

Modellierung und Nutzung von Relationen zwischen Mehrfachrepräsentationen in Geo-Informationssystemen

Von der Fakultät
Luft- und Raumfahrttechnik und Geodäsie
der Universität Stuttgart
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung

vorgelegt von
Diplom-Geograph Steffen Volz
aus Bietigheim-Bissingen

Hauptberichter: Prof. Dr.-Ing. habil. Dieter Fritsch
Mitberichter: Prof. Dr.-Ing. habil. Monika Sester
Mitberichter: Prof. Dr.-Ing. habil. Bernhard Mitschang

Tag der mündlichen Prüfung:
28. Juli 2006

Institut für Photogrammetrie
Universität Stuttgart

2006

Zusammenfassung

Die wachsende Bedeutung von raumbezogenen Daten in den verschiedensten Anwendungsbereichen hat dazu geführt, dass zahlreiche Firmen und öffentliche Institutionen die Erfassung von Geodaten vorantreiben. Dabei werden ein und dieselben Realweltobjekte häufig mehrfach und aus unterschiedlichen Blickwinkeln bzw. Anwendungsperspektiven erfasst und im Computer gespeichert. Auf diese Weise entstehen widersprüchliche bzw. inkonsistente Repräsentationen eines Objektes der Realwelt, so genannte Mehrfachrepräsentationen (MRep).

Sollen diese Mehrfachrepräsentationen innerhalb von offenen Geo-Informationssystemen bzw. Geodateninfrastrukturen, wie beispielsweise der an der Universität Stuttgart entwickelten Nexus-Plattform, bereitgestellt werden, so müssen einerseits die Schemas der heterogenen Ausgangsdaten zusammengeführt werden. Andererseits ist es notwendig, die Inkonsistenzen zwischen Mehrfachrepräsentationen – also auf Instanzebene – adäquat behandeln zu können, um eine gemeinsame Datenverarbeitung zu ermöglichen. Die vorliegende Arbeit hat zum Ziel, einen Beitrag zur Forschung auf diesen Gebieten zu leisten.

Im Verlauf der Untersuchung wird zunächst der Zusammenhang zwischen Interoperabilitätsbestrebungen bzw. Standardisierungsmaßnahmen internationaler und nationaler Institutionen und der Thematik der Mehrfachrepräsentationen aufgezeigt und es wird erläutert, weshalb die vorgeschlagenen Konzepte für das Problemfeld der MRep nur bedingt geeignet sind. Anschließend wird die Nexus-Plattform, eine global angelegte, föderierte Infrastruktur zur Unterstützung orts- und kontextbezogener Systeme, vorgestellt, da sie den Rahmen für die vorliegende Arbeit bildet. Zum Zwecke der Vermittlung von Grundlagen über die Thematik der Mehrfachrepräsentationen erfolgt eine umfassende Begriffsklärung und eine Aufarbeitung des Standes der Forschung.

In der Folge wird eine Vorgehensweise präsentiert, um auf der Basis einer Untersuchung der konzeptionellen Geodatenschemas von Geographic Data Files (GDF), des Amtlichen Topographisch-Kartographischen Informationssystems (ATKIS) und der Automatisierten Liegenschaftskarte (ALK) ein übergeordnetes, globales Schema für die Nexus-Systemplattform ableiten zu können. Dabei werden auch die Applikationsanforderungen, denen das globale Schema zu genügen hat, berücksichtigt. Darüber hinaus werden Abbildungsregeln aufgestellt, die eine Transformation von Daten der Ausgangsschemas in das übergeordnete Schema der Infrastruktur erlauben.

Das grundlegende Konzept dieser Arbeit besteht darin, die innerhalb eines Geo-Informationssystems auftretenden Mehrfachrepräsentationen über explizite Relationen miteinander zu verknüpfen und somit eine neue Form der Modellierung in GIS einzuführen. Zu diesem Zweck wird ein formales Modell für Relationen zwischen Mehrfachrepräsentationen (MRep-Relationen) aufgestellt. Zentraler Aspekt ist es dabei, die Ähnlichkeit von raumbezogenen Objekten über entsprechende geometrische, topologische und thematische Ähnlichkeitsmaße zu beschreiben. Anhand von linearen Testdaten aus dem Bereich des Straßenverkehrs, die aus den konzeptionellen Geodatenschemas von GDF und ATKIS stammen, werden MRep-Relationen zwischen korrespondierenden Instanzen mit Hilfe eines semi-automatischen Verfahrens generiert. Die Probleme, welche bei diesem Zuordnungsprozess auftreten, werden beleuchtet und die Zuordnungsergebnisse werden präsentiert.

Im Mittelpunkt der vorliegenden Untersuchung steht dann die Auswertung der erzeugten MRep-Relationen. Zum einen wird ein datengetriebenes Schema-Matching-Verfahren vorgestellt, das aus der Analyse von Relationen zwischen Mehrfachrepräsentationen automatisch semantische Korrelationen zwischen unterschiedlichen Schemas abzuleiten vermag. Zum anderen wird ein Verfahren zur Netzwerkanalyse entwickelt, das MRep-Relationen bei der Suche von kürzesten Wegen in mehrfach repräsentierten Graphen benutzt.

Das instanzbasierte Schema-Matching-Verfahren untersucht die MRep-Relationen, welche zwischen korrespondierenden Repräsentationen aus GDF und ATKIS generiert wurden. Es betrachtet dabei die Klassenzu-

gehörigkeiten der zugeordneten Objekte in ihren Ausgangsschemas. Treten dabei Signifikanzen in der Hinsicht auf, dass Instanzen einer bestimmten Objektklasse aus GDF überwiegend Instanzen einer bestimmten Objektklasse aus ATKIS zugeordnet werden, so bedeutet dies, dass die entsprechenden Objektklassen aus den unterschiedlichen Ausgangsschemas semantisch korrelieren. Der Grad der Entsprechung zwischen den betrachteten Objektklassen aus GDF und ATKIS wird in Form prozentualer Korrelationsmaße ausgedrückt. Das Verfahren erlaubt es somit, ausschließlich aus der Zuordnung von Instanzen und damit ohne Kenntnis der jeweiligen Schemasemantik, Korrespondenzen zwischen Objektklassen unterschiedlicher Schemas ableiten zu können.

Der Ansatz der Netzwerkanalyse in mehrfach repräsentierten Straßennetzwerken bzw. Graphen entstand vor dem Hintergrund, dass Geodaten häufig in verschiedene Teilstücke zergliedert sind, die miteinander verknüpft werden müssen, um eine gemeinsame Bearbeitung zu ermöglichen. Die Verknüpfung der Teilstücke wird im hier dargestellten Verfahren über MRep-Relationen zwischen korrespondierenden Kanten in den Ausgangsgraphen ermöglicht. Dabei werden zwischen den Graphen Übergangskanten eingeführt, mittels derer ihre Adjazenzmatrizen integriert werden können. Werden die Start- und Endpunkte einer Netzwerkanalyse angegeben, so ermittelt eine Tiefensuche sämtliche potentiellen Wege. Der kürzeste Weg ist dabei jener, dessen Kosten aus der Traversierung der realen Kanten und der Übergangskanten minimal ist. Je nach Gewichtung der Übergangskanten können sich demnach unterschiedliche kürzeste Wege ergeben. Das entwickelte Verfahren erbringt den Nachweis, dass Netzwerkanalysen unter Anwendung des erarbeiteten MRep-Relationenkonzeptes auf mehrfach repräsentierten Graphen ohne komplette Verschmelzung der Ausgangsdaten möglich sind.

Summary

The growing awareness concerning the relevance of spatial information for various fields of application has led to a continuously increasing number of companies and public institutions that are capturing geospatial data from different, application-specific perspectives on reality. Thus, as a result of the data acquisition process, one and the same real world object is available in inconsistent and conflicting representations, so-called multiple representations (MRep), in different spatial databases.

If these multiple representations shall be provided within open geographic information systems or geospatial data infrastructures like the Nexus platform that is currently being developed at the University of Stuttgart, the data schemas of the heterogeneous source data have to be integrated on the one hand and on the other hand adequate techniques must be available in order to deal with inconsistencies of multiple representations – on the instance level – in geospatial data processing. This work aims at contributing to the research in these domains.

In the course of the study, the correlations between efforts of international and national institutions to achieve interoperability in GIS through standardization and the problems of multiple representations are identified first. It is pointed out that these interoperability approaches are necessary but only partially suitable for solving the problems in the field of MRep. Then, the Nexus platform, a global and federated infrastructure supporting location and context aware information systems, is introduced since it provides the framework for this research. In order to impart basic knowledge about the problem of multiple representations, a detailed explanation of the term “multiple representations” is given and related work in this field is described.

Subsequently, an approach is presented that allows deriving a global schema for the Nexus infrastructure on the basis of an analysis of existing geospatial schemas (namely Geographic Data Files (GDF), the German Authoritative Topographic Cartographic Information System (ATKIS) and the German Automatic Real Estate Map (ALK)). This process also includes an investigation of the requirements the global schema has to meet. In a further step, mapping rules are defined in order to transfer data of the available schemas into the global schema of the infrastructure.

The specific concept of this research is based on the idea of linking all multiple representations available within a geographic information system by explicit relations, thus introducing a new type of modeling in GIS. For this purpose a formal model to express the relations between multiple representations (MRep relations) is proposed. The central aspect here is to express the similarity of geospatial objects by appropriate geometric, topological and thematic similarity indicators. Then, using a semiautomatic approach, MRep relations between corresponding representations are generated for test data sets consisting of linear street data stemming from GDF and ATKIS. The problems arising during this matching process are described and the matching results are presented in detail.

In the focus of this work, the relations that have been generated between multiple representations are exploited for different purposes. On the one hand a data-driven schema matching approach is presented which allows for deriving semantic correlations between different schemas by analyzing MRep relations. On the other hand, a network analysis technique for graphs available in multiple representations is developed that uses MRep relations during shortest path search.

The instance-based schema matching approach investigates the MRep relations which have been set up between corresponding representations of ATKIS and GDF. It takes into account the affiliations of the matched objects to object classes in their source schemas. If significant characteristics appear, so that instances of a specific object class in GDF are predominantly matched to instances which belong to a certain object class in ATKIS, a semantic correlation between the respective object classes of the different schemas can be detected. The degree of correlation between the investigated object classes of GDF and ATKIS can be

expressed by percentage values. The method thus only uses the matching results on the instance level and does not need any knowledge about the schemas to be matched in order to derive semantic correspondences between them.

The network analysis approach for street networks or graphs that are available in multiple representations emerged from the fact that geospatial data are often scattered into different pieces which have to be linked to enable a combined processing. Concerning the proposed method, the links between these different pieces can be derived from MRep relations that have been generated for corresponding edges in the source graphs. Between the different source graphs, transition edges are introduced by which their adjacency matrices can be integrated. If the start and end points of the network analysis are given, a depth-first search calculates all potential paths. The shortest path is then determined as the one which shows the minimal costs for the traversal of real edges plus transition edges. Depending of the weight that is attached to the transitions edges, different shortest paths are possible. The method demonstrates that a network analysis can be performed on multiple representations of road data without merging the source data if explicit relations between corresponding instances have been set up.

Danksagung

Zunächst möchte ich mich ganz besonders bei Herrn Prof. Dr. Dieter Fritsch bedanken, der mir die Möglichkeit gegeben hat, an seinem Lehrstuhl, dem Institut für Photogrammetrie (IfP) der Universität Stuttgart, zu arbeiten und zu promovieren und der mich als Hauptreferent betreute. Trotz seiner knapp bemessenen Zeit als Rektor der Universität Stuttgart konnte ich stets auf seine wertvollen Hinweise und seine ermutigende Unterstützung vertrauen.

Mein besonders herzlicher Dank gilt auch der Mitberichterin dieser Arbeit, Frau Prof. Dr. Monika Sester, die bereits meine Diplomarbeit betreute und durch die ich ans Institut für Photogrammetrie gekommen bin. Sie hat mich über Jahre hinweg als Leiterin der Forschungsgruppe Geo-Informationssysteme am IfP begleitet und gefördert. Obwohl sie inzwischen als Professorin an der Universität Hannover tätig ist, hat sie sich immer wieder Zeit für mich genommen und mir wichtige inhaltliche Impulse gegeben.

Ebenso danke ich Herrn Prof. Bernhard Mitschang von der Abteilung „Anwendersoftware“ des Instituts für Parallele und Verteilte Systeme der Universität Stuttgart für seine freundliche Bereitschaft, diese Arbeit als Mitberichter zu begutachten.

Zu größtem Dank bin ich Herrn Dr. Volker Walter verpflichtet. Er ist aktueller Leiter der Forschungsgruppe Geo-Informationssysteme am IfP und hat durch seine vielen Anregungen in wesentlichem Maße zum Gelingen der Arbeit beigetragen. Geduldig vermittelte er mir seine hervorragenden Fachkenntnisse, diskutierte stundenlang inhaltliche Details mit mir und gab auch wertvolle Hinweise bei der schriftlichen Umsetzung der Arbeit.

Ich möchte mich zudem bei Herrn cand. Geogr. Lars Timm, langjähriger Hilfswissenschaftler am Institut für Photogrammetrie, bedanken. Er hat mir durch enormen Einsatz bei der Datenaufbereitung geholfen und mich durch sein selbständiges Arbeiten entlastet.

Für die Bereitstellung des Datenmaterials danke ich dem Landesvermessungsamt Baden-Württemberg, dem Stadtmessungsamt Stuttgart und der Firma NAVTEQ. Die vorliegende Arbeit entstand im Rahmen des Nexus-Projektes an der Universität Stuttgart. Dieses Projekt wurde anfangs vom Ministerium für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg und im weiteren Verlauf von der Deutschen Forschungsgemeinschaft finanziert. Bei beiden Institutionen möchte ich mich daher vielmals bedanken.

Dem gesamten Kollegium am Institut für Photogrammetrie gebührt ebenfalls mein Dank, denn die kollegiale und freundschaftliche Atmosphäre hat mir seit Beginn meiner Tätigkeit am IfP ein angenehmes und produktives Arbeiten ermöglicht. Innerhalb des Nexus-Projektes konnte ich darüber hinaus in zahlreichen Sitzungen einen Einblick in andere Fachbereiche gewinnen, weshalb ich mich an dieser Stelle auch bei allen Projektmitarbeitern bedanken will, die mir durch ihre Anregungen und Erläuterungen weitergeholfen haben.

Außerdem richte ich meinen Dank an Karl-Heinrich Anders, Martin Kada, Markus Müller, meinen Onkel Hans Rapp (†) und Wilhelm Stangl, welche die Rohfassung dieser Arbeit sorgfältig durchgelesen und mir wertvolle Verbesserungsvorschläge gegeben haben.

Schließlich möchte ich mich zutiefst bei meiner Familie bedanken. Meine geliebte Ehefrau Dusida hat mich immer wieder ermutigt und bestärkt und ist die größte Quelle meiner Kraft. Meine Eltern Elly und Paul Volz und mein Bruder Rüdiger haben mir mein Leben lang in allen Belangen geholfen, mich gefördert und mir ihre stetige Zuneigung entgegengebracht.

Inhaltsverzeichnis

1	EINFÜHRUNG	1
1.1	MOTIVATION.....	1
1.2	PROBLEMSTELLUNGEN BEI DER INTEGRATION VON SCHEMAS UND INSTANZEN	2
1.3	ZIELSETZUNGEN UND FRAGESTELLUNGEN DER ARBEIT	4
1.4	AUFBAU DER ARBEIT	6
2	INTEROPERABILITÄT UND STANDARDS IN GIS	8
2.1	ENTWICKLUNG DER BESTREBUNGEN ZUR INTEROPERABILITÄT	8
2.2	BEMERKUNGEN ZUR INTEROPERABILITÄT.....	9
2.3	ÜBERBLICK ÜBER DIE STANDARDISIERUNGSORGANISATIONEN	10
2.4	ORGANISATION UND ANSATZ DES OPENGIS KONSORTIUMS (OGC)	11
2.4.1	<i>Das Spezifikationsprogramm des OGC.....</i>	<i>12</i>
2.4.1.1	Abstrakte Spezifikationen.....	12
2.4.1.2	Implementierungs-Spezifikationen.....	16
2.4.2	<i>Das Interoperabilitätsprogramm des OGC.....</i>	<i>17</i>
2.4.2.1	Das „Information Community Enablement Testbed“	18
2.4.2.2	Das „Geospatial Fusion Services Testbed“	18
2.5	DAS STANDARD-BASIERTE AFIS-ALKIS-ATKIS-REFERENZMODELL	19
2.5.1	<i>Das konzeptuelle Modell.....</i>	<i>19</i>
2.5.1.1	Das AAA-Basisschema	20
2.5.1.2	Das AAA-Versionierungsschema	21
2.5.1.3	Das AFIS-ALKIS-ATKIS Fachschema.....	22
2.5.2	<i>Die Objektartenkataloge von ALKIS und ATKIS.....</i>	<i>22</i>
2.5.3	<i>Die Normbasierte Austauschschnittstelle (NAS)</i>	<i>22</i>
2.6	INTEROPERABILITÄT UND MEHRFACHREPRÄSENTATIONEN	23
3	NEXUS - MEHRFACHREPRÄSENTATIONEN IN EINER OFFENEN SYSTEMPLATTFORM.....	25
3.1	DAS GRUNDKONZEPT VON NEXUS.....	25
3.1.1	<i>Kommunikation und Sicherheit</i>	<i>27</i>
3.1.2	<i>Datenmodellierung und Datenmanagement.....</i>	<i>27</i>
3.1.3	<i>Erfassung und Präsentation.....</i>	<i>28</i>
3.1.4	<i>Anwendungen.....</i>	<i>28</i>
3.2	DAS AUGMENTED WORLD SCHEMA VON NEXUS.....	28
3.2.1	<i>Entstehung des AWS</i>	<i>28</i>
3.2.2	<i>Struktur des AWS.....</i>	<i>29</i>
3.3	ARCHITEKTUR DER PLATTFORM	29
3.3.1	<i>Die Applikationsschicht</i>	<i>30</i>
3.3.2	<i>Die Föderationsschicht.....</i>	<i>30</i>
3.3.3	<i>Die Dienstschicht.....</i>	<i>30</i>
3.4	MEHRFACHREPRÄSENTATIONEN UND MEHRFACHKONZEPTIONEN IN NEXUS	31
4	ZUR PROBLEMATIK DER MEHRFACHREPRÄSENTATIONEN	33
4.1	ZUM VERSTÄNDNIS DES BEGRIFFS „MEHRFACHREPRÄSENTATIONEN“	33
4.1.1	<i>Arten von Mehrfachrepräsentationen aus GIS-Sicht</i>	<i>33</i>
4.1.1.1	Schemabedingte Mehrfachrepräsentationen.....	33
4.1.1.2	Erfassungsbedingte Mehrfachrepräsentationen	34
4.1.1.3	Zeitlich bedingte Mehrfachrepräsentationen	34
4.1.1.4	Verwaltungsbedingte Mehrfachrepräsentationen	35
4.1.1.5	Mehrfachrepräsentationen durch Veredelung	35
4.1.1.6	Formatbedingte Mehrfachrepräsentationen.....	35
4.1.2	<i>Abstraktere Betrachtungen zum Thema „Mehrfachrepräsentationen“</i>	<i>35</i>
4.1.2.1	Ebenen der Repräsentation	35
4.1.2.2	Realweltobjekte und deren Repräsentationen	36
4.1.2.3	Mehrfachrepräsentationen und Konsistenz	37
4.1.2.4	Mehrfachrepräsentationen und Redundanz.....	38
4.2	INTEGRATION VON MEHRFACHREPRÄSENTATIONEN.....	38
4.2.1	<i>Zum Verständnis des Begriffs „Integration“ innerhalb dieser Arbeit</i>	<i>38</i>

4.2.2	<i>Sinn und Zweck der Integration</i>	39
4.2.2.1	Grund 1: Analyse von Mehrfachrepräsentationen	39
4.2.2.2	Grund 2: Fortführung und Konsistenz von Mehrfachrepräsentationen	40
4.2.2.3	Grund 3: Optimierung von nachfolgenden Integrationsprozessen.....	40
4.3	MEHRFACHREPRÄSENTATIONEN IN GIS – STAND DER FORSCHUNG	41
4.3.1	<i>Semantische Integration: Schemaintegration und Ontologien</i>	41
4.3.1.1	Schema-Matching und Schemaintegration.....	42
4.3.1.2	Ontologie-basierte Konzepte	43
4.3.2	<i>Datenverwaltung von mehrfach repräsentierten Objekten</i>	46
4.3.3	<i>Ansätze zur Zuordnung und Verschmelzung von Mehrfachrepräsentationen</i>	47
4.3.3.1	Zuordnungsansätze	48
4.3.3.2	Conflation-Techniken.....	50
4.3.4	<i>Anwendungen: Fortführung und Analyse mehrfach repräsentierter Objekte</i>	51
4.3.4.1	Fortführung.....	51
4.3.4.2	Analyse	52
5	ABBILDUNG BESTEHENDER GEODATENSCHEMAS AUF DAS AWS	54
5.1	UNTERSUCHUNG EXISTIERENDER SCHEMAS FÜR GEODATEN	54
5.1.1	<i>Das Datenschema von GDF</i>	55
5.1.2	<i>Das Datenschema von ATKIS</i>	57
5.1.3	<i>Das Datenschema der ALK</i>	58
5.1.4	<i>Gegenüberstellung der Konzepte zur Modellierung von Straßendaten</i>	59
5.1.5	<i>Erkenntnisse aus der Schema-Analyse</i>	60
5.2	ANFORDERUNGEN AN DAS STRAßEN-DATENSHEMA AUS DER SICHT VON NEXUS	61
5.2.1	<i>Allgemeine Anforderungen</i>	61
5.2.2	<i>Use Cases für Straßenverkehrsdaten</i>	61
5.3	ENTWICKLUNG EINES GLOBALEN SCHEMAS	62
5.3.1	<i>Anforderungskriterien für die Klassen</i>	62
5.3.2	<i>Realisierung der Klassen</i>	63
5.4	ABBILDUNGSVORSCHRIFTEN BESTEHENDER SCHEMAS AUF DAS AWS	65
5.4.1	<i>Abbildung von GDF auf das AWS</i>	65
5.4.2	<i>Abbildung von ATKIS auf das AWS</i>	65
5.4.3	<i>Abbildung von ALK auf das AWS</i>	66
6	MODELLIERUNG VON RELATIONEN ZWISCHEN MEHRFACHREPRÄSENTATIONEN	69
6.1	MOTIVATION FÜR DIE SPEICHERUNG VON MREP-RELATIONEN	69
6.2	ZUM BEGRIFF „RELATION“	70
6.3	RAHMENBEDINGUNGEN FÜR DIE MODELLIERUNG VON MREP-RELATIONEN.....	71
6.3.1	<i>Anforderungen durch Nexus</i>	71
6.3.1.1	Allgemeine Vorgaben durch das AWS.....	71
6.3.1.2	Anforderungen durch das Relationenmodell von Nexus.....	73
6.3.2	<i>Anforderungen durch die Anwendungen</i>	76
6.3.2.1	Anforderungen durch Analyse, Verschmelzung und Fortführung.....	77
6.3.2.2	Anforderungen aus der Sicht des automatischen Schema-Matchings.....	77
6.4	ÄHNLICHKEITSMÄßE FÜR RAUMBEZOGENE OBJEKTE.....	77
6.4.1	<i>Geometrische Ähnlichkeitsbestimmung</i>	78
6.4.2	<i>Topologische Ähnlichkeitsbestimmung</i>	78
6.4.3	<i>Ähnlichkeitsbestimmung von Attributen</i>	79
6.4.4	<i>Gemischte Ähnlichkeitsmaße</i>	79
6.4.5	<i>Bestimmung eines Gesamtmaßes für die Ähnlichkeit von Repräsentationen</i>	80
6.5	MODELLIERUNG VON MREP-RELATIONEN.....	80
6.6	TYPIFIZIERUNG DES IDENTITÄTSGRADES VON MEHRFACHREPRÄSENTATIONEN	82
6.7	UMSETZUNG DES MREP-RELATIONSKONZEPTES IN NEXUS.....	83
7	ERZEUGUNG VON MREP-RELATIONEN	85
7.1	DIE SOFTWARE-UMGEBUNG ZUR ERZEUGUNG UND VISUALISIERUNG VON MREP-RELATIONEN	85
7.1.1	<i>Die RelationBuilder-Toolbox</i>	85
7.1.1.1	Erzeugung von MRep-Relationen	86
7.1.1.2	Speicherung von MRep-Relationen.....	88
7.1.1.3	Richtlinien bei der Zuordnung	89
7.1.2	<i>Die RelationViewer-Toolbox</i>	92
7.2	ABLEITUNG VON MREP-RELATIONEN ANHAND VERSCHIEDENER BEISPIELDATENSÄTZE	93
7.2.1	<i>Beschreibung der verwendeten Daten</i>	93

7.2.2	<i>Probleme bei der Erzeugung von MRep-Relationen</i>	95
7.2.3	<i>Zuordnungsergebnisse für die Testgebiete</i>	97
8	AUSWERTUNG VON MREP-RELATIONEN	104
8.1	AUTOMATISCHE ERZEUGUNG VON SCHEMARELATIONEN	104
8.1.1	<i>Theoretischer Ansatz</i>	104
8.1.2	<i>Ein System zur automatischen Ableitung von Schemarelationen</i>	105
8.1.2.1	Das RelationAnalyzer-Tool	107
8.1.2.2	Ergebnisse der Analyse	108
8.1.3	<i>Diskussion der Ergebnisse</i>	113
8.2	AUSWERTUNG VON MREP-RELATIONEN ZUR NAVIGATION IN MEHRFACH REPRÄSENTIERTEN GRAPHEN	116
8.2.1	<i>Mögliche Strategien zum Aufbau eines integrierten Navigationsgraphen</i>	116
8.2.2	<i>Transformation des Navigationsgraphen in eine Baumstruktur</i>	118
8.2.3	<i>Finden von kürzesten Wegen über eine Tiefensuche</i>	119
8.2.4	<i>Berechnung der Wegekosten</i>	119
8.2.5	<i>Optimierungsmöglichkeiten bei der Baumsuche</i>	121
8.2.6	<i>Realisierung des Ansatzes</i>	121
8.2.7	<i>Diskussion der Ergebnisse</i>	122
9	ZUSAMMENFASSUNG UND DISKUSSION DER ERGEBNISSE	125
10	AUSBLICK	129
11	LITERATUR	130

ANHANG

I	DARSTELLUNG DER TESTGEBIETE	141
	TESTGEBIET 1: ATKIS	141
	TESTGEBIET 1: GDF.....	142
	PARALLELDARSTELLUNG DER DATEN AUS TESTGEBIET 1	143
	TESTGEBIET 2: ATKIS	144
	TESTGEBIET 2: GDF.....	145
	PARALLELDARSTELLUNG DER DATEN AUS TESTGEBIET 2	146
II	OBJEKTKLASSEN-KORRESPONDENZEN VON GDF UND ATKIS	147
II.A	KORRESPONDENZEN DER GDF-KLASSEN	147
II.B	KORRESPONDENZEN DER ATKIS-KLASSEN	149

LEBENS LAUF

Abkürzungsverzeichnis

AAA	AFIS-ALKIS-ATKIS
AdV	Arbeitsgemeinschaft der Vermessungsverwaltungen der Bundesrepublik Deutschland
AFIS	Automatisiertes Festpunkt-Informationssystem
ALB	Automatisiertes Liegenschaftsbuch
ALK	Automatisierte Liegenschaftskarte
ALKIS	Automatisiertes Liegenschaftskataster-Informationssystem
AS	Abteilung Anwendersoftware des IPVS
ATKIS	Amtliches Topographisch-Kartographisches Informationssystem
AWM	Augmented World Model
AWML	Augmented World Modeling Language
AWQL	Augmented World Query Language
AWS	Augmented World Schema
CEN	Comité Européen de Normalisation
COMA	Combining Match Algorithms
DB	Datenbank
DBMS	Datenbankmanagementsystem(e)
DFG	Deutsche Forschungsgemeinschaft
DIN	Deutsches Institut für Normung
DL	Description Logic
DLM	Digitales Landschaftsmodell
DSGK	Deutsche Stadtgrundkarte
EVW	Evaluationswert
FC	Functional Class
FRB	Fahrbahn
GDF	Geographic Data Files
GIS	Geo-Informationssystem(e)
GM	Gesamtmaß der Ähnlichkeit zweier Mehrfachrepräsentationen
GML	Geography Markup Language
GPS	Global Positioning System
ICP	Iterative Closest Point Algorithmus
IM	Individuelles Ähnlichkeitsmaß
ID	Identifikator
IfP	Institut für Photogrammetrie der Universität Stuttgart
IKR	Institut für Kommunikationsnetze und Rechnersysteme der Universität Stuttgart
IP	Interoperabilitätsprogramm des OGC
IPVS	Institut für Parallele und Verteile Systeme der Universität Stuttgart
ISEC	Intersection
ISO	International Standards Organisation
JCS	Java Conflation Suite
JTS	Java Topology Suite
JUMP	Java Unified Mapping Platform
KI	Künstliche Intelligenz

LOD	Level of Detail
MRDB	Multirepräsentations-Datenbank
MRep	Mehrfachrepräsentation(en)
MRR	MRep-Relation(en)
MRRL	Multiple Representation Relation Language
MUR	Minimal umschließendes Rechteck
NAS	Normbasierte Austauschschnittstelle
NOL	Nexus Object Locator
NREO	Nicht Raumbezogene Elementarobjekte
OBAK	Objektabbildungskatalog
OGC	OpenGIS Konsortium
OSKA	Objektschlüsselkatalog
PMO	Punktmengenobjekte
RDEL	Road Element
REO	Raumbezogene Elementarobjekte
RFP	Request for Proposal
SFB	Sonderforschungsbereich
SP	Spezifikationsprogramm des OGC
SQL	Structured Query Language
STR	Straße
TC 204	Technical Commission 204 der ISO
TC 211	Technical Commission 211 der ISO
TC 278	Technical Commission 278 des CEN
TC 287	Technical Commission 287 des CEN
ÜK	Übergangskante
UML	Unified Modeling Language
UMTS	Universal Mobile Telecommunications System
URI	Uniform Resource Identifier
VPF	Vector Product Format
VS	Abteilung Verteilte Systeme des IPVS
WDM	Widmung
WLAN	Wireless Local Area Network
WWW	World Wide Web
XML	Extended Markup Language
ZUSO	Zusammengesetzte Objekte

1 Einführung

1.1 Motivation

Das World Wide Web (WWW) hat die Funktionsfähigkeit einer globalen und offenen Computer-Infrastruktur unter Beweis gestellt. Es besteht aus einer Vielzahl verteilter Rechnersysteme, die untereinander vernetzt sind und verschiedenste Daten und Dienste anbieten. Den Paradigmen des WWWs folgend, wurde an der Universität Stuttgart das Nexus-Projekt [Hohl et al. 99] initiiert, welches eine ähnlich ausgerichtete Plattform für den Bereich der orts- und kontextbezogenen Anwendungen entwickeln soll. Die vorliegende Arbeit entstand im Rahmen dieses Projekts.

Da ortsbezogene Applikationen in erster Linie geographische Daten benötigen, ist es für Nexus von zentraler Bedeutung, existierende Geodatenbestände in die Infrastruktur einbringen zu können. Gegenüber dem World Wide Web besteht für Nexus eine zusätzliche Anforderung darin, Daten aus unterschiedlichen Quellen gemeinsam verarbeiten zu können. Daher wurde die Plattform, anders als das WWW, als föderiertes System konzipiert, das sich den Benutzern gegenüber wie ein einzelnes homogenes Datenbanksystem verhält [Conrad 02b]. Eine gemeinsame Verarbeitung von heterogenen Ausgangsdaten wird in solchen Systemen u.a. dadurch erreicht, dass Daten aus bestehenden, lokalen Datenschemas auf ein übergeordnetes, globales Datenschema abgebildet werden können.

