwandtschaft, welcher Verwandtschaftsgrad, welcher Tumor, Geschlecht
- Standardmeldungen: Geschlecht, Staatsangehörigkeit, Geburtsdatum, Verstorben (?), Mehrling, Wohnort, Raucherstatus
- Standardtumormeldungen: Diagnoseanlass, Diagnosedatum, Diagnose, Histologie, Lokalisation, Dignität, Seite, Diagnosesicherung, Ausbreitung, Grading, Tumorfolge, Diagnosegültigkeit
- Meldungen: Meldestatus, Meldemodus, Eingangsdatum, Erfassungsdatum, Ausgangsdatum, Abrechnungsstatus

Genauerer zu den Datenfeldern und den Relationen findet man in [RW99].

Der Ablauf auf dem Standardmeldeweg stellt sich wie folgt dar:

1. Der Melder schickt in getrennten Dateien die personenidentifizierenden Daten (PID) und die epidemiologischen Daten (ED) an die Vertrauensstelle.
2. Softwarewerkzeuge (z. Zt. CAMEL und CARTRUST) sind nun hilfreich bei der Prüfung der Meldungen auf Schlüssigkeit und Vollständigkeit. Bei Unklarheiten wird beim Melder nachgefragt.
3. Chiffrierung der personenidentifizierenden Daten durch asymmetrische RSA-Verschlüsselung mit dem öffentlichen Schlüssel der Vertrauensstelle. Es muss eine spätere Dechiffrierung auf Klartexte möglich sein, da sonst für Forschungsvorhaben entsprechende Daten der Patienten für einen konkreten Bezug fehlen. Verschlüsselt wird mit einem hybriden IDEA-RSA-Verfahren, wobei ein einziges Schlüsselpaar verwendet wird. Die personenidentifizierenden Daten werden dazu mit einem, nach einem Zufallsverfahren erzeugten, IDEA-Schlüssel chiffriert. Der IDEA-Schlüssel wird RSA-verschlüsselt mit abgespeichert. Mindestlänge des RSA-Schlüssels sollte 640 Bit sein. Der öffentliche Schlüssel wird dabei innerhalb der Vertrauensstelle geheim gehalten, der private bzw. geheime Schlüssel wird außerhalb der Vertrauensstelle sicher verwahrt [AMST96].

4. Generierung der Kontrollnummern mit UNICON für eine spätere Zusammenführung von zusammengehörenden Datensätzen und bundesweiten Abgleich zwischen den Landeskrebsregistern. Jedes Attribut (i.w. Name, Vorname, Geburtstag und Titel) wird bundeseinheitlich mit dem Einwegverfahren MD5 verschlüsselt und anschließend symmetrisch nach dem IDEA-Verfahren chiffriert. Ergebnis sind die Kontrollnummern, deren Format als „Linkage Format“ bezeichnet wird [AMST96].
5. Übermittlung der chiffrierten personenidentifizierenden Daten, der Kontrollnummern und der epidemiologischen Daten im Klartext an die Registerstelle.
6. Löschung der gemeldeten Daten durch die Vertrauensstelle nach der Übermittlung, spätestens nach 3 Monaten.
7. In der Registerstelle wird nun auf der Basis der Kontrollnummern, mit Hilfe des Softwarewerkzeuges CARELIS, so genanntes „Record Linkage“ betrieben, um zusammengehörende Datensätze zusammenzuführen. Für die Zusammenführung werden die Kontrollnummern, das Geschlecht, Geburtsmonat und -jahr sowie die Gemeindekennziffer verwendet [RW98]. Ziel ist es, den komplettesten und qualitativ hochwertigsten Datensatz zu einem Krebsfall zu bekommen, der aus allen gemeldeten Daten möglich ist.

Im Datenschema entsprechen die personenidentifizierenden Daten der Entität Personenidentifizierende Daten, die epidemiologischen Daten entsprechen den Entitäten Standardtumorermeldung, Tätigkeiten, Wohnort- und Familienanamnese. In der Entität Standardmeldung finden sich sowohl personenidentifizierende als auch epidemiologische Daten.

Probleme bei diesem Meldeweg:

Zwischen klinischen Registern und Nachsorgeleitstellen gibt es in Niedersachsen historisch gewachsene Zusammenarbeit. Dabei werden Daten über Krebspatienten ausgetauscht und liegen zum Teil identisch in beiden Institutionen vor. Bei den Meldungen an das epidemiologische Krebsregister kommt es also zu Mehrfachmeldungen ein und desselben Patientenfalles. Auch die Zusammenarbeit von Pathologie und Klinik kann unter Umständen zu Doppelmeldungen von Daten führen [RW98]. Außer der Datenredundanz gilt es auch die mehrfache Vergütung dieser Meldung zu verhindern.

Es sind zur Zeit keine Mechanismen vorhanden, die in der Lage sind ohne Hilfe der Melder derartige institutionalisierte Mehrfachmeldungen zu erkennen und zu verhindern. Laut [RW98] ist es auch nicht möglich einen Kontrollmechanismus in CARAMEL zu implementieren, der ohne eine Markierung durch den Melder auskäme. Der Melder müsste angeben, ob er die Meldung aus einem anderen Melderdatenbestand übernommen hat oder nicht.

Ein weiteres nicht unerhebliches Problem beim Standardmeldeweg ist, dass die Software CARAMEL, mit der die Daten bei der Vertrauensstelle von den Meldern übernommen wird, automatisch davon ausgeht, dass von den jeweiligen Patienten die Einwilligungen

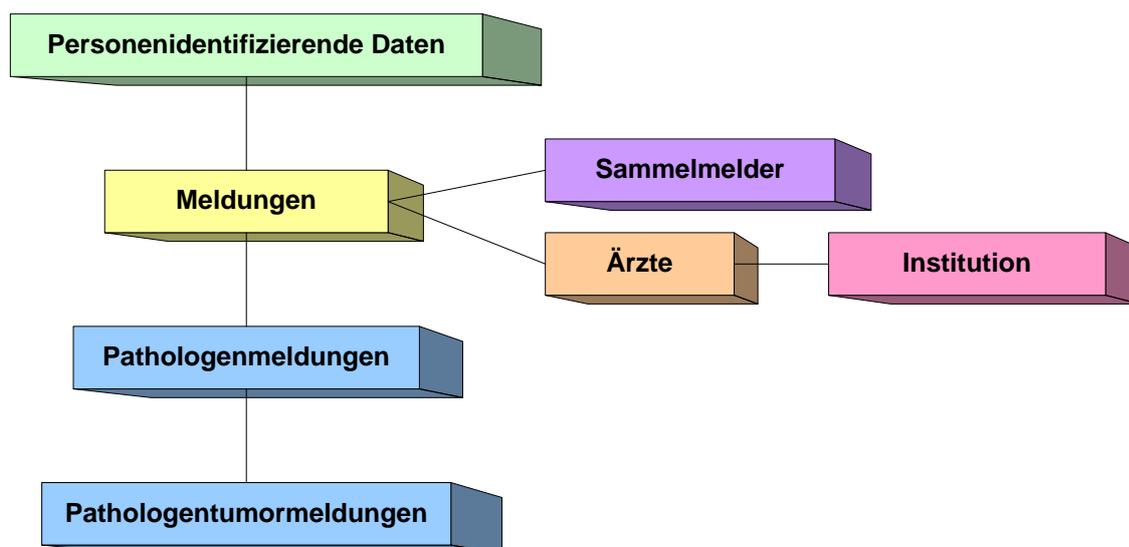


Abbildung D.3: Datenschema Pathologienmeldeweg an das EKN [RW99]

vorliegen, ohne dies zu prüfen. Sicher ist der Melder per Gesetz verpflichtet die Patientenzustimmung einzuholen, allerdings kann er auf die Ausnahmeregelung ausweichen, die aus medizinischen Gründen auf eine Aufklärung des Patienten verzichtet [RW98]. Auch bei diesem Problem müsste laut [RW98] eine Kennzeichnung durch den Melder erfolgen, damit CAMEL Meldungen ohne Einwilligungen herausfiltern könnte, um diese dann auf dem Pathologienmeldeweg faktisch anonymisiert weiter zu behandeln.

Pathologienmeldeweg Dieser Weg wird von den Pathologischen Instituten, niedergelassenen Pathologen u.a. wie Laborärzten oder Hämatologen ausgefüllt [ABH99]. Im Entitäten-Relationen-Schema der Vertrauensstelle bzw. Registerstelle ist er abgebildet wie die Abbildung D.3 zeigt [RW99].

Die Entitäten erfassen im wesentlichen jeweils folgende Daten:

- Pathologienmeldungen: Geschlecht, Geburtsjahr, Wohnort
- Pathologentumormeldungen: Diagnosejahr, Diagnosemonat, Diagnose, ICD-Version, Histologie, ICD-0-Version, Dignität, Seite, Lokalisationsschlüsselversion, Lokalisation, Grading, Diagnosesicherung, Ausbreitung, Tumorfolge

Die Besonderheit beim Meldeweg für die Pathologen ist die so genannte teilanonymisierte Verschlüsselung. Grund ist, dass der Pathologe in der Regel selbst keinen Patientenkontakt hat. Es werden die personenidentifizierenden Daten und die übrigen, epidemiologischen Daten, getrennt auf den Weg gebracht. Die personenidentifizierenden Daten gelangen zur Vertrauensstelle, die epidemiologischen Daten werden direkt der Registerstelle gemeldet.

Nachdem die Vertrauensstelle die personenidentifizierenden Daten verschlüsselt hat und die entsprechenden Kontrollnummern generiert sind, werden sie nach Übertragung zur Registerstelle dort anhand einer eindeutigen Kennung, die bereits beim Pathologen vergeben wurde, mit den zugehörigen epidemiologischen Daten des Patienten zusammengeführt. Die personenidentifizierenden Daten werden nicht für spätere Verwendung in Forschungsvorhaben verschlüsselt gespeichert, sondern nach der Generierung der Kontrollnummern und Geocodes gelöscht. Dadurch entsteht die faktische Anonymisierung.

Im Datenschema entsprechen die personenidentifizierenden Daten den Entitäten Personenidentifizierende Daten und Pathologenmeldungen, die epidemiologischen Daten entsprechen der Entität Pathologentumormeldungen.

Vorteil dieser Meldungsart ist, dass ohne Einholung der Patienteneinwilligung die Datenschutzbelange ausreichend berücksichtigt sind. Nachteil ist, dass eine eventuelle Rückverschlüsselung der Daten z.B. für Forschungsvorhaben unmöglich ist. Auch findet keine Plausibilitätsprüfung der medizinischen Daten in der Vertrauensstelle statt, wie dies beim Standardmeldeweg der Fall ist [RW98].

Probleme bei diesem Meldeweg

Der Melder muss stets mit zwei Stellen kommunizieren, nämlich mit der Vertrauens- und der Registerstelle. Die Registerstelle muss die gleichen Möglichkeiten bereitstellen wie die Vertrauensstelle, um für die direkt an sie eingegangenen epidemiologischen Daten die Plausibilität prüfen zu können.

Der Aufwand an Rückfragen für Melder, Vertrauensstelle und Registerstelle wird als hoch erwartet, so dass man auf die Bearbeitung von Nachfragen beim Melder, bezüglich der medizinischen Daten, verzichtet. Eine Plausibilitätsprüfung sollte am besten schon beim Melder stattfinden. In [RW98] lautet die dazu vorgeschlagene Lösung, einen Katalog von seltenen und / oder schwierigen Kodierungsfällen in Zusammenarbeit mit dem IARC / IACR zusammenzustellen und dieses Wissen den Meldern zur Verfügung zu stellen. Auch die Bestrebungen der Arbeitsgemeinschaft epidemiologischer Krebsregister Deutschland (ABKD) gehen in diese Richtung [RW98].