Die lokalen Schemas betrachten dabei häufig dieselben Sachverhalte der Realität aus unterschiedlichen Anwendungsperspektiven heraus. So können z.B. Straßen in unterschiedlichen Geodatenschemas einerseits für Navigationszwecke modelliert werden und sämtliche entsprechenden Attribute wie Abbiegeverbote oder Geschwindigkeitsbeschränkungen aufweisen. Zudem müssen hierfür sowohl die linienförmigen Geometrien als auch die topologischen Beziehungen der jeweiligen Fahrbahnen detailliert abgebildet werden. Andererseits können Straßen aber auch aus stadt- bzw. verkehrsplanerischer Sicht als Quelle für die Ausbreitung von Lärm- oder Schadstoffemissionen aufgefasst werden, die im Schema als Objekte mit flächenhafter Ausprägung anzulegen sind und Attribute wie beispielsweise den täglichen Kraftfahrzeug-Durchsatz enthalten. Beschreiben lokale Schemas die gleichen Objekttypen der Realwelt auf unterschiedliche Weise, so werden sie im Rahmen dieser Arbeit als *Mehrfachkonzeptionen* bezeichnet. Die Integration von Mehrfachkonzeptionen in ein globales Schema ist als Voraussetzung für die Realisierung der Nexus-Plattform anzusehen. Dementsprechend werden zentrale Aspekte dieses Prozesses im Rahmen der vorliegenden Arbeit behandelt.

Die angestrebte Offenheit des Nexus-Systems erfordert es, verschiedensten Anbietern zu ermöglichen, ihr Geodatenmaterial unabhängig voneinander in die Plattform einzuspeisen. Dabei ist zu beachten, dass ein und dieselben Realweltobjekte, wie z.B. die Geschwister-Scholl-Straße in 70174 Stuttgart/Deutschland, bei den Datenanbietern jeweils in unterschiedlichen digitalen Repräsentationen vorliegen können. Dies ergibt sich in erster Linie daraus, dass die Dateninstanzen auf der Basis der jeweiligen lokalen Schemas (Mehrfachkonzeptionen) erfasst wurden. Unterschiede zwischen mehrfach repräsentierten Objekten müssen jedoch nicht notwendigerweise auf Mehrfachkonzeptionen zurückzuführen sein, sondern können auch aus anderen Gründen entstehen, beispielsweise dadurch, dass Objekte zu unterschiedlichen Zeitpunkten digitalisiert wurden.

Sobald nun voneinander abweichende Repräsentationen innerhalb der Nexus-Plattform bereitgestellt werden, kommt es im Datenbestand der Infrastruktur zu Inkonsistenzen bzw. zur Problematik der *Mehrfachrepräsentationen*. Mehrfachrepräsentationen können entweder nahezu identisch sein oder aber komplett differieren, so dass allein aufgrund der Betrachtung der digitalen Instanzen nicht mehr entschieden werden kann, ob sie dasselbe Objekt der Realwelt beschreiben oder nicht. Dazwischen existieren Übergangsformen jeglicher Art.

Solange Mehrfachrepräsentationen lediglich im Datenbestand von Nexus vorhanden sind, ist dies an sich noch nicht problematisch. Erst wenn sie innerhalb der Infrastruktur gemeinsam verarbeitet werden müssen und die Widersprüche zwischen ihnen offen zu Tage treten, entstehen Probleme. Diese sind umso komplexer, je stärker die Inkonsistenzen in den Daten sind. Man versucht, die auftretenden Schwierigkeiten im Zuge von Integrationsprozessen zu lösen. Der Integrationsprozess besteht dabei aus zwei Schritten: Mehrfachrepräsentierte Objekte müssen zunächst identifiziert bzw. einander zugeordnet werden können. Daraufhin

sind sie auf der Basis des Zuordnungsergebnisses in eine konsistente Sicht zu transformieren. Die Integration von Mehrfachrepräsentationen steht im Zentrum der Forschungsbemühungen dieser Arbeit.

Es ist wichtig zu betonen, dass der soeben anhand von Nexus motivierte Forschungsbedarf auf den Gebieten „Mehrfachkonzeptionen“ und „Mehrfachrepräsentationen“ nicht Nexus-spezifisch ist. Vielmehr spielen die genannten Fragestellungen generell eine wichtige Rolle beim Aufbau von Geodateninfrastrukturen bzw. von interoperablen Geo-Informationssystemen (GIS). Insofern sind auch die Lösungsansätze, die im Rahmen dieser Arbeit entwickelt werden sollen, nicht auf die Verwendung in Nexus beschränkt, sondern können auf GIS im Allgemeinen übertragen werden.

1.2 Problemstellungen bei der Integration von Schemas und Instanzen

Bei der Integration von Mehrfachkonzeptionen bzw. von Schemas, die den gleichen Ausschnitt der Realwelt formal beschreiben, und bei der Integration von Mehrfachrepräsentationen bzw. von Instanzen, die dieselben Objekte der Realwelt im Computer darstellen, treten vielfach komplexe Probleme auf.

Auf *Schemaebene* besteht die Problematik darin, die unterschiedlichen Semantiken der einzelnen Schemas aufeinander abbilden zu können. Dies ist deshalb so komplex, weil sich die Auffassung von einem bestimmten Ding oder Sachverhalt der Realwelt in der Regel nur schwer formalisieren lässt und stark von der individuellen Betrachtungsweise des einzelnen bzw. eines bestimmten Fachkreises abhängt. Zudem können Bezeichnungen unterschiedlich interpretiert werden, da Symbole bzw. Wörter oft nicht eindeutig bzw. kontextabhängig deutbar sind. So können beispielsweise Attributnamen identisch sein, aber eine unterschiedliche Bedeutung haben. Im umgekehrten Fall können Feldbezeichnungen voneinander abweichen, aber semantisch übereinstimmen. Hinter einer Objektbeschreibung steckt häufig auch sehr viel implizites Wissen, das im fachspezifischen Erfahrungshintergrund desjenigen vorhanden ist, der das Schema erstellt, im Schema selbst aber nicht repräsentiert wird. Abweichungen in Schemas können zusätzlich aus der Tatsache herrühren, dass es verschiedene Techniken gibt, um Schemas zu erstellen. So kann ein Sachverhalt beispielsweise im einen Schema lediglich als Attribut eines Objektes festgehalten werden, während er in einem anderen als Objektklasse ausgewiesen wird. Außerdem können in Schemas auch Generalisierungen auftreten, wobei z.B. eine Thematik im einen Schema sehr detailliert über mehrere Attribute spezifiziert werden kann, während sie im anderen nur als ein allgemeines Attribut repräsentiert ist.

In den vergangenen Jahren versuchte man im Bereich der Schemaintegration durch den Einsatz von ontologiebasierten Systemen [z.B. Fonseca et al. 02], mittels derer semantische Konzepte auf einer Meta-Ebene beschrieben werden können, Fortschritte zu erzielen. Begrifflichkeiten müssen dabei zunächst definiert werden, um den Interpretationsspielraum zu minimieren, und implizites Wissen muss explizit gemacht werden. Daraufhin können Ähnlichkeiten von Schemas aufgedeckt werden. Auch im Datenbankbereich existieren verschiedene Ansätze zur Schemaintegration [Conrad 02b]. Dennoch sind viele Fragestellungen in diesem Forschungsbereich noch ungelöst.

Auf *Instanzebene* ergibt sich zunächst die Schwierigkeit, innerhalb der gegebenen Datenbestände jene Repräsentationen zu identifizieren, die dasselbe Realweltobjekt abbilden. In diesem Prozess, der als Zuordnung bezeichnet wird, müssen die Eigenschaften von Mehrfachrepräsentationen miteinander verglichen werden. Daraufhin kann der Ähnlichkeitsgrad zweier Objekte bestimmt werden. Vektorbasierte Geodaten bestehen in jedem Fall aus einer geometrischen Komponente. Der Objektgeometrie sind meist bestimmte thematische Informationen bzw. Attribute zugeordnet. In manchen Datenschemas wird zudem die Topologie zusätzlich abgespeichert. Das Problem besteht nun darin, dass zwischen vielen Mehrfachrepräsentationen – ähnlich wie auf Schemaebene – keine eindeutigen Beziehungen existieren, sondern die geometrischen, thematischen und topologischen Eigenschaften der verschiedenen Objektinstanzen mehr oder weniger stark voneinander abweichen.

Der wichtigste Faktor bei der Zuordnung von raumbezogenen Instanzen ist naturgemäß die Geometrie – jedes Realweltobjekt befindet sich an einer bestimmten geographischen Position. Alle Computerrepräsentationen dieses Objektes, die dessen Geometrie mit abspeichern, sind daher vor allen Dingen über ihre Lage im Raum miteinander in Beziehung zu bringen. Auch wenn die Koordinaten nicht identisch sind, so müssen die Repräsentationen doch zumindest räumlich benachbart sein, falls eine Korrespondenz besteht. Dieser Sach-

verhalt ermöglicht eine deutliche Einschränkung des Suchraumes (siehe Abbildung 1-1), um eine Effizienzsteigerung der rechenintensiven Zuordnungsprozesse zu bewirken. Allerdings ist eine genaue Abgrenzung dieses Suchfensters ebenfalls mit Problemen verbunden. Sie resultieren daraus, dass man im Vorfeld einer Identifikation mehrfach repräsentierter Objekte nicht wissen kann, wie stark die Positionen der potentiell zuordenbaren Objekte tatsächlich voneinander abweichen und man daher Gefahr läuft, bei der Beschränkung des Suchraumes mögliche Zuordnungskandidaten auszuschließen.

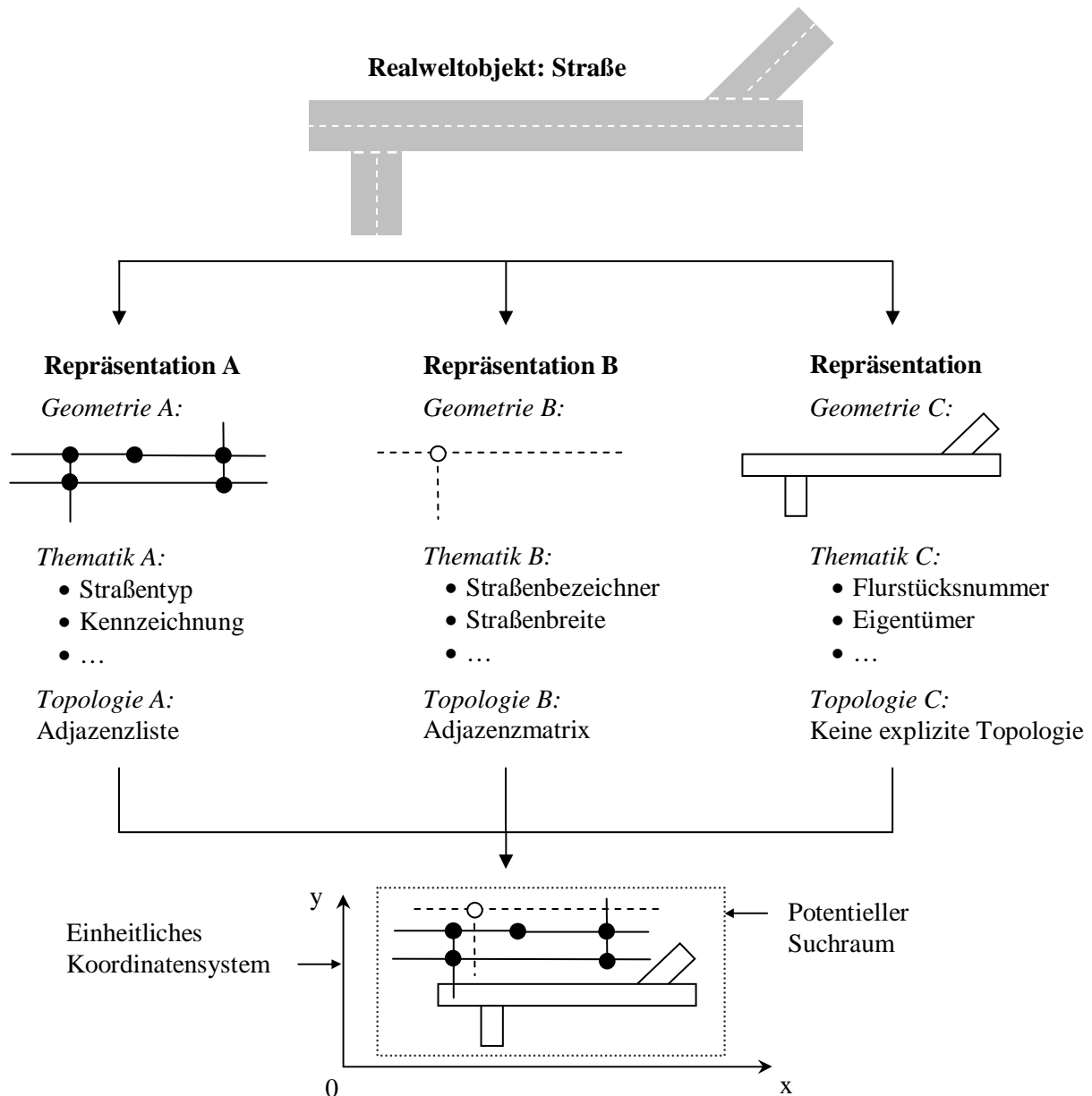


Abbildung 1-1: Mehrfache Repräsentationen eines Realweltobjektes in einem Geo-Informationssystem

Bei der Zuordnung von Mehrfachrepräsentationen anhand der geometrischen Ausprägung können unterschiedliche Probleme auftreten. Einige seien hier genannt:

1. Die Repräsentationen besitzen nicht den gleichen Objekttyp. Das Beispiel aus Abbildung 1-1 illustriert diese Problematik. So haben die Repräsentationen A und B eine linienhafte Geometrie, während Repräsentation C flächenförmig beschrieben ist.
2. Zwischen den Geometrien der verschiedenen Repräsentationen können Mehrfachzuordnungen auftreten. Abbildung 1-1 deutet eine solche Situation an, da alle Repräsentationen eine unterschiedliche Anzahl geometrischer Bausteine besitzen.
3. Die Repräsentationen weisen eine unterschiedliche geometrische Detailstufe auf.

Neben der Geometrie werden im Rahmen eines Zuordnungsprozesses die Attributwerte mehrfach repräsentierter Objekte miteinander verglichen. Die unterschiedliche Semantik der Sachdaten birgt dabei ein hohes Konfliktpotential. Prinzipiell entstehen auf der Ebene der Attributwerte, sofern es sich nicht um eindeutige numerische Ausdrücke handelt, naturgemäß die gleichen Probleme wie beim Vergleich der Semantik verschiedener Datenschemas. Dies sei hier noch einmal aufgegriffen, da die Problematik auch in Abbildung 1-1 illustriert wird: Repräsentation A besitzt ein Attribut „Kennzeichnung“, innerhalb dessen die Kennnummer der Straße eingetragen werden soll. Man könnte sich nun vorstellen, dass sich das Attribut „Straßenbezeichner“ der Repräsentation B ebenfalls auf diese Kennnummer bezieht. Um eine solche Vermutung verifizieren zu können, müsste man beispielsweise einen Vergleich konkreter Attributwerte verschiedener Instanzen durchführen. Wäre sowohl bei Repräsentation A als auch bei Repräsentation B aus obigem Beispiel der Eintrag bei „Kennzeichnung“ bzw. „Straßenbezeichner“ identisch (z.B. „B 27“), so lieferte dies einen Anhaltspunkt dafür, dass die beiden Feldbezeichnungen in semantischer Hinsicht identisch sind.

Für die Zuordnung korrespondierender Mehrfachrepräsentationen können auch topologische Relationen herangezogen werden. Probleme entstehen dann, wenn die Repräsentationen eine unterschiedliche Anzahl von Nachbarn haben oder die benachbarten Objekte einer Instanz von denen einer zweiten abweichen. Neben den topologischen Beziehungen bestehen auch weitere Relationen eines Objektes zu anderen Objekten im selben Datensatz. Häufig sind solche Beziehungen aber nicht explizit abgespeichert, weshalb man sie durch geeignete Methoden aus dem Forschungsgebiet des Data Mining [Anders 01] ableiten muss. Die Gewinnung solcher Informationen ist jedoch nicht trivial.

Sind Zuordnungsbeziehungen zwischen mehrfach repräsentierten Objekten einmal abgeleitet, so kann aus den sich überlappenden Ausgangsdaten *ein* topologisch kohärenter, blattschnittfreier Datenbestand erzeugt werden, auf dem die Anwendungen operieren und ihre Analysen und Abfragen durchführen können [Laurini 98]. Zu diesem Zweck werden die verschiedenen Repräsentationen eines Realweltobjektes zu einer resultierenden Repräsentation verschmolzen. Dabei ist es häufig schwierig zu entscheiden, welche Objekteigenschaften von welcher Repräsentation übernommen werden sollen. Insbesondere ist die Verschmelzung der Geometrie mit Problemen behaftet, wobei z.B. zu untersuchen ist, ob neue Geometrien aus den bestehenden erzeugt werden oder eine Referenzgeometrie übernommen werden soll. Im Zusammenhang mit der Verschmelzung von multiplen Instanzen spielen oft Fragen der Datenqualität eine Rolle, da beispielsweise die geometrische Genauigkeit oder die Aktualität einer Repräsentation Hinweise dafür liefern, welche Repräsentation die verlässlichsten Eigenschaften aufweist.

Erhebliche Probleme ergeben sich auch bei der Fortführung von Mehrfachrepräsentationen [z.B. Uitermark et al. 99a und 99b]. Es stellt sich die Frage, welche Auswirkungen die Änderung einer Repräsentation auf eine korrespondierende Repräsentation hat. Hier konnten nur Teillösungen gefunden werden, allgemein gültige Konzepte existieren bislang nicht. Die Problematik der Analyse von Mehrfachrepräsentationen in GIS gilt als Neuland der Forschung.

1.3 Zielsetzungen und Fragestellungen der Arbeit

Die hier vorgelegte Untersuchung steht in direktem Zusammenhang mit dem Nexus-Projekt der Universität Stuttgart, das die Entwicklung einer offenen und globalen Plattform für orts- und kontextbezogene Informationssysteme zum Ziel hat. Unter dieser Plattform ist eine föderierte, hochgradig verteilte Infrastruktur zu verstehen, die Daten und Dienste integriert, welche speziell für orts- und kontextbezogene Anwendungen von Bedeutung sind. Um den Applikationen eine einheitliche Sicht auf die Daten zu verschaffen und die Föderation der Daten zu ermöglichen, liegt der Plattform ein globales, generisches Datenschema, das so genannte Augmented World Schema (Angereichertes Weltschema, AWS), zugrunde.

Grundlegendes Ziel der vorliegenden Arbeit ist es, Teilaspekte für die Behandlung von Mehrfachkonzeptionen und Mehrfachrepräsentationen in Nexus zu untersuchen und Lösungen für diese zu finden. Die entwickelten Ansätze sollen dann prinzipiell auch auf andere raumbezogene Dateninfrastrukturen übertragbar sein. Die konkreten Fragestellungen, die im Zuge der Arbeit betrachtet werden, werden an dieser Stelle benannt:

Eine erste Aufgabenstellung besteht darin, zu untersuchen, wie man aus bestehenden Schemas zur Beschreibung von Geodaten, also aus Mehrfachkonzeptionen, ein globales Nexus-Schema ableiten kann. Daraufhin

müssen entsprechende Abbildungsvorschriften von Daten in Ursprungsschemas auf die Objektklassen des Nexus-spezifischen Schemas bzw. des Augmented World Schemas definiert werden. Es wird auch erörtert, welche Anforderungen aus der Sicht der Nexus-Anwendungen an das Augmented World Schema bestehen. Da eine umfassende Betrachtung sämtlicher Geoobjekte für das AWS den Rahmen der Untersuchung sprengen würde, erfolgt eine gezielte Beschränkung auf den Bereich des Straßenverkehrswesens. Dabei werden die folgenden Geodatenschemas untersucht:

- Geographic Data Files (GDF)
- Amtliches Topographisch-Kartographisches Informationssystem (ATKIS)
- Automatisierte Liegenschaftskarte (ALK)

Dieser Arbeitsschritt geschieht vor dem Hintergrund, dass eine Übertragung von bereits existierenden Daten in das gemeinsame, Nexus-spezifische Weltmodell möglichst einfach zu realisieren sein soll, um eine schnelle und umfassende Anreicherung der Plattform mit Datenmaterial zu gewährleisten.

Ist ein globales, generisches Datenschema vorhanden und können bereits vorhandene Daten in dieses transformiert werden, so wird es sich notwendigerweise ergeben, dass verschiedene Repräsentationen eines Realweltobjektes von unterschiedlichen Anbietern innerhalb von Nexus bereitgestellt werden und somit Mehrfachrepräsentationen auftreten. Bei Applikationsanfragen über einen bestimmten geographischen Ausschnitt kann es nun zu der Situation kommen, dass innerhalb des Anfragegebietes oder zumindest in Teil- bzw. Überlappungsbereichen eine gemeinsame Verarbeitung mehrfach repräsentierter Objekte notwendig ist (siehe Abbildung 1-2).

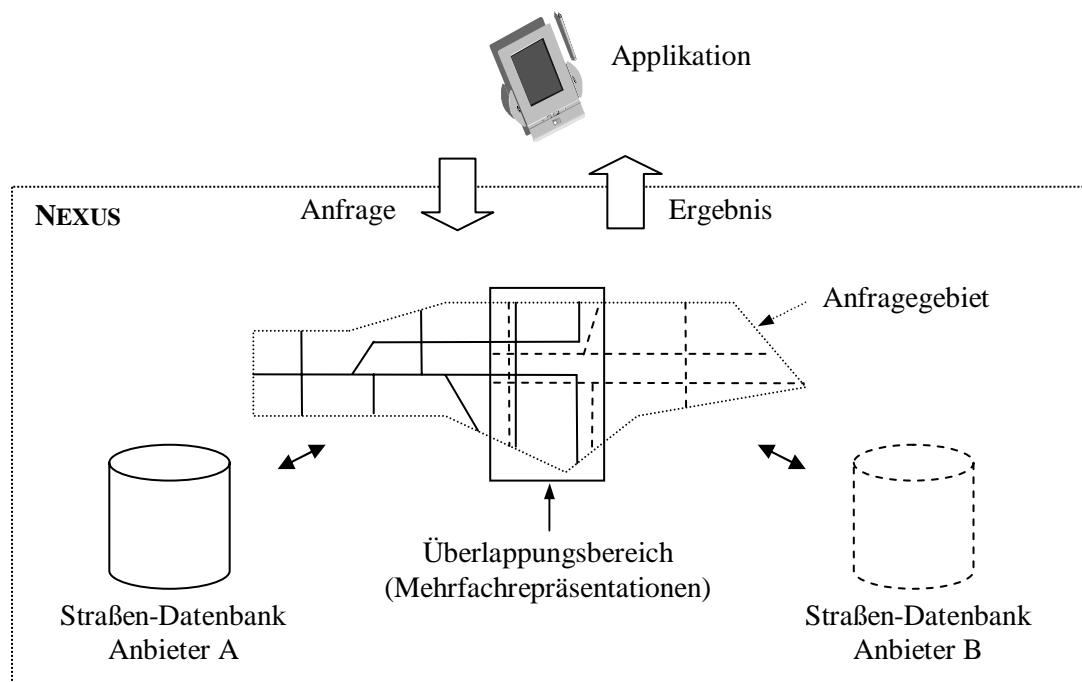


Abbildung 1-2: Innerhalb eines Anfragegebietes überlappen sich Daten verschiedener Anbieter; Objekte liegen mehrfach repräsentiert vor.

Ein zweiter Arbeitspunkt greift daher die Frage auf, welche Beziehungen zwischen Mehrfachrepräsentationen generell existieren können. Diese Beziehungen gilt es dann in einem Modell explizit und formalisiert auszudrücken, d.h. es soll eine zusätzliche Form der Geodatenmodellierung eingeführt werden: Neben der bisher bekannten Modellierung geometrischer, topologischer und thematischer Eigenschaften von Objekten sollen auch deren Relationen zu korrespondierenden Instanzen repräsentiert werden können. Dabei ist insbesondere zu untersuchen, welche Maße für die Beschreibung der Ähnlichkeit von raumbezogenen Objekten verwendet werden können. Anhand konkreter Daten aus dem Bereich des Straßenverkehrswesens soll der Ansatz auf seine Funktionsfähigkeit hin überprüft und somit validiert werden.

In einem dritten Arbeitspunkt, der den Schwerpunkt dieser Arbeit darstellt, wird untersucht, für welche Zwecke sich formalisierte Beschreibungen von Beziehungen zwischen Mehrfachrepräsentationen nutzen lassen. Dabei soll einerseits herausgefunden werden, wie eine automatisierte Vorgehensweise zur Ableitung von Korrespondenzen zwischen Objektklassen verschiedener Schemas (bzw. zwischen Mehrfachkonzeptionen) unter Verwendung der Beziehungen auf Instanzebene entwickelt werden kann (siehe Abbildung 1-3) und inwiefern diese in der Lage ist, die manuelle Vorgehensweise zur Schemaintegration zu ergänzen bzw. zu optimieren.

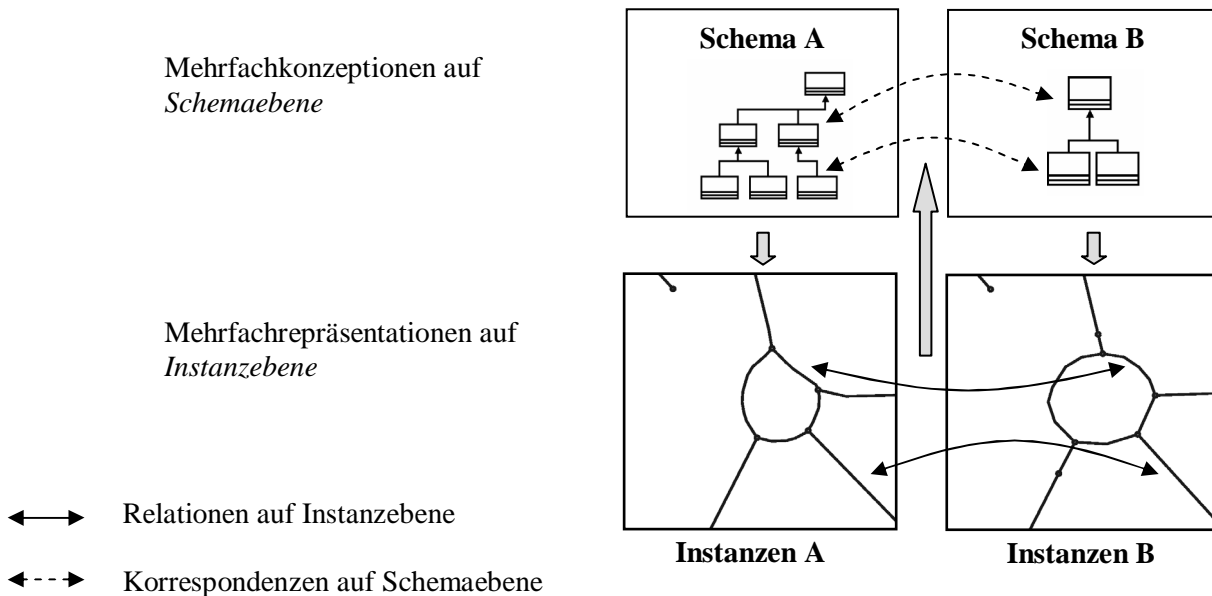


Abbildung 1-3: Aus Relationen zwischen Mehrfachrepräsentationen auf Instanzebene sollen Korrespondenzen zwischen Mehrfachkonzeptionen auf Schemaebene abgeleitet werden (nach [Volz & Walter 04]).

Andererseits liefern Relationen zwischen mehrfach repräsentierten Objekten die Grundlage dafür, eine Analyse auf mehrfach repräsentierten Daten sowie eine automatisierte Verschmelzung und Fortführung von Mehrfachrepräsentationen durchführen zu können. Diese Aspekte sind für die Effizienz der Nexus-Plattform von großer Bedeutung. Im Rahmen der Arbeit sollen die Möglichkeiten der Analyse von mehrfach repräsentierten Daten unter Nutzung ihrer Beziehungen näher betrachtet werden. Dabei wird am Beispiel einer Netzwerkanalyse untersucht, wie Navigation in mehrfach repräsentierten, aus Straßendaten abgeleiteten Graphen funktionieren kann und welche Probleme hier auftreten.

1.4 Aufbau der Arbeit

Im weiteren Verlauf der Arbeit werden zunächst die bislang existierenden Bestrebungen zur Interoperabilität bzw. zur Standardisierung von Geo-Informationssystemen beleuchtet und es werden die Zusammenhänge zwischen den Themenbereichen Interoperabilität und Mehrfachrepräsentationen aufgezeigt (Kapitel 2). Im Anschluss daran wird das Nexus-Projekt vorgestellt, da es den Rahmen für die vorliegenden Untersuchungen bildet (Kapitel 3).

In Kapitel 4 wird zuerst der Begriff „Mehrfachrepräsentationen“ näher beleuchtet, da dessen Verständnis von grundlegender Bedeutung für die vorliegende Arbeit ist. Danach erfolgt eine Aufarbeitung des aktuellen Stands der Forschung bezüglich der Integration heterogener und mehrfach repräsentierter Geodatenbestände.

Kapitel 5 setzt sich mit den Fragestellungen auseinander, wie ein globales Schema für Nexus erzeugt und wie eine Transformation bestehender Daten in dieses realisiert werden kann. Die Entwicklung des Modells zur Abbildung von Relationen zwischen Mehrfachrepräsentationen ist Gegenstand von Kapitel 6. In Kapitel 7 wird das entwickelte Modell auf zwei Testgebiete angewandt, in denen Mehrfachrepräsentationen von Straßenobjekten vorliegen. Im Zuge dessen werden Relationen zwischen den mehrfach repräsentierten Daten erzeugt und analysiert.

Kapitel 8 beschäftigt sich mit der Auswertung der erzeugten Relationen. Zum einen wird untersucht, wie aus diesen Hinweise bezüglich der Ähnlichkeit von Schemas extrahiert werden können. Zum anderen wird die Frage erörtert, wie eine Analyse von Mehrfachrepräsentationen unter Nutzung der Relationen durchgeführt werden kann. Dies geschieht am Beispiel einer Netzwerkanalyse.

Abschließend werden die erhaltenen Ergebnisse diskutiert und in Beziehung zu den Anforderungen von Nexus gesetzt (Kapitel 9). Am Ende der Arbeit steht ein Ausblick auf zukünftige Forschungsaufgaben (Kapitel 10).

2 Interoperabilität und Standards in GIS

Unter „Interoperabilität“ kann u.a. die Fähigkeit eines Geo-Informationssystems verstanden werden, heterogene Daten integrieren und gemeinsam verarbeiten zu können. Insofern ist die Problematik der Mehrfachrepräsentationen als ein Teilgebiet des Forschungsbereiches der Interoperabilität zu betrachten. In internationalen Gremien wird versucht, die Schaffung interoperabler Systeme über die Einführung von Standards zu realisieren. Das Thema Mehrfachrepräsentationen findet dabei allerdings noch wenig Beachtung. Wie gezeigt wird besitzen die bisherigen Standardisierungsbestrebungen dennoch eine gewisse Relevanz bezüglich der Integration mehrfach repräsentierter Datenbestände.

Dieses Kapitel soll erläutern, aus welchen Gründen sich die Bemühungen zur Interoperabilität von GIS entwickelt haben und was Interoperabilität genau bedeutet. Außerdem wird ein Überblick über die zentralen Akteure im Umfeld der Interoperabilitätsbemühungen vermittelt, und es werden die für die vorliegende Arbeit relevanten Standardisierungsansätze vorgestellt. Schließlich wird nochmals resümierend auf den Zusammenhang zwischen „Interoperabilität“ und „Mehrfachrepräsentationen“ eingegangen und es wird erläutert, weshalb die Standardisierungen offizieller Institutionen zwar notwendig, für das Forschungsgebiet der Mehrfachrepräsentationen jedoch nicht ausreichend sind.

2.1 Entwicklung der Bestrebungen zur Interoperabilität

Ginge man von idealen Bedingungen im Bereich der Geoinformatik aus, so existierte ein globaler, einheitlicher, konsistenter und hochdetaillierter (d.h. möglichst im Maßstab 1:1 vorliegender) Gesamtdatensatz als Abbildung der physischen Welt, innerhalb dessen Änderungen der Realwelt stets automatisiert nachgeführt würden. Dieser Datensatz wäre für sämtliche raum- bzw. ortsbezogenen Anwendungen geeignet bzw. über standardisierte Transformationsregeln an die Applikationsanforderungen anpassbar, d.h. jeder könnte sich seine gewünschte Repräsentation individuell generieren und es gäbe keine Mehrfachrepräsentationen im eigentlichen Sinne. Die Daten des Idealdatensatzes könnten über ein Standardformat ausgetauscht werden und stünden somit allen potentiellen Benutzern zur Verfügung. Zu dessen Bearbeitung würden Operationen mit standardisierten Schnittstellen angeboten. Ein solch ideales Szenario ist jedoch als vollkommen unrealistisch einzustufen.

Vielmehr entwickelten sich aus den verschiedensten GIS-Anwendungsbereichen heraus ganz unterschiedliche Datenschemas und Datenformate sowie entsprechend unterschiedliche Softwaresysteme mit unterschiedlichen Methoden zur Speicherung, Verwaltung, Analyse und Präsentation von Geodaten. Folglich kam es mit der Zeit zu einer sehr heterogenen Geoinformations-Landschaft, die von starken Barrieren zwischen den einzelnen Systemen geprägt war [Reinhardt et al. 04]. In früheren Jahren, als Geo-Informationssysteme lediglich Insellösungen bzw. in sich abgeschlossene, monolithische Systeme darstellten [Bishr 98, Timm et al. 98], spielten die Anstrengungen zur Überwindung dieser Barrieren bzw. zur Schaffung von Interoperabilität eine untergeordnete Rolle. Damals erfasste jede Anwendung die für sie relevanten Daten selbst, ein Austausch von Daten kam nicht oder nur in seltenen Fällen zustande. Zudem waren generell nur wenig raumbezogene Daten verfügbar. Aufgrund der Tatsache, dass die Datenerfassung für GIS mit hohen zeitlichen und somit finanziellen Aufwendungen verbunden ist und Geo-Informationssysteme inzwischen in vielen Anwendungsbereichen von Bedeutung sind, ergibt sich aber zunehmend der Wunsch, existierende Datenbestände sowie funktionale Komponenten austauschen und Geoinformationen gemeinsam nutzen und fortführen zu können [Bishr 98]. Dieser Wunsch hat vor dem Hintergrund der stetig steigenden Vernetzung digitaler Systeme und der Tatsache, dass die Aufgaben, die mit Hilfe von Geo-Informationssystemen bearbeitet werden sollen, immer interdisziplinärer werden und daher oft nur unter Einbeziehung unterschiedlicher Geodaten und verschiedenster Verarbeitungsmethoden zu lösen sind, eine wachsende Bedeutung bekommen.

Aus dieser Situation heraus entstanden in den vergangenen Jahren vermehrt Bestrebungen, die die Standardisierung bzw. die Interoperabilität von GIS zum Ziel haben. Schließlich ist es volkswirtschaftlich betrachtet wenig sinnvoll, dass einzelne Informations-Mosaiksteine zwar existieren, sie aufgrund technischer Gegebenheiten aber nicht zu einem Gesamtbild zusammengefügt werden können. Die Notwendigkeit, in diesen Bereichen Abhilfe zu schaffen und eine Effizienzsteigerung im Anwendungsbereich zu bewirken, war für

alle Beteiligten ersichtlich. Das zentrale Gremium, das sich zur Umsetzung dieser Zielsetzung auf internationaler Ebene gebildet hat, ist das OpenGIS Konsortium (OGC). Daneben gibt es die offiziellen Standardisierungsorganisationen wie beispielsweise die International Standards Organisation (ISO), die sich in speziellen Kommissionen derselben Aufgabe verschrieben haben. Für die Standardisierung der amtlichen Daten im nationalen Bereich ist die Arbeitsgemeinschaft der Vermessungsverwaltungen der BRD (AdV) verantwortlich.

2.2 Bemerkungen zur Interoperabilität

Die Realisierung von Interoperabilität soll dem Anwender von Geo-Informationssystemen einen deutlichen wirtschaftlichen Nutzen einbringen, da hierdurch die Effizienz des Zugriffs auf und der Verarbeitung von raumbezogenen Daten erhöht wird. Was bedeutet nun Interoperabilität genau? Die ISO legt in der Norm 19118 folgende Definition vor [Bartelme 00]:

„Interoperabilität ist die Fähigkeit zur Kommunikation, zur Ausführung von Programmen und zum Austausch von Daten zwischen verschiedenen funktionalen Einheiten in einer Art und Weise, die von Anwendern wenige oder gar keine Kenntnisse über die Besonderheiten dieser Einheiten erfordert.“

[Bill 99b] definiert Interoperabilität folgendermaßen:

„Interoperabilität bezeichnet die Möglichkeit, verschiedenartige Daten in einen einzelnen Arbeitsablauf zu integrieren. Dies setzt voraus, dass Syntax und Semantik der Daten dem Anwender in einheitlicher Form zur Verfügung gestellt werden. Interoperabilität erlaubt den transparenten Zugang zu mehreren raumbezogenen Daten- und Verarbeitungsressourcen innerhalb eines einzigen Arbeitsablaufes, ohne sie in einen Datenbestand zu überführen.“

Das OGC orientiert sich prinzipiell an der Definition der ISO. In den über die Webseite zugänglichen Dokumenten bzw. Spezifikationen [OpenGIS 04a] werden darüber hinaus drei grundlegende Charakteristika interoperabler Systeme bzw. Systemkomponenten betont:

1. Geodaten können über gemeinsame Schnittstellen systemübergreifend und möglichst ohne Informationsverlust ausgetauscht werden (Einheitlichkeit der Syntax),
2. Geodaten können über standardisierte Funktionen bearbeitet und analysiert werden. Auch die Funktionen können zwischen verschiedenen Systemen bzw. Plattformen ausgetauscht werden (Einheitlichkeit der Funktion),
3. Geodaten können zwischen verschiedenen Nutzergemeinden trotz unterschiedlicher Vokabulare für die Datenbeschreibung transferiert werden (Einheitlichkeit der Semantik).

Aus den obigen Definitionen wird ersichtlich, dass grundsätzlich zwischen technischer und semantischer Interoperabilität zu unterscheiden ist. Die technische Interoperabilität setzt sich dabei mit Daten- und Funktionsschnittstellen auseinander, während die semantische Interoperabilität eine Übersetzbarkeit des Dateninhalts einfordert. Standardisierungsbestrebungen kümmern sich traditionell vornehmlich um die technische Seite von Interoperabilität, also um Betrachtungen zu offenen Funktions- und Datenschnittstellen und Austauschsprachen. Immer mehr gewinnt jedoch der semantische Aspekt der Interoperabilität an Gewicht, bei dem die unterschiedliche Bedeutung bzw. der unterschiedliche Sinn von Objekteigenschaften im Mittelpunkt steht. Das OGC hat hierzu lediglich erste Studien in die Wege geleitet (vgl. Kapitel 2.4). Es wird dabei hauptsächlich versucht, den Abgleich von Begrifflichkeiten zwischen verschiedenen Anwendungsbereichen zu realisieren.

Nach Ansicht des Verfassers kann unter semantischer Interoperabilität im engeren Sinne durchaus auch der Sachverhalt verstanden werden, dass mehrfache digitale Repräsentationen, die durch Abweichungen im Datenschema bzw. in der Objektbildung zustande kommen und insofern semantisch unterschiedlich sind, gemeinsam verarbeitet werden können. Der Aspekt der semantischen Interoperabilität wird von den Standardisierungsorganisationen aber bislang nur ansatzweise aufgegriffen. Dementsprechend äußern sich auch [Harvey et al. 99]: *„While the OpenGIS consortium has made important progress on technical interoperabil-*

ity, semantic interoperability still remains an unpassed hurdle for efforts to share geographic information across organizational and institutional boundaries at the local, regional, and other levels.”

2.3 Überblick über die Standardisierungsorganisationen

Beim OpenGIS Konsortium (OGC) handelt es sich um ein internationales Industrie-Konsortium, dem Firmen, Regierungsbehörden sowie Universitäten angehören. Die Vorschläge, die vom OGC erarbeitet werden, werden in der Regel von der International Standards Organization (ISO) als offizielle bzw. als De-jure-Standards umgesetzt. Umgekehrt übernimmt das OGC auch die Inhalte der ISO. Die ISO widmet sich dabei vor allem in ihrer Technical Commission 211 GIS-relevanten Fragestellungen. Die Technical Commission ist in diverse Arbeitsgruppen, so genannte Working Groups, unterteilt, die sich mit speziellen Themen auseinandersetzen; z.B. gibt es die Working Group 4 „Geospatial Services“ oder die Working Group 8 „Location-Based Services“. Auch die ISO TC 204, die den Titel “Intelligent Transport Systems” trägt und sich mit der Normierung von Daten aus dem Bereich Fahrzeugnavigation beschäftigt, ist inhaltlich gesehen für das GIS-Umfeld von Bedeutung. ISO und OGC arbeiten seit 1997 sehr eng zusammen und sind auch personell verflochten.

Auf europäischer Ebene spielt neben der ISO das Comité Européen de Normalisation (CEN) eine Rolle auf dem Gebiet der Geodaten-Normierung. Das CEN kooperiert auf der Basis des Wiener Abkommens mit der ISO [Reinhardt et al. 04] und orientiert sich stark an den dort realisierten Standards. Die Erstellung von Geodatennormen findet hier in der TC 287 statt. Die Entsprechung der ISO TC 204 ist die TC 278 „Road Transport and Traffic Telematics“ der CEN. Die Festsetzungen der CEN werden wiederum vom Deutschen Institut für Normung (DIN) übernommen, welches also selbst keine inhaltliche Arbeit durchführt.

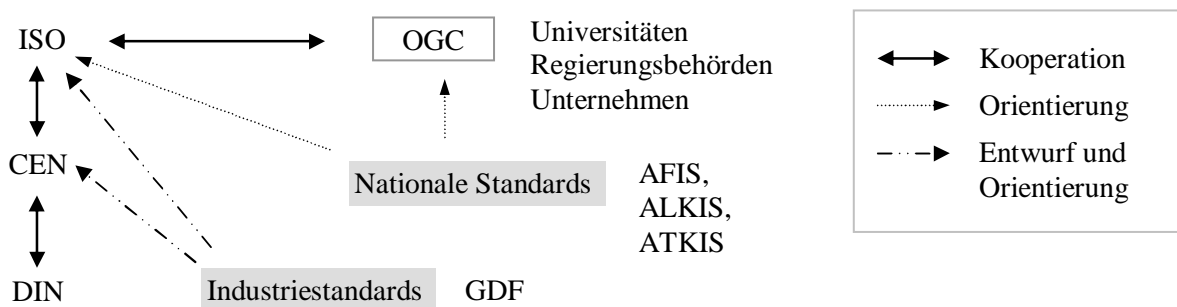


Abbildung 2-1: Zusammenspiel verschiedener Standardisierungsorganisationen

In Deutschland haben darüber hinaus die Vermessungs- und Katasterverwaltungen der Bundesländer die Aufgabe, Geobasisdaten für Verwaltung, Wirtschaft und private Nutzer zu liefern. Durch die Entwicklung der Automatisierten Liegenschaftskarte (ALK), des Automatisierten Liegenschaftsbuchs (ALB) und des Amtlichen Topographisch-Kartographischen Informationssystems (ATKIS) wurden somit ebenfalls Standards für amtliche Geodaten geschaffen. Diese Konzepte stammen jedoch noch aus den 70er bzw. 80er Jahren und haben keinen Bezug zu den Standards der offiziellen Organisationen. Aktuell wird aber versucht, ein gemeinsames Referenzmodell zu realisieren, um die Standards der Geobasisdaten mit jenen der ISO bzw. des OGC in Einklang zu bringen (vgl. Kapitel 2.5).

Neben den genannten Normierungsansätzen wird auch aus bestimmten Anwendungsbereichen heraus in Form von industriellen Kooperationen versucht, so genannte De-facto-Standards zu schaffen, die allein durch ihren hohen Verbreitungsgrad die Stellung eines Quasi-Standards erreichen. So entstand beispielsweise das Geographic Data Files (GDF) Format als Grundlage für die Fahrzeugnavigation. Es wurde in einem ersten Entwurf u.a. von den beiden Firmen Bosch (Deutschland) und Philipps (Niederlande) im Rahmen des EUREKA-Projekts DEMETER entwickelt [Walter 97] und inzwischen auch als CEN- bzw. ISO-Standard umgesetzt. In Abbildung 2-1 werden die Zusammenhänge zwischen den Standardisierungsorganisationen skizziert.

2.4 Organisation und Ansatz des OpenGIS Konsortiums (OGC)

Das OpenGIS Konsortium (OGC) wurde im Jahre 1994 gegründet [OpenGIS 01b]. Seine Mitgliederzahl wuchs seither kontinuierlich von zunächst 8 Gründungsmitgliedern auf heute 320 Mitglieder (Stand: August 2006) an. Die oberste Instanz des OGC ist das „OGC Board of Directors“. Dieses Führungsgremium bespricht sich zur Festlegung der längerfristigen OGC-Strategie mit dem „Strategic Member Advisory Committee“, einem Beraterkreis, der sich aus den so genannten strategischen, also besonders einflussreichen, Mitgliedern zusammensetzt. Das strategische Beraterkomitee kooperiert seinerseits vor allem mit dem Planungs-Komitee, welches gemeinsam mit dem Technischen Komitee im Rahmen des OGC Spezifikationsprogramms (vgl. Kapitel 2.4.1) an einem Konsens arbeitet, der sich letztlich in der Realisierung von Standardisierungsrichtlinien manifestiert. Neben dem Spezifikationsprogramm (SP) existiert seit 1999 das Interoperabilitätsprogramm (IP, vgl. Kapitel 2.4.2) des OGC. Es wurde mit dem Ziel eingerichtet, die Entwicklung von Spezifikationsentwürfen zu beschleunigen, indem ein entsprechendes technisches und organisatorisches Umfeld geschaffen wird. Die verschiedenen Initiativen werden dabei von einem so genannten IP-Management-Team koordiniert, das seine Vorgaben an die Initiative Management Teams (Init. Mgmt. Teams), welche die einzelnen Initiativen leiten, richtet. Das OGC regt seine Mitglieder darüber hinaus dazu an, weitere Anforderungen für die Realisierung von Interoperabilität zu spezifizieren und diese innerhalb des Forums zur Diskussion zu stellen. Im Rahmen des „Outreach and Community Adoption“-Programms wird ferner versucht, die Konzepte des OGC publik zu machen und Aufklärungs- bzw. Ausbildungsarbeit zu leisten, um eine weitreichende Akzeptanz und Unterstützung der Spezifikationen zu gewährleisten. Jeder der drei Programmsäulen steht ein „Executive Director“ vor, der seine Weisungen vom Board of Directors empfängt (siehe Abbildung 2-2). Die hier gemachten Angaben sind alle den Webseiten des OGC [OpenGIS 04b] entnommen.

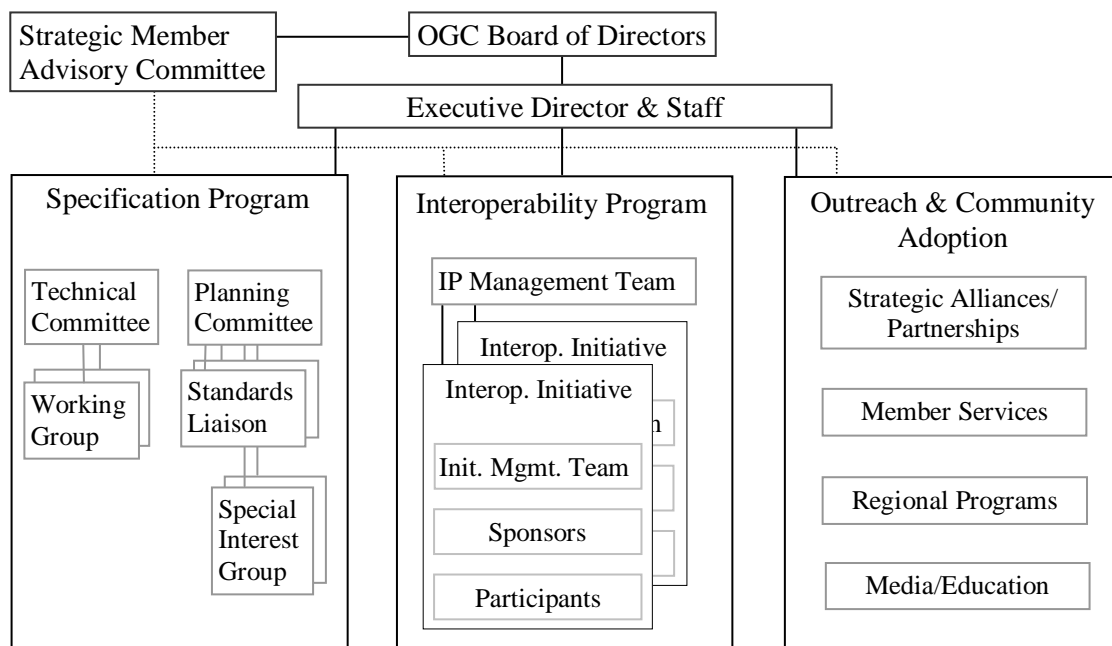


Abbildung 2-2: Organisation des OGC (nach [OpenGIS 04b])

Die Vision des OGC besteht darin, dass jeder über verschiedenste Netzwerke und Plattformen hinweg auf geographische Informationen zugreifen und sie für seine individuellen Zwecke nutzen kann. Der Beitrag des Konsortiums soll dabei vor allem in der Bereitstellung offener Schnittstellen für den Zugriff auf und die Verarbeitung von Geodaten liegen [OpenGIS 98].

Im weiteren Verlauf werden einige wichtige Konzepte des OGC stellvertretend für sämtliche Standardisierungsorganisationen beleuchtet, da innerhalb dieses Gremiums ein wesentlicher Teil der inhaltlichen Arbeit bezüglich der Schaffung interoperabler Systeme im GIS-Sektor geleistet wird und sich die Standardisierungsbestrebungen der verschiedenen Organisationen ohnehin gleichen bzw. parallel und in gegenseitiger Abstimmung ablaufen. Zunächst wird dabei ein kurzer Überblick über jene Arbeiten innerhalb des Spezifika-

tions- und des Interoperabilitätsprogramms des OGC gegeben werden, die einen Bezug zur vorliegenden Untersuchung haben. Eine komplette und umfassende Darstellung der Aktivitäten des OGC erscheint für den Kontext dieser Arbeit jedoch nicht notwendig, da die Lösungen, die durch bisherige Standardisierungen angeboten werden, nur in begrenztem Maße für die Forschungsproblematik der Mehrfachrepräsentationen von Bedeutung sind.

2.4.1 Das Spezifikationsprogramm des OGC

Für das OGC Spezifikationsprogramm [OpenGIS 02] sind in erster Linie das Technische und das Planerische Komitee verantwortlich (vgl. Abbildung 2-2). Das Technische Komitee wurde bereits bei der Gründung des OGC im Jahre 1994 eingerichtet. Bis zum Jahre 1999, in dem das Interoperabilitätsprogramm ins Leben gerufen wurde, entwickelte das Technische Komitee die OpenGIS-Spezifikationen in so genannten „Special Interest Groups“ und in „Working Groups“. Zunächst wurden dabei die Abstrakten Spezifikationen („Abstract Specifications“), die eher ein konzeptionelles Modell liefern, entworfen. Gegen Ende der 90er Jahre wurden auf der Basis der „Requests for Proposal (RFP)“, also Anfragen zur Konkretisierung, dann auch Implementierungsspezifikationen („Implementation Specifications“) zu bestimmten Bereichen bereit gestellt. Die Implementierungsspezifikationen legen konkrete Realisierungsrichtlinien fest. Seit 1999 werden die Spezifikationen nun hauptsächlich durch das Interoperabilitätsprogramm entworfen.

Im Rahmen des Spezifikations-Prozesses hat das Planungs-Komitee darauf zu achten, dass die Entwicklung der Spezifikationen auch stets den Anforderungen des Marktes gerecht wird. Das Komitee legt die weitere Vorgehensweise innerhalb des Programms daher in einer so genannten „Technology Roadmap“ fest, einer Art Fahrplan, anhand dessen die zukünftigen Schritte ersichtlich und für Außenstehende nachvollziehbar sind.

Sowohl die Abstrakten als auch die Implementierungs-Spezifikationen werden über das Internet öffentlich verfügbar gemacht [OpenGIS 04a]. Einige der im Kontext dieser Arbeit interessanten Spezifikationen sollen in der Folge präsentiert werden. Dies geschieht, um zu zeigen, welche standardisierte Struktur sämtliche Objekte, also auch alle potentiellen Mehrfachrepräsentationen sowie die Relationen zwischen ihnen, gemäß den Auffassungen des OGC aufweisen müssen.

2.4.1.1 Abstrakte Spezifikationen

Die abstrakten Spezifikationen des OGC versuchen, konzeptionelle Modelle zu entwickeln, auf deren Basis Implementierungs-Spezifikationen definiert werden können. Eine abstrakte Spezifikation besteht grundsätzlich aus zwei Modellen (siehe z.B. [OpenGIS 99b]):

1. dem „Essential Model“, das eine Verknüpfung von Realwelt und Software-System bzw. Systemdesign herstellen soll und beschreibt, wie sich die Vorgänge in der Realwelt abspielen;
2. dem „Abstract Model“, das das zu entwickelnde Softwaresystem auf eine implementierungs-neutrale Art und Weise definiert und beschreibt, wie die Software funktionieren soll.

Derzeit (Stand: September 2004) existieren 17 abstrakte Spezifikationen zu verschiedenen so genannten „Topics“, die jedoch verschiedene Reifegrade aufweisen. Manche befinden sich erst in einem Anfangsstadium, andere sind bereits sehr fundiert und können als Basis für Implementierungs-Spezifikationen herangezogen werden. Im Folgenden sollen die für diese Arbeit relevanten abstrakten Spezifikationen kurz vorgestellt werden.

Das OpenGIS Feature

Das zentrale Element innerhalb der OGC Gedankenwelt ist das so genannte „Feature“ [OpenGIS 99a]. Ein Feature im Sinne der OpenGIS Spezifikation ist eine Abstraktion eines Realweltphänomens. Weist ein Feature eine Lokation auf, die relativ zur Erde definiert werden kann, so handelt es sich um ein geographisches Feature. Der Begriff „Feature“ wird im Deutschen (von der Arbeitsgemeinschaft der Vermessungsverwaltungen der BRD, AdV) auch als „Fachobjekt“ übersetzt. Innerhalb dieser Arbeit werden die beiden Bezeichnungen ebenfalls synonym benutzt.

Eine Abbildung der Realwelt kann über eine Menge von Features erfolgen. Der Zustand eines Features ist mittels einer gewissen Anzahl an Eigenschaften festzulegen, wobei jede Eigenschaft als {Name, Typ, Wert}-Tripel ausgedrückt werden kann. Die Anzahl von Eigenschaften sowie deren Typen und Namen wird durch die Typdefinition eines Features, den so genannten Feature-Type, reglementiert. Raumbezogene Features haben geometrische Eigenschaften und können zusätzlich auch topologisch strukturiert sein sowie andersartige Relationen unterhalten. Wenn mehrere Features zu einem neuen Feature zusammengefasst werden, spricht man von Feature Collections. Auch Feature Collections haben einen bestimmten Feature-Typ und entsprechende Eigenschaften, zusätzlich zu jenen der enthaltenen Features. Features zeichnen sich ferner dadurch aus, dass sie über gewisse Qualitätskriterien hinsichtlich ihrer Genauigkeit kategorisierbar sind. Sie können über Objektidentifikatoren eindeutig referenziert werden.

Features können grundsätzlich eine unterschiedliche Granularität aufweisen. Ein Feature kann beispielsweise ein einzelnes Pixel eines georeferenzierten Satellitenbildes oder aber das gesamte Satellitenbild selbst sein. Auf Vektorebene kann ein einzelnes Straßensegment oder aber ein Straßenobjekt aus beliebig vielen Teilabschnitten als Feature interpretiert werden. Die inhaltliche Beschreibung bzw. die Semantik von raumbezogenen Features ist stark von der Informationsgemeinschaft abhängig, die die Features im Rahmen ihrer Anwendungen benützen möchte. Eine Informationsgemeinschaft zeichnet sich dadurch aus, dass sie die Realwelt aus einer ganz bestimmten Perspektive betrachtet. Sie wird in der OGC-Terminologie als „Geospatial Information Community“ bezeichnet. Die unterschiedliche semantische Auffassung von Features durch verschiedene Informationsgemeinschaften stellt eines der Hauptprobleme auf dem Gebiet der Interoperabilität bzw. der Mehrfachrepräsentationen dar.

Feature Geometrie und Feature Topologie

Vektorbasierte Features bestehen aus geometrischen und topologischen Primitiven, die getrennt oder in Kombination abgespeichert werden. In der abstrakten Spezifikation zur Feature Geometrie [OpenGIS 01a] werden die raumbezogenen Eigenschaften von Features sowohl in 2D als auch in 3D näher beschrieben und es werden relevante räumliche Operatoren zur Verarbeitung von Features festgelegt. Die Spezifikation ist identisch mit der ISO-Norm 19107 „Spatial Schema“. Sie ist in die folgenden drei Teile untergliedert:

- das Geometrie-Paket (Geometry Package),
- das Topologie-Paket (Topology-Package),
- die Topologischen Relationen (Topologic Relations)

Das Geometrie-Paket enthält die verschiedenen Klassen zur Abspeicherung der Koordinaten-Geometrien. Sie sind alle von der abstrakten Stammklasse *GM_Object* abgeleitet. Über die Stammklasse erben alle Unterklassen eine Assoziation zu einem Referenz-Koordinatensystem. Das Geometrie-Paket hat weitere interne Subpakete für primitive Geometrien (*GM_Primitive*), Aggregate (*GM_Aggregate*) und Komplexe (*GM_Complex*).

Instanzen von Subklassen der abstrakten Oberklasse *GM_Primitive*, so z.B. *GM_Point*, können nicht weiter in Teilobjekte zerlegt werden und stellen somit Basisobjekte dar. Geometrische Aggregate (z.B. *GM_MultiSurface*) werden benutzt, um gleichartige geometrische Primitive zu gruppieren, ohne dass diese Gruppierung eine interne Struktur aufweist. Geometrische Komplexe hingegen bestehen aus einer Ansammlung von strukturierten Primitiven. So sind z.B. Instanzen der Klasse *Composite_Curve* so definiert, dass sie die gleiche Richtung haben und aneinander anschließen müssen (vgl. Abbildung 2-3).

Für sämtliche Geometrie-Klassen werden auch die entsprechenden geometrischen Operationen als Schnittstellen definiert. So liefert beispielsweise die Methode *GM_Object::envelope()* : *GM_Envelope* das minimal umschließende Rechteck zurück. Abbildung 2-3 zeigt einen Teil der Klassen des Geometrie-Pakets mit deren Spezialisierungs-Relationen.

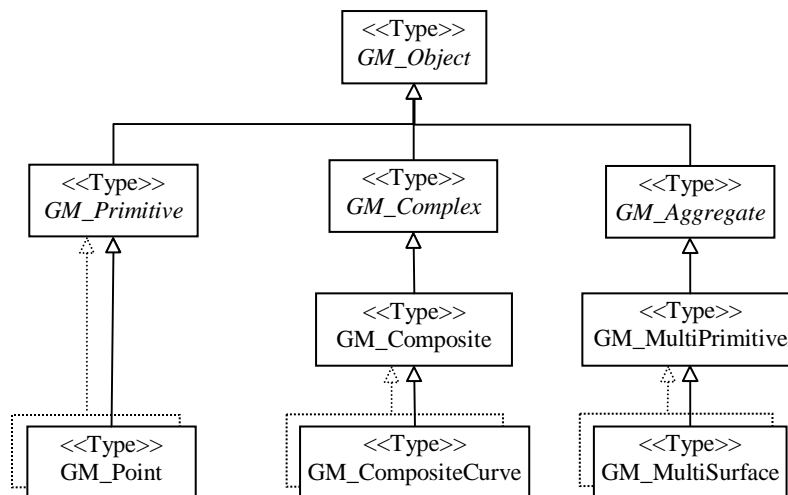


Abbildung 2-3: Ein Ausschnitt aus dem Klassengefüge des Geometrie-Pakets mit Spezialisierungs-Relationen

Das Topologie-Paket hat als Stammklasse `TP_Object`. Direkte Subklassen hiervon sind `TP_Primitive` und `TP_Complex`, wobei sich `TP_Complex`-Objekte aus `TP_Primitive`-Objekten aufbauen. `TP_Primitive` vererbt an `TP_Node`, `TP_Edge`, `TP_Face` und `TP_Solid`. Die Topologie-Klassen können unabhängig von den Geometrie-Klassen benutzt werden. Auch bei den jeweiligen Topologie-Klassen werden die notwendigen Operationen definiert. So liefert beispielsweise die Methode `TP_Face::boundary() : TP_FaceBoundary` die gerichteten Kanten zurück, die den Rand einer „`TP_Face`“ (topologische Masche) beschreiben.

Bei den Topologischen Relationen werden Operatoren definiert, die eine Abfrage topologischer Relationen ermöglichen. Zielsetzung ist es, dass unterschiedliche Implementierungen der Operatoren bei Datenbeständen gleichen Informationsgehalts auch identische Abfrageergebnisse liefern. Es werden

- Boolesche Operatoren,
- Egenhofer Operatoren (gemäß der 9-Intersection-Matrix) [Egenhofer 91],
- Clementini Operatoren [Clementini & Di Felice 94] und
- Kombinationen der genannten Operatoren

unterschieden.

Boolesche Operatoren drücken aus, ob die Schnittmenge zweier Objekte leer oder nicht leer ist. Es kann jedoch nicht zwischen Objektinnerem und Objektgrenze differenziert werden. Folgende Schnittbedingungen können für zwei Objekte A und B untersucht werden: Schnitt zwischen den Objekten (also Grenze plus Inneres, wird als „Closure“ bezeichnet) A und B, Schnitt zwischen Objekt A und Objektäußerem von Objekt B, Schnitt zwischen Objekt B und Objektäußerem von Objekt A, Schnitt zwischen dem Objektäußerem von Objekt A und B. Die Egenhofer-Operatoren erweitern dieses Modell auf 9 Schnittmöglichkeiten, da auch ein Test zwischen den Objektgrenzen möglich ist, wodurch 2^9 (also 512) potentiell mögliche topologische Relationen entstehen. Die Clementini-Operatoren erlauben eine feinere Differenzierung der möglichen Ergebnismenge. So kann man beispielsweise unterscheiden, ob ein Schnitt zweier Objekte eine Punktmenge, Punkte und Linien, etc. liefert.

Relationen zwischen Features

Die Objekte der Realwelt bzw. die Features existieren nicht isoliert voneinander. Oft sind die Beziehungen zwischen raumbezogenen Objekten bei der Analyse von besonderem Interesse. Dem gemäß hat das OGC auch eine abstrakte Spezifikation für Relationen zwischen Features herausgegeben [OpenGIS 99b], deren wichtigste Inhalte im Folgenden dargelegt werden. Das Relationskonzept von Nexus und damit auch die im Rahmen dieser Arbeit angestrebte Realisierung von Relationen zwischen Mehrfachrepräsentationen beruhen auf dieser OGC-Spezifikation (vgl. Kapitel 6.3.1.2).

Grundsätzlich können Relationen auf unterschiedliche Weise abgeleitet werden. Einerseits kann durch die Analyse von Attributen herausgefunden werden, wie Features miteinander in Beziehung stehen. Beispielsweise könnte man anhand eines identischen Attributes „Höhe über N.N.“ (bzw. über einen z-Wert) zweier sich kreuzender Objekte „Straße“ und „Bahnlinie“ herausfinden, welches Objekt das andere Objekt über- bzw. unterquert. Diese Art der Ableitung von Relationen würde auf der Interpretation von Attributen beruhen. Die Interpretationsregeln hierfür lassen sich im Allgemeinen jedoch nicht aus dem Datenschema ableiten und sind somit nicht automatisierbar. Aus diesem Grund müssen Relationen explizit formuliert werden. Dieser Aspekt ist auch von zentraler Relevanz für die Entwicklung des in dieser Arbeit verfolgten Konzeptes zur expliziten Modellierung von Relationen zwischen Mehrfachrepräsentationen (vgl. Kapitel 6).

Gemäß der OGC-Spezifikation geschieht eine explizite Formulierung von Objektbeziehungen über so genannte Relationships oder Relationen, die einen gewissen „Relationship Type“ (bzw. Relations-Typ) aufweisen. Features bzw. Instanzen der beiden Feature-Typen („Feature Types“) „Bahnlinie“ und „Straße“ könnte man z.B. über eine Relation des Relationstyps „Straße kreuzt Bahnlinie“ in Beziehung setzen. Die an der Relation teilnehmenden Objekte würden in diesem Fall als Attribute innerhalb des Relationsobjekts abgespeichert, d.h. die Objektinstanzen der Feature-Typen Bahnlinie und Straße würden über Attribute des betreffenden Relationsobjekts referenziert. Werden lediglich zwei Objekte über eine Relation verknüpft, so spricht man von einer binären Relation, deren Stelligkeit bzw. Grad („Degree“) 2 beträgt.