Sterbemeldeweg Dieser Meldeweg wird von den Gesundheitsämtern genutzt. Im Entitäten-Relationen-Schema der Vertrauensstelle ist er, wie Abbildung D.4 zeigt, abgebildet [RW99]:

Die Entitäten erfassen im wesentlichen jeweils folgende Daten:

- Sterbemeldungen: Todesjahr, Todesmonat, Gauß-Krüger-Koordinaten des Ortes, Grundleidensversion, ICD-Kodierung der unmittelbaren Todesursache, ICD-Kodierung des Grundleidens, ICD-Version der unmittelbaren Todesursache, Angaben über Autopsie, Verweis auf Arzt, Geschlecht, Wohnort, Version der Länderkodierung, Geburtsjahr, kodierte Staatsangehörigkeit, Geburtsmonat

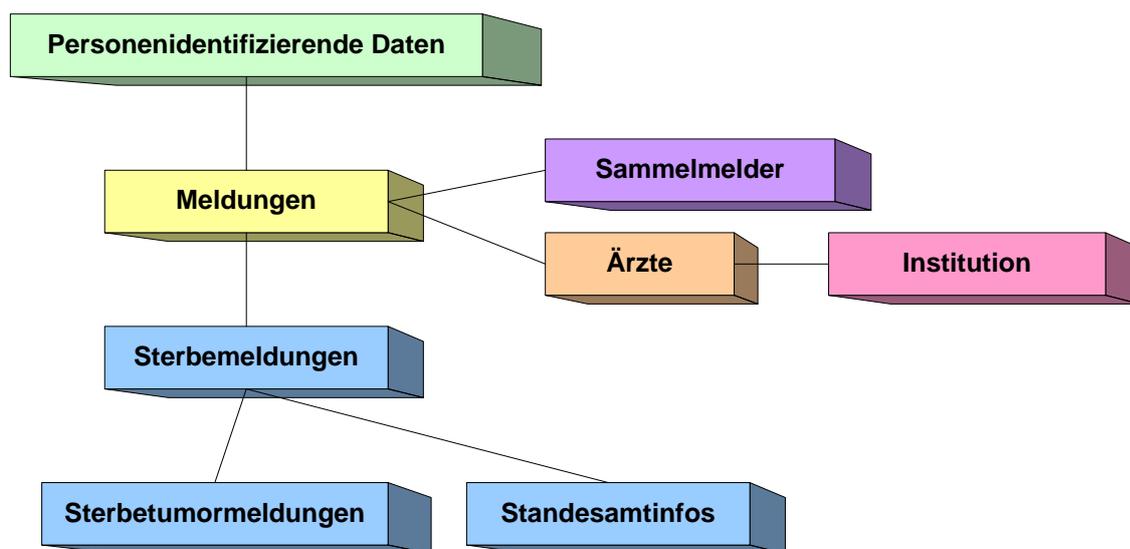


Abbildung D.4: Datenschema Sterbemeldeweg an das EKN [RW99]

- Sterbetumormeldungen: Art der Todesursache, Diagnosejahr, Diagnosemonat, Diagnoseanlass, Diagnose, ICD-Version, Histologie, ICD-O-Version, Lokalisation, Lokalisationsschlüsselversion, Dignität, Seite, Diagnosesicherung, Ausbreitung, Grading, Tumorfolge, Tod tumorbedingt?
- Standesamtinfos: Standesamtnummer, Sterbebuchnummer, Jahrgang, Buchart, Doppelbuchnummer

Dieser Meldeweg ist als dreistufiger Meldeweg unter Einbeziehung der Einwohnermeldeämter geplant. Sie sollen zum einen die personenidentifizierenden Daten und das Todesdatum der Patienten liefern, zum anderen Daten über Umzüge aus Niedersachsen heraus, damit auch diese Daten mit dem Registerbestand verglichen werden können. Die Meldungen der Gesundheitsämter (Todesbescheinigungen) erfolgen selbstverständlich ohne Einwilligung des Patienten, es ist allerdings auch keine Zustimmung z.B. von Angehörigen erforderlich [RW98].

Anfragemeldeweg Dieser Meldeweg dient den Anfragen der Patienten, die über den behandelnden Arzt oder Zahnarzt angestoßen werden können. Es können laut § 4 KRG Anträge auf Auskunft oder Löschung der Daten bei der Vertrauensstelle gestellt werden [ABH99]. Im Entitäten-Relationen-Schema der Vertrauensstelle ist er abgebildet wie Abbildung D.5 zeigt [RW99]:

Die Entität Anfragemeldungen erfasst im wesentlichen folgende Daten: Geschlecht, Geburtsjahr, Geburtsmonat und Wohnort.

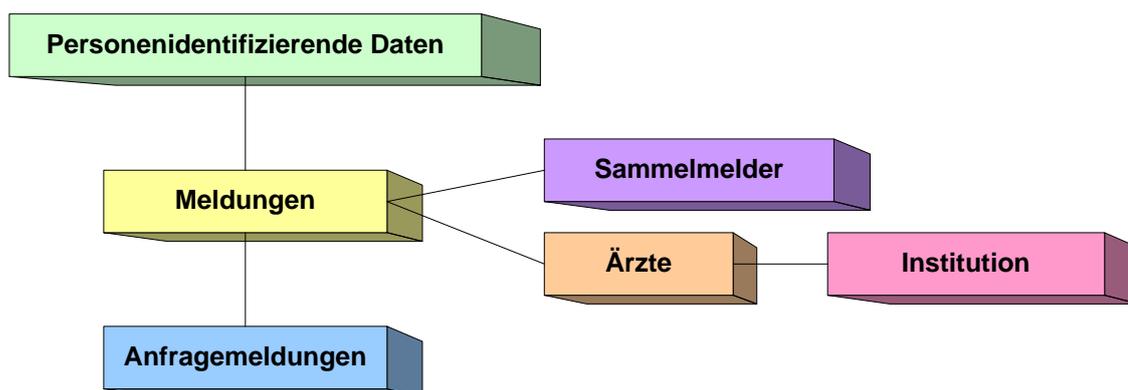


Abbildung D.5: Datenschema Anfragemeldeweg an das EKN [RW99]

D.6.2 Technischer Stand der Melder

Allgemein Die Melder kommen aus unterschiedlichen Institutionen mit zum Teil völlig differierenden Aufgaben. Demgemäß ist auch die Softwareunterstützung der Anwender sehr verschieden. Auch innerhalb einer Meldergruppe können stark abweichende Hard- und Softwareausstattungen vorliegen. Das hängt immer davon ab, welche Bedeutung einer EDV-Unterstützung beigemessen wird und wie sie von den Verantwortlichen angenommen und umgesetzt werden kann.

Ein niedergelassener Arzt zum Beispiel könnte ganz ohne EDV auskommen. Die meisten Behandler verwenden zumindest ein computergestütztes Abrechnungssystem. Andere haben ein modernes Praxisnetzwerk mit Terminals in jedem Behandlungsraum und am Empfang, einschließlich integrierten Erfassungs- und Auswertungsprogrammen für ihre Medizingeräte (z.B. Elektrokardiograph), sowie andere unterstützende Informationsprogramme rund um die Praxis.

Abbildung D.6 zeigt eine mitunter an Statistiken der Kassenärztlichen Bundesvereinigung [KBV03a] orientierte Aufstellung der verwendeten Software der meldenden Institutionen in Niedersachsen (Stand 31.12.2002).

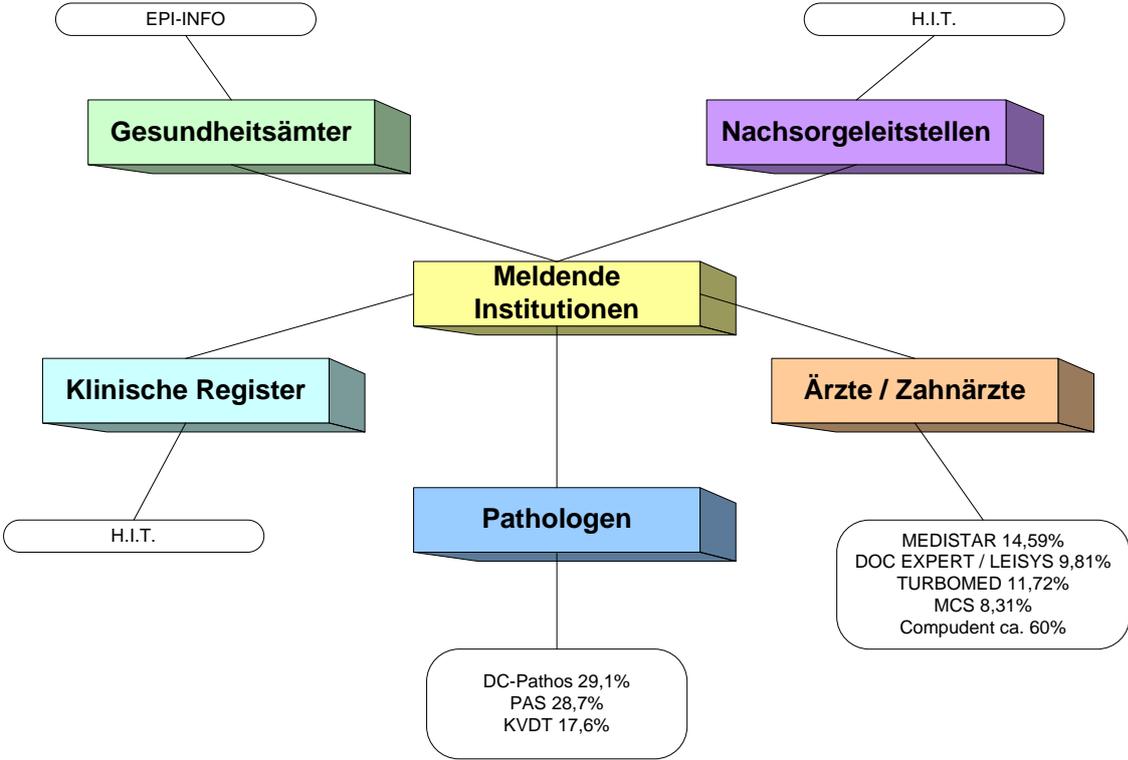


Abbildung D.6: Wesentlichste Softwareausstattung der meldenden Institutionen in Niedersachsen

E Inhaltsverzeichnis der CD

Auf der beigefügten Dissertations-CD befinden sich Dateien aus der prototypischen Implementierung bzw. Dateien, die im Rahmen der Implementierung von Interesse sind. Die Dateien befinden sich alle in einem übergeordneten Ordner mit dem Namen `Evaluation`. Darin verzweigt sich dann die Inhaltsstruktur wie folgt:

1. Basisdokumentation für Tumorkranke: UML-Diagramme für die Abschnitte D1 bis E21
2. Domänenmodelle: zur Verfügung gestellte BDT DTD, die erstellten Basisdokumentation für Tumorkranke DTD und HL7 DTD
 - Basisdokumentation für Tumorkranke DTD
 - BDT DTD
 - HL7 DTD
3. OnkoDok: zur Verfügung gestellten XML-Schemata von OnkoDok ([Wol02])
4. Semantic Web Spezifikationen: erstellte Spezifikationen in RDF, RDFS und OWL
 - Klassendefinitionen
 - OWL-Klassenäquivalenzen
 - RDF-Statements
 - Spezifikationen in Notation 3
5. Teil der eingesetzten Software: verwendete Software von Altova mit der Möglichkeit eines 30 tägigen Tests
 - Altova XMLSPY 2004 Enterprise Edition
 - MapForceEnt2005_DE
6. XSLT: mit MapForce erstellte XSL-Transformationen
 - Basisdokumentation-OnkoDok
 - OnkoDok-Basisdokumentation
 - BDT-HL7
 - HL7-BDT

Abbildungsverzeichnis

1.1	Architekturebenen von Organisationseinheiten nach [Has00a]	3
2.1	Grundsätzliches Kommunikationsprinzip	8
2.2	Grundzüge Interoperabilität	9
2.3	Applikationsinteroperabilität nach [Pol01]	11
3.1	Grundzüge eines föderierten Informationssystems von [Wil02]	16
3.2	Die drei Schichten einer Mediator-Architektur nach [Wie92]	19
3.3	Eine mediatorbasierte Architektur nach [Has02]	21
4.1	Notation für Potenz und Ebene für ein Modellierungselement	32
4.2	Das Modell für mehrfache Metaebenen (MoMM) nach [AK01]	33
5.1	Ein Feature-Diagramm nach [CH03], das auf oberster Ebene die Hauptvariationensgebiete darstellt	36
6.1	Ausschnitt BDT-Struktur als UML-Diagramm	50
6.2	Ausschnitt HL7 Struktur als UML-Diagramm	54
6.3	Ausschnitt DICOM-Struktur als UML-Diagramm	55
6.4	Struktur von DICOM SR	56
6.5	Ausschnitt EHCR-Struktur des CEN als UML-Diagramm	58
6.6	Ausschnitt CDA-Struktur als UML-Diagramm	60
6.7	Trennung von Information und Wissen im openEHR Modell	62
6.8	Ausschnitt medizinische Begriffssysteme als UML-Diagramm	64
6.9	Ausschnitt Begriffsbeziehungen und Codierprinzipien als UML-Diagramm [Ber95, PH02]	66
6.10	Zusammenhang der Standards im Gesundheitswesen	68
7.1	Allgemeine Mediator-basierte Architektur	77
7.2	Mediator-basierte Architektur für das Gesundheitswesen	79
7.3	Metamodell der vorgestellten Mediatorbasierten-Architektur (Abbildung 7.2 stellt eine Instanz dieses Modells dar)	81
7.4	Schemaarchitektur	82
7.5	Zusammenhang der Metadaten als UML-Diagramm	84
7.6	Tiefe Instanziierung, bezogen auf die Begriffssysteme	85
7.7	Ausschnitt aus dem Metamodell der Kommunikationsstandards in der Medizin als UML-Diagramm	86

7.8	Die drei Ebenen der unterteilenden Abschnitte als UML-Diagramm	87
7.9	Details zu drei Klassen des Begriffssystems als UML-Diagramm	88
7.10	An die Meta-Object Facility Architektur angelehnte Architektur	89
7.11	Ausschnitt aus dem Metamodell der Dokumentenstandards in der Medizin als UML-Diagramm	90
7.12	Ausschnitt medizinische Begriffssysteme als UML-Diagramm	92
7.13	Ausschnitt Begriffsbeziehungen und Codierprinzipien als UML-Diagramm	94
7.14	Strukturanalyse top-down vom Begriffssystem für Kommunikation zur generischen Struktur von HL7 und BDT	96
7.15	Transformationen bezogen auf die Ebenen, angepasst auf die analysierte Domäne Medizin	97
7.16	Abbildung von Klassen aus der generischen Struktur mit einer UML-basierten Abbildungssprache	97
8.1	Evaluation der Architektur am Beispiel des EKN: Spezialisierung der allgemeinen Architektur in Abbildung 7.2 für die Anforderungen des EKN mit den Standards HL7, BDT und ICD-O.	107
8.2	Schema-Architektur: Eine Sicht der statischen Abhängigkeiten zwischen den Modellen	108
8.3	UML-Objektdiagramm als Instanz des UML-Klassendiagramms aus Abbildung 7.3 auf Seite 81 für das Fallbeispiel EKN	110
8.4	Auszug aus HL7-DTD für Tumormeldungen	111
8.5	Top-Down vom Begriffssystem zum konkreten Einsatz des Standards HL7	112
8.6	Top-Down vom Begriffssystem zum konkreten Einsatz des Standards BDT	113
8.7	Klassendefinitionen für BDT in N3	114
8.8	Klassendefinition von BDT in RDF/XML	115
8.9	RDF-Statements für BDT in N3	116
8.10	RDF-Statements für BDT in RDF/XML	116
8.11	Transformationen bezogen auf die Ebenen für das Anwendungsbeispiel EKN	117
8.12	Modellkorrespondenzen der UML-Diagramme von HL7 und BDT	118
8.13	Prinzipielle Klassenäquivalenzen von HL7- und BDT-Klassen in OWL	119
9.1	Komponenten-Editor-Sicht des e*Gate Schema Designers	123
9.2	Übergebene XSL-Datei als Collaboration Rules in e*Gate	124
9.3	UML-Aktivitätsdiagramm zur Übergabe einer xsl-Datei an den e*Gate Collaboration Service	125
9.4	Abbildungen zwischen HL7-Segmenten und BDT-Sätzen nach dem vorgestellten Leitfaden	127
9.5	Abbildungen zwischen HL7-Felder und BDT-Datenfelder, beispielhaft für HL7-Segment ZST und BDT-Satz Tumorbeurteilung	128
9.6	Komplette Zuordnung zwischen den BDT-Datenfeldern und den HL7-Feldern für das Beispiel Basistumordokumentation	129
9.7	Importierte DTDs von BDT und HL7 in MapForce	130
9.8	Abbildungsdefinition mittels Verbindungslinien zwischen BDT und HL7 in MapForce	130

9.9	Generierte XSL-Transformationen von MapForce zwischen BDT und HL7	131
9.10	Transformation einer BDT-Beispielnachricht in HL7 in der MapForce-Vorschau	131
9.11	Erste grobe Abbildungen zwischen Strukturierungselementen der Basisdokumentation und OnkoDok nach dem vorgestellten Leitfaden	133
9.12	Abbildungen zwischen Strukturierungselementen von Basisdokumentation und OnkoDok, beispielhaft für Abschnitt D und OnkoDok-Header	134
9.13	Komplette Zuordnung zwischen den Strukturierungselementen der Basistumordokumentation und OnkoDok	135
9.14	Abbildungsdefinition mittels Verbindungslinien zwischen Basisdokumentation D und OnkoDok-Header in MapForce	136
10.1	Eine tabellarische Übersicht der verwandten Ansätze aus der Informatik einschließlich des eigenen Ansatzes	144
11.1	Eine tabellarische Übersicht der verwandten Ansätze aus der Medizininformatik einschließlich des eigenen Ansatzes	150
13.1	Eine Matching-Taxonomie aus [vdHR05]	158
A.1	UML-Diagramm einer Klassifikation für Informationssysteme	164
A.2	Schichten-Architektur eines Föderierten Informationssystems aus [Wil02]	165
A.3	UML-Diagramm von Klassifikationskriterien für Föderierte Informationssysteme	167
A.4	Schemata in eng integrierten FIS	169
A.5	Lose integrierte föderierte Informationssysteme	170
A.6	Arten von föderierten Informationssystemen	176
A.7	Klassifikation föderierter Informationssysteme	177
A.8	Fünf-Schichten-Schema-Architektur von föderierten Datenbanksystemen (FDBS) nach [SL90]	178
A.9	Schichten-Architektur von mediatorbasierten Informationssystemen (MBIS) nach [Wil02]	181
B.1	Ein RDF-Tripel als Graph	183
C.1	Semiotische Triade	191
C.2	Ausschnitt medizinische Begriffssysteme als UML-Diagramm	201
C.3	Ausschnitt Begriffsbeziehungen und Codierprinzipien als UML-Diagramm	202
D.1	Meldende Institutionen an das EKN [RW98]	212
D.2	Datenschema Standardmeldeweg an das EKN [RW99]	213
D.3	Datenschema Pathologenmeldeweg an das EKN [RW99]	216
D.4	Datenschema Sterbemeldeweg an das EKN [RW99]	218
D.5	Datenschema Anfragemeldeweg an das EKN [RW99]	219
D.6	Wesentlichste Softwareausstattung der meldenden Institutionen in Niersachsen	220

Tabellenverzeichnis

6.1	Auswahlschema für geeignete Kombinationen von Standards	68
8.1	Dimensionen zur Einordnung von Standards im Gesundheitswesen nach [LS01a]	103
8.2	Beispiele von Standards im Gesundheitswesen nach [LS01a]	103
8.3	Das EKN im Dimensionenschema nach [LS01a]	104
8.4	Auswahlschema zur Auswahl geeigneter Kombinationen von Kommunikations- und Dokumentationsstandards für den Anwendungsbereich EKN	105

Literaturverzeichnis

- [Abe95] ABEL, U.: *Chemotherapie fortgeschrittener Karzinome: eine kritische Bestandsaufnahme*. Hippokrates Verlag, Stuttgart, 2. Auflage, 1995. (ZITIERT AUF SEITE 205.)
- [ABH99] APPELRATH, H.-J., M. BEYER und H. HINRICHS: *Tätigkeitsbericht für den Zeitraum 1.1.-31.12.1998*. Registerstelle des Epidemiologischen Krebsregisters Niedersachsen, 1999. (ZITIERT AUF DEN SEITEN 208, 216 UND 218.)
- [ABK97] *Krebs in Deutschland - Häufigkeiten und Trends*, Saarbrücken Dezember 1997. (ZITIERT AUF SEITE 205.)
- [AEZ02] *Ärztliche Zentralstelle Qualitätssicherung: Leitlinien: Einteilung, Entwicklung, Implementierung, Evaluation*. URL: <http://www.aeqq.de/publikationen/0index/pdfpraesentationen/follrehafreiburg041202go>, 2002. Retrieved: 06.08.2003. (ZITIERT AUF SEITE 73.)
- [AK01] ATKINSON, C. und T. KÜHNE: *The Essence of Multilevel Metamodeling*. In: GOGOLLA, M. und C. KOBRYN (Herausgeber): *Proc. UML 2001 – The Unified Modeling Language. Modeling Languages, Concepts and Tools*, Seiten 19–33. Springer, 2001. (ZITIERT AUF DEN SEITEN 3, 30, 31, 32, 33, 34 UND 223.)
- [AMST96] APPELRATH, H.-J., J. MICHAELIS, I. SCHMIDTMANN und W. THOBEN: *Empfehlung an die Bundesländer zur technischen Umsetzung der Verfahrensweisen gemäss Gesetz über Krebsregister (KRG)*. Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 27(2):101–110, Februar 1996. (ZITIERT AUF DEN SEITEN 206, 214 UND 215.)
- [ANS02] *American National Standards Institute (ANSI) Online*. URL: <http://www.ansi.org>, 2002. Retrieved: 07.03.2002. (ZITIERT AUF SEITE 71.)
- [Bea02] BEALE, T.: *Archetypes: Constraint-based Domain Models for Future-proof Information Systems*. In: *OOPSLA 2002 Workshop on behavioural semantics*, 2002. (ZITIERT AUF SEITE 61.)

- [Ber95] BERNAUER, J.: *Zur semantischen Rekonstruktion medizinischer Begriffssysteme*. In: *Habilitationsschrift*. Universität Hildesheim, Fachgebiet Medizinische Informatik, 1995. (ZITIERT AUF DEN SEITEN 63, 65, 66, 67, 93, 95, 194, 204 UND 223.)
- [Ber00] BERNAUER, J.: *Vorlesungsskript Studiengang Medizinische Dokumentation und Informatik: Dokumentations- und Ordnungslehre*. Fachhochschule Ulm, 2000. (ZITIERT AUF DEN SEITEN 190, 191, 192, 193, 194, 199 UND 200.)
- [bIT05a] *Chronologie der elektronischen Gesundheitskarte (Stand 1. März 2005)*. URL: <http://www.dimdi.de/static/de/ehealth/karte/download/chronologie.pdf>, 2005. Retrieved: 27.4.2005. (ZITIERT AUF SEITE 48.)
- [bIT05b] *Technik der elektronischen Gesundheitskarte*. URL: <http://www.dimdi.de/static/de/ehealth/karte/technik.index.htm>, 2005. Retrieved: 27.4.2005. (ZITIERT AUF SEITE 48.)
- [BKLW99] BUSSE, SUSANNE, RALF-DETLEF KUTSCHE, ULF LESER und HERBERT WEBER: *Federated Information Systems: Concepts, Terminology and Architectures*. Technischer Bericht Forschungsberichte des Fachbereichs Informatik 99-9, Technische Universität Berlin, 1999. (ZITIERT AUF DEN SEITEN 15, 17, 18, 83, 161, 162, 166, 168 UND 176.)
- [Blo02] BLOBEL, B.: *Was bringt uns der HL7 Kommunikationsstandard Version 3?* Forum der Medizin Dokumentation und der Medizin Informatik, 1:23–25, 2002. (ZITIERT AUF SEITE 51.)
- [BM00] BEHME, H. und S. MINTERT: *XML in der Praxis*. Addison-Wesley-Verlag, München, 2. erweiterte Auflage, 2000. (ZITIERT AUF DEN SEITEN 27 UND 29.)
- [BMR⁺01] BUSCHMANN, F., R. MEUNIER, H. ROHNERT, P. SOMMERLAD und M. STAL: *Pattern-Oriented System Architecture — A System of Patterns*, Band 1 der Reihe *Wiley-Series in Software Design Patterns*. Wiley-Verlag, 2001. (ZITIERT AUF SEITE 12.)
- [BP03] BLOBEL, B. und P. PHAROW: *Zukunftssichere Lösung für virtuelle, interoperable und portable EHR-Systeme*. Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 34/3:576–577, 2003. (ZITIERT AUF SEITE 61.)
- [Bun02] *Förderprogramm Informations- und Kommunikationstechnik; Glossar der verwendeten Begriffe*. URL: <http://www.it2006.de/kapitel8.html>, 2002. Retrieved: 17.04.2002. (ZITIERT AUF SEITE 9.)
- [Bus02] BUSSE, S.: *Modellkorrespondenzen für die kontinuierliche Entwicklung mediatorbasierter Informationssysteme*. Doktorarbeit, Technische Universität Berlin, Fakultät Elektrotechnik und Informatik, 2002. (ZITIERT AUF SEITE 20.)