Die Art und Weise, wie die Straße die Bahnlinie kreuzt, ist in den Rollen (oder „Roles“) spezifiziert, z.B. „Straße führt über Bahnlinie“ und „Bahnlinie verläuft unter Straße“, die als Attribute der Relation festgehalten werden. Rollen drücken also die Funktionen aus, die Objekte innerhalb einer Relation haben. Die Rollen zweier Objekte, die in Beziehung stehen, können jedoch auch identisch sein, wie das z.B. bei Adjazenzbeziehungen der Fall ist. In diesem Fall spricht man von symmetrischen Relationen. Relationen zwischen Mehrfachrepräsentationen sind stets symmetrisch.

Bei den Relationen sind laut OGC so genannte leichtgewichtige (*lightweight*) und schwergewichtige (*heavyweight*) Exemplare zu unterscheiden. Die schwergewichtigen Relationen existieren ohne Abhängigkeit von den Features, die sie in Beziehung setzen, und haben folglich eine eigene Identität, während leichtgewichtige nur so lange bestehen, wie ihre Rollen nicht leer sind. Außerdem können schwergewichtige Relationen im Gegensatz zu leichtgewichtigen eigene Attribute haben. Schließlich sind leichtgewichtige Relationen auf einen Grad von 2 beschränkt, wogegen schwergewichtige einen Grad von n haben können. Sowohl bei leicht- als auch bei schwergewichtigen Relationen muss deren Gültigkeit über die Einführung von Integritätsbedingungen gesichert werden.

Datenqualität

Oft können Applikationen nur mit Daten arbeiten, die gewisse Qualitätskriterien erfüllen, um verlässliche Analyseergebnisse zu erzielen. Daher ist es notwendig, Daten Metainformationen über ihre Genauigkeit mitzugeben. Diese Metainformationen können dann von der Anwendung interpretiert werden und sie kann entscheiden, ob eine ausreichende Qualität gewährleistet werden kann. Beispielsweise wird bei Vektorgeometrien häufig die Positionsgenauigkeit der Koordinaten abgespeichert. Diesem so genannten „Truth in Labeling“-Ansatz folgt auch die OGC Spezifikation zur Datenqualität [OpenGIS 99c].

Im Essential Model werden zunächst die verschiedenen Bezeichnungen für Genauigkeit („Accuracy“) aufgeführt. Als absolute Genauigkeit wird beispielsweise die Fehlerschätzung für einen einzelnen Punkt relativ zum angegebenen räumlichen Referenzsystem definiert, als relative Genauigkeit der Fehler bei der Distanzmessung zweier Punkte bzw. die Genauigkeit von einem Punkt im Verhältnis zu einem anderen, usw.

Grundsätzlich erweitert die Spezifikation das Feature bzw. die Feature Collection um eine weitere Relation zu einem so genannten „Metadata Set“-Objekt. Innerhalb dieses Objekts werden mehrere Metadaten-Entitäts-Objekte abgespeichert, welche wiederum aus Metadaten-Elementen aufgebaut sind. Es gibt verschiedene Metadaten-Entitäts-Objekte für die verschiedenen Positionsgenauigkeiten wie absolute und relative Genauigkeit. Zusätzlich dazu kann die Genauigkeit von numerischen und nicht-metrischen (z.B. Zeichenketten) Attributwerten in Metadaten-Entitäts-Objekten abgelegt werden.

Die Abstrakte Spezifikation zur Datenqualität befindet sich in einem Frühstadium. Man hat sich dabei lediglich auf einzelne Aspekte wie die Positionsgenauigkeit bzw. die Genauigkeit der Geometrie beschränkt und weniger auf eine komplette Betrachtung des gesamten Themas Wert gelegt. Es wird angemerkt, dass man sich in diesem Gebiet Weiterentwicklungen durch die ISO TC 211 erhofft. Bei der Integration von Mehrfachrepräsentationen können Angaben über die Qualität der korrespondierenden Objekte hilfreich sein, weshalb an dieser Stelle kurz erwähnt wurde, welche Standardisierungsansätze in diesem Bereich bislang existieren.

Semantik und Informationsgemeinschaften

Informationsgemeinschaften können als Anwendergruppen aufgefasst werden, deren semantisches Verständnis von Realweltobjekten als weitgehend ähnlich oder identisch zu bezeichnen ist. Um auch Daten zwischen unterschiedlichen Informationsgemeinschaften austauschen zu können, wurde die Spezifikation zu „Semantics and Information Communities“ [OpenGIS 99d] aufgesetzt.

Die individuellen Sichten unterschiedlicher Informationsgemeinschaften auf die Realwelt werden „Projektwelten“ (Project Worlds) genannt. Jede Projektwelt hat ein bestimmtes Vokabular, eine gewisse Zusammensetzung von Schemas, eine bestimmte räumliche Ausdehnung und einen räumlichen Bezug über ein definiertes Referenzsystem. Sie besteht aus einer eindeutigen Ansammlung von Feature-Instanzen. Sobald eine Änderung an einem Schema durchgeführt wird, resultiert daraus auch eine unterschiedliche Projektwelt. Ein Beispiel hierfür wäre eine Modifikation eines Geometrieschemas in der Weise, dass eine Kurve, die zunächst über einen Polygonzug repräsentiert wurde, nun als Spline modelliert wird. Ähnliches könnte man sich auch für die Thematik von Objekten vorstellen, wenn beispielsweise Attributnamen geändert werden.

Das generelle Ziel des OGC besteht darin, den Informationsverlust beim Austausch von Features unterschiedlicher Projektwelten möglichst gering zu halten. Im Rahmen der Spezifikation wird daher ein abstraktes Modell entworfen, das die Funktionen für den Übergang zwischen unterschiedlichen Projektwelten definiert. Diese Funktionen sind in ihrem jetzigen Stadium jedoch so abstrakt, dass eine Relevanz für die vorliegende Arbeit nicht ersichtlich ist und folglich auf deren Erläuterung verzichtet werden kann. Grundsätzlich wären die Spezifikationen in diesem Bereich für Nexus aber von großem Interesse, da die Plattform Daten von unterschiedlichen Anbietern bzw. Informationsgemeinschaften und damit aus unterschiedlichen Projektwelten aufnehmen muss. Somit könnten diese Standardisierungen auch die Datenintegration in Nexus erleichtern.

2.4.1.2 Implementierungs-Spezifikationen

Die Implementierungs-Spezifikationen des OGC stellen spezifische Programmierregeln und Vorgaben für die Realisierung von Schnittstellen und Protokollen zur Verfügung, die die Interoperabilität zwischen raumbezogenen Informationssystemen ermöglichen. Es existieren bereits einige solcher Implementierungsspezifikationen. An dieser Stelle soll jedoch lediglich die Geography Markup Language dargestellt werden, da sie als Grundlage für die Abspeicherung von Objektgeometrien in Nexus dient und somit ein Teil des Nexus-Datenformates ist.

Geography Markup Language (GML)

Die Geography Markup Language (GML) ist eine auf der Grundlage der XML-Schema-Spezifikation [W3C 04] definierte Grammatik für die Modellierung, den Austausch und die Speicherung von raumbezogenen Daten [OpenGIS 03a]. Derzeit liegt sie in der Version 3.1.1 (Stand: August 2006) vor. Ihre Struktur basiert auf den Abstrakten Spezifikationen des OGC. Mittels GML kann alles repräsentiert werden, was ein Feature ausmacht, also z.B. dessen Geometrie, dessen topologische Beziehungen und sonstige Relationen, dessen Koordinatenreferenzsystem, etc. Auf der Grundlage von GML können raumbezogene Applikationsschemas abgeleitet werden, um anwendungsbezogene Objekte speichern und austauschen zu können (vgl. Abbildung 2-4).

GML besteht aus verschiedenen GML-Schemas, die jeweils spezielle Typen für bestimmte Eigenschaften eines Features liefern. So gibt es beispielsweise Schemas zur Topologie und zur Repräsentation von temporalen Aspekten eines Features sowie zur 3D-Modellierung etc. Eine Übersicht sämtlicher GML-Schemas ist der Spezifikation zu entnehmen [OpenGIS 03a].

Wie XML-Schema Dokumente im Allgemeinen definieren auch die GML-Schemas XML-Typen und XML-Elemente. Ein GML-Objekt wird dabei stets als XML-Element eines Typs repräsentiert, der entweder direkt oder indirekt von der abstrakten Klasse *AbstractGMLType* abgeleitet wurde. Auch die Eigenschaften der GML-Features werden in Form von Elementen beschrieben. Sie können ihrerseits wieder über XML-Attribute konkretisiert werden. Alle Elemente, die innerhalb von GML definiert werden, befinden sich im GML-Namensraum, der mit dem Präfix *gml* gekennzeichnet wird.

Wird eine Applikation entwickelt, welche GML-konform sein soll, so müssen die hierfür benötigten Features in einem Applikationsschema spezifiziert werden. Sämtliche Feature-Typen sind von dem abstrakten GML-Element *gml:AbstractFeatureType* abzuleiten, der im GML-Feature-Schema (feature.xsd) definiert wurde. Bei der Aufstellung von Applikationsschemas muss also in jedem Fall das Feature-Schema eingebunden werden. Ansonsten sind nur jene Schemas zu referenzieren, die von der jeweiligen Applikation verwendet werden. Sind nur einfache 0-, 1- oder 2-dimensionale Geometrien von Bedeutung, genügen folglich die beiden Schemas *geometryBasic0d1d.xsd* und *geometryBasic2d.xsd*. Nur für nicht-lineare oder 3-dimensionale Geometrien müssen weitere Geometrie-Schemas verwendet werden.

Außerdem gilt, dass die Typen, die in den GML-Schemas definiert sind, durch die Applikationsschemas auch eingeschränkt oder erweitert werden können. Wenn aus Applikationssicht keine Änderungen an den GML-Schemas notwendig sind, werden die instanzierbaren Elemente und Attribute direkt aus den Basis-schemas übernommen. Abbildung 2-4 zeigt einen Auszug aus einem GML-Instanzdokument, das applikationsspezifische Straßen-Features enthält. Es wird ersichtlich, dass neben GML-Features applikationsspezifische Features des Namensraumes „streetapplication“ vorhanden sind, die auch über eigene Attribute verfügen. Die Geometrie eines Straßenfeatures wird mittels GML-LineStrings repräsentiert.

```
<?xml version="1.0" encoding="UTF-8"?>
<streetapplication:ApplicationStreetModel xmlns:streetapplication="http://www.application.de/streetapplication"
xmlns:gml="http://www.opengis.net/gml" xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.application.de/streetapplication/
StreetApplication.xsd http://www.w3.org/1999/xlink schemas.opengis.net/gml/3.0.0/xlink/xlinks.xsd">

  <gml:description>Ein einfaches Beispiel für die Repräsentation von Straßen in GML</gml:description>
  <gml:name>Straßenszene</gml:name>
  <!-- Definition des geographischen Ausschnitts der Szene -->
  <gml:boundedBy>
    <gml:Envelope>
      <gml:pos>0.0 0.0</gml:pos>
      <gml:pos>1000.0 1000.0</gml:pos>
    </gml:Envelope>
  </gml:boundedBy>
  <!-- Beschreibung der Features -->
  <gml:featureMember>
    <streetapplication:Street gml:id="S1">
      <streetapplication:Street_name> Geschwister-Scholl-Straße </streetapplication:Street_name>
      <gml:centerLineOf>
        <gml:LineString gml:id="L1">
          <gml:coordinates>30.0, 40.0 20.0,90.0 40.0,80.0</gml:coordinates>
        </gml:LineString>
      </gml:centerLineOf>
    </streetapplication:Street>
  </gml:featureMember>
  <gml:featureMember>
    ...
  </gml:featureMember>
</streetapplication:ApplicationStreetModel>
```

Abbildung 2-4: Auszug aus einem GML-Instanzdokument. Es zeigt die Repräsentation eines GML-konformen, applikationsspezifischen Straßenfeatures.

2.4.2 Das Interoperabilitätsprogramm des OGC

Das Interoperabilitätsprogramm des OGC [OpenGIS 03b] stellt eine Prozesskette für den Entwurf, die Entwicklung, das Testen, die Demonstration und die Förderung des Einsatzes von Schnittstellen und Protokollen dar, die eine interoperable Datenverarbeitung ermöglichen. Es besteht aus Interoperabilitäts-Initiativen, die sich folgenden Aufgaben widmen:

- Einrichtung von so genannten Testbeds: Schaffung von Umgebungen, in denen sich die Schnittstellen- und Protokoll-Entwürfe testen lassen. Auf der Basis von Testbeds werden dann Vorschläge für konkrete Spezifikationen erarbeitet. Beispiele hierfür sind die Web Mapping Testbeds oder das Location Service Testbed.
- Realisierung von Pilotprojekten: Durch zusammenarbeitende Informationsgemeinschaften werden die Spezifikationen des OGC in einer realen Umgebung untersucht.
- Planungs- und Machbarkeitsstudien sowie Integrationsprojekte: Das OGC steht Organisationen zur Seite, die die Standards in ihre Produkte und Umgebungen integrieren möchten.
- Information und Ausbildung: Es wird ein Fundus von Informationen angeboten, die dem Entwickler helfen sollen, die OGC-Standards zu implementieren. Dazu gehören auch Diskussions-Foren im Internet. Die Vermittlung der OGC-Inhalte geschieht darüber hinaus im Rahmen von Seminaren und Workshops sowie auf Konferenzen.
- Anregung zur Mitarbeit: die Mitglieder des OGC werden über so genannte Requests for Technology, Requests for Proposals und Calls for Participation zur Auseinandersetzung mit den Konzepten des OGC aufgefordert.
- Abstimmung mit anderen Gremien: es muss eine Abstimmung und Konsensfindung mit anderen industriellen Konsortien sowie Standardisierungs-Institutionen wie ISO und CEN erfolgen.

Einige der bisher ins Leben gerufenen Initiativen befassen sich auch mit der semantischen Interoperabilität. Insbesondere sind hier das „Information Community Enablement Testbed“ und das „Geospatial Fusion Services Testbed“ zu nennen, die an dieser Stelle kurz vorgestellt werden sollen. Es ist aber darauf hinzuweisen, dass bisher lediglich erste Konzepte vorhanden sind.

2.4.2.1 Das „Information Community Enablement Testbed“

Innerhalb der Web Services Initiative hatte das OGC das so genannte „Information Community Enablement Testbed“ eingerichtet, welches sich die Abstimmung der Begriffsauffassungen zwischen verschiedenen Informationsgemeinschaften zum Ziel setzte. Hierzu wurde das Konzept der Semantic Translators, also der Semantik-Übersetzer, entworfen (siehe auch [McKee & Kottman 99]). Es führt Meta-Beschreibungen ein, die die Semantik von Objekten einer bestimmten Informationsgemeinschaft auf einer übergeordneten Ebene beschreiben. Sie werden in einem speziellen Katalog abgelegt. Schließlich müssen auf der Basis dieser Metadaten Transformationsregeln für die Überführung von Instanzen einer Informationsgemeinschaft in jene einer anderen festgelegt werden. Diese Transformationsregeln legen folglich fest, inwiefern zwischen den unterschiedlichen semantischen Vokabularen eine Beziehung besteht. Dabei gibt es von der exakten Entsprechung bis zu der Möglichkeit, dass keine Beziehung existiert, verschiedene Übergänge. Grundsätzlich ist dieses Konzept ein erster Ansatz, um das Problem der semantischen Interoperabilität anzugehen. Allerdings findet man auf den Webseiten des OGC keine aktuellen Informationen über das Testbed oder über dessen Ergebnisse.

2.4.2.2 Das „Geospatial Fusion Services Testbed“

Das „Geospatial Fusion Services Testbed“ beabsichtigt, raumbezogene Informationen, die sich ähneln bzw. sich auf das gleiche Realweltobjekt beziehen, zu fusionieren. Unter Fusion wird hierbei die Erzeugung expliziter Relationen zwischen geographischen Objekten verstanden. Zunächst könnte man vermuten, dass der Ansatz ähnlich gelagert ist wie jener, der der vorliegenden Arbeit zugrunde liegt. Es sollen hier aber nicht die Beziehungen zwischen vektorbasierten geographischen Objekten detailliert beschrieben werden, vielmehr sollen sämtliche Formate wie Digitalfotografien, Videosequenzen, Namen von Objekten, deren Koordinaten, etc. in ein gemeinsames Rahmenwerk integriert werden. Ein Dienst, der hier entwickelt wird, ist z.B. der OpenGIS Geocoding Service, der Lokationen einer sekundären Metrik (z.B. Adressangaben) in Koordinaten bzw. primäre Metriken umsetzen, usw.

2.5 Das Standard-basierte AFIS-ALKIS-ATKIS-Referenzmodell

Die Verfahrenslösungen zur Erfassung amtlicher Geodatenbestände für die Bundesrepublik Deutschland wurden in den 70er und 80er Jahren von der Arbeitsgemeinschaft der Vermessungsverwaltungen der BRD (AdV) konzipiert. Es handelt sich dabei um drei Basisinformationssysteme. Für den Bereich der Katastervermessung existiert zum einen die ALK, zum anderen das ALB. Beide Verfahren werden derzeit im Automatisierten Liegenschaftskataster-Informationssystem (ALKIS) zusammengeführt. Der Bereich der Topographie wird durch ATKIS abgedeckt. Zwar wurden die genannten Verfahren kontinuierlich fortentwickelt, dennoch sind die sich bei der Datenerfassung ergebenden Datenstrukturen nicht konform mit den internationalen GIS-Standards. Diese Situation hat man bei der AdV erkannt und es konnte inzwischen ein gemeinsames Referenzmodell für ALKIS und ATKIS sowie das neu entstandene Automatisierte Festpunkt-Informationssystem AFIS geschaffen werden, das den internationalen Standards gerecht wird. Im Zuge des Aufbaus dieses AFIS-ALKIS-ATKIS (AAA)-Referenzmodells geschah auch eine semantische Harmonisierung der Objektartenkataloge von ALKIS und ATKIS. Des Weiteren sind Konzepte erarbeitet worden, um den existierenden Datenbestand aus den Basisinformationssystemen in das neue Modell mittels eines Stufenkonzeptes zu migrieren. So soll ein bundesweit einheitlicher Grunddatenbestand der Daten des amtlichen Vermessungswesens entstehen.

Im Folgenden wird zunächst das konzeptuelle Modell des Applikationsschemas vorgestellt. Anschließend werden die neu gestalteten Objektartenkataloge des AAA-Referenzmodells präsentiert und schließlich wird die Austauschchnittstelle erläutert. Alle Ausführungen sind aus [AdV 03] und [AdV 04a] entnommen.

2.5.1 Das konzeptuelle Modell

„Das gemeinsame AFIS-ALKIS-ATKIS Anwendungsschema bietet einen einheitlichen und objektorientierten Modellansatz für AFIS, ALKIS und ATKIS“ [AdV 03]. Es ist unter Verwendung existierender Schemas aus der Normfamilie ISO 19100 der TC 211 entwickelt worden. Für jene Bereiche, in denen noch keine Normen vorliegen, werden die Spezifikationen des OGC herangezogen.

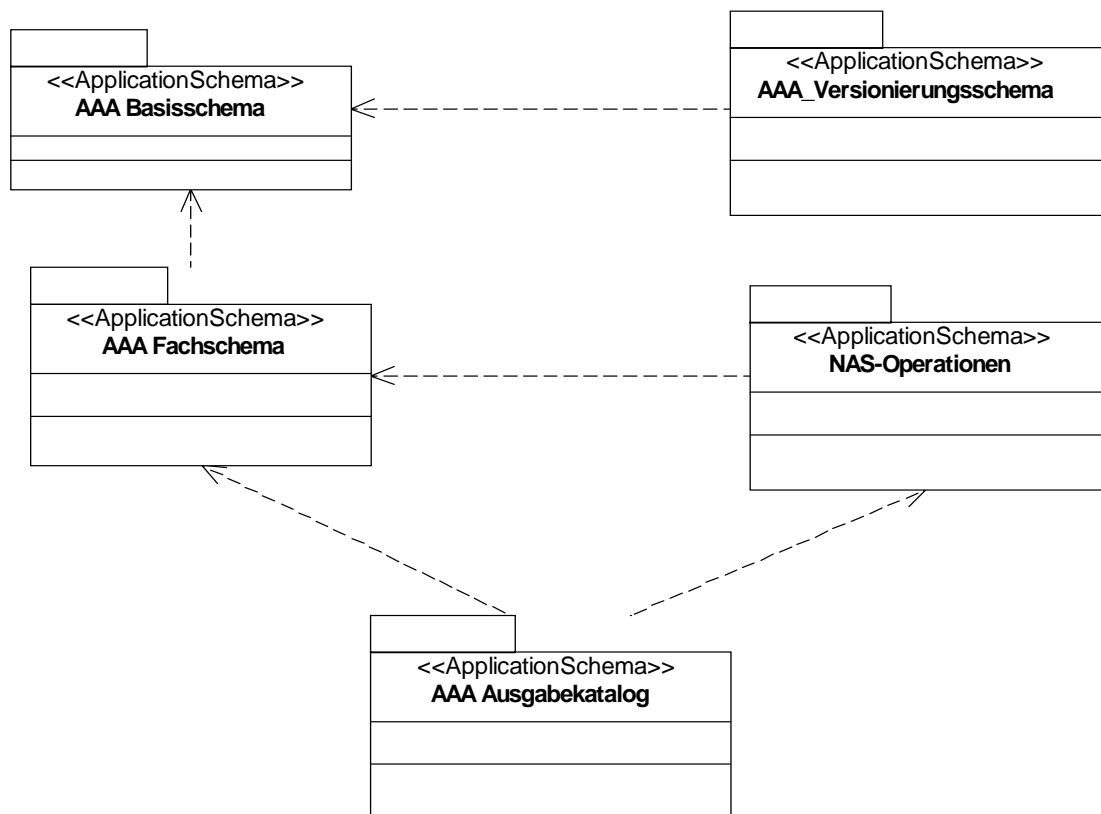


Abbildung 2-5: Die Bestandteile des AFIS-ALKIS-ATKIS (AAA)-Anwendungsschemas [AdV 04a]

Grundsätzlich wird das AAA-Anwendungsschema in das Basisschema, das Versionierungsschema, das Fachschema, die Operationen der Normbasierten Austauschchnittstelle (NAS) und den Ausgabekatalog unterteilt. Auf der Grundlage des Basisschemas können die jeweiligen Fachobjekte in den Fachschemas modelliert werden. Das Versionierungsschema dient der Historienverwaltung von Fachobjekten. Die NAS-Operationen definieren die Prozesse zum Datenaustausch und der Ausgabekatalog bestimmt das Ausgabeformat. Abbildung 2-5 zeigt die Bestandteile des Gesamtschemas. Die Modellierung geschah mittels der Unified Modeling Language (UML). Alle Dateien sind auf den Internetseiten des AdV verfügbar [AdV 04b].

2.5.1.1 Das AAA-Basisschema

Das Basisschema liefert die Grundlage für eine gemeinsame Modellierung von AFIS-, ALKIS- und ATKIS-Objekten. Es wurde unter Berücksichtigung der ISO/TC 211 Norm 19109 „Rules for Application Schema“ entwickelt, welche ein allgemeines Modell zur Beschreibung und Bildung von Fachobjekten bzw. Features, das so genannte General Feature Model, beinhaltet. Um innerhalb des Basisschemas auch Objekte ohne Raumbezug abbilden zu können, wurde das Basisschema gegenüber dem General Feature Model um die Klasse „AA_ObjektOhneRaumbezug“ erweitert.

Das Basisschema besteht aus insgesamt 11 Paketen. Innerhalb dieser Pakete werden Objektklassen definiert, die entweder direkt verwendet oder aus denen in den Fachschemas instanziiierbare Subklassen mittels Vererbung abgeleitet werden. Außerdem werden bereits einige Relationen, die zwischen Objekten existieren können, innerhalb des Basisschemas vorgegeben. So werden z.B. Unter- und Überführungsrelationen eingeführt. An dieser Stelle sollen einige wichtige Pakete des Basisschemas kurz skizziert werden.

AAA_Basisklassen

Das Basisschema gibt 4 allgemeine Arten von Objektausprägungen, die so genannten Elementarobjekte, vor. Die Elementarobjekte sind die kleinstmögliche Einheit und können nicht weiter unterteilt werden. Somit entfällt die Bildung von Objektteilen, wie sie bei ALK oder ATKIS üblich war:

1. Raumbezogene Elementarobjekte (AA_REO): ein REO besteht aus geometrischen, topologischen und fachlichen Eigenschaften;
2. Nicht raumbezogene Elementarobjekte (AA_NREO): ein NREO besteht ausschließlich aus fachlichen Eigenschaften;
3. Zusammengesetzte Objekte (AA_ZUSO): ein ZUSO besteht aus einer beliebigen Zahl und Mischung semantisch zusammengehörender REOs, NREOs oder ZUSOs;
4. Punktmengenobjekte (AA_PMO): Für bestimmte Fachobjekttypen wie Digitale Geländemodelle, die aus vielen Punkten mit jeweils gleichen Attributen bestehen, wurden Punktmengenobjekte eingeführt.

AAA_SpatialSchema

Die geometrische Modellierung der AAA-Features basiert auf der ISO-Norm 19107 „Spatial Schema“, die identisch mit der abstrakten Spezifikation zur Feature Geometry des OGC ist (vgl. Kapitel 2.4.1.1). Das AAA-Anwendungsschema beschränkt sich dabei aus Gründen der Komplexitätsreduzierung auf folgende geometrische und topologische Objekte (siehe Tabelle 2-1):

Geometrische Objekte (GM_Object)			Topologische Objekte (TP_Object)	
Geometrische Primitive	Geometrische Komplexe	Geometrische Aggregate	Topologische Primitive	Topologische Komplexe
GM_Point GM_Curve GM_PolyhedralSurface	GM_CompositeCurve GM_CompositeSurface	GM_MultiPoint GM_MultiCurve GM_MultiSurface	TS_PointComponent TS_CurveComponent TS_SurfaceComponent TS_Face	TP_Complex

Tabelle 2-1: Raumbezugsgrundformen des AFIS-ALKIS-ATKIS Applikationsschemas [AdV 04a]

Jedes raumbezogene AFIS-ALKIS-ATKIS-Fachobjekt verweist auf maximal eine Geometrie. Müssen mehrere Geometrien zu einem Realweltobjekt gespeichert werden, so muss auch jedes Mal ein eigenes Fachobjekt angelegt werden. Für jede Geometrie kann gemäß der ISO-Norm 19111 „Spatial Referencing by Coordinates“ das zugehörige Koordinatenreferenzsystem angegeben werden.

AAA_GemeinsameGeometrie

Das Paket AAA_GemeinsameGeometrie enthält Basisklassen für Features, die aus Punkten, Linien und Flächen bestehen und sich ihre Geometrie teilen. So können die geometrietragenden Primitive GM_Point und GM_Curve relational mit Fachobjekten verbunden und gleichzeitig von mehreren Fachobjekten gemeinsam benutzt werden. Dies gilt allerdings nicht für flächenförmige Objekte.

AAA_UnabhaengigeGeometrie

Die Klassen des Pakets AAA_UnabhaengigeGeometrie dienen dazu, Fachobjekte mit voneinander unabhängiger Geometrie modellieren zu können. Objekte, die aus diesen Klassen abgeleitet wurden, dürfen sich keine Geometrien mit anderen Objekten teilen. Objektarten, deren Geometrie als unabhängig angesehen wird, sind beispielsweise die Präsentationsobjekte (siehe nächster Abschnitt).

AAA_Praesentationsobjekte

Die Präsentation von Objekten in Form von Kartensignaturen oder Texten ist abhängig vom Modell. In der Regel wird die Präsentation vollautomatisch aus der Auswertung der so genannten Bestandsdaten gewonnen, die eine vollständige Beschreibung von Fachobjekten einschließlich ihrer kartographischen oder textgebundenen Darstellung in verschiedenen Maßstäben enthalten. Für sehr komplexe Fälle, in denen die Auswertung der Bestandsdaten einen unverhältnismäßig hohen Aufwand erfordern würde, müssen die Präsentationsobjekte jedoch explizit angelegt werden. Als Präsentationsobjekte werden folglich alle Texte und Kartensignaturen definiert, die nicht vollautomatisch für einen bestimmten Zielmaßstab erzeugt und platziert werden können.

AAA_Katalog

Die Struktur der AAA-Objektartenkataloge basiert ebenfalls auf einem Standard, der so genannten Feature Cataloguing Methodology-Norm 19110 der ISO. Neben den Objekteigenschaften werden auch Methoden von Objekten im Objektartenkatalog beschrieben. Innerhalb des AAA_Katalog-Pakets werden die Strukturen der ISO-Katalogvorschriften erweitert. So können beispielsweise für ALKIS-ATKIS-Attributarten zusätzlich deren Bildungsregeln festgelegt werden.

2.5.1.2 Das AAA-Versionierungsschema

Es kann notwendig sein, für AAA-Objekte Historien zu verwalten, d.h. neben den aktuellen müssen auch historische Informationen zu einem Objekt gespeichert werden können. Als Voraussetzung für die Versionierung gilt, dass Features bzw. Fachobjekte einen systemweit (also bundesweit) eindeutigen Objektidentifikator haben müssen. Ebenso ist es notwendig, ein Lebenszeitintervall zu führen, das ein Entstehungs- und Untergangsdatum für eine Instanz nachweist.

Wenn ein Objekt erstmals digitalisiert und in einen Datenbestand eingetragen wird, wird es in einem Objektbehälter abgelegt. Wenn sich dann aufgrund von Nachführungsoperationen eine nicht objektbildende Eigenschaft ändert, so entsteht eine neue Version des Objekts im Objektbehälter, die den gleichen Objektidentifikator wie Version 1 führt. Es bleibt jedoch auch die erste Version erhalten. Sie erhält ein Untergangsdatum, das mit dem Entstehungsdatum der zweiten Version identisch ist. Werden bei der Fortführung auch objektbildende Eigenschaften geändert, so geht das Objekt unter. Das Untergangsdatum der letzten Objektversion ist damit auch gleich dem Untergangsdatum des Gesamtobjektes. Das Objekt bleibt dennoch im Datenbestand erhalten. Auch die Änderungen von Objektrelationen führen zur Versionierung von Objekten. Eine Änderung von Relationen tritt dann auf, wenn eines der beiden in Beziehung stehenden Objekte im Hinblick auf die Relation geändert oder gelöscht wird.

2.5.1.3 Das AFIS-ALKIS-ATKIS Fachschema

Im AAA-Fachschema werden die selbstbezogenen Eigenschaften (Attribute) von AFIS-, ALKIS- und ATKIS-Objekten festgelegt. Attribute werden über einen Namen und einen Wertebereich definiert. Der Wertebereich kann sowohl aus Basisdatentypen (Zahlen, Zeichenketten, etc.) als auch aus komplexen Datentypen wie Geometrien bestehen. Auf der Ebene des Fachschemas können zusätzlich zu den bereits im Basischema angelegten Relationen weitere Beziehungen (fremdbezogene Eigenschaften) beschrieben werden.

2.5.2 Die Objektartenkataloge von ALKIS und ATKIS

Die Objektartenkataloge (OKs) von AFIS, ALKIS und ATKIS basieren auf dem gemeinsamen AFIS-ALKIS-ATKIS-Fachschema, wodurch eine semantische Harmonisierung der Realweltmodellierung erreicht werden konnte. Im ATKIS-OK sind die jeweiligen Fachobjekte des Digitalen Landschaftsmodells aufgeführt, im ALKIS-OK jene des Liegenschaftskatasters und im AFIS-OK die des Festpunktmodells.

Die Struktur der Objektartenkataloge für ATKIS und ALKIS ist identisch: Sie sind nach Objektbereichen gegliedert, die wiederum aus Objektartengruppen aufgebaut sind. Den Objektartengruppen sind die Objektarten nachgeordnet. Die Objektartengruppen haben folgende Struktur:

- Bezeichnung, Definition der Objektartengruppe
- Definition der Objektarten, weitere Angaben, z.B. zu
 - Attributarten
 - Relationsarten

Der schematische Aufbau des neuen, auf dem AAA-Fachschema basierenden ATKIS-Objektartenkataloges wird in Abbildung 2-6 anhand eines Ausschnitts dargestellt.

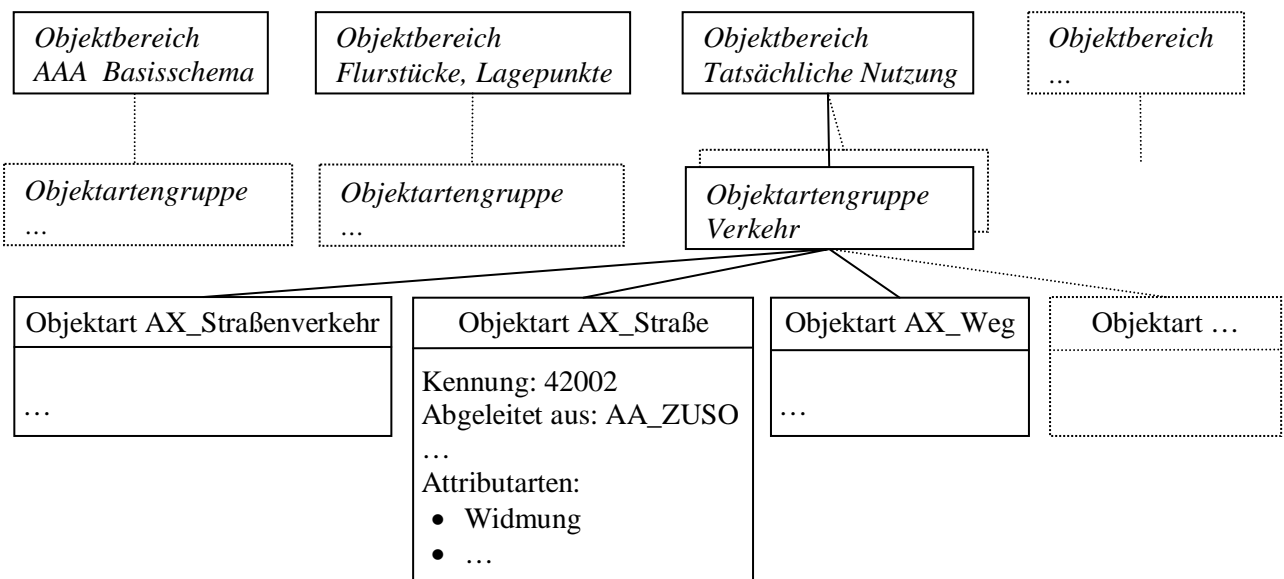


Abbildung 2-6: Ausschnitt aus dem auf dem AAA-Fachschema beruhenden ATKIS-Objektartenkatalog

2.5.3 Die Normbasierte Austauschchnittstelle (NAS)

Um Fachobjekte, die auf der Grundlage des AAA-Anwendungsschemas erfasst wurden, austauschen zu können, wurde die Normbasierte Austauschchnittstelle NAS entwickelt. Die NAS basiert auf der ISO-Norm 19118 „Encoding“, die so genannte „Encoding Rules“ für die Erzeugung von XML-Schemadefinitionen aus UML-Anwendungsschemas definiert. Diese Regeln gewähren jedoch noch einen gewissen Spielraum und so würde deren alleinige Berücksichtigung zu AdV-spezifischen Schnittstellen führen. Die NAS sollte jedoch

so konzipiert werden, dass ein möglichst generischer und standardbasierter Datenaustausch stattfinden kann. Daher lehnt man sich soweit wie möglich an den OGC-Standard GML an [AdV 04a].

2.6 Interoperabilität und Mehrfachrepräsentationen

Betrachtet man nochmals die Definitionen aus Kapitel 2.2 für den Begriff der Interoperabilität, so spiegeln diese u.a. die Anforderungen wider, dass bei interoperablen Prozessen eine Kombination von unterschiedlichen Daten möglich sein muss. Die Anwendungen sollen zudem in der Lage sein, auf den kombinierten Daten arbeiten zu können. Außerdem wird erwähnt, dass Interoperabilität zu einem transparenten Zugang zu Daten führen soll, wobei diese innerhalb eines Arbeitsablaufes nicht in einen Datenbestand überführt werden müssen. Dem letztgenannten Aspekt kommt im Rahmen der vorliegenden Arbeit eine besondere Bedeutung zu, da es nicht das vornehmliche Ziel ist, Mehrfachrepräsentationen in einen resultierenden Datenbestand zu verschmelzen, sondern Relationen zwischen ihnen aufzubauen, die dann im Zuge einer Datenanalyse ausgewertet werden können. Dem Gesagten ist zu entnehmen, dass zwischen den Forschungsbereichen „Mehrfachrepräsentationen“ und „Interoperabilität“ durchaus Überlappungen in den Zielsetzungen existieren.

Traditionell gesehen bestehen aber natürlich Unterschiede zwischen den beiden Problematiken. Die Bemühungen zur Standardisierung von Geodaten und zur Schaffung von Interoperabilität sind aus der Situation heraus entstanden, dass Anwender *verschiedene* raumbezogene Informationen, die in diversen Formaten vorlagen, kombinieren und gemeinsam nutzen wollten, um die Effizienz bei der Informationsgewinnung zu erhöhen. Die Problematik der Harmonisierung von *gleichen*, mehrfach im Computer repräsentierten Realweltobjekten wurde und wird dabei von den Standardisierungsorganisationen – wenn überhaupt – nur am Rande tangiert. Dies konnte durch die Untersuchung der aktuellen Interoperabilitätsbemühungen des OGC aufgezeigt werden.

Der größte Teil der Arbeit bei der Standardisierung widmet sich weiterhin hauptsächlich der technischen Seite der Interoperabilität. Diese technische Seite ist sicherlich auch für den Bereich der Mehrfachrepräsentationen relevant, da – wenn Geodaten standardisiert vorliegen – auch ein besserer Vergleich ihrer Eigenschaften möglich ist und somit die Beziehungen zwischen Mehrfachrepräsentationen leichter geklärt werden können. Gibt es beispielsweise Vorgaben für die Kriterien zur Beschreibung der Datenqualität eines Objektes, z.B. zur geometrischen Genauigkeit, so kann aufgrund der definierten Rahmenbedingungen festgestellt werden, welche Repräsentation eine höhere Genauigkeit aufweist und wie ähnlich die Genauigkeitsangaben einander sind.

Mit dem Aufkommen der Fragestellungen zur semantischen Interoperabilität kann sich die Situation im Bereich der Standardisierung in Zukunft jedoch ändern und es können auch Ansätze zur Lösung der semantischen Integration von Geoinformationen entwickelt werden. So ist es denkbar, zunächst einmal die Ausprägung von Objektklassen für bestimmte Anwendungsbereiche auf standardisierte Weise zu spezifizieren und somit eine allgemeingültige Semantik für die entsprechenden Benutzerkreise zu schaffen. Dabei könnten auch allgemeingültige Erfassungsvorschriften bzw. Objektbildungsregeln definiert werden. Somit erhielte jeder ausgewiesene Anwendungsbereich quasi ein einheitliches Datenschema. Dadurch würde zumindest die potentielle Unterschiedlichkeit von mehrfach repräsentierten Objekten innerhalb eines Anwendungsbereiches eingeschränkt, weshalb sich eine ebenfalls standardisierte Vorgehensweise zum Bestimmen der Ähnlichkeit korrespondierender Instanzen und damit verbesserte Möglichkeiten zu deren Integration ergeben könnten.

Im Ansatz wurde dieses Konzept bereits bei der Aufstellung des AAA-Referenzmodells verfolgt. Es orientiert sich an existierenden Standards und ist letztlich mit dem Ziel angelegt worden, eine einheitliche Sicht auf die Daten des Vermessungswesens in Deutschland zu schaffen. So geschah beispielsweise eine Harmonisierung der Objektartenkataloge von ATKIS und ALKIS. Es ist aber zu betonen, dass der Bereich der amtlichen Daten in sich relativ homogen ist, weil eine spezifische Applikationsausrichtung nicht angestrebt wird und eher Geobasisdaten geliefert werden sollen. Ferner muss darauf hingewiesen werden, dass die Standardisierungsbemühungen auf dem Gebiet der Semantik durchaus auch an ihre Grenzen stoßen können. Denn ob eine Straße als Fläche oder als Linienzug repräsentiert werden soll, ob bei mehrspurigen Straßen die einzelnen Fahrbahnen erfasst werden oder das Straßenobjekt insgesamt digitalisiert wird, ob lediglich das Gesamtobjekt oder sämtliche konstituierenden Liniensegmente zu attributieren sind, auf Basis welchen Maßstabs die

Erfassung geschieht, welche Bezeichnungen für Attribute und welche Wertebereiche für Attributwerte gewählt werden, usw., ist in vielen Fällen nur von den jeweiligen Anwendungen, jedoch nur schwer anwendungsübergreifend in Form eines Standards reglementierbar. Selbst innerhalb nahezu identischer oder zumindest stark ähnlicher Applikationsbereiche können hier bereits Komplikationen bei der Vereinheitlichung auftauchen [siehe z.B. Walter 97]. Dementsprechend äußern [Harvey et al. 99] ihre Skepsis bezüglich einer weitreichenden Vereinheitlichung von Geodaten, da sie davon ausgehen, dass die Semantik geographischer Phänomene zu vielfältig ist, um einer Standardisierung unterworfen zu werden.

Dennoch müssen mehrfach repräsentierte Geodaten aus heterogenen Fachbereichen gemeinsam verarbeitet werden können, wobei im Sinne von [Kuhn 94] eine Minimierung des Informationsverlustes bei der Integration erreicht werden soll. Sobald daher einheitliche Sichten auf Geodaten innerhalb verschiedener Anwendungsbereiche geschaffen werden können, wird auch der Prozess der Erkennung und Integration von gleichen semantischen Inhalten verschiedener Anwendungsbereiche bzw. von Mehrfachkonzeptionen auf Schemaebene und von Mehrfachrepräsentationen auf Datenebene vereinfacht. Ein theoretischer Ansatz für diese Integration wird vom OGC in Form der Semantischen Übersetzer vorgeschlagen. Konkrete Umsetzungen fehlen bislang jedoch. Außerdem können weitere Verfahren bei der semantischen Integration zum Einsatz kommen, wie sie derzeit intensiv in der Forschung untersucht werden (vgl. Kapitel 4.3.1). Ansätze zur Lösung der Problematik bietet auch die vorliegende Arbeit.

3 Nexus - Mehrfachrepräsentationen in einer offenen Systemplattform

Das Nexus-Projekt wurde im Jahr 1998 initiiert und zunächst vom Land Baden-Württemberg im Rahmen einer Anschubfinanzierung unterstützt. Die Zielsetzung von Nexus besteht darin, eine generische Plattform für mobile, orts- und kontextbezogene Systeme zu entwickeln, welche auf die Applikationsanforderungen hin zugeschnittene Daten und Dienste anbietet [Hohl et al. 99]. Die Anwendungen können dabei über drahtlose Kommunikation auf die Nexus-Infrastruktur zugreifen (siehe Abbildung 3-1). Ein klassisches Beispiel eines ortsbezogenen Systems ist ein mobiler Stadtführer, der dem Nutzer auf einem mobilen Endgerät seine Position in einer Karte darzustellen vermag, Informationen über seine Umgebung, beispielsweise über historisch interessante Gebäude, zur Verfügung stellt und bestimmte relevante Dienste, wie z.B. die Navigation von einem Ausgangs- zu einem Zielpunkt, bereitstellt.

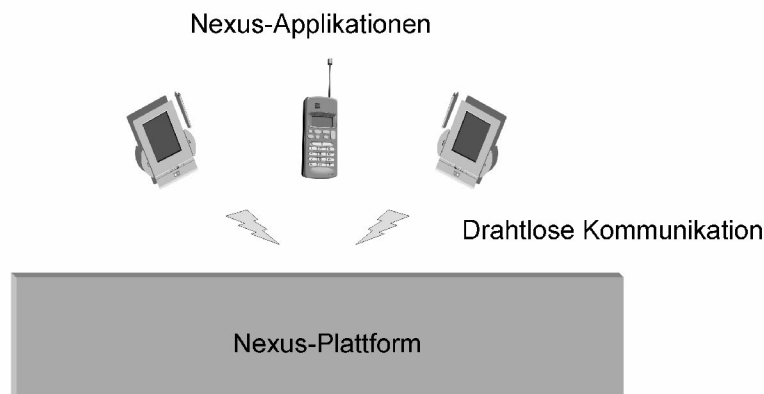


Abbildung 3-1: Nexus stellt eine Plattform für orts- und kontextbezogene Dienste bereit, auf die mobile Applikationen über drahtlose Kommunikation zugreifen können.

In der Anfangsphase des Projektes entstanden erste Prototypen ortsbezogener Systeme (u.a. [Volz et al. 99], [Klinger 00]) in Zusammenarbeit zwischen den damaligen Projektpartnern, der Abteilung Verteilte Systeme (VS) des Instituts für Parallele und Verteilte Systeme (IPVS) und dem Institut für Photogrammetrie (IfP) der Universität Stuttgart. Im Jahre 2000 mündete Nexus in eine Forschergruppe der Deutschen Forschungsgemeinschaft (DFG), an der nun zusätzlich zu den beiden ursprünglichen Partnern das Institut für Kommunikationsnetze und Rechnersysteme (IKR) sowie die Abteilung Anwendersoftware des IPVS teilnahmen. Im Rahmen einer dreijährigen Förderperiode wurde das Konzept der Nexus-Plattform entworfen und ein erster Prototyp der Plattform entwickelt. Seit Beginn des Jahres 2003 hat Nexus den Status eines Sonderforschungsbereiches (SFB) der DFG. Der Titel des stark interdisziplinär ausgerichteten SFBs lautet: „Umgebungsmodelle für mobile, kontextbezogene Systeme“. Derzeit (Stand: September 2004) arbeiten ca. 30 wissenschaftliche Mitarbeiter aus den Bereichen Informatik, Elektrotechnik, Verkehrswissenschaften, Photogrammetrie, Fertigungstechnik und Philosophie gemeinsam an der Umsetzung der angestrebten Zielsetzungen [Nexus 04].

In den folgenden Abschnitten wird das Grundkonzept von Nexus grob umrissen und es werden die tangierten Forschungsbereiche kurz skizziert. Anschließend wird auf das Datenschema des Systems, das so genannte Augmented World Schema, und die Architektur der Plattform eingegangen. Schließlich wird dargelegt, inwiefern die vorliegende Arbeit in den Nexus-Kontext einzuordnen ist und an welcher Stelle ein Beitrag zum Gesamtvorhaben geleistet werden soll.

3.1 Das Grundkonzept von Nexus

Der Grundgedanke von Nexus besteht darin, jenen Ausschnitt der Realwelt, der für orts- und kontextbezogene Anwendungen von Interesse ist, digital in Form so genannter Umgebungsmodelle zu repräsentieren und entsprechende Dienste anzubieten, die die digitalen Daten dieser Umgebungsmodelle nützen können. Am Anfang der Forschungsarbeit hatte man sich daher im Rahmen einer Use-Case-Analyse [Nicklas 01] über-

legt, welche Anwendungsfälle für Nexus von Bedeutung sind und welche Entitäten der Realwelt zu deren Realisierung im Datenschema repräsentiert werden sollten. In [Messmer 01] wurde ein erster Ansatz für dieses Datenschema, das so genannte Augmented World Schema, erarbeitet, das seither kontinuierlich weiter entwickelt wurde. Es wird deshalb als „Augmented“ bzw. angereichert bezeichnet, weil es nicht nur reale, sondern auch virtuelle Objekte enthält, die dazu dienen, zusätzliche Informationen in Nexus einzubringen. Ein typisches Beispiel ist die Metapher der virtuellen Litfasssäulen [Leonhardi et al. 99, Bauer & Leonhardi 00], die im Schema abgebildet bzw. im Datenbestand vorhanden sind und beispielsweise in einer Kartendarstellung eingeblendet werden können, jedoch keine Entsprechung in der Realität haben. Ihre Aufgabe besteht darin, Daten externer Informationsräume wie beispielsweise des World Wide Web ortsbezogen zu strukturieren und für Nexus-Anwendungen zugänglich zu machen (siehe Abbildung 3-2).

Wird das Augmented World Schema mit Daten gefüllt, so spricht man vom Augmented World Model (AWM) oder auch vom Globalen Umgebungsmodell. Es enthält die konkreten Instanzen aller in Nexus verfügbaren Daten. Natürlich besteht dieses AWM aber nicht aus einem einzigen Datentopf, sondern ist hochgradig in einzelne räumlich begrenzte Datenbestände, so genannte Augmented Areas bzw. lokale Umgebungsmodelle, unterteilt (siehe Abbildung 3-2). Das Problem der Mehrfachrepräsentationen in Nexus resultiert daraus, dass verschiedene Augmented Areas ein und dasselbe Realweltobjekt abbilden können. Innerhalb einer Augmented Area dürfen allerdings keine Mehrfachrepräsentationen existieren.

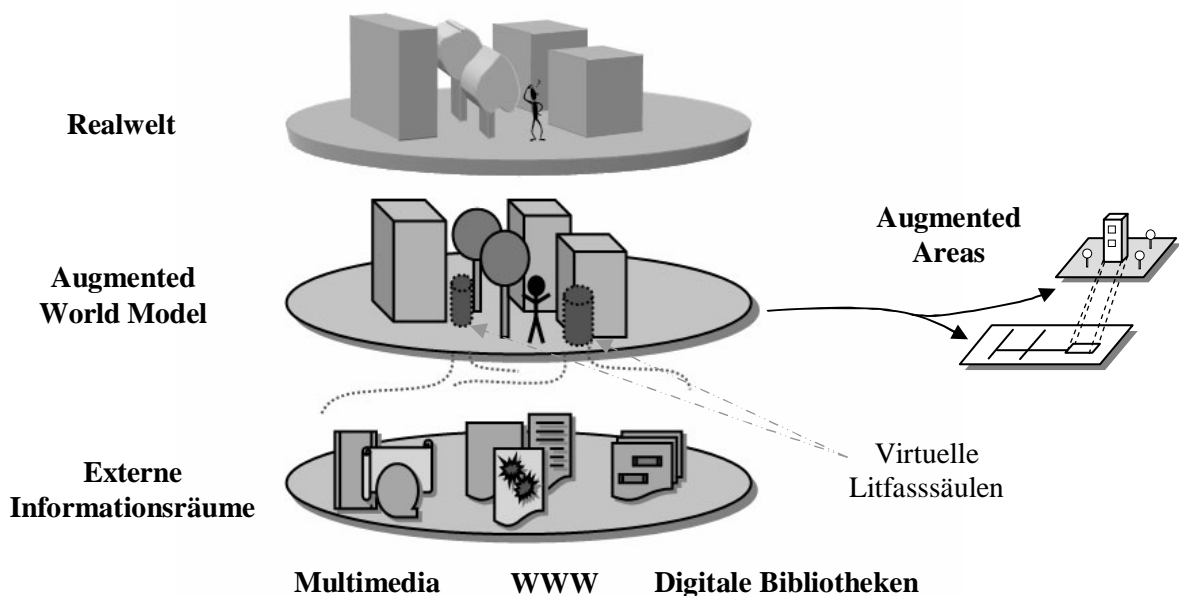


Abbildung 3-2: Das Prinzip von Nexus: die Realwelt soll in Form des Augmented World Models (AWM) digital repräsentiert werden. Die Daten des AWM können durch Informationen aus externen Quellen angereichert werden (nach [Nicklas et al. 01]).

Neben der Entwicklung eines Datenschemas für orts- und kontextbezogene Dienste wurde auch die Architektur der Infrastruktur spezifiziert [Nicklas et al. 01]. Die Prämisse hierbei war es, ein offenes System entwickeln zu wollen, das hochgradig verteilt ist [Volz & Sester 00] und an dem jeder teilnehmen und für welches jeder Daten und Dienste bereitstellen kann. Dementsprechend wurde die Architektur prinzipiell ähnlich zu der des World Wide Webs angelegt, wobei es jedoch einen entscheidenden Unterschied gibt: Zwischen der Anwendungsebene einerseits und der Daten- und Dienstebene andererseits existiert eine intelligente Schicht, die die Anfragen der Applikationen an die verschiedenen Datenserver weiterleitet und deren Ergebnisse zu einer für die Anwendung konsistenten Sicht aufbereiten kann. Nexus ist demnach im Sinne von [Conrad 02a] als föderiertes System zu charakterisieren, da es sich den Benutzern gegenüber wie ein einzelnes, homogenes Datenbanksystem verhalten soll, das die Verteilung und Inkonsistenz der Daten vor den Anwendungen verbirgt. Daraus ergibt sich die Anforderung der Integration von Mehrfachrepräsentationen.

Aufbauend auf den gemeinsamen Grundlagen Augmented World Schema und Augmented World Model sowie einer föderierten Architektur ergeben sich im Rahmen von Nexus verschiedene Forschungsbereiche [Nexus 04]. Sie sind in vier verschiedene Säulen zu unterteilen, die im Folgenden kurz skizziert werden.

3.1.1 Kommunikation und Sicherheit

Damit mobile Nutzer bzw. Applikationen mit der Nexus-Plattform in Kontakt treten und Daten austauschen können, ist es notwendig, eine robuste Kommunikationsinfrastruktur für mobile Anwendungen aufzubauen, die verschiedene Netztechnologien integriert [Gloss & Kühn 02]. So darf es beispielsweise nicht zu Problemen auf Applikationsseite kommen, wenn ein Benutzer von einem WLAN in ein Mobilfunknetz wie UMTS wechselt. Die Kommunikation muss außerdem möglichst effizient sein. Daher müssen entsprechende Methoden entwickelt werden, um in jenen Bereichen, wo eine hohe Bandbreite verfügbar und damit ein kostengünstiger Informationszugriff möglich ist (so genannte Datentankstellen), die Daten zu laden, die der mobile Benutzer auf seinem weiteren Weg mit großer Wahrscheinlichkeit benötigen wird. In diesem Zusammenhang spricht man auch von Prefetching oder Hoarding [Kubach & Rothermel 00, Bürklen et al. 04]. Um voraussagen zu können, wohin sich ein Benutzer von einem gegebenen Standort aus bewegt, bedarf es entsprechender Modelle, die das Mobilitätsverhalten der Benutzer simulieren [Ressel 04].

Die zu entwickelnde Kommunikationsinfrastruktur muss ferner Kommunikationsdienste anbieten, die speziell für ortsbezogene Anwendungen relevant sind. Ein Beispiel hierfür ist das GeoCast, das es erlaubt, elektronische Nachrichten in geographisch adressierte Gebiete, also z.B. an alle Benutzer in einem bestimmten Gebäude, zu schicken. Die Forschungsarbeiten hierzu sind u.a. dokumentiert in [Dürr & Rothermel 03]. Ein weiterer Kommunikationsansatz, der durch die Verwendung von Umgebungsmodellen möglich wird, ist der so genannte Ereignisdienst. Ein mobiler Benutzer des Nexus-Systems kann sich über die Plattform informieren lassen, wenn eine bestimmte raumbezogene Situation aufgetreten ist [Bauer & Rothermel 04]. So ist es z.B. möglich, einen Nexus-Teilnehmer, der sich durch eine ihm unbekannte Stadt bewegt und kulturell interessiert ist, immer dann zu benachrichtigen, wenn dieser in die Nähe eines Museums kommt.

Ein wichtiger und hinsichtlich der Akzeptanz des Nexus-Gesamtkonzeptes kritischer Punkt ist die Frage der Sicherheit von persönlichen Informationen, die in der Plattform zur Verfügung gestellt werden. So verwaltet Nexus die Positionsinformationen seiner mobilen Benutzer und deren Anfragen, woraus prinzipiell Nutzerprofile abgeleitet werden könnten. Daher sind geeignete Konzepte zur Pseudonymisierung der Identitäten mobiler Benutzer zu entwickeln, um so den erforderlichen Datenschutz gewährleisten zu können [Hauser et al. 02].

3.1.2 Datenmodellierung und Datenmanagement

Das Kernstück der Nexus-Plattform ist die Datenmanagementkomponente. Ihre Entwicklung bzw. Fortentwicklung schließt die Erweiterung des Augmented World Schemas und damit die Datenmodellierung, z.B. zur Repräsentation von zeitlichen oder topologischen Zusammenhängen von Nexus-Objekten, die effiziente Verwaltung des Umgebungsmodells bzw. des Augmented World Models und die Realisierung eines optimierten Föderationskonzeptes ein. Die Zielsetzung besteht darin, den Nexus-Anwendungen eine einheitliche Sicht auf die Informationen und Dienste der Plattform zu liefern. Die Betrachtung der Konsistenz von Umgebungsmodellen und somit die Problematik von Mehrfachrepräsentationen fällt ebenfalls in diesen Aufgabenbereich. Wichtige Aspekte dieses Forschungsbereiches werden in den untenstehenden Abschnitten beleuchtet.

An dieser Stelle seien der Vollständigkeit halber jene Aufgaben des Forschungsbereiches „Datenmodellierung und Datenmanagement“ angeführt, auf die im weiteren Verlauf nicht mehr eingegangen wird. Hierzu zählt die Datenverwaltungskomponente für die Positionsdaten, der so genannte Lokationsdienst. Bei dessen Entwicklung wird die Frage untersucht, wie hochdynamische Daten bzw. Positionsangaben einer großen Zahl von Nutzern effizient und skalierbar in Nexus verarbeitet werden können. Wichtige Arbeit wurde in diesem Bereich von [Leonhardi 03] geleistet. Darüber hinaus werden Konzepte zur Realisierung von kontextbezogenen Applikationen in infrastrukturlosen Umgebungen bzw. in mobilen Ad-Hoc Netzwerken untersucht. Dabei werden keine Umgebungsinformationen von einer Infrastruktur wie Nexus zur Verfügung

gestellt, sondern die mobilen Klienten erzeugen sich selbständig über entsprechende Sensorik eigene Umgebungsmodelle und tauschen diese untereinander aus [Bauer et al. 03].

3.1.3 Erfassung und Präsentation

Die Arbeiten in diesem Forschungsbereich konzentrieren sich insbesondere auf die Erzeugung von Umgebungsinformation durch den Einsatz von Sensorsystemen. Dabei wird die Gewinnung von Positionsinformationen mobiler Nutzer durch die Kombination verschiedener Sensoren untersucht [Klinec & Fritsch 01]. Insbesondere wird ein Ansatz entwickelt, mittels dessen Fußgänger über bildgebende Sensoren unter Ausnutzung des Umgebungsmodells positioniert werden können, um einen erleichterten Zugriff auf Nexus-Informationen zu erhalten [Klinec & Fritsch 03]. Auch die Integration weiterer Sensorik zur Erfassung des aktuellen Zustandes der Realwelt und die Einbindung dieser Informationen in das Umgebungsmodell ist Forschungsgegenstand. Zudem wird die Problematik der Widersprüchlichkeit von Sensorinformationen betrachtet [Nexus C3 04].

Neben der Thematik der Erfassung von Umgebungsmodell-Informationen wird analysiert, wie die Daten des Augmented World Models für den Benutzer aufbereitet werden sollen. So ist es notwendig, unnötige Details in Kartendarstellungen zu abstrahieren, um dem Benutzer eine einfache Interpretation der bereitgestellten Informationen zu erlauben. Abhängig vom Kontext soll Information daher semantisch generalisiert werden. Auf diese Weise können beispielsweise bei einer Navigationsanfrage bestimmte Landmarken der Umgebung, die für die Orientierung des Benutzers von Bedeutung sind (z.B. ein Kirchturm), betont werden, während unwichtige Sachverhalte ausgeblendet werden. Die Vereinfachung von Daten ist auch notwendig, um die Datenübertragungsraten im mobilen Kommunikationsumfeld in Grenzen zu halten [Kada 03]. Nexus soll darüber hinaus „Augmented Reality“ Szenarien unterstützen, um den Nutzer mit visuellen Zusatzinformationen, die in der Realwelt nicht sichtbar sind, versorgen zu können [Nexus C5 04].

3.1.4 Anwendungen

Die Funktionsfähigkeit der Nexus-Plattform soll anhand zweier Anwendungsprojekte evaluiert werden. Einerseits wird die Realisierung einer „Smart Factory“ (also einer intelligenten Fabrik) angestrebt, innerhalb derer Betriebsmittel unter Nutzung des Nexus-Umgebungsmodells intelligent verwaltet werden können [Bauer et al. 04a]. Andererseits werden die Informationen und Dienste der Nexus-Plattform dazu verwendet, eine bessere Orientierung und Navigation für blinde Menschen anbieten zu können [Nexus D2 04].

Schließlich wird das Nexus-Vorhaben von einem Teilprojekt begleitet, das die gesellschaftliche Akzeptanz von Nexus und seiner Implikationen für die Umwelt bewerten soll [Hubig 04, Siemoneit 04]. In dessen Rahmen werden auch neue Anwendungsszenarien für die Plattform betrachtet.

3.2 Das Augmented World Schema von Nexus

Die Einführung eines globalen Schemas ist deshalb notwendig, weil es den Nexus-Anwendungen eine einheitliche Sicht auf die innerhalb der Plattform gespeicherten Daten erlaubt und eine Föderation der unterschiedlichen Datenquellen ermöglicht [Nicklas & Mitschang 01]. Die folgenden Abschnitte skizzieren kurz die Entstehung des AWS sowie dessen Struktur.

3.2.1 Entstehung des AWS

Das Nexus Augmented World Schema entstand aus einer Szenarien- bzw. einer Use-Case-Analyse heraus und wurde speziell für die Anforderungen von orts- und kontextbezogenen Anwendungen konzipiert, d.h. es spezifiziert sämtliche für diese Art von Applikationen relevanten Objektklassen. Typische Use-Cases, die bei der Entwicklung des AWS zugrunde gelegt wurden, ergeben sich v.a. aus der Tatsache, dass die Benutzer der Nexus-Plattform mobil sind – so z.B. die Use-Cases Navigation, raumbezogene Ereignisse, ortsbezogene Abfragen, mobile Kommunikation oder Geo-Nachrichten (bzw. GeoCast), um hier nur einige zu nennen. Zusätzlich zur Szenarienanalyse wurden auch existierende Schemas von Daten betrachtet, die für Nexus relevant sind, um eine möglichst einfache Übertragung bestehender Bestände in die Plattform zu gewährleisten. Das AWS befindet sich in einem Prozess kontinuierlicher Weiterentwicklung.

3.2.2 Struktur des AWS

Das Augmented World Schema ist objektorientiert strukturiert und besteht aus einem Standard-Klassenschema, das die Basis-Objektklassen von Nexus enthält. Entstehen neuartige Anwendungen, die mit dem angebotenen Klassenspektrum nicht auskommen, so können auch Erweiterte Klassenschemas von den existierenden Standard-Klassen über Vererbungsmechanismen abgeleitet werden (siehe Abbildung 3-3). Um eine reiche Semantik der Objektklassen zu gewährleisten, wurden entsprechend viele Attribute spezifiziert, von denen die meisten aber optional angelegt sind. Dadurch kann eine unkomplizierte und flexible Bereitstellung von Daten innerhalb der Plattform gewährleistet werden.

Das Standard-Klassenschema enthält entsprechend seiner Ausrichtung vorwiegend Spezifikationen raumbezogener Objekte (*SpatialObjects*), die in statische und mobile Objektklassen zu unterteilen sind. Bei den statischen Objektklassen (*StaticObjects*) handelt es sich im Wesentlichen um Objekttypen, wie sie für GIS typisch sind (Straßen, Gebäude, etc.), während die mobilen Objektklassen (*MobileObjects*) alle dynamischen bzw. sich bewegenden Objekte wie Personen, Fahrzeuge, etc. definieren. Sowohl bei den statischen als auch bei den mobilen Objektklassen existieren Spezifikationen für virtuelle Objekte. Ein Beispiel für eine statische virtuelle Objektklasse wurde bereits oben erwähnt: die virtuelle Litfasssäule. Mobile virtuelle Objekte sind z.B. so genannte „Post-Its“. Dabei handelt es sich um Annotationen, die Benutzer an anderen Modellobjekten anbringen können und die potentiell mobil sein können.

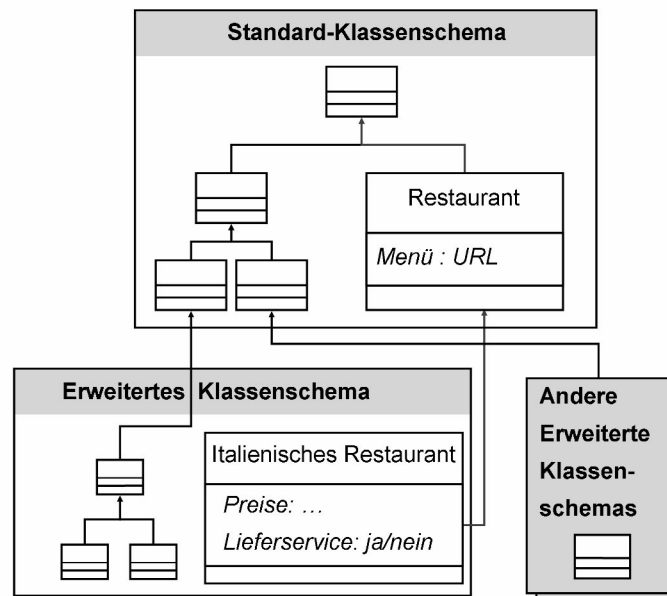


Abbildung 3-3: Das Standard-Klassenschema kann gemäß der Anforderungen von Applikationen erweitert werden (nach [Nicklas et al. 01]).