-
- [CCO01] *Overview of HL7's CCOW Standard.* URL: http://www.hl7.org/special/Committees/ccow_sigvi.htm, 2001. Retrieved: 28.4.2005. (ZITIERT AUF SEITE 72.)
- [CDI03a] *Introduction to the CDISC Operational Data Model Version 1.1 (Final).* URL: <http://www.cdisc.org>, 2003. Retrieved: 03.02.2003. (ZITIERT AUF SEITE 70.)
- [CDI03b] *Overview of Techniques for Reading and Writing ODM Data.* URL: <http://www.cdisc.org>, 2003. Retrieved: 03.02.2003. (ZITIERT AUF SEITE 70.)
- [CDI03c] *Overview of the CDISC Operational Data Model for Clinical Data Acquisition and Archive.* URL: <http://www.cdisc.org>, 2003. Retrieved: 03.02.2003. (ZITIERT AUF SEITE 70.)
- [CEN02] *CEN/TC251.* URL: <http://www.centc251.org>, 2002. Retrieved: 06.06.2002. (ZITIERT AUF DEN SEITEN 57, 58 UND 59.)
- [CEN03] *CEN/TC 251: Draft initial report version 0.4 of Task Force EHRCOM - Revision of ENV 13606: Electronic Health Record Communication, Stand 25.6.2002.* URL: <http://www.centc251.org/tcmeet/doclist/TCdoc02/N02-032rev.pdf>, 2003. Retrieved: 06.10.2003. (ZITIERT AUF SEITE 59.)
- [CH03] CZARNECKI, K. und SIMON HELSEN: *Classification of Model Transformation Approaches.* In: *OOPSLA 2003 Workshop*, Anaheim, California, October 2003. (ZITIERT AUF DEN SEITEN 35, 36, 41 UND 223.)
- [CHKT05] CONRAD, S., W. HASSELBRING, A. KOSCHEL und R. TRITSCH: *Enterprise Application Integration.* Spektrum-Verlag, Heidelberg, im Druck Auflage, 2005. (ZITIERT AUF DEN SEITEN 20 UND 121.)
- [CNM02] COLLINS, S., S. NAVATHE und L. MARK: *XML schema mappings for heterogeneous database access.* Information and Software Technology, 44:251–257, 2002. (ZITIERT AUF SEITE 109.)
- [Con97] CONRAD, S.: *Föderierte Datenbanksysteme: Konzepte der Datenintegration.* Springer-Verlag, Berlin, 1997. (ZITIERT AUF SEITE 18.)
- [Con02] CONRAD, S.: *Schemaintegration — Integrationskonflikte, Lösungsansätze, aktuelle Herausforderungen.* Informatik, Forschung und Entwicklung, Band 17(Heft 3):101–111, September 2002. (ZITIERT AUF SEITE 141.)
- [Cor97] *Patient Identification Service (PIDS).* URL: <http://www.omg.org/docs/corbamed/97-06-04.pdf>, 1997. Retrieved: 4.10.2003. (ZITIERT AUF SEITE 71.)
- [Cor03] *CORBA BASICS.* URL: <http://www.omg.org/gettingstarted/corbafaq.htm>, 2003. Retrieved: 4.10.2003. (ZITIERT AUF SEITE 71.)
-

- [CS04] CHARI, K. und S. SESHADRI: *Demystifying Integration*. Communications of the ACM, 47(7):59–63, Juli 2004. (ZITIERT AUF SEITE 71.)
- [D2D02] *D2D Telematik-Initiative der KV Nordrhein*. URL: <http://www.kvno.de/texte/aktuell/meldungn/d2dflyer.htm>, 2002. Retrieved: 09.04.2002. (ZITIERT AUF SEITE 73.)
- [DIC02] *Digital Imaging and Communications in Medicine*. URL: <http://medical.nema.org/dicom.html>, 2002. Retrieved: 14.05.2002. (ZITIERT AUF SEITE 53.)
- [DJK04] DOSTAL, W., M. JECKLE und W. KRIECHBAUM: *Semantik und Web Services: Beschreibung von Semantik*. JavaSPEKTRUM, 4:45–49, 2004. (ZITIERT AUF DEN SEITEN 7 UND 9.)
- [DLK⁺05] DOGAC, A., G. LALECI, S. KIRBAS, Y. KABAK, S. SINIR und A. YILDIZ: *Artemis: Deploying Semantically Enriched Web Services in the Healthcare Domain*. akzeptiert für: Information System Journal, April 2005. (ZITIERT AUF SEITE 142.)
- [Dub04] *Dublin Core Metadata Initiative*. URL: <http://www.dublincore.org/>, 2004. Retrieved: 27.10.2004. (ZITIERT AUF DEN SEITEN 119 UND 185.)
- [Dud03] DUDECK, J.: *10 Jahre HL7-Benutzergruppe in Deutschland*. HL7-Mitteilungen, 16:5–11, 2003. (ZITIERT AUF SEITE 51.)
- [DZM03] *Das DIMDI richtet das Deutsche Zentrum für medizinische Klassifikation (DZMK) ein*. URL: <http://www.dimdi.de/dynamic/de/dimdi/presse/pm/newsarticle.html?newsId=7111&channelID=74>, 2003. Retrieved: 03.10.2003. (ZITIERT AUF SEITE 63.)
- [eBT03] *Electronic Business Transition Working Group (eBTWG)*. URL: <http://www.ebtwg.org>, 2003. Retrieved: 20.08.2003. (ZITIERT AUF SEITE 29.)
- [ebX03] *UN/CEFACT and OASIS Meeting Showcases ebXML for Healthcare and B2B*. URL: http://www.ebxml.org/news/pr_20010509.htm, 2003. Retrieved: 20.08.2003. (ZITIERT AUF SEITE 29.)
- [Eic01] EICHELBERG, M.: *Ein Verfahren zur Bewertung der Interoperabilität medizinischer Bildkommunikationssysteme*. Doktorarbeit, Carl von Ossietzky Universität Oldenburg, Fachbereich Informatik, 2001. (ZITIERT AUF DEN SEITEN 53, 146 UND 155.)
- [eMa05] *eMammaAkte: Pilotregion Essen - Stand der Arbeiten*. URL: <http://www.inf.fh-dortmund.de/concute/personen/mitarbeiter/gerullat/cont.htm>, 2005. Retrieved: 25.3.2005. (ZITIERT AUF SEITE 149.)

- [ETH02] *Eidgenössische Technische Hochschule Zürich, Institute of Information Systems, Database Research Group, Realizing the Hyperdatabase Vision.* URL: <http://www-dbs.inf.ethz.ch/vision/index.html>, 2002. Retrieved: 03.05.2002. (ZITIERT AUF SEITE 17.)
- [Fri01] FRIEBE, J.: *Architekturen für komponentenbasierte Geographische Informationssysteme im Internet.* Doktorarbeit, Carl von Ossietzky Universität Oldenburg, Fachbereich Informatik, 2001. (ZITIERT AUF SEITE 20.)
- [GEH02] *Good Electronic Healthcare Record.* URL: <http://www.gehr.org>, 2002. Retrieved: 06.06.2002. (ZITIERT AUF DEN SEITEN 57 UND 61.)
- [GH03] GÖRKE, H. J. und K.U. HEITMANN: *ArGe SCIPHOX GbR mbH: Die Arbeitsgemeinschaft SCIPHOX.* URL: <http://www.hl7.de/veranstaltungen/kongress/sciphox2003/prents/arge-sciphox.pdf>, Juli 2003. Retrieved: 16.08.2003. (ZITIERT AUF SEITE 105.)
- [Gra05] GRAWUNDER, M.: *DYNAQUEST - Dynamische und adaptive Anfragebearbeitung in virtuellen Datenbanksystemen.* Dissertation, Universität Oldenburg, Department Informatik, 2005. (ZITIERT AUF SEITE 28.)
- [Hae99] HAEBERLIN, V.: *Schnittstellenkonzepte in Tumordokumentationssystemen.* Doktorarbeit, Justus-Liebig-Universität Gießen, Fachbereich Medizinische Informatik, 1999. (ZITIERT AUF DEN SEITEN 49, 51, 109, 111, 147 UND 153.)
- [Has00a] HASSELBRING, W.: *Information System Integration.* Communications of the ACM, 43:33–38, 2000. (ZITIERT AUF DEN SEITEN 2, 3 UND 223.)
- [Has00b] HASSELBRING, W.: *The Role of Standards for Interoperating Information Systems.* In: JAKOBS, K. (Herausgeber): *Information Technology Standards and Standardization: A Global Perspective*, Seiten 116–130. Idea Group Publishing, Hershey, PA, 2000. (ZITIERT AUF DEN SEITEN 48, 80 UND 155.)
- [Has02] HASSELBRING, W.: *Web Data Integration for E-Commerce Applications.* IEEE MultiMedia, 9(1):16–25, 2002. (ZITIERT AUF DEN SEITEN 20, 21, 28, 48, 76, 80 UND 223.)
- [HBD99] HEITMANN, K., B. BLOBEL und J. DUDECK: *HL7 Kommunikationsstandard in der Medizin: Kurzeinführung und Information.* Verlag Alexander Mönch, 1. Auflage, 1999. (ZITIERT AUF DEN SEITEN 51 UND 53.)
- [HCP00] *Health Care Professionals Protocol vom Projektbüro Telemedizin der Kassenärztlichen Vereinigung Bayerns und der Bayerischen Landesärztekammer.* URL: <http://www.hcp-protokoll.de/index.htm>, 2000. Retrieved: 26.7.2000. (ZITIERT AUF SEITE 72.)

- [Hei03] HEITMANN, K.U.: *Clinical Document Architecture und SCIPHOX: CDA in Deutschland*. URL: <http://www.hl7.de/veranstaltungen/kongress/sciphox2003/prents/cdasciphox.pdf>, Juli 2003. Retrieved: 16.08.2003. (ZITIERT AUF SEITE 105.)
- [hil01] R. Hildebrand: *Leitartikel Klinische Pfade*. URL: <http://www.hmanage.de/pdf/T100-Klinische-Pfade-uebersicht-01.pdf>, 2001. Retrieved: 06.08.2003. (ZITIERT AUF SEITE 73.)
- [HL701a] HL7 JAHRESTAGUNG IN GÖTTINGEN: *Tutorial Clinical Document Architecture*, 2001. (ZITIERT AUF SEITE 59.)
- [HL701b] HL7 JAHRESTAGUNG IN GÖTTINGEN: *Tutorial Einführung in die Electronic Healthcare Record Architecture*, 2001. (ZITIERT AUF DEN SEITEN 57, 58 UND 61.)
- [HL702a] *Health Level Seven*. URL: <http://www.hl7.org>, 2002. Retrieved: 07.03.2002. (ZITIERT AUF DEN SEITEN 51, 57 UND 59.)
- [HL702b] *Health Level Seven Vocabulary*. URL: <http://www.hl7.org/special/Committees/Vocab/vocab.htm>, 2002. Retrieved: 28.04.2002. (ZITIERT AUF SEITE 52.)
- [HSD01] HEITMANN, K.U., R. SCHWEIGER und J. DUDECK: *Das SCIPHOX-Projekt: Kommunikation zwischen Arztpraxis und Krankenhaus - ein Schritt weiter..* Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 32(2-3):169–170, 2001. (ZITIERT AUF SEITE 105.)
- [HW01] HASSELBRING, W. und H. WEIGAND: *Languages for Electronic Business Communication: State of the Art*. Industrial Management & Data Systems, 101(5):217–227, 2001. (ZITIERT AUF SEITE 29.)
- [ICD02a] *International Statistical Classification of Diseases and Related Health Problems (ICD)*. URL: <http://www.dimdi.de/de/klassi/index.htm>, 2002. Retrieved: 31.08.2002. (ZITIERT AUF DEN SEITEN 63, 65 UND 93.)
- [ICD02b] *International Statistical Classification of Diseases and Related Health Problems - Oncology (ICD-O)*. URL: <http://www.dimdi.de/de/klassi/index.htm>, 2002. Retrieved: 31.08.2002. (ZITIERT AUF SEITE 104.)
- [IEE90] *Institute of Electrical and Electronics Engineers (IEEE)*. Standard Computer Dictionary - A Compilation of IEEE Standard Computer Glossaries. URL: <http://standards.ieee.org/catalog/dict.html>, 1990. 1990. (ZITIERT AUF SEITE 7.)
- [IHE03] *Integrating the Healthcare Enterprise*. URL: <http://www.ihe-d.org>, 2003. Retrieved: 20.02.2003. (ZITIERT AUF SEITE 71.)