Neben den raumbezogenen Spezifikationen gibt es auch weitere Klassen, z.B. für Objekte, die sich im World Wide Web befinden bzw. die über einen URI (Uniform Resource Identifier) adressiert werden können. Im Kontext dieser Arbeit sind insbesondere die Nexus-Relationen relevant. Sie werden detailliert in Kapitel 6 erläutert, wenn es um die Definition von Relationen zwischen Mehrfachrepräsentationen geht.

Die Wurzel-Klasse des Augmented World Schemas ist die Klasse *NexusObject*. Alle instanzierbaren Objekte der Nexus-Plattform werden von einer direkten Subklasse von *NexusObject*, der Klasse *NexusDataObject*, abgeleitet. Alle Instanzen von *NexusDataObject* benötigen einen innerhalb von Nexus eindeutigen Identifikator, den so genannten Nexus Object Locator (NOL).

3.3 Architektur der Plattform

Die Komponenten der Nexus-Architektur sind in drei verschiedenen Schichten angeordnet: der Applikationsschicht, der Föderationsschicht und der Dienstschicht (siehe Abbildung 3-4). Die drei Schichten sollen in

der Folge näher betrachtet werden, wobei sich die Ausführungen in erster Linie auf eine aktuelle Dokumentation zum Informationsmanagement innerhalb der Nexus-Plattform von [Bauer et al. 04b] stützen.

3.3.1 Die Applikationsschicht

Nexus-Anwendungen laufen typischerweise auf kleinen, tragbaren Endgeräten, die über entsprechende Sensoren ihre Position bestimmen und drahtlos kommunizieren können. Sie richten ihre Anfragen über standardisierte, offene Schnittstellen an die Nexus-Knoten innerhalb der Plattform. Die wichtigsten Schnittstellen sind hierbei die Anfragesprache „Augmented World Query Language“ (AWQL) sowie das Format zum Austausch von Daten des Nexus-Umgebungsmodells, die „Augmented World Modeling Language“ (AWML) [Bauer et al. 04b]. Darüber hinaus gibt es weitere Schnittstellendefinitionen für die einzelnen Komponenten. Sämtliche Spezifikationen sind XML-basiert.

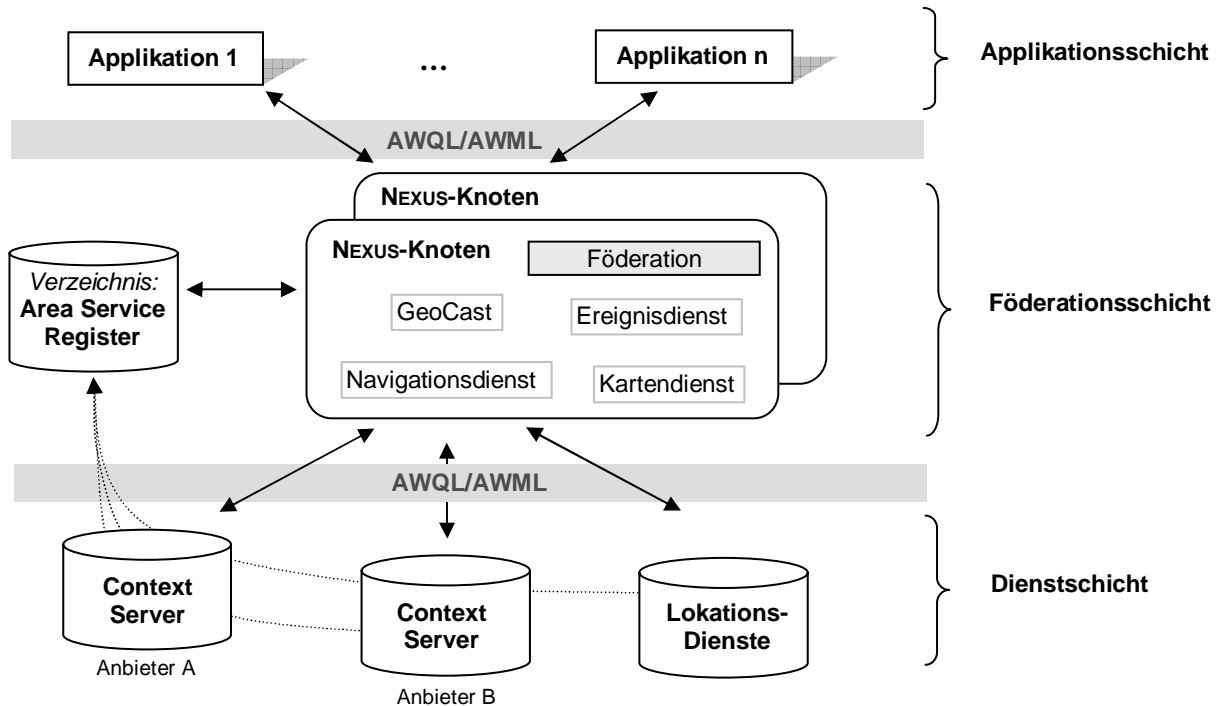


Abbildung 3-4: Die Architektur der Nexus-Plattform (verändert nach [Bauer et al. 04b])

3.3.2 Die Föderationsschicht

Die Föderationsschicht besteht aus den Nexus-Knoten und einem (bzw. mehreren replizierten) räumlichen Verzeichnisdienst(en), dem (bzw. den) so genannten Area Service Register(n). Jeder Nexus-Knoten verfügt seinerseits über eine Föderationskomponente, die die Strategie zur Beantwortung der Client-Anfragen entwirft. Über das Area Service Register macht sie jene Server der Dienstschicht ausfindig, die für die Anfragebearbeitung relevante Daten speichern. Anschließend kontaktiert sie die gefundenen Server und weist sie an, die benötigten Daten zur Verfügung zu stellen. Schließlich erzeugt die Föderation aus den erhaltenen Informationen eine konsistente Sicht (vgl. Kapitel 3.4 bzw. Kapitel 8.2). Auf dieser können dann weitere Dienste eines Nexus-Knotens wie z.B. der Navigations- oder der Kartendienst ihre Operationen aufsetzen.

Neben den genannten Diensten gehören auch der Ereignisdienst, der das Eintreten einer von Anwendungsseite definierten Situation im verteilten Umfeld überwacht, und die GeoCast-Komponente, welche die Versendung von Nachrichten in geographische Räume erlaubt, zu den Komponenten eines Nexus-Knotens.

3.3.3 Die Dienstschicht

Auf der Dienstschicht sind die Datenhaltungskomponenten der Nexus-Plattform angesiedelt. Sie werden als Context-Server bezeichnet. Dabei kann es sich beispielsweise um Geo-Informationssysteme oder um raumbezogene Datenbanken handeln. Ein Sonderfall der Context-Server ist der Lokationsdienst, der die Positions-

informationen mobiler Objekte verwaltet. Herkömmliche Datenbanken sind für diese Zwecke ungeeignet, daher hat [Leonhardi 03] einen Hauptspeicherbasierten Ansatz entwickelt, mit dem die hochfrequenten Positionsaktualisierungen sehr performant zu verwalten sind.

Alle Context-Server müssen die Standardschnittstellen AWQL/AWML bedienen. Als weitere Voraussetzung zur Teilnahme an Nexus gilt lediglich, dass sich die Datenanbieter beim Area Service Register registrieren müssen, um von der Föderation aufgefunden werden zu können. Dieser Prozess beinhaltet im Wesentlichen die Angabe der räumlichen Ausdehnungen aller von einem Context-Server verwalteten Augmented Area Modelle und die Nennung der von ihm angebotenen Objektklassen.

3.4 Mehrfachrepräsentationen und Mehrfachkonzeptionen in Nexus

Aufgrund der Tatsache, dass Datenanbieter ihr Material unabhängig voneinander in die Nexus-Plattform einbringen können und es analog zum World Wide Web keine Instanz gibt, die die Konsistenz der eingespeisten Daten überwacht, kommt es innerhalb von Nexus notwendigerweise zu Mehrfachrepräsentationen desselben Realweltobjektes. Die mehrfach repräsentierten Daten sind dabei in unterschiedlichen Augmented Areas bzw. Umgebungsmodellen auf einem oder verschiedenen Context-Servern gespeichert.

In der Folge wird das bisherige Konzept zur Behandlung von Mehrfachrepräsentationen innerhalb der Nexus-Plattform vorgestellt. Durch die im Rahmen dieser Arbeit entwickelten Ansätze kommt es allerdings zu einer Modifikation dieses Konzeptes, das zu einer Steigerung der Effizienz beim Umgang mit mehrfach repräsentierten Daten führen soll. Darauf wird in den Kapiteln 6 bis 8 eingegangen.

Nach den bisherigen Vorstellungen existiert innerhalb von Nexus zunächst keine Kenntnis darüber, welche Repräsentationen das gleiche Realweltobjekt beschreiben. Sobald aber für eine Anfrage Datensätze, die Mehrfachrepräsentationen enthalten, auf Föderationsebene miteinander kombiniert werden müssen, muss eine Integration auf Objektebene stattfinden. Dabei müssen zunächst die korrespondierenden Instanzen identifiziert werden. Nach diesem Prozess, der als Zuordnung bezeichnet wird, kann eine konsistente Sicht auf die Daten erzeugt werden, indem die Mehrfachrepräsentationen miteinander verschmolzen werden. So entsteht eine Mono-Repräsentation, auf der herkömmliche GIS-Operatoren arbeiten können.

Zusätzlich zu der Problematik der Mehrfachrepräsentationen stellt sich im Rahmen der Nexus-Plattform auch die Aufgabe, unterschiedliche, einander ähnliche Schemas bzw. Mehrfachkonzeptionen integrieren zu können. Dies ist deshalb notwendig, weil außerhalb der Infrastruktur verschiedene Datenschemas existieren, welche die gleichen Objekttypen der Realwelt auf unterschiedliche Weise spezifizieren und diese in das globale Augmented World Schema überführt werden müssen. Die Situation der Mehrfachkonzeptionen kann allerdings auch innerhalb von Nexus auftreten, wenn zwei Datenanbieter unterschiedliche Erweiterte Klassenschemas für dieselben Objekttypen definieren.

Um Mehrfachkonzeptionen auflösen und ein integriertes Schema – also das Augmented World Schema – aufstellen zu können, gibt es verschiedene Ansätze. Einerseits kann der Prozess manuell auf der Basis einer Analyse existierender Schemas sowie der Anforderungen des globalen Zielschemas durchgeführt werden. Diese Vorgehensweise wird in Kapitel 5 für Datenschemas aus dem Bereich des Straßenverkehrswesens illustriert. Andererseits ist es auch möglich, Hinweise auf korrespondierende Objektklassen automatisch abzuleiten. Ein Ansatz hierfür wird in Kapitel 8.1 präsentiert.

Die Situation der Mehrfachrepräsentationen und der Mehrfachkonzeptionen in Nexus soll im folgenden Schaubild (siehe Abbildung 3-5) verdeutlicht werden. Es existieren Mehrfachkonzeptionen über Objekttypen der Realwelt. Im Beispiel sind hier die beiden Datenschemas ATKIS und GDF, die in Kapitel 5.1 näher erläutert werden, als Mehrfachkonzeptionen zu betrachten. Sie sollen im einheitlichen Schema von Nexus bzw. im AWS abgebildet werden. Dadurch kann eine Transformation von Daten der Ausgangsschemas ins AWS-Schema und somit eine Bereitstellung dieser Daten im AWML-Format realisiert werden. Nach Abschluß dieses Prozesses kommt es zu Mehrfachrepräsentationen im Datenbestand von Nexus. Die Inkonsistenzen werden offensichtlich, wenn die mehrfach repräsentierten Daten auf Föderationsebene miteinander verarbeitet werden müssen. Die auftretenden Konflikte müssen durch eine Objektintegration gelöst werden,

in deren Rahmen eine Mono-Repräsentation generiert wird. Diese kann als Ausgangsbestand für die Datenanalyse verwendet oder der Applikation direkt zur Verfügung gestellt werden.

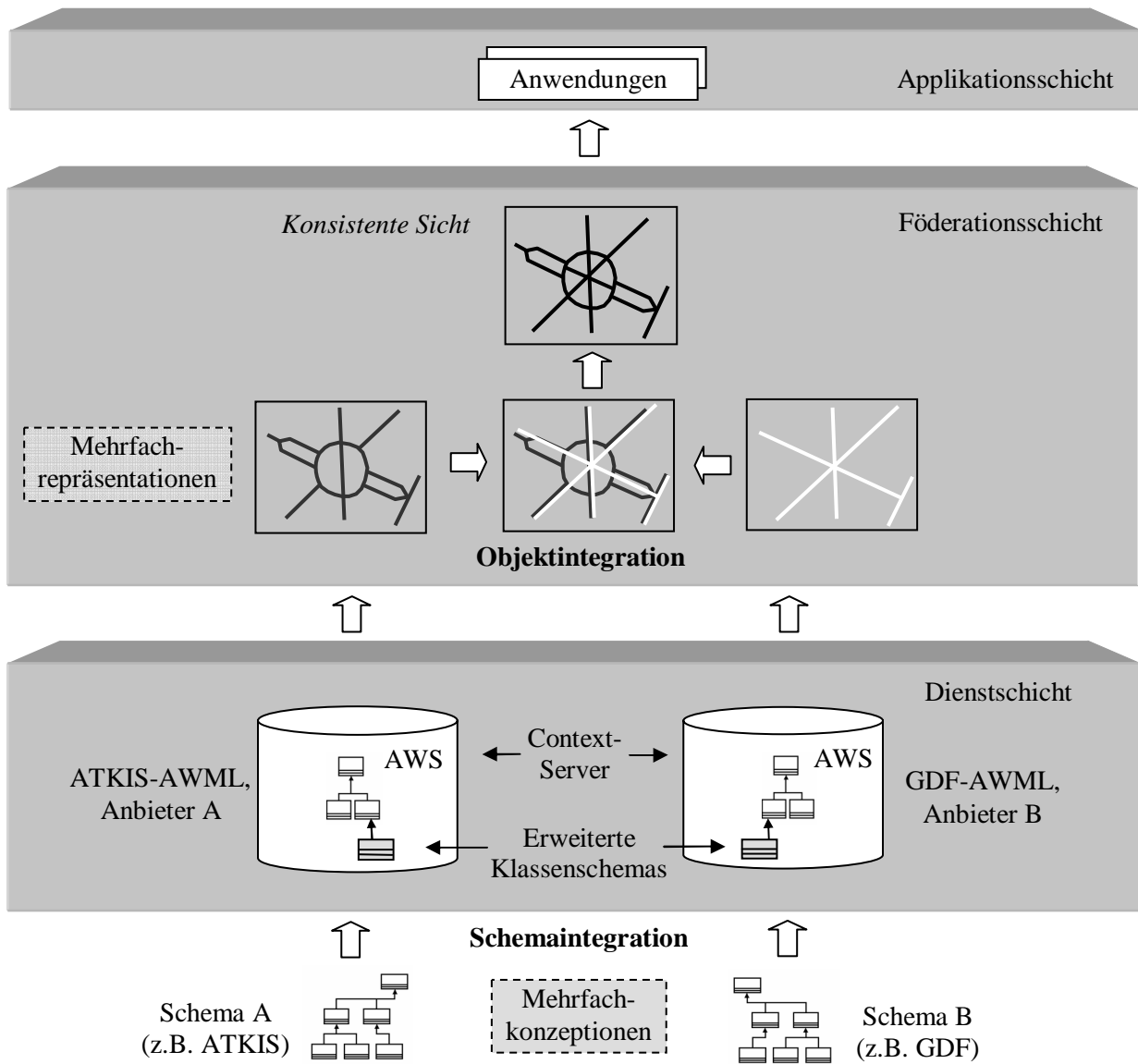


Abbildung 3-5: Mehrfachrepräsentationen und Mehrfachkonzeptionen in Nexus

4 Zur Problematik der Mehrfachrepräsentationen

Die Problematik der Mehrfachrepräsentationen (MRep) ist für diese Arbeit von zentraler Bedeutung. Daher wird sie in diesem Kapitel aufgearbeitet. Dabei wird auch auf die Schwierigkeiten im Umgang mit Mehrfachkonzeptionen eingegangen, da die Forschungsbereiche „Mehrfachrepräsentationen“ und „Mehrfachkonzeptionen“ oft miteinander verwoben sind und in dieser Arbeit beide Themengebiete eine wesentliche Rolle spielen.

In der Folge wird der Begriff „Mehrfachrepräsentationen“ zunächst aus GIS-Sicht betrachtet. Anschließend soll dies auch auf einer abstrakteren Ebene geschehen, indem der Zusammenhang von Daten und Realität allgemein untersucht wird, was mitunter auch philosophische Aspekte enthält. Dabei muss deutlich gemacht werden, was unter einem Realweltobjekt überhaupt zu verstehen ist. Außerdem wirft die Thematik der Mehrfachrepräsentationen die Frage nach der Konsistenz auf, worauf ebenfalls eingegangen wird. Anschließend wird erläutert, weshalb eine Integration von Mehrfachrepräsentationen sinnvoll und notwendig ist. Schließlich wird der aktuelle Stand der Forschung auf dem Gebiet der Mehrfachrepräsentationen aus der Perspektive der Geo-Informationssysteme aufgezeigt.

4.1 Zum Verständnis des Begriffs „Mehrfachrepräsentationen“

In dieser Arbeit ist unter dem Begriff „Mehrfachrepräsentationen“ der Sachverhalt zu verstehen, dass mehrere digitale Abbilder ein und desselben realen Objektes existieren, die entweder komplett oder partiell korrespondieren. Abbilder entstehen dadurch, dass reale Entitäten über Beschreibungen approximiert werden. Konzeptionelle Schemas (im GIS-Umfeld häufig auch als Datenmodelle bezeichnet) sind sozusagen formalisierte Beschreibungen. Mit Hilfe von Beschreibungen bzw. Schemas kann es jedoch niemals gelingen, die Realität gänzlich zu erfassen, da sie stets subjektiv wahrgenommen wird und zudem eine inhärente Unschärfe und unkalkulierbare Dynamik aufweist. Insofern ist es auch unmöglich, ein ideales Modell für ein reales Objekt zu entwickeln.

Der Mensch verwendet Wörter, Symbole oder Bilder, um sein Wissen zu repräsentieren, sprich seine Umgebung in den Kopf zu bekommen und im Gehirn zu „speichern“. Damit kann er deren Komplexität auf ein Maß reduzieren, das von ihm verarbeitbar ist, d.h. die Schaffung von Modellen dient in erster Linie einer Komplexitätsreduzierung. Die Ausprägung eines jeden Modells ist davon abhängig, wie ein Individuum seine Umwelt wahrnimmt, d.h. über welche Sensorik es verfügt, und welchen Erfahrungshintergrund es aufweist. Dementsprechend entstehen unterschiedliche interne Vorstellungen von der Welt, die im Laufe des Lebens erworben werden [Neisser 96]. Beispielsweise hat ein Hund ein anderes Modell von einem Stuhl als ein Mensch, ein Pilot ein anderes Modell von einem Flugzeug als ein normaler Passagier. Letzteres Beispiel impliziert, dass Modelle auf verschiedenen Abstraktionsstufen vorgehalten werden können, d.h. jeder Mensch hat ein eigenes Modell eines Realweltobjektes in seinem Gehirn, das um so detaillierter ist, je besser er dieses kennt bzw. je spezialisierter er ist.

4.1.1 Arten von Mehrfachrepräsentationen aus GIS-Sicht

Grundsätzlich existieren verschiedene Arten von Mehrfachrepräsentationen in Geo-Informationssystemen, die sich nach der Ursache für deren Entstehung unterscheiden lassen. Dabei sind schemabedingte, erfassungsbedingte, zeitlich bedingte, verwaltungsbedingte, durch Veredelung von Daten bedingte sowie letztlich formatbedingte Gründe auszumachen. Aus den Unterschieden zwischen Mehrfachrepräsentationen erwachsen die Konflikte, die später im Rahmen der Integration gelöst werden müssen.

4.1.1.1 Schemabedingte Mehrfachrepräsentationen

Bei der Modellierung raumbezogener Objekte wird von Fachleuten versucht, jenes Wissen, welches sie über einen bestimmten Ausschnitt der Realwelt haben, zu strukturieren und in digitaler Form als Schemas (bzw. Datenmodelle) abzubilden. Je nach Anwendungszweck werden mit den Schemas aber unterschiedliche Zielsetzungen verfolgt, so dass sie notwendigerweise voneinander abweichen. Werden nun Daten auf der

Basis unterschiedlicher Schemas bzw. auf der Basis von Mehrfachkonzeptionen von ein und demselben Realweltobjekt erfasst, entstehen Mehrfachrepräsentationen.

Einen Sonderfall der schemabedingten Mehrfachrepräsentationen stellen die maßstabsabhängigen Mehrfachrepräsentationen dar. Der Erfassungsmaßstab eines Objektes ist von der jeweiligen Aufgabenstellung abhängig. Um folglich verschiedene Arten von Aufgabenstellungen abdecken zu können, muss ein Objekt entweder in unterschiedlichen Maßstäben digitalisiert werden oder es besteht die Möglichkeit, dass man über einen hochdetaillierten Datensatz verfügt, von dem man dann über Generalisierungsoperatoren die jeweils adäquate Abstraktionsstufe eines Objektes ableiten kann [Mackaness et al. 97]. Solche Generalisierungsoperatoren stehen aber derzeit noch nicht zur Verfügung, so dass die unterschiedlich detaillierten Repräsentationen auf separaten, diskreten Auflösungsebenen vorgehalten werden müssen [Sester 00]. Die Objektrepräsentationen auf den verschiedenen Ebenen müssen dabei gegenseitig verzeigert sein. Diese Verzeigerungen bzw. Verbindungen zwischen den einzelnen Instanzen können auf verschiedene Art und Weise generiert werden [Sester et al. 98]. Im einfachsten Fall geschieht die Zuweisung manuell. Außerdem ist die Zuordnung über ein geometrisch-topologisches Matching möglich [Walter 97]. Das Matching funktioniert jedoch nur, wenn sich die Objekte auf ähnlichen Maßstabsebenen befinden. Daher schlagen [Anders & Sester 97] vor, zunächst den detaillierteren Datensatz in einem Vorverarbeitungsschritt einer Objektgeneralisierung bzw. einer Objekttaggregation zu unterziehen und so die Repräsentationen in einen ähnlichen Maßstab zu transformieren, um Links bzw. Verbindungen zwischen ihnen herstellen zu können. Sie versuchen, dies über die Vorgabe eines expliziten Regelwerks zu ermöglichen. Daraus ergibt sich der generelle Ansatz, Regeln vorzugeben, um Objekte von einem diskreten Zustand in einen anderen überführen zu können. [Sester 99] hat diese Übergangsregeln zwischen Repräsentationen in unterschiedlichen Maßstäben mit Hilfe von Verfahren aus dem Bereich des Maschinellen Lernens automatisch generiert. Die Regeln werden dabei anhand von Beispielen, die ein Tutor interaktiv am System eingibt, abgeleitet.

4.1.1.2 Erfassungsbedingte Mehrfachrepräsentationen

Abweichungen zwischen Mehrfachrepräsentationen können auch dann entstehen, wenn Daten auf der Basis identischer Datenschemas erfasst werden, da die Digitalisierung von raumbezogenen Objekten von der individuellen Interpretation verschiedener Operateure abhängig ist. [Glemser 01] wies dies beispielsweise dadurch nach, dass er ein Waldgebiet aus einem Luftbild von verschiedenen Bearbeitern als Vektorobjekt digitalisieren ließ und danach feststellen konnte, dass teilweise erhebliche Abweichungen zwischen den erfassten Objekten auftraten. Dies resultiert zum einen aus dem differierenden Erfahrungshintergrund der Erfasser, andererseits sind physische Objektgrenzen bei der Digitalisierung oft nicht eindeutig determinierbar [Burrough 96]. Auch die Vorschriften für die Erfassung können voneinander abweichen, so dass identische Szenen ganz unterschiedlich repräsentiert werden [Bofinger 01].

Raumbezogene Daten werden häufig auch automatisch mit Hilfe unterschiedlicher Sensoren (z.B. GPS, optische Sensoren, etc.) erfasst, die wiederum unterschiedliche Ergebnisse liefern können. Außerdem können Unterschiede von digitalen Repräsentationen eines Realweltobjektes auch darauf zurückzuführen sein, dass sie in unterschiedlichen Koordinatensystemen oder unter Verwendung unterschiedlicher Qualitätskriterien aufgenommen wurden.

4.1.1.3 Zeitlich bedingte Mehrfachrepräsentationen

Neben den bereits aufgeführten Gründen können Mehrfachrepräsentationen auch noch dadurch entstehen, dass dasselbe Realweltobjekt zu unterschiedlichen Zeitpunkten erfasst wurde. Gegebenenfalls werden also Historien eines Objektes in Form von verschiedenen Repräsentationen verwaltet, wie es z.B. beim Versionierungsschema des AFIS-ALKIS-ATKIS-Anwendungsschemas (vgl. Kapitel 2.5.1.2) der Fall ist. Die Schwierigkeit besteht darin zu sagen, wann eine Repräsentation ihre Identität ändert und als neue Version der Repräsentation abgespeichert werden muss und wann lediglich eine Fortführung stattfindet, die ein Objekt in gewissem Umfang modifiziert [Kent 00].

Einen Sonderfall der zeitlich bedingten Mehrfachrepräsentationen stellen die mobilen Objekte wie z.B. Fahrzeuge oder Fußgänger dar. Im einfachsten Fall ändert sich mit der Zeit ausschließlich die Position eines Objektes. Derzeit setzt sich die Forschungsarbeit in diesem Bereich damit auseinander, entsprechende Datenmodelle und Anfragesprachen für sich bewegende Objekte zu entwickeln (u.a. [Vazirgiannis & Wolfson

01, Böhlen et al. 00]). Eine besondere Problematik ergibt sich aus der Fragestellung, wie sich kontinuierlich bewegende flächenhafte Objekte (z.B. Waldbrände, Ölteppiche, etc.) in diesem Zusammenhang behandelt werden müssen [Tøssebro & Güting 01].

4.1.1.4 Verwaltungsbedingte Mehrfachrepräsentationen

Im Hinblick auf eine effiziente Bereitstellung von Geodatenätzen über Netzwerke hinweg werden vielfach Objektkopien angefertigt bzw. Daten repliziert. Somit entstehen ebenfalls mehrfache Repräsentationen von Objekten der Realwelt. Die Duplikate bleiben hier in der Regel identisch, d.h. wenn die Originalrepräsentation geändert wird, wird diese erneut repliziert und die älteren Versionen werden gelöscht.

4.1.1.5 Mehrfachrepräsentationen durch Veredelung

Wird ein und derselbe Datenbestand an verschiedene Abnehmer weitergegeben – wie das in der Regel bei Geobasisdaten wie ATKIS oder ALK der Fall ist – so können diese gemäß ihrer Applikationsanforderungen und Objektauffassungen die Eigenschaften von Objekten erweitern bzw. verändern. Diesen Vorgang bezeichnet man als Datenveredelung. Die Daten stammen also ursprünglich vom gleichen Bestand, durch die Veredelung entstehen aber nicht nur verschiedene Objektkopien, sondern auch voneinander abweichende Repräsentationen.

4.1.1.6 Formatbedingte Mehrfachrepräsentationen

Je nachdem, welches Geo-Informationssystem eingesetzt wird, kann dieses evtl. nur bestimmte Vektorformate lesen. Ein und dasselbe Datenobjekt kann folglich in unterschiedlichen Vektorformaten repräsentiert werden. Bei der Konvertierung von einem Format ins andere können im günstigsten Fall alle Objektinformationen erhalten bleiben, der Prozess kann aber auch zu Informationsverlust führen. Ebenso können die gleichen Realweltobjekte nicht nur in verschiedenen Vektorformaten, sondern auch im Rasterformat erfasst werden. In den genannten Fällen spricht man von formatbedingten Mehrfachrepräsentationen.

4.1.2 Abstraktere Betrachtungen zum Thema „Mehrfachrepräsentationen“

Die folgenden Ausführungen orientieren sich in erster Linie an dem Werk von [Kent 00]. Er setzt sich in seiner Abhandlung über „Daten und Realität“ intensiv damit auseinander, wie Informationen in Computern repräsentiert werden können. Die von dem Autor gemachten Bemerkungen vermitteln differenziertere Einsichten in die Thematik der Repräsentation von Informationen im Allgemeinen und damit ebenso in Bezug auf Mehrfachrepräsentationen, auch wenn die Betrachtung der Thematik nicht aus dem Blickwinkel des Raumbezugs geschieht.

Auf der Basis der Kentschen Ausführungen werden in den beiden folgenden Abschnitten verschiedene Stufen der Repräsentation von Objekten und damit auch der Mehrfachrepräsentationen abgeleitet. Darüber hinaus wird erklärt, was im Rahmen dieser Arbeit unter dem Begriff Realweltobjekt zu verstehen ist, und es wird die Problematik der Konsistenz und Redundanz von Mehrfachrepräsentationen skizziert.

4.1.2.1 Ebenen der Repräsentation

[Kent 00] stellt sich die Frage, wie Informationen überhaupt im Computer repräsentiert werden können, und vergleicht die existierenden Konzepte wie relationale oder objektorientierte Modelle mit unterschiedlichen Arten von Karten, die jeweils ihre Stärken und Schwächen bei der Abbildung eines bestimmten Sachverhaltes haben. Er ist jedoch der Auffassung, dass keines dieser formalen Modellierungssysteme in der Lage ist, Information adäquat zu beschreiben:

„Information in its real essence is probably too amorphous, too ambiguous, too subjective, too slippery and elusive to be ever pinned down precisely by the objective and deterministic processes embodied in a computer.“

Laut [Metaxides 75] ist es sogar überhaupt nicht möglich, die Realität, in welcher Form auch immer, objektiv zu repräsentieren:

„Entities are a state of mind. No two people agree on what the real world is.“

Aus dieser Erkenntnis folgt, dass, sobald mehrere Menschen ein Realweltobjekt wahrnehmen, auf mentaler Ebene auch mehrfache Repräsentationen davon entstehen, und zwar abhängig davon, welche Konzeptionen sie sich im Laufe der Zeit über Objekte der Realwelt aus ihren Erfahrungen heraus gebildet haben. Dementsprechend wird also davon ausgegangen, dass es *eine* Realität gibt, die genauso oft repräsentiert wird, wie sie wahrgenommen wird. Werden die individuellen Sichtweisen auf die Realität formalisiert beschrieben, so entstehen unterschiedliche Repräsentationen eines Objektes auf Schemaebene (bzw. Mehrfachkonzeptionen). Innerhalb des objektorientierten Paradigmas bedeutet dies, dass verschiedene Objektklassen, Attribute, Relationen, etc. eingesetzt werden, um Entitäten der Realwelt computergerecht zu definieren. Wird nun auf Basis unterschiedlicher Schemas eine Erfassung durchgeführt, werden dabei die konkreten Instanzen von Objekten bzw. die Daten generiert. Da auf der Basis *eines* Schemas verschiedene Erfassungen stattfinden können, kann es auch hier zu Mehrfachrepräsentationen kommen. Die verschiedenen Ebenen, auf denen multiple Repräsentationen entstehen können, zeigt Abbildung 4-1.

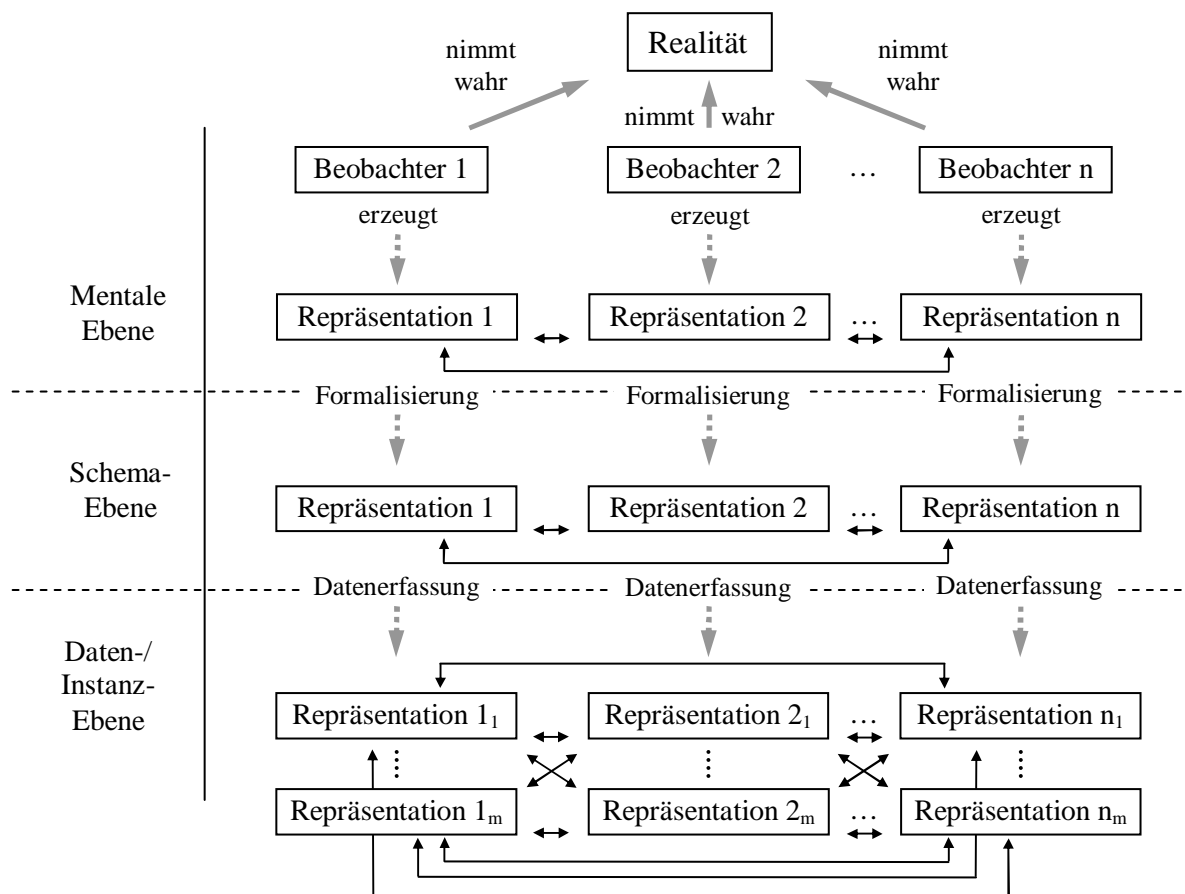


Abbildung 4-1: Verschiedene Ebenen der Repräsentation (mögliche Korrespondenzen zwischen den Repräsentationen einer Ebene sind als schwarze Doppelpfeile angedeutet)

Geht man vom gleichen Realweltobjekt aus, so existieren in der Regel zumindest Korrespondenzen zwischen dessen Repräsentationen im Geiste zweier Personen sowie zwischen dessen Repräsentationen in verschiedenen Schemas und zwischen dessen Repräsentationen in Form von Daten innerhalb eines Informationssystems. Die Korrespondenzen auf Schema- und vor allem auf Datenebene aufzudecken und in Form von Relationen explizit auszudrücken, ist das Ziel dieser Arbeit.

4.1.2.2 Realweltobjekte und deren Repräsentationen

Die Realwelt ist grundsätzlich nicht unterteilt. Erst durch das Bedürfnis des Menschen, die Realität zu vereinfachen, sie „fassbar“ und beschreibbar zu machen und sich über diese austauschen zu wollen, tritt eine Objektbildung auf. Was aber als Objekt identifiziert wird, bleibt gänzlich dem Betrachter überlassen, d.h. ein

Realweltobjekt kann jeder Sachverhalt der Realwelt sein, der durch einen Betrachter als solcher wahrgenommen wird. Es gibt folglich keine Vorgabe hinsichtlich der Definition von Realweltobjekten. Allerdings existieren zwei entscheidende Faktoren, die beeinflussen, was als Realweltobjekt aufgefasst bzw. was von einem Menschen als *ein Ding* erkannt wird.

Zum einen ist dies der Kontext, innerhalb dessen sich ein Beobachter befindet und aus dem heraus er die Realwelt wahrnimmt. Dieser Kontext ist auf eine Vorprägung bzw. Konditionierung des Beobachters zurückzuführen, der die Realwelt mit einem bestimmten Fachwissen, einer bestimmten Weltanschauung oder kulturellen Tradition, usw. interpretiert. Entsprechend des Beobachterkontexts können z.B. die Abgrenzung sowie die Detailliertheit eines Realweltobjektes variieren. [Kent 00] erläutert dies am Beispiel eines Warenhauses und stellt die Frage, was denn genau ein Warenhaus ist? Sind es alle Läden in einem Gebäude oder, wenn es sich um mehrere Gebäude handelt, alle Läden in allen Gebäuden? Verschiedene Interpretationen eines Objektes können dabei alle stimmig sein, ein objektiver Objektbegriff ist folglich nicht möglich. Je unschärfer die natürlichen Grenzen eines Objektes sind, desto mehr unterliegt die Objektbegrenzung der Willkür eines Betrachters. Ein Ding wird hier also zu einem diskretisierten Ausschnitt eines Kontinuums – und der Prozess der Diskretisierung ist oft nicht nachvollziehbar.

Zum anderen gibt es den Kontext des Objekts, der bei der Wahrnehmung des Betrachters von Bedeutung ist. Dies ist vor dem Hintergrund zu sehen, dass Realweltobjekte entweder gleichzeitig oder in einer zeitlichen Abfolge verschiedene Rollen annehmen können. So kann ein Mensch beispielsweise sowohl als Objekt „Sportler“ als auch als Objekt „Student“, etc. wahrgenommen werden. Aus dem Studenten kann über die Zeit hinweg ein wissenschaftlicher Angestellter werden, usw.

Wird eine Repräsentation eines Realweltobjekts erzeugt, so ist bei diesem Vorgang also der Kontext des Beobachters und der Kontext des Objekts ausschlaggebend dafür, wie das Objekt interpretiert wird und welche Merkmale des Objektes in die Repräsentation einfließen bzw. symbolisiert werden. Beide Arten von Kontext fließen zusammen und schaffen somit die Semantik der Repräsentation.

4.1.2.3 Mehrfachrepräsentationen und Konsistenz

Der Begriff der Konsistenz wird in vielen Bereichen verwendet. Im normalen Sprachgebrauch gilt er als synonym mit dem Wort Widerspruchsfreiheit. Im Umkehrschluss tritt Inkonsistenz dann auf, wenn Widersprüche auftreten. Es ist jedoch wichtig zu betonen, dass zwei Dinge nur dann widerspruchsfrei/konsistent oder widersprüchlich/inkonsistent sein können, wenn sie in irgendeiner Weise etwas miteinander zu tun haben bzw. miteinander verbunden – also kohärent – sind.

[Kemper und Eickler 99] definieren Konsistenz vor dem Hintergrund einer Transaktion in einer Datenbank:

„Eine Transaktion hinterlässt nach Beendigung einen konsistenten Datenbasiszustand ... Zwischenzustände, die während der Transaktionsbearbeitung entstehen, dürfen inkonsistent sein, aber der resultierende Endzustand muss die im Schema definierten Konsistenzbedingungen erfüllen.“

Unter Konsistenzbedingungen versteht [Bill 99a] in Bezug auf GIS in erster Linie geometrische bzw. topologische Beziehungen, um Daten auf Eindeutigkeit überprüfen zu können. In diesem Zusammenhang wird ein maßgebliches Konsistenztheorem, der Satz von Euler für zusammenhängende, planare Graphen, vorgestellt:

Für jede zusammenhängende Abbildung (planarer Graph) mit p Knoten, l Kanten und f Flächen gilt:

$$C = p - l + f = 2$$

[Bartelme 00] führt semantische Konsistenzbedingungen auf, z.B. dürfen Straßen nicht durch Häuser führen, Gebäudeobjekte dürfen sich nicht überlappen, etc. Aus seiner Sicht ist die Datenkonsistenz eine der wichtigsten Forderungen, die an ein Informationssystem gestellt werden muss.

Den obigen Ausführungen ist zu entnehmen, dass Konsistenz sowohl im Datenbank- als auch im GIS-Bereich bedeutet, dass die Datenablage mit der internen Schemabeschreibung eines Informationssystems übereinstimmt (vertikale Konsistenz). Konsistenz kann darüber hinaus aber auch innerhalb verschiedener

Ebenen hergestellt werden (horizontale Konsistenz). Im Rahmen dieser Arbeit wird sowohl die horizontale Konsistenz/Inkonsistenz zwischen verschiedenen Schemas als auch die horizontale Konsistenz/Inkonsistenz auf Datenebene betrachtet (vgl. Abbildung 4-2). Dabei ist die horizontale Konsistenz (bzw. Inkonsistenz) insbesondere als Widerspruchsfreiheit (bzw. Widersprüchlichkeit) zwischen mehrfachen Repräsentationen eines Realweltobjektes oder zwischen Mehrfachkonzeptionen eines Objekttyps der Realwelt aufzufassen.

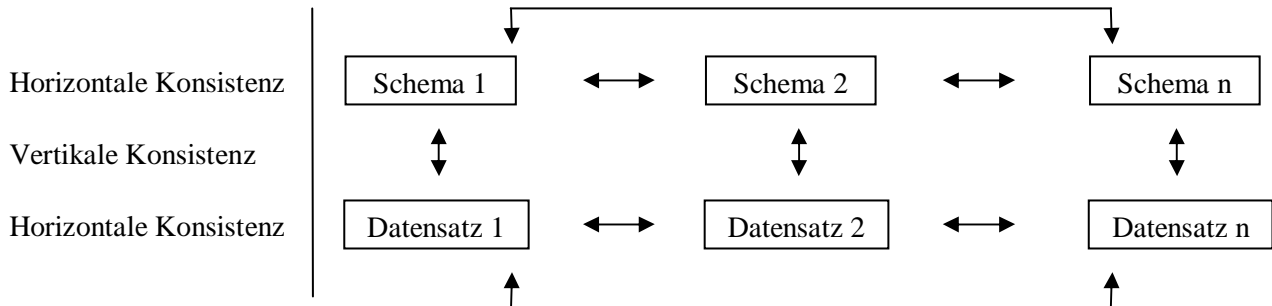


Abbildung 4-2: Verschiedene Arten der Konsistenz

4.1.2.4 Mehrfachrepräsentationen und Redundanz

Sobald Mehrfachrepräsentationen auftreten, entsteht auch stets ein gewisses Maß an Redundanz. Handelt es sich bei den Mehrfachrepräsentationen um Duplikate, so ist die Redundanz vollständig. Sobald aber nur partielle Übereinstimmungen bzw. Redundanzen zwischen verschiedenen Repräsentationen existieren, führt dies in der Regel zu Inkonsistenzen, sofern die korrespondierenden Instanzen innerhalb eines Informationssystems verwaltet werden.

4.2 Integration von Mehrfachrepräsentationen

Die folgenden Abschnitte sollen aufzeigen, was unter dem Begriff „Integration“ im Rahmen dieser Arbeit zu verstehen ist und weshalb eine Integration von Mehrfachrepräsentationen im GIS-Bereich von großer Bedeutung ist.

4.2.1 Zum Verständnis des Begriffs „Integration“ innerhalb dieser Arbeit

Zunächst soll noch einmal betont werden, dass die Integration von raumbezogenen Datenbanken auf 3 verschiedenen Ebenen geschehen kann (siehe Abbildung 4-3). Die oberste Ebene ist die Schema-Ebene (1), als zweites ist die Objekt- bzw. Instanzebene zu nennen (2) und schließlich kann eine Integration auch nur auf geometrischer Ebene (3) stattfinden.

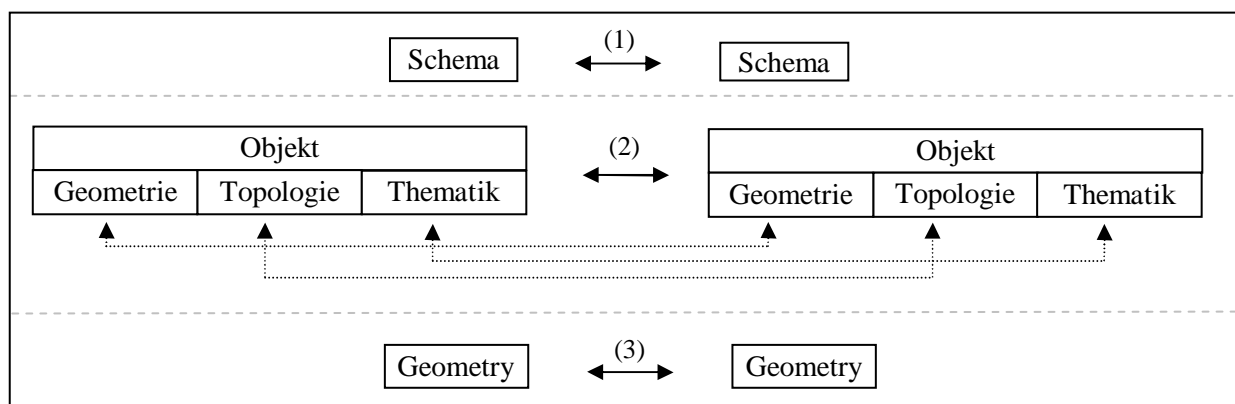


Abbildung 4-3: Verschiedene Ebenen der Integration in GIS

Integration wird im Rahmen dieser Arbeit in Bezug auf Mehrfachrepräsentationen auf Objektebene als Überbegriff für zwei verschiedene Vorgänge verwendet: Zum einen ist unter Integration die komplette Verschmelzung (Conflation) mehrerer Instanzen zu einem resultierenden Objekt (Mono-Repräsentation) zu

verstehen. Zum anderen steht der Terminus auch für die Abspeicherung von Beziehungen bzw. Relationen zwischen Mehrfachrepräsentationen (siehe Abbildung 4-4). Im einfachsten Fall können diese Beziehungen in Form einer Liste ausgedrückt werden. In dieser Arbeit sollen Relationen zwischen mehrfach repräsentierten Objekten jedoch detaillierter in Form von expliziten Relations-Objekten beschrieben werden.

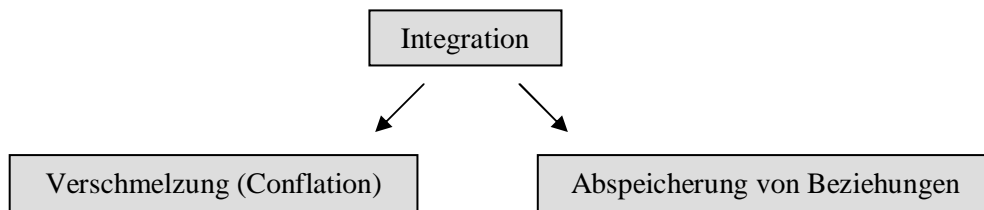


Abbildung 4-4: Integration als Überbegriff für zwei verschiedene Prozesse

Grundsätzlich wird in dieser Arbeit zwischen horizontaler und vertikaler Integration von Mehrfachrepräsentationen unterschieden. Bei der *horizontalen* Integration sollen Daten, die in ähnlichen Maßstabsbereichen erfasst wurden, einander zugeordnet bzw. fusioniert werden. Andererseits versucht die *vertikale* Integration, Geodaten aus verschiedenen Maßstabsbereichen aufeinander abzubilden und zu verschmelzen.

Die Aussagen zur Integration von Mehrfachrepräsentationen auf Instanzebene können prinzipiell auch auf die Integration von Mehrfachkonzeptionen übertragen werden. Allerdings ist unter Schemaintegration in der Literatur ausschließlich die Zusammenführung (bzw. die Verschmelzung) bestehender lokaler Schemas in ein globales Schema zu verstehen, während die Erzeugung von Korrespondenzbeziehungen zwischen den Objektklassen verschiedener Schemas, das so genannte Schema-Matching, lediglich als ein Teilschritt des Integrationsprozesses betrachtet wird [Conrad 02b].

4.2.2 Sinn und Zweck der Integration

Weshalb muss eine Integration von mehrfach repräsentierten Objekten erfolgen? Grundsätzlich lassen sich hierfür verschiedene Gründe anführen. [Sester et al. 03] nennen u.a. die folgenden Aspekte:

1. Die Informationen aus integrierten Datensätzen können sich ergänzen und so für ein breiteres Spektrum von Anwendungen bzw. Analysen zur Verfügung stehen. So können beispielsweise Attribute und Attributwerte von einem Bestand in einen anderen übertragen werden.
2. Während des Integrationsprozesses können Verlässlichkeit und Qualität von Datenbeständen verifiziert bzw. Inkonsistenzen und damit auch Datenfehler aufgedeckt werden.
3. Durch den Integrationsprozess kann eine gemeinsame und automatisierte Fortführung ursprünglich getrennter Datenbestände erfolgen, wodurch eine erhebliche zeitliche und damit auch finanzielle Einsparung beim Aktualisierungsvorgang zu erzielen ist.
4. Bereits vorhandene Integrationsergebnisse können auch dazu verwendet werden, zukünftige Integrationsverfahren zu optimieren.

Zusammenfassend bemerken [Sester et al. 03]: „Basically, this means that new data acquisition – typically the most expensive part of spatial analysis tasks – can be largely reduced and is only required if no data are available or changes in the reality have occurred. Consequently, a considerable saving of cost and labour is obtained by adding significant value to the existing data.“

Im Folgenden soll erläutert werden, welche Integrationsszenarien für die Nexus-Plattform eine besonders wichtige Rolle spielen.

4.2.2.1 Grund 1: Analyse von Mehrfachrepräsentationen

Die Integration von Mehrfachrepräsentationen ist dann notwendig, wenn für ein geographisches Anfragegebiet kein einzelner Datensatz existiert, der den Anforderungen einer Analyse genügt. Dieser Fall kann einerseits dann eintreten, wenn das Anfragegebiet nicht komplett durch *einen* Datensatz bzw. eine Augmented Area abgedeckt wird. Zur Lösung dieser Problematik muss – analog zur Anfelderung bzw. zum Mosaiking

von Bilddaten – ein einheitliches Anfragegebiet generiert werden. Hierbei sind nur die Objekte, die im Angrenzungs- bzw. im Überlappungsgebiet zweier Datensätze mehrfach repräsentiert sind, zu integrieren. Insofern handelt es sich in diesem Fall gemäß der Terminologie dieser Arbeit lediglich um eine *partielle Integration* (vgl. Abbildung 4-5, Fall 1). Andererseits könnte zwar ein Datenbestand für das gesamte Anfragegebiet verfügbar sein, wenn dieser aber nicht vollständig ist oder nicht alle für die Analyse notwendigen Informationen enthält, ergibt sich ebenfalls die Notwendigkeit zur Datenintegration. In diesem Fall wird in dieser Arbeit von *kompletter Integration* gesprochen. (siehe Abbildung 4-5, Fall 2).

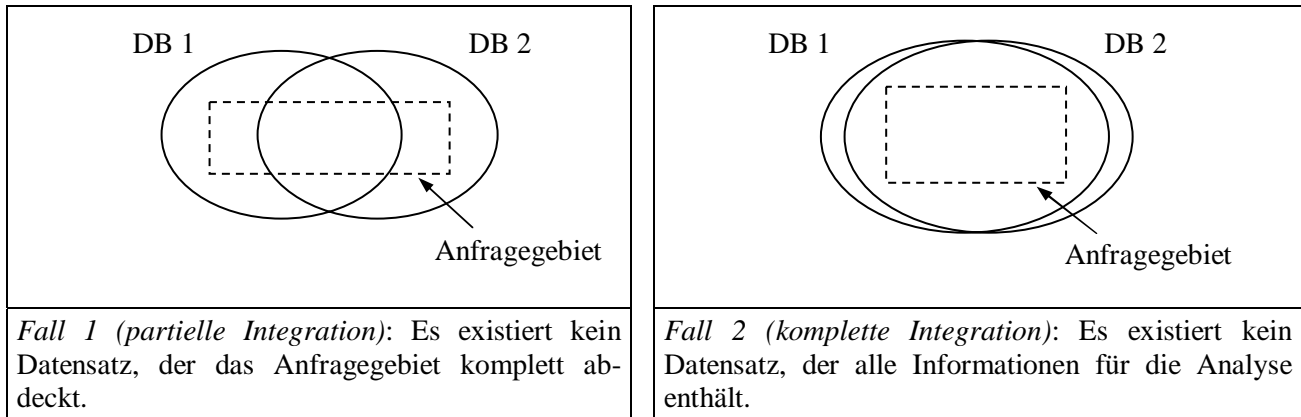


Abbildung 4-5: Verschiedene Gründe für die Datenintegration zum Zwecke der Analyse

Ein Szenario für Fall 2 aus Abbildung 4-5 wäre z.B. folgendes: Man kann sich vorstellen, dass für eine Analyseaufgabe sämtliche Straßen und Wege eines geographischen Gebietes benötigt würden und ebenfalls auf die Straßennamen zurückgegriffen werden soll. Stünde nun sowohl ein GDF- als auch ein ATKIS-Datensatz für das gewünschte Gebiet zur Verfügung, wobei die ATKIS-Straßendaten noch keine Straßennamen enthielten und in der GDF-Repräsentation die Feldwege fehlten, so könnte durch eine Kombination beider Quellen ein optimaler Ausgangsbestand für die Analyse generiert werden.

Wenn für einen Teilbereich sehr viele Datensätze vorliegen, die Mehrfachrepräsentationen enthalten, können die Konflikte zwischen den Mehrfachrepräsentationen evtl. nicht mehr adäquat behandelt und aufgelöst werden. Dies führt zu einer Situation, die als „Information Overload“ [Wiederhold 99] charakterisiert werden kann.

4.2.2.2 Grund 2: Fortführung und Konsistenz von Mehrfachrepräsentationen

Wenn ein Realweltobjekt mehrfach innerhalb einer Dateninfrastruktur wie Nexus vorliegt und an einer der betreffenden Repräsentationen eine Änderung durchgeführt wird, kann diese an die korrespondierenden Instanzen weitergegeben werden. Dies funktioniert selbstverständlich nur, wenn die geänderte Repräsentation von der Existenz weiterer Repräsentationen weiß. Dieses Wissen kann nur dann vorhanden sein, wenn im Vorfeld der Fortführung eine Abspeicherung der Links bzw. Relationen zwischen unterschiedlichen Repräsentationen stattgefunden hat.

Außerdem können unter Zuhilfenahme expliziter Relationen Mechanismen eingeführt werden, die die Überprüfung der Stimmigkeit von Repräsentationen innerhalb eines Gesamtdatenbestandes zum Ziel haben. Ist beispielsweise in zwei Datensätzen A und B neben einem Gebäudeobjekt noch eine Garage vorhanden und existiert diese in Datenbank C nicht, so hat man (in der Regel) ein Indiz dafür gefunden, dass Datenbank C nicht mit dem Zustand der Realwelt konsistent ist.

4.2.2.3 Grund 3: Optimierung von nachfolgenden Integrationsprozessen

Das Systemkonzept der Nexus-Plattform sieht vor, dass unterschiedlichste Datenanbieter ihre Geodaten unabhängig voneinander an die Plattform übergeben können. Unter der Voraussetzung, dass bereits Zuordnungsbeziehungen zwischen zwei oder mehr Datensätzen innerhalb der Plattform bestehen, kann bei Hinzukommen eines neuen, das gleiche Gebiet abdeckenden und die gleichen Objekte enthaltenden Datensatzes zum Nexus-Datenbestand, ein erweiterter Ansatz verfolgt werden: Der neu hinzukommende Datensatz wird

zunächst einem Zuordnungsprozess zu *einem* bereits in der Plattform vorhandenen Nexus-Datensatz unterzogen. Ist dieser abgeschlossen, so ist es möglich, dessen Ergebnisse bei den Zuordnungsprozessen mit den anderen in Nexus registrierten Datensätzen auszunützen. Dieser Ansatz erfordert also ebenfalls eine explizite Abspeicherung von Links bzw. Beziehungen zwischen Mehrfachrepräsentationen.

In Abbildung 4-6 wird die Idee der Vorgehensweise noch einmal verdeutlicht: Ein Objekt A1 aus Datenbank A hat bereits eine Zuordnung zu Objekt B1 aus Datenbank B und zu Objekt C1 aus Datenbank C erfahren. Soll nun eine Integration von Datenbank B und Datenbank C durchgeführt werden, können die bereits existierenden Integrationsergebnisse verwendet werden, um die neuen Zuordnungen zu fundieren. Dies geschieht folgendermaßen: Da die Beziehungen von A1 zu B1 und C1 feststehen, ergibt sich bei einer Integration der Datenbanken B und C aufgrund der zu erwartenden Transitivität der Zuordnungen eine höhere Wahrscheinlichkeit dafür, dass die Objekte B1 und C1 zuzuordnen sind, als dass B1 und C2 korrespondierende Instanzen darstellen.

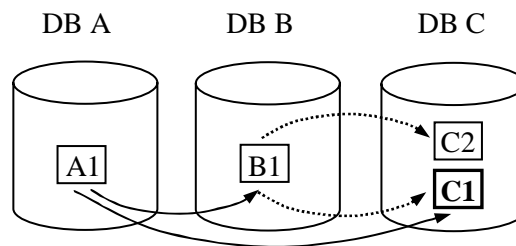


Abbildung 4-6: Ausnützung bestehender Integrationsergebnisse, um nachfolgende Prozesse verlässlicher zu machen

4.3 Mehrfachrepräsentationen in GIS – Stand der Forschung

Das Forschungsthema „Mehrfachrepräsentationen“ spielt auf dem Gebiet der Geoinformatik seit einigen Jahren eine wichtige Rolle. Mit dem Aufkommen des Internets und den dadurch bedingten Möglichkeiten zum einfachen Austausch von Geodaten sowie aufgrund der wachsenden Interdisziplinarität von GIS-Projekten erhofft man sich durch die gemeinsame Fortführung und Analyse von Mehrfachrepräsentationen eine erhebliche Effizienzsteigerung. Die Thematik ist jedoch sehr komplex und deshalb gelingen Fortschritte nur langsam. Bisher wurde der Bereich der Zuordnung von Mehrfachrepräsentationen am intensivsten untersucht. Mehr und mehr spielen auch Gesichtspunkte der semantischen Interoperabilität sowie der Schemaintegration eine Rolle. Auf den Gebieten Analyse und Fortführung von Mehrfachrepräsentationen, die im engeren Sinne als Anwendungen des Forschungsbereiches der Mehrfachrepräsentationen aufgefasst werden können, existieren hingegen nur erste Ansätze. Die folgenden Abschnitte sollen einen Überblick über den bisherigen Stand der Forschung geben.

4.3.1 Semantische Integration: Schemaintegration und Ontologien

Die Forschung im GIS-Bereich setzt sich mehr und mehr mit dem Problem der unterschiedlichen Semantik von Geodaten auseinander. Viele Autoren messen der semantischen Interoperabilität eine Schlüsselfunktion auf dem Weg zur Erreichung eines integrierten Geo-Informationskosmos bei [z.B. Sheth 99, Bishr 98, Harvey 99]. Meist werden auf dem Gebiet der Semantischen Geodatenintegration Formalismen verwendet oder entworfen, um die Semantik von Objekten auf einer Metaebene zu beschreiben, auf der diese dann einem Vergleich unterzogen werden können. [Uitermark et al. 99a] bezeichnen die Integration der Semantik von Geodaten als Kommunikationsprozess, da zwei Kommunikationspartner ein gemeinsames Verständnis von einem Sachverhalt haben müssen, um sich über diesen unterhalten zu können.

Für die Auflösung der semantischen Heterogenität gibt es verschiedene Ansätze. Im Anschluss werden zunächst die aus Datenbank-Perspektive entwickelten Konzepte des Schema-Matchings und der Schemaintegration vorgestellt. Das Problem der Integration von Semantik wird in vielen Fachdisziplinen auch mit Hilfe von Ontologie-basierten Verfahren zu lösen versucht. Im Forschungsbereich Geo-Informationssysteme existieren ebenfalls solche Ansätze, die in der Folge beleuchtet werden.

4.3.1.1 Schema-Matching und Schemaintegration

[Do & Rahm 02] definieren den Prozess des Schema-Matchings als die Aufgabe, semantische Korrespondenzen zwischen Elementen zweier Datenbank-Schemas zu finden. Sie stellen eine Methodik vor, um diesen Prozess so weitgehend wie möglich zu automatisieren. Diese Methodik wurde innerhalb einer generischen Plattform namens COMA (Combining Match Algorithms) implementiert. Sie erlaubt es, verschiedene Matching-Algorithmen zu kombinieren, um so optimale Ergebnisse zu erzielen. COMA unterstützt sowohl verschiedene Applikationen als auch verschiedene Schema-Typen wie z.B. relationale Schemas und XML-Schemas. Die Plattform basiert auf einer erweiterbaren Bibliothek von Matching-Algorithmen. Diese werten die verschiedenen Arten von Schemainformationen wie Elementnamen, Datentypen sowie weitere strukturelle Eigenschaften entweder einzeln oder in Kombination aus. COMA beinhaltet eine interaktive Komponente, da ein vollautomatisches Schema-Matching aufgrund der semantischen Diversität von Schemas laut Auffassung der Autoren nicht erzielt werden kann. Außerdem erlaubt das System die Wiederverwendung von Matching-Ergebnissen bei nachfolgenden Schema-Matching-Verfahren.

[Rahm & Bernstein 01] stellen u.a. die Anwendungsdomänen für das Schema-Matching dar. Als wichtigste gilt die Schemaintegration. Aus Datenbanksicht geht es bei der Schemaintegration darum, eine globale Sicht auf unabhängig voneinander entwickelte Schemas herzuleiten. Es wird darauf hingewiesen, dass dies im Wesentlichen mit dem Problem der Integration von heterogenen Ontologien in eine einzelne Ontologie innerhalb eines KI (Künstliche Intelligenz)-Umfeldes gleichzusetzen ist. Im Zuge der Schemaintegration müssen in einem ersten Schritt die Beziehungen zwischen den verschiedenen Ausgangsschemas aufgedeckt werden. Diese Aufgabe kommt dem Schema-Matching zu. Im Rahmen der erwähnten Arbeit werden außerdem die verschiedenen bislang verfügbaren Verfahren des automatisierten Schema-Matchings erläutert. Dabei wird u.a. zwischen instanz- und schemabasierten Ansätzen unterschieden. Während die *instanzbasierten* Ansätze die konkreten Daten berücksichtigen, werden bei den *schemabasierten* Lösungen nur die Informationen auf Schemaebene ausgewertet. Außerdem können innerhalb von Schemas entweder nur einzelne Elemente oder auch komplexere Strukturen wie z.B. Elementkombinationen zugeordnet werden. Schema-Matching Ansätze unterscheiden sich ferner darin, welche Kardinalitäten sie unterstützen und ob während des Matching-Vorgangs auch Ergebnisse früherer Matchings, Benutzereingaben oder weitere Informationen einbezogen werden können.

Laut [Conrad 02b] besteht die Schemaintegration darin, unterschiedliche Beschreibungen vorhandener Datenbestände bzw. unterschiedliche Datenbankschemas in *ein* konzeptionelles Schema des Gesamtdatenbestandes zusammenzuführen. Er unterscheidet dabei zwischen physischer und logischer Integration. Bei der physischen Integration werden Daten kopiert und in einer neuen Datenbank abgelegt. Diese Vorgehensweise wird beim Aufbau von so genannten Data Warehouses verfolgt. „Bei einer logischen Integration bleiben die Daten in ihren ursprünglichen Quellen erhalten und werden nur bei Anfragen – soweit zur Beantwortung der Anfragen notwendig – zusammengeführt“ [Conrad 02a]. Die Schemaintegration wird in [Conrad 02b] anhand der 5-Ebenen-Schemaarchitektur nach [Sheth & Larson 90] beschrieben. Auf der untersten Ebene sind dabei die lokalen Schemas der verschiedenen Ausgangsdatenbanken angesiedelt. Sie werden über verschiedene Übergangsstadien in das globale, föderierte Schema integriert, wobei zunächst die Wahl eines geeigneten Datenmodells (z.B. relational oder objektorientiert) erfolgen muss. Der Autor weist außerdem auf die Konflikte hin, die bei der Schemaintegration auftreten (siehe unten bei [Spaccapietra et al. 96]). Schließlich wird in [Conrad 02b] u.a. die Technik der *zusicherungsbasierten* Schemaintegration erläutert. Hier versucht man, paarweise Beziehungen zwischen den Elementen zweier Schemas, so genannte Zusicherungen oder auch Inter-Schema-Korrespondenzen, zu finden und über bestimmte Integrationsregeln das integrierte Schema aufzubauen. Dabei existieren folgende Arten von Korrespondenzen:

- Element-Korrespondenzen (zwischen zwei Objekttypen),
- Element-und-Attribut-Korrespondenzen (zwischen Objekttypen und Attributen),
- Pfad-Korrespondenzen (zwischen mehreren Objekttypen).