-
- [IRS01] INGENERF, J., J. REINER und B. SEIK: *Standardized terminological services enabling semantic interoperability between distributed and heterogeneous systems*. International Journal of Medical Informatics, 64:223–240, 2001. (ZITIERT AUF DEN SEITEN 7, 47 UND 95.)
- [Jun03] JUNG, B.: *Semantic Health Records*. Doktorarbeit, University of Dublin, Trinity College, Department of Computer Science, 2003. (ZITIERT AUF SEITE 143.)
- [KBV03a] *Installationsstatistik*. URL: <http://www.kbv.de/it/2815.htmund.../2820.htm>, 2003. Retrieved: 30.09.2003. (ZITIERT AUF SEITE 219.)
- [KBV03b] *Updates: Datensatzbeschreibungen Bewegungsdaten*. URL: <http://kbv.de/it/2410.htm>, 2003. Retrieved: 16.08.2003. (ZITIERT AUF DEN SEITEN 51 UND 155.)
- [KG01] KUHN, K. A. und D. A. GIUSE: *From Hospital Information Systems to Health Information Systems - Problems, Challenges, Perspectives*. Yearbook of Medical Informatics, Seiten 63–76, 2001. (ZITIERT AUF SEITE 71.)
- [Kom05] *Wikipedia - Die freie Enzyklopädie*. URL: <http://de.wikipedia.org/wiki/Hauptseite>, 2005. Retrieved: 27.3.2005. (ZITIERT AUF SEITE 138.)
- [Koo05] *Deutlich positiver Impuls erwartet: Fraunhofer IBMT und VDAP vereinbaren Kooperation*. URL: <http://www.vdap.de/html/images/pm-FhG-Kooperation-03-11-17.pdf>, 2005. Retrieved: 24.4.2005. (ZITIERT AUF SEITE 73.)
- [Kor01] KORZENIOWSKI, PAUL: *Standards Emerge to Ease Metadata Integration*. eAI Journal, Seiten 52–53, April 2001. (ZITIERT AUF SEITE 183.)
- [Küh01] KÜHN, K.: *Wie viel kosten das Krankenhaus die DRG's? - Simulation sichert Budgetplanungen*. Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 32(2–3):206–207, 2001. (ZITIERT AUF SEITE 74.)
- [LBM⁺05] LENZ, R., M. BEYER, C. MEILER, S. JABLONSKI und K.A. KUHN: *Informationsintegration in Gesundheitsversorgungsnetzen — Herausforderungen an die Informatik*. Informatik Spektrum, Band 28(Heft 2):105–119, April 2005. (ZITIERT AUF DEN SEITEN 47, 59 UND 73.)
- [Len05] LENZ, R.: *Evolutionäre Informationssysteme*. In: *Habilitationsschrift*. Universität Marburg, Fachgebiet Mathematik/Informatik, 2005. (ZITIERT AUF DEN SEITEN 12 UND 34.)
- [Lex02] *Reformgesetz zum Risikostrukturausgleich ab 01.01.2002*. URL: <http://www.lexsoft.de/aktuelles/2742>, 2002. Retrieved: 06.08.2003. (ZITIERT AUF SEITE 73.)
-

- [LGHKG99] LEINER, F., W. GAUS, R. HAUX und P. KNAUP-GREGORI: *Medizinische Dokumentation: Lehrbuch und Leitfaden für die Praxis*. Schattauer Verlagsgesellschaft mbG, Stuttgart, 3. neubearbeitete und erweiterte Auflage Auflage, 1999. (ZITIERT AUF DEN SEITEN 57, 189 UND 203.)
- [LK01] LENZ, R. und K.A. KUHN: *Intranet meets Hospital Information Systems: The Solution to the Integration Problem?* *Methods of Information in Medicine*, 40:99–105, 2001. (ZITIERT AUF SEITE 67.)
- [LK03] LENZ, R. und K. KUHN: *Zur Architektur und Evolution von Krankenhausinformationssystemen*. In: *Tagungsband der Informatik 2003 - Jahrestagung der Gesellschaft für Informatik*, Frankfurt/M., 2003. Gesellschaft für Informatik. (ZITIERT AUF SEITE 71.)
- [LPH99] LANGE, M., H.-U. PROKOSCH und W. HASSELBRING: *Eine Taxonomie für Kommunikationsserver im Krankenhaus*. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie*, 30(1):21–43, 1999. (ZITIERT AUF SEITE 121.)
- [LS01a] LAGENDIJK, P. und R. STEGWEE: *Healthcare Information and Communication Standards Framework*. In: STEGWEE, R. und T. SPILL (Herausgeber): *Strategies for Healthcare Information Systems*, Seiten 66–77. Idea Group Publishing, Hershey, PA, 2001. (ZITIERT AUF DEN SEITEN 102, 103, 104 UND 227.)
- [LS01b] LAUTERBACH, K. W. und S. STOCK: *Reform des Risikostrukturausgleichs: Disease Management wird aktiviert*. *Deutsches Ärzteblatt*, 98(30):A1935–A1937, Juli 2001. (ZITIERT AUF SEITE 73.)
- [Lut00] LUTZ, JEFFREY C.: *EAI Architecture Patterns*. *eAI Journal*, Seiten 64–73, March 2000. (ZITIERT AUF DEN SEITEN 12 UND 13.)
- [Map05] *Altova MapForce 2005 Enterprise Edition*. URL: http://www.altova.com/download_mapforce.html, 2005. Retrieved: 9.2.2005. (ZITIERT AUF SEITE 123.)
- [MED98] MEDNET: *Zwölf Kriterien für Systeme zum Austausch von patientenbezogenen Daten zwischen Einrichtungen der gesundheitlichen Versorgung*. www.mednet.de/daten/thesen.htm, November 1998. (ZITIERT AUF SEITE 47.)
- [MeS02] *Medical Subject Headings (MeSH)*. URL: <http://www.dimdi.de/de/klassi/index.htm>, 2002. Retrieved: 31.08.2002. (ZITIERT AUF DEN SEITEN 65 UND 93.)
- [MIR93] MILLER, R.J., Y. IOANNIDIS und R. RAMAKRISHNAN: *The Use of Information Capacity in Schema Integration and Translation*. In: *Proceedings of the Nineteenth International Conference on Very Large Data Bases (VLDB)*, Seiten 120–133, Dublin, Ireland, August 1993. Retrieved: 16.3.2005. (ZITIERT AUF SEITE 42.)

-
- [MIR04] *MIRApplus - Medizinisches Informations-, Retrieval- und Archivsystem.* URL: <http://www.imbi.uni-freiburg.de/medinf/mirapplus>, 2004. Retrieved: 21.4.2004. (ZITIERT AUF SEITE 146.)
- [MM03] MAEDCHE, A. und B. MOTIK: *Repräsentations- und Anfragesprachen für Ontologien - eine Übersicht.* Datenbank-Spektrum, 3(6):43–53, Juni 2003. (ZITIERT AUF DEN SEITEN 85, 142 UND 187.)
- [MOF03] *Meta-Object Facility (MOF) Spezifikation.* URL: <http://www.omg.org/technology/documents/formal/mof.htm>, 2003. Retrieved: 4.12.2003. (ZITIERT AUF DEN SEITEN 3 UND 89.)
- [Moh03] MOHR, G.: *Die elektronische MammaAkte.* Deutsches Ärzteblatt/Praxis Computer, 4:4–6, 2003. (ZITIERT AUF SEITE 149.)
- [MRI02] *Medical Record Institute.* URL: <http://www.medrecinst.com>, 2002. Retrieved: 06.06.2002. (ZITIERT AUF SEITE 58.)
- [MSS01] MÄDCHE, A., S. STAAB und R. STUDER: *Ontologien.* Wirtschaftsinformatik, 43:393–395, 2001. (ZITIERT AUF SEITE 67.)
- [Mut96] MUTSCHLER, E.: *Arzneimittelwirkungen: Lehrbuch der Pharmakologie und Toxikologie.* Wissenschaftliche Verlagsgesellschaft, Stuttgart, 7. Auflage, 1996. (ZITIERT AUF SEITE 205.)
- [Nat05] *IT-Unternehmen fördern Datenvereinheitlichung im US-Gesundheitssystem.* URL: <http://www.heise.de/newsticker/meldung/55565>, 2005. Retrieved: 30.1.2005. (ZITIERT AUF SEITE 47.)
- [Nie01] NIEDERSTADT, C. J.: *Problematik der allgemeinmedizinischen Leitlinien-Entwicklung am Beispiel der Leitlinie Harninkontinenz der DEGAM.* Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 32(2–3):238–239, 2001. (ZITIERT AUF SEITE 73.)
- [Not00] *Primer: Getting into RDF & Semantic Web using N3.* URL: <http://www.w3.org/2000/10/swap/Primer.html>, 2000. Retrieved: 27.10.2004. (ZITIERT AUF SEITE 184.)
- [OAS03] *Organization for the Advancement of Structured Information Standards (OASIS).* URL: <http://www.oasis-open.org>, 2003. Retrieved: 20.08.2003. (ZITIERT AUF SEITE 29.)
- [Oes01] OESTEREICH, BERND: *Objektorientierte Softwareentwicklung: Analyse und Design mit der Unified Modeling Language.* R. Oldenbourg Verlag, München, 5. völlig überarbeitete Auflage, 2001. (ZITIERT AUF SEITE 2.)
- [Ohl03] *C. Ohlms: The Future of Semantic Web - A Perspective on the Market Adoption of Semantic Web Technologies.* URL: <http://www.aifb.uni-karlsruhe.de/AIK/veranstaltungen/aik10/veranstaltungen/aik9/presentations/slides/>
-

- 020419FutureSemanticWeb.pdf, 2003. Retrieved: 20.08.2003. (ZITIERT AUF SEITE 186.)
- [ope02] *openEHR*. URL: <http://www.openehr.org>, 2002. Retrieved: 06.06.2002. (ZITIERT AUF DEN SEITEN 57 UND 61.)
- [OS97] OLDEROG, E.-R. und B. STEFFEN: *Formale Semantik und Programmverifikation*. In: RECHENBERG, P. (Herausgeber): *Informatik-Handbuch*, Seiten 129–148. Hanser Fachbuchverlag, München, 1997. (ZITIERT AUF SEITE 47.)
- [OWL03] *Ontology Web Language-basierte Serviceontologie*. URL: <http://www.daml.org/services/owl-s/1.0/>, 2003. Retrieved: 22.12.2003. (ZITIERT AUF SEITE 187.)
- [OWL04] *OWL Web Ontology Language*. URL: <http://www.w3.org/TR/2004/REC-owl-features-20040210/>, 2004. Retrieved: 2.11.2004. (ZITIERT AUF SEITE 28.)
- [PEH⁺01] PROKOSCH, H. U., U. ENGELMANN, P. HAAS, H. HANDELS, S. H. SCHUG, G. STEYER und M. WALZ: *GMDs-Thesenpapier zur telematischen Vernetzung von Versorgungseinrichtungen im deutschen Gesundheitswesen*. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie*, 32(4):385–394, 2001. (ZITIERT AUF SEITE 2.)
- [PH02] PEDERSEN, S. und W. HASSELBRING: *Begriffssysteme für die medizinische Dokumentation*. In: *Tagungsband der 7. Fachtagung*, Seiten 47–52, Berlin, 2002. Deutscher Verband Medizinischer Dokumentare e.V. (ZITIERT AUF DEN SEITEN 10, 62, 63, 65, 66, 67, 93, 95 UND 223.)
- [PH03] PEDERSEN, SUSANNE und WILHELM HASSELBRING: *Standards für die medizinische Kommunikation und Dokumentation*. Bericht 04/2003, Universität Oldenburg, Department für Informatik, August 2003. (ZITIERT AUF DEN SEITEN 48, 78 UND 153.)
- [Pol01] POLLOCK, JEFFREY T.: *The Big Issue: Interoperability vs. Integration*. *eAI Journal*, Seiten 48–52, October 2001. (ZITIERT AUF DEN SEITEN 10, 11, 12 UND 223.)
- [Pri97] PRIVIGHITORITA, R.: *Konzeption, Entwurf, Implementierung und Validierung der OERR-Funktionalität in einer Standard KIS-Architektur mit Hilfe von DHE*. Diplomarbeit, Universität Dortmund, Fachbereich Informatik, 1997. (ZITIERT AUF SEITE 70.)
- [Pro01] PROKOSCH, H. U.: *KAS, KIS, EKA, EPA, EGA, E-Health: Ein Plädoyer gegen die babylonische Begriffsverwirrung in der Medizinischen Informatik*. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie*, 32/4:371–382, 2001. (ZITIERT AUF SEITE 58.)

-
- [RDF04a] *Resource Description Framework (RDF)*. URL: <http://www.w3.org/RDF/>, 2004. Retrieved: 27.10.2004. (ZITIERT AUF SEITE 185.)
- [RDF04b] *RDF-Primer - W3C Recommendation 10 February 2004*. URL: <http://www.w3.org/TR/2004/REC-rdf-primer-20040210>, 2004. Retrieved: 18.10.2004. (ZITIERT AUF SEITE 28.)
- [RDF04c] *RDF Vocabulary Description Language 1.0: RDF Schema - W3C Recommendation 10 February 2004*. URL: <http://www.w3.org/TR/2004/REC-rdf-schema-20040210>, 2004. Retrieved: 18.10.2004. (ZITIERT AUF SEITE 28.)
- [RDF04d] *RDF/XML Syntax Specification*. URL: <http://www.w3.org/TR/rdf-syntax-grammar/>, 2004. Retrieved: 30.10.2004. (ZITIERT AUF SEITE 114.)
- [Rea02] *Read Clinical Code (RCC)*. URL: <http://www.nhsia.nhs.uk>, 2002. Retrieved: 31.08.2002. (ZITIERT AUF DEN SEITEN 63, 65 UND 93.)
- [REvGJ04] RIESMEIER, J., M. EICHELBERG, S. VON GEHLEN und P. JENSCH: *Standardisierter Austausch strukturierter medizinischer Befundberichte*. In: 49. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (gmds), Innsbruck, September 2004. (ZITIERT AUF SEITE 147.)
- [RW98] RETTIG, J. und S. WEIDLICH: *Integration der EDV-Melder des Epidemiologischen Krebsregisters Niedersachsen mit CAMEL*. URL: http://www.offis.de/projekte/carlos/projekt_carlos.htm, Oktober 1998. (ZITIERT AUF DEN SEITEN 101, 205, 209, 211, 212, 213, 215, 216, 217, 218 UND 225.)
- [RW99] ROHDE, M. und F. WIETEK: *Das Datenschema für das Epidemiologische Krebsregister Niedersachsen*. URL: http://www.offis.de/projekte/carlos/projekt_carlos.htm, Juli 1999. (ZITIERT AUF DEN SEITEN 213, 214, 216, 217, 218, 219 UND 225.)
- [RWT01] *Online-Skript RWTH Aachen: Medizinische Terminologie- und Codierungssysteme*. URL: <http://www.klinikum.rwth-aachen.de/cbt/ok3/mtc/Kapitel1/index.html>, 2001. Retrieved: 27.03.2001. (ZITIERT AUF DEN SEITEN 196 UND 200.)
- [Sau98] SAUTER, GÜNTER: *Interoperabilität von Datenbanksystemen bei struktureller Heterogenität*. Doktorarbeit, Technische Universität Kaiserslautern, Fachbereich Informatik, 1998. (ZITIERT AUF DEN SEITEN 9 UND 10.)
- [Sch96] SCHEK, H.-J.: *Improving the Role of Future Database Systems*. ACM Computing Surveys, 28(4):88, Dezember 1996. (ZITIERT AUF SEITE 17.)

- [SCI02] *Standardisation of Communication between Information Systems in Physician's Offices and Hospitals using XML*. URL: <http://www.sciphox.de>, 2002. Retrieved: 03.04.2002. (ZITIERT AUF DEN SEITEN 52, 72, 88, 105 UND 145.)
- [SCI03] *Pressemitteilung: HL7 Benutzergruppe Deutschland und Qualitätsring Medizinischer Software gründen Arbeitsgemeinschaft SCIPHOX zur Weiterentwicklung des CDA-Standards in Deutschland*. URL: <http://www.sciphox.de/presse/qms200305html>, 2003. Retrieved: 16.08.2003. (ZITIERT AUF DEN SEITEN 51 UND 155.)
- [SCS03] SATTLER, K.-U., S. CONRAD und G. SAAKE: *Datenintegration und Mediatoren*. In: RAHM, E. und G. VOSSEN (Herausgeber): *Web & Datenbanken. Konzepte, Architekturen, Anwendungen*, Seiten 163–190. dpunkt, 2003. (ZITIERT AUF SEITE 13.)
- [See04] SEEBEYOND, <http://www.seebeyond.com/software/ican.asp>: *e*Gate*, 2004. Retrieved 30.08.2004. (ZITIERT AUF SEITE 121.)
- [See05] SEEBEYOND, Dokumentation zur ICAN Suite: *XML-Toolkit Release 5.0.4 for Schema Run-Time Environment (SRE)*, 2005. Retrieved 10.01.2005. (ZITIERT AUF SEITE 124.)
- [SG96] SHAW, M. und D. GARLAN: *Software architecture: perspectives on an emerging discipline*. Prentice Hall, 1996. (ZITIERT AUF SEITE 23.)
- [SL90] SHETH, A.P. und J.A. LARSON: *Federated Database Systems for Managing Distributed, Heterogenous and Autonomous Databases*. ACM Computing Surveys, 22(3):183–236, September 1990. (ZITIERT AUF DEN SEITEN 18, 106, 176, 178, 179 UND 225.)
- [SL03] SATTLER, K. und F. LEYMAN: *Schwerpunktthema: Information Integration & Semantic Web*. Datenbank-Spektrum, 3(6):5–6, Juni 2003. (ZITIERT AUF SEITE 186.)
- [SLM02] SLUIS, D., K.P. LEE und N. MANKOVICH: *DICOM SR - integrating structured data into clinical information systems*. MEDICAMUNDI, 46(2):31–36, 2002. (ZITIERT AUF DEN SEITEN 56 UND 59.)
- [SNO02] *Systematized Nomenclature of Human and Veterinary Medicine (SNOMED)*. URL: <http://www.snomed.org>, 2002. Retrieved: 31.08.2002. (ZITIERT AUF DEN SEITEN 63, 65 UND 93.)
- [SS05] SCHMITT, INGO und GUNTER SAAKE: *A comprehensive database schema integration method based on the theory of formal concepts*. Acta Inf., 41(7):475–524, 2005. (ZITIERT AUF SEITE 141.)
- [Tes03] TESCHKE, T.: *Semantische Komponentensuche auf Basis von Geschäftsprozessmodellen*. Doktorarbeit, Carl von Ossietzky Universität Oldenburg, Fachbereich Informatik, 2003. (ZITIERT AUF SEITE 143.)

-
- [TNM02] *Tumor, Nodes, Metastasis (TNM-Klassifikation)*. URL: <http://www.uicc.org>, 2002. Retrieved: 31.08.2002. (ZITIERT AUF DEN SEITEN 65 UND 93.)
- [TRHL02] TIRADO-RAMOS, A., J. HU und K.P. LEE: *A Case Study of Transcoding DICOM to XML*. Journal of the American Medical Informatics Association, 6:63–72, 2002. (ZITIERT AUF SEITE 56.)
- [Tum98] *Jahresbericht des Tumorregisters Magdeburg 1997, 1998*. (ZITIERT AUF SEITE 211.)
- [UML02] *Unified Medical Language (UMLS)*. URL: <http://www.nlm.nih.gov/research/umls>, 2002. Retrieved: 31.08.2002. (ZITIERT AUF DEN SEITEN 63, 65 UND 93.)
- [URI04] *Unified Resource Identifier*. URL: <http://www.ietf.org/rfc/rfc2396.txt>, 2004. Retrieved: 27.10.2004. (ZITIERT AUF SEITE 183.)
- [VDA02] *Verband der Deutschen Arztpraxis-Softwarehersteller e.V.* URL: <http://www.vdap.de>, 2002. Retrieved: 29.04.2002. (ZITIERT AUF SEITE 72.)
- [vdH02] HEUVEL, W.-J. VAN DEN: *Integrating Modern Business Applications with Objectified Legacy Systems*. Doktorarbeit, Tilburg University, Faculty of Economics and Business Administration, 2002. (ZITIERT AUF SEITE 143.)
- [vdHR05] HEUVEL, W.J. VAN DEN und R. REUSSNER: *Integrating Modern Business Applications with Legacy Systems: A Software Component Perspective*. MIT Press, im Druck Auflage, 2005. (ZITIERT AUF DEN SEITEN 157, 158 UND 225.)
- [WBW01a] WENDT, T., B. BRIGL und A. WINTER: *Modellierung von Integrations-typen*. Informatik, Biometrie und Epidemiologie in Medizin und Biologie, 32(2–3):331, September 2001. (ZITIERT AUF SEITE 149.)
- [WBW01b] WINTER, A., B. BRIGL und T. WENDT: *A UML-based ontology for describing hospital information system architectures*. Medinfo, 10(Pt 1):778–782, 2001. (ZITIERT AUF SEITE 149.)
- [Wee03] WEERS, M.: *UML-basierte Abbildungssprache zwischen Datenschemata*. Diplomarbeit, Carl von Ossietzky Universität Oldenburg, Fachbereich Informatik, 2003. (ZITIERT AUF DEN SEITEN 96 UND 123.)
- [Wen05] WENDT, T.: *Modellierung von Integration in Krankenhausinformationssystemen*. Doktorarbeit, Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, 2005. (ZITIERT AUF SEITE 149.)
- [WG97] WIEDERHOLD, G. und M.R. GENESERETH: *The Conceptual Basis for Mediation Services*. IEEE Expert, 12, No.5, 1997. (ZITIERT AUF SEITE 19.)
-

- [Wie92] WIEDERHOLD, G.: *Mediators in the Architecture of Future Information Systems*. IEEE Computers, 25(3):38–49, 1992. (ZITIERT AUF DEN SEITEN 19, 76 UND 223.)
- [Wie95] WIEDERHOLD, G.: *Mediation in Information Systems*. ACM Computing Surveys, 27(2):265–267, 1995. (ZITIERT AUF DEN SEITEN 19 UND 76.)
- [Wil02] WILLMS, W.: *Eine Abbildung des HL7 Referenzinformationsmodells auf die Datenstruktur im Epidemiologischen Krebsregister Niedersachsen*. Diplomarbeit, Carl von Ossietzky Universität Oldenburg, Fachbereich Informatik, 2002. (ZITIERT AUF DEN SEITEN 16, 105, 109, 165, 181, 223 UND 225.)
- [Win79] WINGERT, F.: *Medizinische Informatik*. Teubner-Verlag, Stuttgart, 1979. (ZITIERT AUF SEITE 200.)
- [WK94] W. KOWALK, M. BURKE: *Rechnernetze: Konzepte und Techniken der Datenübertragung*. B.G. Teubner, Stuttgart, 1994. (ZITIERT AUF SEITE 48.)
- [Wol02] WOLFF, A.: *Patientenzentrierte Dokumentation onkologischer Erkrankungen: Ein generisches XML-basiertes Informationsmodell zur syntaktischen und semantischen Strukturierung einrichtungsübergreifender elektronischer Patientenakten*. Doktorarbeit, Ruprecht-Karls-Universität Heidelberg, Abteilung Medizinische Informatik, 2002. (ZITIERT AUF DEN SEITEN 74, 91, 121, 132, 148 UND 221.)
- [XML04] *Altova XMLSPY Enterprise Edition*. URL: <http://www.xmlspy.com>, 2004. Retrieved: 22.12.2004. (ZITIERT AUF SEITE 109.)
- [XSL99] *XSL Transformations (XSLT) Version 1.0, W3C Recommendation 16 November 1999*. URL: <http://www.w3.org/TR/1999/REC-xslt-19991116>, 1999. Retrieved: 22.2.2005. (ZITIERT AUF SEITE 123.)
- [ZI 02] *Zentralinstitut für die kassenärztliche Versorgung in der Bundesrepublik Deutschland*. URL: <http://zi-koeln.de>, 2002. Retrieved: 02.04.2002. (ZITIERT AUF DEN SEITEN 49 UND 105.)

Index

- A**
- Abbildung
 - bidirektionale 41, 128, 136
 - bijektive 138
 - Abbildung Dokumentation 114
 - Abbildung Kommunikation 112
 - Abbildungsregel 20, 41, 78
 - Abbildungsrichtung 126, 133
 - Abbildungsspezifikation .. 76, 78, 109, 112, 153
 - Abbildungssprache
 - graphische 123
 - UML-basierte 96
 - Abstract Message Definition 52
 - Abstraktion 171
 - Achsen im Begriffssystem 195
 - ACORD 29
 - Adaptor 24
 - Alt-Anwendungen 161
 - Alt-Datenbanken 161
 - Altsystem 143
 - Anfrage-Paradigmen 172
 - Anfragesprache
 - einheitliche 169
 - Ansatz
 - Ein-Modell- 58
 - generischer 58
 - Zwei-Modell- 61
 - Anwendungsarchitektur 2
 - Anwendungsschicht 48
 - Applikationsintegration 10
 - Applikationsinteroperabilität 10
 - Applikationsmediator 20, 76, 80
 - Archetype 61
 - Architektur
 - CCOW- 72
 - Drei-Ebenen- 2, 13, 76
 - Drei-Schichten- 19
 - Drei-Schichten-Schema- 178
 - Fünf-Schichten- 162
 - Integrations- 145
 - Krankenhausinformationssystem- 149
 - mediator-basierte 76, 78, 102, 106, 153, 154
 - mediatorbasierte Integrations- 20
 - Mehrebenen-Modellierungs- 30
 - Metamodellierungs- 33
 - modellgetriebene 25, 26, 35, 62
 - MOF- 83, 95
 - Pipes-and-filters- 23
 - Service-Orientierte 121
 - UML-Objektdiagramm 106, 109
 - vermittelte 19
 - Architekturframework
 - Metamodellierungs- 30
 - Architekturmuster 12
 - EAI- 12
 - ARTEMIS 142
 - Attribut
 - DICOM- 53
 - Auswahlschema 68
 - Auswahlverfahren 102
 - Autonomie 15, 161, 178
 - Ausführungs- 15
 - Entwurfs- 15
 - Kommunikations- 15
- B**
- Basisdokumentation für Tumorkranke ... 83, 91, 96, 104, 105, 109, 111, 112, 114, 121, 124, 126, 132, 136, 137, 153
 - Basisdokumentations-DTD 133
 - Basisdokumentationskomponentenmediator . 112
 - Basisschicht 19, 76
 - BDT . 2, 49, 52, 83, 95, 102, 104, 105, 109, 112, 117, 123, 124, 155
 - BDT-Datenfeld 88, 126, 132
 - BDT-Datenpaket 87
 - BDT-Domänenmodell 109, 153
 - BDT-DTD 111, 126
 - BDT-Satz 88, 126, 128
 - Begriffsbeziehung 67
 - Begriff 7, 65, 91, 93, 190
 - Begriffsbeziehung 91, 194
 - Abstraktions- 194
 - Assoziations- 195
 - Bestands- 194
 - generische 67, 95, 194

- hierarchische 67, 95, 194
 Komplementär- 195
 nichthierarchische 67, 95
 Oppositions- 195
 partitive 67, 95, 194
 pragmatische 67, 95, 195
 sequentielle 67, 95, 195
 Begriffssordnung 12, 65, 93
 Begriffssystem 12, 114, 142, 190
 einheitliches 132, 137
 enumeratives 63, 65, 91, 93
 kompositionelles 63, 65, 91, 93
 medizinisches . 1, 62, 63, 65, 67, 80, 83, 89,
 93, 95, 137, 200
 monohierarchisches 67, 95
 multiaxiales 67, 95
 multihierarchisches 67, 95
 uniaxiales 67, 95
 Begriffssystem für Begriffssysteme .. 63, 80, 91,
 200
 Begriffssystem für Dokumentation 89
 Begriffssystem für Dokumentationsstandards 80,
 83, 89, 109, 114, 137, 153, 155
 Begriffssystem für Kommunikation 89, 95
 Begriffssystem für Kommunikationsstandards 78,
 83, 85, 96, 109, 111, 112, 114, 117,
 123, 137, 153
 Begriffssysteme
 enumerative 63
 kompositionelle 63
 medizinische 193
 standardisierte 57
 Behandlungsdatenträger 49, 51, 115
 Bezeichnung 65, 91, 93, 115, 190
 Bidirektionalität 37
 Bijektivität 137, 138
 Bildinformation
 digitale 53
 BIPS 29
 Bottom-Up-Integration 173
 Bundeskrebsregistergesetz 206
- C**
- CCM 27
 CCOW 71
 CCOW-Architektur 72
 CDA 57–59, 83, 91, 105, 109, 112, 114, 121,
 132, 145, 153, 155
 CDA-Komponentenmediator 78, 112
 CEN 13606 57–59, 91
 Channels 23
 Clinical Context Object Workgroup 71
 Clinical Document Architecture 57–59
 Clinical Pathways 73
 Cluster Original Component Complex 59
- Code
- Additions- 199
 angeordnet sequentieller 198
 bedeutungstragender 198
 gruppensequentieller 67, 95, 198
 hierarchischer 67, 95, 198
 inkrementeller 198
 Juxtapositions- 198
 Kombinations- 198
 mnemonischer 198
 nichtbedeutungstragender 198
 sequentieller 198
 Zufalls- 198
 zusammengesetzter 198
 Code als Codier-Schema 197
 Code als Schlüssel 197
 Code-Tabelle 53
 Code-Typen 197
 Codierprinzip 67, 91, 95, 197
 Collaboration Rules 122
 Common Warehouse Metamodel 27, 41
 Composition Original Component Complex .. 59
 CORBA Komponentenmodell 27
 CWM 27, 41
- D**
- D2D 73, 149
 DAML+OIL 187
 DAML-S 187
 Data Dictionary 53
 Data Item 59
 Data Warehousing 174
 DataGate 121
 Datenübersetzung 11
 Datenbank
 förderierte 141
 Datenbank-Integration 141
 Datenbanksystem
 förderiertes 18, 176, 177
 Datenfeld 51, 115
 Datenformat
 einheitliches 29
 universelles 29
 Datenkapselung 11
 Datenmapping-Tool 123
 Datenmodell
 einheitliches 106
 kanonisches 81, 106, 109, 170, 178
 Datenpaket 51
 Datentyp-Deklaration 111
 DCMR 56
 Deutsches Institut für Medizinische Dokumenta-
 tion und Information 63
 Deutsches Zentrum für Medizinische Klassifika-
 tion 63

- DHE 71
 Diagnosis Related Groups 63
 DICOM 53, 83, 155
 DICOM Attribut 88
 DICOM Content Mapping Resource 56
 DICOM Informationsentität 88
 DICOM Modul 88
 DICOM SR 53, 147
 DICOM Structured Reporting 53, 147
 DICOM-kompatibles System 146
 Digital Imaging and Communications in Medicine 53
 DIMDI 63
 Dimensionenschema 69, 102
 DIS 164
 Disease Management-Programm 73
 Distributed Healthcare Environment 71
 Distributed Information System 164
 Doctor to Doctor 73
 Dokumentation
 medizinische 1, 57, 65, 93, 189
 Dokumentations-Komponentenmediator . 78, 89, 109
 Dokumentationsstandard... 3, 78, 101, 121, 132, 153
 Dokumentenspezifikation
 hierarchische 59
 Dokumentenstruktur
 standardisierte 57
 Domänenfacilitator 20
 Domäneninformationsmodell
 anwendungsbereichsspezifisches 61
 Domänenmodell 106, 109, 124, 132
 BDT- 112
 HL7- 112
 Domänenstandard 76
 Domain Information Model 104
 Drei-Ebenen-Architektur 2, 13
 Drei-Schichten-Architektur 19
 DRG 63, 74
 DTD 106, 124
 Dublin Core 119, 183, 185
 DZMK 63
- E**
- e*Gate 121, 123
 e*Gate Collaboration Rule 124
 e*Gate Collaboration Service 124
 e*Gate Integrator 122
 e*Gate Schema 122
 e*Gate Schema Designer 122
 EAI 2, 12
 Ebene 31, 34, 85
 EHCR 57
 EHCR Root Architectural Components 59
 EHCR von CEN 57–59, 91
 EHCR-Modell
 globales 57
 EHR-Standard
 globaler 61
 EHRcom 59
 Einwohnermeldeamt 101, 205, 213
 EJB 27
 EKN ... 3, 96, 102, 104–106, 112, 117, 153, 208
 Electronic Healthcare Record 2, 57, 83
 Electronic Healthcare Record Architecture ... 57
 Electronic Healthcare Record System 57
 Electronic Healthcare Records 143
 eMamma-Akte 73, 149
 Encoding Rules 52
 Enterprise Application Integration 2, 12
 Enterprise JavaBeans 27
 Entkopplung 13
 Entwurfsmuster 26
 Transformations- 36
 EPA-Stufenmodell 58
 ePA-Tumor 148
 Epidemiologische Krebsregister 101, 205
 Epidemiologisches Krebsregister Niedersachsen
 3, 96, 111, 153, 205, 208
 ETD 122
 ETD builder Wizard 122
 Event Description 53
 Event Type Definition 122
 Evolution 75, 153
 Exportschema 81, 112
 Extensible Markup Language 29
 Extensible Stylesheet Language Transformation
 41, 43, 123
- F**
- Föderation 15
 enge 168, 170
 lose 168–170
 Föderationsschicht 17, 164, 168, 180
 Föderiertes Schema 81
 Facilitator 19, 76, 78, 109, 112
 domänenspezifischer 76
 FDBS 18, 177
 Feature-Diagramm 35
 Feld 53
 Feld-Ende-Markierung 51
 Felddefinition 53
 Feldidentifikation 51
 Feldinhalt 51
 Feldlänge 51
 Feldseparator 53
 File Transfer 20, 22
 FIS 1, 15, 164, 166
 FIS-Klassifikationskriterien 166

Flexibilität 75, 153
 Folder Original Component Complex 59
 Fusion 171

G

GALEN 63, 65, 93, 204
 Gegenstand 65, 91, 93, 190
 GEHR 57, 58, 61
 GEHR Archetypen 61
 GEHR Object Model 61
 Generalisierung 67, 95
 Generic Integration Model 141
 Geschäftsarchitektur 2
 Gesundheitsamt 205, 213
 Gesundheitsdatennetz 47
 Gesundheitswesen 78, 83, 153
 GIM 141
 GLIF 61
 GOM 61
 Good Electronic Healthcare Record .. 57, 58, 61
 GuideLine Interchange Format 61

H

Hardwareausstattung 219
 Headed Section Original Component Complex 59
 header DTD 59
 Health Level Seven 51
 Heterogeneous Information System 164
 Heterogenität ... 1, 2, 15, 83, 101, 161, 162, 178
 Datenmodell- 81
 datenmodellbasierte 15, 162, 180
 logische 15, 163
 schemabasierte 15
 Schnittstellen- 15
 semantische 10, 15, 47, 179
 strukturelle 10, 15
 syntaktische 15, 162
 technische 15, 180
 HIS 164
 HISA 71
 HL7. 1, 51, 83, 95, 104, 105, 109, 111, 112, 117,
 123, 124, 155
 HL7 V3 data type DTD 59
 HL7 Version 3 51
 HL7-DIM 78
 HL7-Domänenmodell 153
 HL7-DTD 89, 109, 126, 153
 HL7-Facilitator 78
 HL7-Feld 88, 126, 132
 HL7-Nachricht 53, 87
 HL7-RIM 59, 109, 153
 HL7-Segment 88, 126, 128
 HL7-Vokabular 51, 53
 Homonym 10, 62, 193

Homonymie 193
 Hyperdatabase 17

I

ICD 1, 52, 63, 65, 67, 93, 95, 102, 137, 203
 ICD-O 104, 137
 IHE 71
 Industriestandard 53
 Information Resource Dictionary Systems ... 25
 Informationsentität 53
 Informationsfluss
 effizienterer 2
 Informationsintegration 10, 12
 Informationsobjekt 53
 normalisiertes 53
 zusammengesetztes 53
 Informationsserver
 HL7- 146
 Informationssystem
 eng gekoppeltes 179
 förderiertes 1, 15, 83, 161, 164
 lose gekoppeltes 176, 177
 mediator-basiertes 18, 20, 76, 162, 176, 179
 medizinisches 57
 Informationssysteme
 heterogene 7, 48
 lose gekoppelt 18
 Informationstransformation 11
 Instanziierung
 flache 30, 31
 tiefe 3, 30, 31, 85
 Instanziierungsmechanismus 30
 Integrating the Healthcare Enterprise 71
 Integration
 semantische 20, 171
 Integrations-Architektur 145
 Integrationsadapter 13
 Integrationsbote 13
 Integrationsfassade 13
 Integrationsmediator 13
 Integrationsplattform 121
 Integrationsschicht 17, 164
 Integrationsschnittstelle 11
 Integrationsstil 20
 Integrationstechnik 1, 2
 Integrierende Schicht 15
 Interoperabilität ... 1, 2, 7, 9, 25, 56, 61, 76, 102,
 154
 institutionsübergreifende . 2, 42, 75, 78, 153
 semantische 7, 9, 12, 48, 102
 technische 9
 Interoperabilitätsbewertung
 automatisierte 147
 Interoperabilitätslösung 10
 Interpretation

- einheitliche 2
 IOD 53
 normalisiertes 53
 zusammengesetztes 53
 IRDS 25, 89
- J**
- Java Collaboration Rules Editor 122
 Java Event Type Definition Editor 122
- K**
- KAON 142
 Klasse 93
 BDT- 119
 HL7- 119
 Klassenäquivalenz 28, 117, 119, 126
 Klassendefinition für BDT 115
 Klassifikation 63, 65, 91, 93, 190, 194, 199
 mehrdeutige 30
 Klinische Pfade 73
 Klinisches Krebsregister 101
 Klinisches Register 205, 212
 Kommunikation 1, 2
 asynchrone 23
 elektronische 47
 medizinische 1
 Kommunikations-Facilitator 78, 85, 109
 Kommunikationsdrehscheibe 57
 Kommunikationsprinzip 7
 Kommunikationsserver 119, 121
 Kommunikationsstandard 3, 63, 78, 101, 121,
 124, 153
 medizinischer 48
 Kommunikationssystem 48
 Komponente
 semi-strukturierte 168
 strukturierte 168
 unstrukturierte 168
 Komponentenmediator 20, 76, 78, 109, 114
 Komponentenmodell 153
 Komponentenschema 81
 Konformitätsprüfung 146
 Kontext-Manager 72
 Konvergenz der Standards 155
 Konzept 7
 Kopplung
 lose 23
 Korrespondenzen 3
 Kostensenkung 2
 Krankenhausinformationssystem 51
 Krebserkrankungen 205
 Krebsfall 101
 Krebsinzidenz 101, 205
 Krebsprävalenz 101
- L**
- Landesamt für Statistik 101, 205, 213
 Landesvermessungsamt 101, 205, 213
 Leitfaden zur Strukturanalyse 3, 34, 85, 95,
 96, 101, 117, 119, 124, 126, 128, 132,
 136, 137, 153, 154
 Leitlinien 73
 Lese- und Schreib-Zugriff 175
 level one body DTD 59
 Link Item 59
 Logik
 Transformations- 36
 Lokales Modell 109
 Lokales Schema 106
- M**
- MapForce 2005 Enterprise Edition 123, 126
 Matching 141
 metamodell-basiertes 143
 Materialisierte Integration 174
 MBIS 20
 integrierende 20
 metainformationsbasiert 20
 MDA 25, 35, 62
 Mediator 13, 19, 142, 153, 179, 180
 Medical Logic Moduls 61
 Medical Subject Headings 65, 93
 Mehrebenen-Modellierungsframework 32, 34, 85
 Meldende Einrichtung 101
 Meldende Institutionen 211
 Melder 101
 heterogene 102
 Melderguppe 101
 Meldeweg
 Anfrage- 218
 Pathologen- 216
 Standard- 213
 Sterbe- 217
 Meldung
 Anfrage- 209
 Pathologen- 209
 Standard- 209
 Sterbe- 209
 Merkmal 192
 Merkmalart 192
 MeSH 204
 Message Development Framework 51
 Message Header Segment 53
 Messages 23
 Messaging 20, 22, 23, 121
 Meta Object Facility 26, 27, 33
 Meta Object Facility Standard 3
 Meta-Facilitator 76, 78
 Meta-Komponentenmediator 76, 78

Metadaten . . .	11, 17, 25, 30, 34, 76, 83, 89, 101, 114, 153, 161, 164, 166, 186
Infrastruktur-	18, 166
logische	17, 166
nutzerbezogene	166
nutzerbezogene-	18
qualitätsbezogene	18, 166
semantische	17, 166
technische	17, 166
Zusammenhang der	83
Metafacilitator	105, 112
Metakomponentenmediator	105, 114
Metametamodell	33
Metamodell for Multiple Metalevels (MoMM)	33
Metamodell	3, 85, 153, 155, 166
Meta-	89
UML-	89
Metamodell der Architektur	80
Metamodelle	17
Metamodellierung	30
Metathesaurus	65, 93
MIRApplus	146
MLM	61
Modell	
domänenspezifisches	20, 76, 78, 80, 153
generisches	57, 61
globales	80
Komponenten	80
MOF-	89, 153
plattformspezifisches	35
plattformunabhängiges	35
standardbasiertes	76
Modellierungsframework	
Mehrebenen-	32, 34, 85
Modellkorrespondenz	117, 123, 124, 126, 132
Modelltransformation	1–3
Modelltransformations-Ansätze	35
Modul	53
Modularitätsmechanismus	38
Moduldefinition	53
Modultabelle	53
MOF	3, 26, 27, 33, 89
MOF-Architektur	83, 95
MOF-Modell	153
Monohierarchie	196
MSH	65
Multidatenbank-Anfragesprache	177
MUSTANG	95

N

N3	114, 184
Nachrichten-Container	11
Nachsorgeleitstelle	101, 205, 211
Nachvollziehbarkeitsverbindungen	39
Namenskonflikt	10

Niedersächsisches Landesgesundheitsamt	208
Nomenklatur	63, 65, 91, 93, 190, 194, 200
Normsprache	143
Notation 3	184
Nur Lese-Zugriff	175
Nutzerschicht	19, 76

O

Objekt	184
Observation ID	52
OFFIS Care GmbH	208
OFX	29
OnkoDok	136
OnkoDok-Modell	91, 121, 132, 153
OnkoDok-Schema	133
OntoBroker	142
Ontologie	12, 67, 85, 141, 142, 186
Ontologierepräsentationssprache	28, 117, 187
Ontologische Informationsrepräsentation	11
Ontology Web Language	28, 187
openEHR	57, 61
Organisationseinheit	2
Organisationsstruktur	38
Original Component Complex	59
OSI-Basisreferenzmodell	48
OWL	28, 117, 119, 126, 187
OWL DL	188
OWL Full	188
OWL Lite	187

P

P2P-Kommunikationsarchitektur	142
PaDok	73, 149
Patient Identification	53
Patient Record Architecture	59
Patientenakte	
elektronische	2
Patientenbegleitende Dokumentation	73
Patientenkennung	115
eindeutige	51
Pauschalentgeltsystem	
diagnosebasiertes	74
Plattformheterogenität	9
Polyhierarchie	196
Portabilität	26
Potenz	31, 34, 85
Prädikat	184
PRA	59
Pragmatik	47
Produktivität	26
Property	184
Prozess-Controller	11
Prozessautomatisierung	13

- Q**
- Quellmodell 36
 - Query, Views, Transformations 26
 - QVT 26
- R**
- RCC 204
 - RDF 28, 119, 183, 187
 - RDF-Aussage 115, 183
 - RDF-Domäne 186
 - RDF-Graph 183
 - RDF-Modell 183
 - RDF-Schema 114, 187
 - RDF-Statement 115, 183, 184
 - RDF-Statements für BDT 115
 - RDF-Tripel 114, 183
 - RDF-Vokabular 185, 186
 - RDF-Zielbereich 186
 - RDF/XML 114
 - RDFS 114, 119, 187
 - RDFS-Klasse 114
 - Read Clinical Code 63, 65, 93
 - Record Components 59
 - Reference Information Model 51
 - Referenzinformationsmodell 62, 105
 - generisches 61
 - Regel-Scheduling 38
 - Regelanwendungsbereich 37
 - Regelanwendungsstrategie 37
 - Regelauswahl 38
 - Regeliteration 38
 - Regelorganisation 38
 - Regelparametrisierung 37
 - Regeltabelle 51
 - Registerstelle 206, 207
 - Remote Procedure Invocation 20, 22
 - Replikation der Konzepte 30
 - Repository 166
 - Resource 183
 - Resource Description Framework .. 28, 183, 187
 - Ressource 183
 - RIM 51
 - Routing 23
- S**
- Sammlung 171
 - Satz 51
 - Satzart 51
 - Satzbeschreibung 51
 - Satzidentifikation 51
 - Satzlänge 51
 - Sattabelle 51
 - Scheduling-Form 38
 - Schema
 - Export 179
 - externes 179
 - föderiertes 164, 168, 179, 180
 - globales 173
 - integriertes 164, 168
 - Komponenten- 178
 - lokales 178
 - vordefiniertes 168
 - Schemaarchitektur 81, 106
 - Fünf-Ebenen- 106
 - Schemaintegrationsverfahren 141
 - Schicht
 - integrierende 161
 - Schichtenarchitektur 76
 - SCIPHOX 52, 88, 105, 145, 149, 155
 - SeeBeyond 121
 - Segment 53
 - Daten- 53
 - Kontroll- 53
 - Segmentseparator 53
 - Segmenttabelle 53
 - Selected Component Complex 59
 - Semantic Web 114, 121, 141, 143, 187
 - Semantik 7, 47
 - gemeinsame 67
 - Semantikvergleich 143
 - Semantischer Konflikt 10
 - Semiotische Triade 65, 93, 190
 - Separation of Concerns 26
 - Shared Database 20, 22
 - Shared-Care 7
 - Single Information System 164
 - SIS 164
 - Skalierbarkeit 20, 75, 153
 - Skalierungskonflikt 10
 - Small Semantic Units 145
 - SNOMED 52, 63, 65, 93, 102, 204
 - SOA 121
 - Software-Komponenten 20
 - Softwarearchitektur
 - flexible 48, 80
 - skalierbare 48, 80
 - weiterentwickelbare 80
 - wiederverwendbare 48
 - Softwareausstattung 219
 - Spezialisierung 67, 95
 - Spezifikation
 - einheitliche 95, 102, 117, 153
 - SSU 145
 - Standard 1, 2, 48, 95, 101, 114, 137, 173
 - Daten- 25, 28
 - domänenspezifisch 42, 48, 153
 - domänenspezifischer 153
 - Framework-/Architektur- 25
 - Metamodell- 25

Nachrichtendefinitions-.....25
 Open Financial Exchange29
 Standardisierung
 terminologische.....7
 Standards
 Metadaten-.....25
 Strukturanalyse95
 Strukturierungselement59, 91, 136
 Sub-Facilitator.....76, 78
 Subjekt.....183
 Suche
 semantische141
 Supplementation171
 Synonym10, 62, 193
 Synonymie193
 Syntaktische Trennung37
 Syntax47
 System
 DICOM-kompatibles.....146

T

Taxonomie194
 Technologiearchitektur2
 Teile-Ganzes-Beziehung67, 95
 Telemedizin.....47
 Terminologie.....10, 187
 medizinische102
 Terminologielehre65, 91, 190
 Thesaurus.....63, 65, 91, 93, 190, 194, 200
 TNM.....137
 TNM-Klassifikation65, 93, 203
 Top-Down-Integration . 48, 80, 81, 95, 101, 111,
 112, 173
 Transformation .. 3, 23, 101, 106, 109, 119, 121,
 153
 Andere Modell-zu-Modell-40
 bidirektional39, 42
 Direktmanipulations-.....39
 Graphen-.....39
 hybride39
 Modell-zu-Code-40
 Modell-zu-Modell-.....40
 Schablonen-basierte39, 40
 strukturgetriebene39
 traversierende39, 40
 unidirektional39
 XML-109
 Transformations-Entwurfsmuster.....36
 Transformations-Logik.....36
 Transformations-Variable36
 Transformationsansatz
 Modell-zu-Code-39
 Modell-zu-Modell-.....39
 Transformationsphasen38
 Transformationsregel.....36, 96, 117, 155

Transparenz172
 Orts-172
 Schema-172
 Sprach-172
 Transport-Service11
 Trigger Events52
 Tumorbasisdokumentation121
 Tumordokumentation49, 52, 147
 Tumorlokalisierungsschlüssel.....203

U

Uebertragungstechnik@Übertragungstechnik.72
 UML2, 12, 26, 30, 48, 49, 58, 63, 78, 85, 91, 95,
 101, 109, 114, 117, 137, 154
 UML-Klassendiagramm
 Architektur80
 UML-Metamodell89
 UML-Metamodellierungs-Framework30
 UMLS52, 63, 65, 93, 95, 204
 Unified Modeling Language2, 12, 26,
 30, 48, 49, 58, 63, 78, 85, 91, 95, 101,
 109, 114, 117, 137
 Uniform Resource Identifiers183
 Uniform Resource Locator184
 Universal Service ID.....52
 Unterklasse115
 Upward Inheritance.....141
 URI115, 183
 URL184
 Use Case52
 Use-Case-Modell.....51

V

Value184
 Variablen
 Transformations-.....36
 VCS.....72, 149
 VDAF Communication Standard72
 Vermittlung
 automatische85
 Vermittlungsschicht.....19, 76, 180
 Verständnis
 gemeinsames12
 Verteilung.....15, 161, 163, 178
 physische15
 Vertrauensstelle206
 Verwechslungskonflikt10
 Virtuelle Integration174
 Vokabular7, 67, 114, 115, 119, 187

W

Wartbarkeit26
 Web-Services142
 Wiederverwendungsmechanismus.....38

Wissen
 deklaratives 12
 prozedurales 12
Wissensrepräsentation 12
Wrapper 20, 24, 76, 80, 81, 143, 162, 180

X

XMI 41, 62
XML 29, 187
XML Metadata Interchange 41, 62
XML-Document Type Definition 27
XML-DTD 27, 122
XML-Namespace-Deklaration 184
XML-Schema 27, 122, 187
XMLSPY Enterprise Edition 109
XSL-Datei 124
XSLT 41, 43, 121, 123, 126, 128, 136

Z

Z-Segment 109, 126, 132, 148, 153
Zielmodell 36
Zugriffsmethoden 175
Zwischenstrukturen 37

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen verwendet habe.

Neuenburg, den 14.6.2005

Lebenslauf

Angaben zur Person

Name: Susanne Pedersen, geb. Müller
Wohnort: Spolsener Moorstr. 3, 26340 Neuenburg
Geburtsdatum, -ort: 1.11.1969, Ulm/Donau
Familienstand: verheiratet
Staatsangehörigkeit: deutsch

Schulbildung

09/1975 - 06/1979: Grundschule Senden-Aufheim
06/1979 - 06/1988: Nikolaus-Kopernikus-Gymnasium Weißenhorn
Abschluss: allgemeine Hochschulreife (Note: 1,8)

Studium

10/1988 - 02/1994: Wirtschaftsmathematik an der Universität Ulm
Abschluss: Diplom-Wirtschaftsmathematiker (Note gut)

Beruf

10/1994 - 04/1996: Angestellte bei Wollert-Elmendorff Deutsche Industrie-Treuhand GmbH in Hamburg (Wirtschaftsprüfungsgesellschaft)
05/1996 - 10/1998: Freiberufliche Tätigkeit für Wollert-Elmendorff Deutsche Industrie-Treuhand GmbH in Hamburg
seit 02/1999: Selbstständig in eigener Heilpraktikerpraxis sowie Halbtags­tätigkeit als zahnmedizinische Fachangestellte
seit 01/2000: Externe Promotion an der Universität Oldenburg, Department für Informatik, Abteilung Software Engineering
09/2003 - 02/2004: Wissenschaftliche Mitarbeiterin an der Universität Oldenburg, Department für Informatik, Abteilung Software Engineering
05/2004 - 09/2004: Textentwurf für einen Teil des Buches „Enterprise Application Integration“, Spektrum Verlag
11/2004 - 09/2005: Lektorierung des Buches „Enterprise Application Integration“ für den Spektrum Verlag