Mit der Aufdeckung der Korrespondenzen ist noch nicht festgelegt, wie die Schemas integriert werden sollen. Daher müssen Integrationsregeln festlegen, wie die korrespondierenden Schemaelemente auf das integrierte Schema abgebildet werden.

Eine häufig zitierte Arbeit auf dem Gebiet der Integration raumbezogener Datenbanken haben [Devoegele et al. 98] veröffentlicht. Sie weisen darauf hin, dass eine Datenbankintegration nicht nur die Schemaintegration, sondern auch die Integration der Dateninstanzen umfasst. Der Integrationsprozess wird hier in drei Schritte unterteilt. Zunächst werden im Rahmen der *Schemavorbereitung* sämtliche Aktivitäten durchgeführt, um vorhandene Datenbeschreibungen in ein einheitliches Format zu bringen, wobei dies auch den geometrischen Abgleich, z.B. die Konvertierung von Raster- in Vektordaten oder die Repräsentation von Daten in einem einheitlichen Referenzsystem betrifft. Als zweiter Schritt wird eine *Untersuchung von Korrespondenzen* sowohl der unterschiedlich modellierten Objektklassen als auch der verschiedenen Instanzen durchgeführt. Schließlich folgt in der letzten Phase die eigentliche *Integration*, innerhalb derer die vorhandenen Konfliktsituationen zwischen den unterschiedlichen Objektklassen einer Entität der Realwelt aufgelöst werden. Aus dieser ergibt sich dann eine integrierte Beschreibung vorhandener Datenquellen in Form eines Meta-Schemas.

[Spaccapietra et al. 96] schlagen in einem ähnlichen Ansatz vor, autonome, heterogene Datenbanken innerhalb einer virtuellen Datenbank logisch zu vereinen und auf diese Weise eine Datenbankintegration durchzuführen. In der Arbeit werden in erster Linie die Konflikte aufgedeckt, die bei der Datenbankintegration auftreten können und es werden Konfliktlösungsstrategien angeboten:

1. **Klassifikationskonflikte:** Gleiche Entitäten der Realwelt werden durch unterschiedliche Objektklassen, die auch einen unterschiedlichen Detaillierungsgrad aufweisen können, beschrieben (z.B. Gehweg und Verkehrsweg). Zur Konfliktlösung kann das integrierte Schema dann nach dem Prinzip der Minimalität (Verschmelzung der Klassen aller Ausgangsschemas in eine generische und globale Klasse) oder nach dem Prinzip der Vollständigkeit (alle Klassen sämtlicher Ausgangsschemas sollen im integrierten Schema erhalten bleiben) angelegt werden.
2. **Beschreibende Konflikte:** Korrespondierende Typen in unterschiedlichen Datenbanken haben unterschiedliche Eigenschaften oder korrespondierende Eigenschaften mit unterschiedlichen Charakteristika (z.B. unterschiedliche Geometrien). Zur Konfliktlösung kann das integrierte Schema einem Objekt erlauben, verschiedene Objekteigenschaften (z.B. verschiedene Geometrien) anzunehmen.
3. **Strukturelle Konflikte:** Gleiche Realweltobjekte werden mittels unterschiedlicher Modellierungskonzepte dargestellt (semantischer Relativismus). Z.B. können Attribute eines Linienzugs entweder dynamisch segmentiert sein oder der Linienzug ist in entsprechende Abschnitte unterteilt, die einen Attributwechsel anzeigen (vgl. Kapitel 5.1). Um eine einheitliche Darstellung zu erreichen, wäre eine Lösungsmöglichkeit denkbar, welche die unterschiedlichen Teilabschnitte aggregiert und deren Attribute an ein resultierendes Gesamtobjekt übergibt, das wiederum dynamisch segmentiert wird.
4. **Fragmentierungskonflikte:** Ein Realweltobjekt wird in einer Datenbank über mehrere Objekte abgebildet, in der anderen als einzelnes Objekt. Um dieses Problem aufzulösen, müssten ebenfalls Aggregationskonzepte in das resultierende Schema integriert werden.

4.3.1.2 Ontologie-basierte Konzepte

Um das Wissen zwischen verschiedenen Anwendungsbereichen bzw. Nutzergemeinden austauschen und Interoperabilität zwischen verschiedenen Geodatenbanken herstellen zu können, ziehen viele Ansätze Ontologien heran. [Guarino 98] definiert den Begriff „Ontologie“ als logische Theorie, die die beabsichtigte Bedeutung eines bestimmten, formalen Vokabulars erläutert. [Rodriguez et al. 96] stellen fest, dass Ontologien für die konzeptionellen Dimensionen geographischer Objekte herstellbar sind, d.h. also für deren Geometrie, deren Topologie, deren Thematik und für deren symbolische Repräsentation. [Fonseca et al. 00 und 03] weisen darauf hin, dass es verschiedene Ebenen der Ontologien geben kann, die sich aufgrund ihrer (semantischen und räumlichen) Granularität differenzieren lassen.

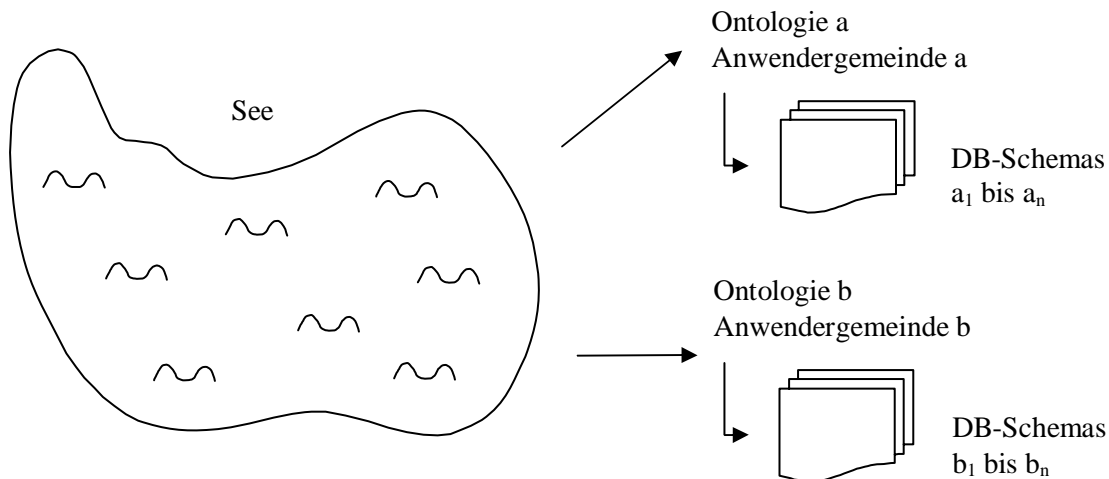


Abbildung 4-7: Ontologien und DB-Schema

Ein Beispiel für die Verwendung von Ontologien ist die Arbeit von [Fonseca et al. 02]. Unter Ontologien werden dabei die verschiedenen Theorien unterschiedlicher Anwenderkreise über einen bestimmten geographisch relevanten Sachverhalt (wie beispielsweise einen See) verstanden, die über ein spezifisches Vokabular in Form von Klassen mit zugehörigen Attributen und Methoden formal beschrieben werden können. Es wird darauf hingewiesen, dass Ontologien Konzepte über bzw. Sichten auf ein Realweltobjekt darstellen, unabhängig davon, ob dieses digital gespeichert ist, wogegen Datenbank-Schemas (DB-Schemas) repräsentieren, was in einer Datenbank abgelegt wird. Pro Anwendungsbereich gibt es daher auch nur eine Ontologie über ein Objekt (die allerdings in unterschiedliche Detailstufen unterteilt werden kann, s.u.), während es jedoch verschiedenste DB-Schemas geben kann (siehe Abbildung 4-7). Ontologien haben somit eine reichere Semantik als Datenbank-Schemas.

Die Arbeit sieht vor, dass die Ontologien wie in einem Browser durchsucht werden können, so dass sich die Benutzer eines Systems über dessen Informationsinhalte informieren können. Ontologien werden dabei auch auf unterschiedlichen Detailebenen angelegt, um den jeweiligen Applikationsbedürfnissen zu genügen. So könnten die in Abbildung 4-7 dargestellten, so genannten Top-Level-Ontologien (Ontologien a und b), von den Nutzergemeinden über Vererbung weiter differenziert werden (z.B. Ontologien a1 bis an), so dass eine Multi-Ebenen-Ontologiestruktur entsteht.

Sobald eine Ontologie spezifiziert ist, kann sie in Klassenstrukturen übersetzt und auf einem Ontologieserver abgelegt werden. Der Ontologieserver kann dann vom Benutzer über eine graphische Oberfläche durchsucht werden. Dabei wird beispielsweise eine Definition eines bestimmten Objektes (im Beispiel also eines Sees) dargestellt, es werden seine Rollen als „geographische Region“ oder als „Oberflächenform“ sowie seine Attribute (z.B. Ort, Säuregehalt, etc.) und Funktionen (z.B. schwimmen, fischen, etc.) angezeigt. Will der Benutzer nun entsprechende See-Objekte eines bestimmten Gebietes zur Verfügung gestellt bekommen, werden so genannte Mediatoren gestartet, die in den angeschlossenen Datenbanken alle Objekte suchen, welche der spezifizierten Ontologie-Klasse entsprechen. Diese werden dann an den Benutzer zurückgeliefert und können von ihm analysiert werden.

[Fonseca et al. 03] erweitern den soeben vorgestellten Ansatz im Hinblick auf die Betrachtung unterschiedlicher semantischer Detailstufen von Ontologien. Sie stellen die naheliegende Aussage unter Beweis, dass die Übereinstimmungen zwischen verschiedenen Informationsgemeinschaften in größeren Detailstufen größer sind als in detaillierteren. Außerdem führen die Autoren das so genannte 5-Universen-Paradigma ein, das die konzeptionellen Charakteristika geographischer Objekte auf 5 Abstraktionsebenen beschreibt und somit die Grundlage für Ontologie-gestützte GIS darstellt. Die 5 Ebenen bzw. Universen sind:

- das physische Universum,
- das kognitive Universum,
- das logische Universum,

- das Repräsentations-Universum und
- das Implementierungs-Universum.

Ein geographisches Objekt in der physischen Welt wird vom kognitiven System eines Menschen erfasst und im menschlichen Gehirn gespeichert. Die Formalisierung der Konzepte, die in einem menschlichen Gehirn über ein Objekt bestehen, geschieht im logischen Universum. Hier werden die Ontologien gebildet und es wird bereits zwischen generischeren High-Level- und spezialisierteren Low-Level-Ontologien unterschieden. Innerhalb des Repräsentations-Universums werden die im logischen Universum abstrahierten Konzepte auf finite symbolische Beschreibungen von Elementen der geographischen Welt abgebildet, d.h. es kommen dann die Besonderheiten des Raumbezugs zur Ontologiebildung hinzu, z.B. das geographische Bezugssystem oder die Differenzierung der Repräsentation in Felder und Objekte [siehe Molenaar 98]. Dementsprechend spiegeln die hier angelegten Ontologien die verschiedenen Dimensionen und Eigenschaften geographischer Phänomene wider. Das Repräsentations-Universum ist mittels „Semantischer Mediatoren“ mit dem logischen Universum verknüpft. Beispielsweise kann ein Objekt „Wasserreservoir“ einerseits auf logischer Ebene als Konzept existieren, eine Repräsentation davon kann andererseits als Luftbild, 2D-Polygon oder als Digitales Oberflächenmodell vorliegen. Die Umsetzung des Repräsentations-Universums in ein Computersystem findet letztlich innerhalb des Implementierungs-Universums statt.

Der Einsatz von Ontologien wird auch in weiteren Arbeiten als Lösungsmöglichkeit für die Aufhebung der semantischen Heterogenität von Geodaten gesehen. [Hakimpour & Timpf 01] machen in ihrer Veröffentlichung eine wichtige Bemerkung, die die Integration bzw. Interoperabilität unterschiedlicher GIS im Allgemeinen betrifft. Es werden hier grundsätzlich zwei Fälle unterschieden:

1. Datenbankschemas basieren auf derselben Ontologie. Hier ist das Problem darauf beschränkt, Synonyme und Homonyme zwischen den Terminologien der unterschiedlichen Schemas zu finden.
2. Datenbankschemas basieren auf unterschiedlichen Ontologien. Dieses Problem ist dadurch lösbar, dass eine gemeinsame Ontologie („Shared Ontology“) entwickelt wird. Hierzu müssen die Ähnlichkeiten zwischen verschiedenen Konzepten aufgedeckt werden. Wie bereits oben erwähnt ist dies laut [Rahm & Bernstein 01] identisch mit der Problematik der Schemaintegration.

Für den zweiten Fall entwerfen die Autoren eine Beschreibungslogik (Description Logic, DL) für die Repräsentation von Ontologien. Jede Nutzergemeinde kann ihre Wahrnehmung von Objekten mittels DL beschreiben. Auf der Basis der Beschreibungslogik kann dann ein Vergleich zweier Ontologien durchgeführt werden. Daraufhin ist es möglich, eine gemeinsame Ontologie ableiten zu können, was anhand verschiedener Konzepte über das Objekt „Straße“ illustriert wird.

[Uitermark 96] macht folgende wichtige Bemerkung zur Geodatenintegration, die auch für die vorliegende Arbeit als Grundlage dient: „*Geographic Data Set Integration (or Map Integration) is the process of establishing relationships between corresponding object instances in different, autonomously produced, geographic data sets of a certain region. The purpose of geographic data set integration is to share information between different geographic information sources.*“ Dieser Ansicht gemäß untersuchen [Uitermark et al. 99b] die semantischen und räumlichen Relationen zwischen voneinander unabhängig erfassten Geodaten des gleichen Gebietes und verwenden hierzu ebenfalls Ontologien. Ihr Ziel besteht darin, den manuellen Eingriff bei der Fortführung korrespondierender, semantisch ähnlicher Objektinstanzen zu minimieren.

Zunächst werden hier zwei verschiedene Arten von Ontologien unterschieden: Zum einen gibt es die *Domänen-Ontologien* für bestimmte Disziplinen, in denen allgemeine Definitionen für die betrachteten Objekte (in der Disziplin der topographischen Informationssysteme z.B. für Straßen, Gebäude, etc.) gegeben sind. Sie werden durch *Applikations-Ontologien* ergänzt, die quasi für jeden untersuchten Datensatz existieren und jeweils semantisch abweichende Konzepte über identische Objekte aus individuellen Anwendungsperspektiven enthalten können. Die Applikations-Ontologien beschreiben somit die spezifischen Konzepte, die in den tatsächlich vorhandenen Datenbeständen realisiert sind. Zwischen den Domänen- und Applikations-Ontologien bestehen nun Relationen, die in Form so genannter Abstraktionsregeln ausgedrückt werden. Dabei kann es Äquivalenz- oder Aggregations-Relationen zwischen Klassen der Domänen-Ontologie und Klassen der Applikationsontologie geben. Die semantische Ähnlichkeit von Applikations-Ontologien wird ebenfalls über Relationen ausgedrückt. So stehen Klassen zweier Applikations-Ontologien beispielsweise

miteinander in einer Äquivalenz-Beziehung, wenn sie zur selben Klasse der Domänen-Ontologie äquivalent sind. Es existieren weitere solcher Ähnlichkeitsrelationen. Korrespondierende Objektinstanzen können letztlich dann identifiziert werden, wenn sie zu Applikations-Ontologien gehören, die eine semantische Ähnlichkeit aufweisen und wenn sie am gleichen geographischen Ort in Erscheinung treten.

Bei [Bishr et al. 99a, 99b] werden die unterschiedlichen Interpretationen der Realwelt durch verschiedene Geoinformations-Anwender ebenfalls als wichtiges Hindernis für die Interoperabilität von Geodaten betrachtet. U.a. wird am Beispiel von Straßenverkehrsdaten aus ATKIS und GDF erläutert, weshalb die unterschiedliche Auffassung über Entitäten wie Kreuzungsobjekte innerhalb der beiden Datenschemas als Ausgangspunkt für Konflikte zu sehen ist. Im Rahmen der Untersuchungen werden zwei Komponenten aufgeführt, der „Semantic Mapper“ und der „Semantic Wrapper“, die diese Konflikte aufzulösen in der Lage sind. Der Wrapper verpackt Objekte aus Geodatenbanken so, dass sie einer vom Mapper angebotenen Schnittstelle entsprechen und an ihn weitergeleitet werden können. Diese Schnittstelle ist für jede Anwendungsdomäne wie z.B. Topographie, Transport, Umwelt, etc. unterschiedlich. Die Objekte verfügen innerhalb des Semantic Mappers über einen systemweit eindeutigen Objektidentifikator und enthalten detaillierte Information über ihre Semantik. Werden nun von einem Klienten Objekte aus Geodatenbanken angefordert, kann der Mapper die semantischen Abweichungen zwischen den angeforderten Objekten und jenen, die bereits in der Datenbank des Klienten vorhanden sind, erkennen und auflösen. Der Ansatz von Bishr steht in engem Zusammenhang mit dem Konzept der „Semantic Translators“ des OpenGIS-Konsortiums (vgl. Kapitel 2.4.2.1).

Auch [Kuhn 03] greift im Prinzip den Ansatz des OGC auf und entwirft die Idee der semantischen Referenzsysteme, die, ähnlich den raumbezogenen Referenzsystemen auf der Ebene geometrischer Daten, einen Rahmen bilden, um thematische Daten zwischen unterschiedlichen Informationsgemeinschaften austauschen zu können. Dabei sollen die Semantiken von Geodaten über standardisierte Transformationen (analog zu Koordinatentransformationen von Geometrien) von der einen zur anderen Informationsgemeinschaft übersetzt werden. Innerhalb dieser Gemeinschaften kann die Komplexität von Semantiken über Projektionen reduziert werden. Die Anforderungen semantischer Referenzsysteme gehen über die Möglichkeiten, wie sie derzeit zur Formulierung von Ontologien gegeben sind, hinaus. Insbesondere müssen nicht nur die Konzepte über einen Sachverhalt formalisiert werden können, sondern auch die Abbildung von einer Ontologie auf eine andere.

Dass die Integration von Geodaten nicht immer unproblematisch ist und daraus durchaus auch ein Semantikverlust resultieren kann, erwähnen [Bernard & Pundt 98]. Sie weisen darauf hin, dass Ausgangsdaten im Zuge der Integration mehrfach verändert und somit auch verfälscht werden, so dass sich deren Semantik ändert. Nur wenn diese Semantikveränderung berücksichtigt wird, sind anschließend Aussagen über die Qualität von Analyseergebnissen, die auf Basis veränderter Ausgangsdaten erzeugt wurden, möglich. Dementsprechend fordern die Autoren, dass Methoden entwickelt werden müssen, um den Semantikverlust zu minimieren oder zumindest, um angeben zu können, wo es zu Semantikverlust kommen kann und welche Konsequenzen dieser für die erzeugten Ergebnisse hat. Die Problematik wird anhand eines Beispiels aus dem Bereich der GIS-gestützten Klimaplanung skizziert.

4.3.2 Datenverwaltung von mehrfach repräsentierten Objekten

Wie in [Parent et al. 00] bemerkt wird, stellen bisher existierende Geo-Informationssysteme bzw. Geodatenbanken keine Datenschemas und Datenverwaltungs-Mechanismen zur Verfügung, die in der Lage sind, Mehrfachrepräsentationen adäquat abzubilden und zu behandeln und somit den Anforderungen verschiedener Nutzer zu genügen. Es wird festgestellt, dass die bisherigen Datenmanagement-Technologien auf einem zentralisierten Repräsentationsparadigma beruhen. Dieses impliziert, dass alle Eigenschaften auf konzeptioneller Ebene einer einzelnen Repräsentation beigegeben werden. Aus dieser Repräsentation können dann zwar über View-Mechanismen jene Sichten auf ein Objekt generiert werden, die von einer speziellen Anwendung gefordert werden. Jedoch bergen diese View-Mechanismen starke systematische Beschränkungen. Wird beispielsweise ein View aktualisiert, kann diese Aktualisierung nicht an die ursprüngliche Repräsentation in der Datenbank weitergegeben werden, da keine 1:1-Abbildung von Tupeln des Views und Tupeln in der Datenbank gewährleistet werden kann. Folgerichtig fordern die Autoren: *„In summary, modern data management requires a new representation paradigm where multiple representations of the same phenome-*

non may coexist in a database, and this should be explicitly described and made known to the system so that it may manage the situation accordingly.“

Um die Limitierungen bestehender Systeme zu überwinden, versuchten die Autoren daher innerhalb des MurMur Projektes [MurMur 02], die Funktionalität von GIS bzw. von raumbezogenen Datenbankmanagement-Systemen (DBMS) so zu erweitern, dass sie flexiblere Datenmodelle und somit mehrfache, koexistierende Repräsentationen desselben Realweltphänomens unterstützen. Es wurde dabei sowohl eine semantische Flexibilität der Objektrepräsentationen als auch eine kartografische Flexibilität (Repräsentationen in verschiedenen Maßstäben) angestrebt. Der Ansatz des Projektes sieht vor, jede Repräsentation mit einem Stempel zu versehen, der angibt, für welchen Maßstabsbereich, welchen zeitlichen Rahmen, etc. sie verwendbar ist. Die so gekennzeichneten Repräsentationen können dann in einem gemeinsamen Objekttyp vereinigt werden, behalten aber ihre Stempel weiterhin bei. Andererseits kann auch pro Repräsentation ein eigener Objekttyp erzeugt werden. Die verschiedenen Typen können dann über so genannte Inter-Repräsentations-Links miteinander verbunden werden. Die Links können unterschiedlichen Kategorien angehören:

1. Multi-Instanzierungs-Links (ist ein, ist vielleicht ein);
2. Aggregations-Links (verbindet eine Repräsentation mit allen sie konstituierenden Repräsentationen);
3. Identitäts-Links (verbindet zwei Repräsentationen mit gleicher Identität);
4. Set-to-Set-Link (verbindet zwei verschiedene Sets von Repräsentationen).

[Spaccapietra et al. 00] legen ein Rahmenwerk für die Behandlung von Mehrfachrepräsentationen vor. Sie machen drei verschiedene Dimensionen für die Unterschiedlichkeit von Mehrfachrepräsentationen aus und bezeichnen die Problematik demgemäß als multidimensional:

1. Die Dimension der räumlichen Auflösung;
2. Die Dimension der Benutzerperspektive;
3. Die Dimension der Klassifikation.

Ein Punkt innerhalb dieses 3-dimensionalen Problemraumes entspricht folglich der Repräsentation einer bestimmten Objektinstanz als Mitglied einer bestimmten Objektklasse, die aus einem bestimmten Blickwinkel (bzw. einer bestimmten Benutzerperspektive) erstellt wurde und eine definierte Auflösung hat. Die Autoren halten fest, dass es in den verschiedenen Dimensionen grundsätzlich ähnliche Varianten zur Verwaltung von Mehrfachrepräsentationen gibt. Entweder wird ein Objekt als einzelne Repräsentation verwaltet, die beispielsweise verschiedene Geometrien mit verschiedener Auflösung haben oder entsprechend der Benutzersicht verschiedene Rollen annehmen kann. Andererseits besteht die Möglichkeit, dass ein Objekt entsprechend der verschiedenen Rollen bzw. Detaillierungsstufen in multiplen, miteinander verbundenen Repräsentationen abgebildet wird. Diese Mehrfachrepräsentationen können einem einzelnen oder mehreren Schemas entstammen.

4.3.3 Ansätze zur Zuordnung und Verschmelzung von Mehrfachrepräsentationen

Bei der Zuordnung von mehrfach repräsentierten Objekten geht es darum, Instanzen, die partiell oder komplett das gleiche Realweltobjekt beschreiben, einander zuzuordnen. Im Zuge des Prozesses müssen also räumliche Zusammenhänge erfasst werden, was durch computergestützte Prozesse jedoch nur eingeschränkt machbar ist. Folglich können menschliche Operateure in diesem Bereich auch nur zum Teil durch Maschinen ersetzt werden, sprich die Qualität der durch Menschen erzeugten Referenzzuordnungen ist durch Computerauswertungen nicht zu erreichen. Doch auch das menschliche Gehirn kann bei der Deutung einer räumlichen Situation durchaus zu unterschiedlichen Schlussfolgerungen kommen (vgl. Kapitel 7.2.2).

Aufgrund der enormen Menge raumbezogener Daten und entsprechend hoher Update-Raten ist es allerdings nicht realistisch, Zuordnungsprozesse allein von Menschen durchführen zu lassen. Selbst semi-automatische Ansätze sind teilweise noch zu aufwendig. Eine möglichst weitgehende Automatisierung der Zuordnungsverfahren muss folglich angestrebt werden. Die Gestaltung der Algorithmen hängt hierbei wesentlich von der Art der Unterschiede der Mehrfachrepräsentationen ab. Handelt es sich beispielsweise um Abweichungen im

Maßstab, so müssen zunächst Generalisierungsoperatoren eingesetzt werden, um den detaillierteren Datensatz auf den Detaillevel des größeren Bestandes zu bringen. Erst dann kann ein Zuordnungsverfahren eingesetzt werden.

Bislang existieren verschiedene Methoden zur Zuordnung von mehrfach repräsentierten Daten. Generell wird versucht, die Elemente einer Menge A über eine Zuordnungsfunktion möglichst optimal auf die Elemente einer Menge B abzubilden. Es wird hier zwischen rein merkmalsbasierten Ansätzen, die lediglich geometrische und thematische Objektinformationen auswerten, und relationalen Methoden, die zusätzlich die Relationen eines Objektes zu den Objekten seiner Umgebung (z.B. topologische Eigenschaften) betrachten, unterschieden. Der relationale Ansatz wird in [Vosselmann 92] beschrieben. Die neueren Verfahren (z.B. [Walter 97]) orientieren sich in der Regel an der relationalen Vorgehensweise.

Zuordnungsverfahren stellen grundsätzlich den ersten Schritt beim Conflation bzw. bei der Fusionierung oder Verschmelzung von heterogenen multiplen Repräsentationen dar. Auf der Basis des Zuordnungsergebnisses kann entschieden werden, welche Instanzen miteinander verschmolzen werden sollen. In der Folge sollen die bisher existierenden Methoden zur Zuordnung und Verschmelzung raumbezogener Daten vorgestellt werden.

4.3.3.1 Zuordnungsansätze

Der erste Schritt bei der Zuordnung besteht in der Überlegung, wie man die Ähnlichkeit von Objekten erfassen soll. [Holt 98, 99] hat sich mit dieser Thematik auseinandergesetzt und die entsprechenden Ansätze auf dem Gebiet der Kognitionspsychologie, der Philosophie sowie der Informationswissenschaft studiert. Er schlägt vor, die Ähnlichkeit der Form von Objekten, also den Isomorphismus, auf dem Gebiet der Geo-Informationssysteme anzuwenden, um ähnliche raumbezogene Objekte zu detektieren. Dabei soll die aus dem Forschungsbereich der Künstlichen Intelligenz bekannte Methode des fallbasierten Schließens zum Einsatz kommen. Sie sieht generell vor, vorhandenes Erfahrungswissen, das in Form von so genannten Fällen in einer Fallbasis gespeichert wird, zur Lösung neuer Probleme heranzuziehen. Der Autor will auf diese Weise die Ähnlichkeit von bekannten und unbekanntem Objekten erfassen können. In [Holt 99] wird die Methodik anhand von Flurparzellen untersucht. Dabei werden unbekannte Objekte anhand verschiedener Kriterien, die je nach Bedeutung für die Bestimmung der Ähnlichkeit gewichtet werden können, mit der Fallbasis verglichen, um so letztlich die ähnlichsten Objekte identifizieren zu können.

Auch [Bruns & Egenhofer 96] haben sich mit der Erfassung der Ähnlichkeit raumbezogener Daten auseinandergesetzt und eine Methode hierzu entwickelt. Sie basiert auf der Auswertung von topologischen, richtungsbezogenen und abstandsbezogenen Relationen von Objekten innerhalb verschiedener räumlicher Szenen. Die Grundidee besteht darin zu untersuchen, wieviele Schritte benötigt werden, um eine Szene in die andere zu überführen. Die Anzahl der Schritte liefert dann ein Maß für die Ähnlichkeit, d.h. je stärker die Ähnlichkeit zwischen zwei Szenen ist, desto weniger Schritte werden benötigt.

Nach [Walter 97] gibt es insgesamt vier unterschiedliche Komplexitätsstufen bei der Zuordnung von Geodaten. Grundsätzlich gilt dabei, dass der Prozess umso einfacher ist, je ähnlicher die Ausgangsschemata und -maßstäbe sind. Stufe 1 enthält all jene Fälle, bei denen unterschiedliche Datensätze zugeordnet werden sollen, die vom selben ursprünglichen Datenbestand abstammen, aber getrennt voneinander fortgeführt wurden. Hierbei ist es relativ leicht möglich, über automatisierte Suchverfahren Abweichungen in den jeweiligen Datenbeständen ausfindig zu machen. In Stufe 2 werden alle Ansätze subsummiert, bei denen Daten aus dem gleichen Datenmodell, die jedoch von unterschiedlichen Operateuren/Institutionen unter Verwendung voneinander abweichender Erfassungsrichtlinien erfasst wurden, zugeordnet werden sollen. Eine entsprechende Zielsetzung hatte das EVIDENCE-Projekt [Pandazis 99], innerhalb dessen eine Methode für die Referenzierung von GDF-Straßendaten der beiden kommerziellen Anbieter TeleAtlas und NavTech durch die Einführung eines Zuordnungscodes entwickelt wurde. Der Algorithmus basiert auf der Idee, Straßenkreuzungen durch einen bestimmten Code darzustellen. Der Code setzt sich aus verschiedenen, verschlüsselten Angaben zusammen: den Koordinaten des Kreuzungsknotens, den Abkürzungen der Namen angrenzender Straßen und der Anzahl inzidierender Kanten. Wenn nun die Codes für alle Kreuzungen in zwei sich überdeckenden raumbezogenen Datensätzen generiert werden, so können über einen Vergleich dieser Codes die einander entsprechenden Objekte gefunden werden. Die Methode wurde von [Bofinger 01] implementiert und evaluiert.

Unter die Stufe 3 fallen all jene Integrationsansätze, die eine Homogenisierung von Geodaten aus ähnlichen Datenmodellen realisieren. Hier hat [Walter 97] eine Vorgehensweise entwickelt, um GDF-Daten auf Straßendaten aus ATKIS abzubilden. Sie stützt sich auf einen relationalen Ansatz. Dabei wird über eine Pufferbildung (Buffer Growing) eine Liste potentieller Zuordnungspaare von Linienobjekten erstellt, die Kardinalitäten von 1:0 bis n:m aufweisen können. Die Liste enthält zunächst sehr viele Zuordnungen. Folglich ist die Anzahl der potentiellen Korrespondenzen im weiteren Verlauf zu reduzieren. Dies geschieht, indem unwahrscheinliche Zuordnungspaare mittels einer Auswertung geometrischer und relationaler Faktoren evaluiert und gegebenenfalls eliminiert werden. Die verbleibenden Zuordnungen werden dann nochmals über eine Leistungsfunktion bewertet. Daraufhin kann die optimale Zuordnung bestimmt werden. Um die Qualität des automatischen Systems beurteilen zu können, wird das Zuordnungsergebnis mit einer manuell durchgeführten Referenzzuordnung verglichen. Auch [Xiong & Sperling 03] haben einen Ansatz für ein rein geometrisches Matching von Straßennetzwerken vorgestellt, wobei hier zwischen der Zuordnung auf Knoten-, Kanten- und Segmentebene unterschieden wird. Zusätzlich werden Teilnetzwerke bzw. Cluster definiert, um Korrespondenzen zwischen Objektgruppen und nicht nur zwischen einzelnen Instanzen abzuleiten.

Schließlich finden sich in Kategorie 4 des Stufenmodells von Walter die komplexesten Integrationsfälle wieder, bei denen identische Objektklassen aus stark voneinander abweichenden Datenmodellen zusammengeführt werden sollen. [Gabay & Sester 02] haben hier eine Methode für die Datenintegration von ATKIS-Daten des Digitalen Landschaftsmodells im Maßstab 1:25.000 und von Daten der Automatisierten Liegenschaftskarte (ALK) im Maßstab von ca. 1:500 bis 1:2.000 vorgestellt. Diese basiert auf einer manuellen Zuordnung von Objektinstanzen in den verschiedenen Datensätzen. In der Arbeit wird insbesondere die Problematik bei der Integration von Geodaten unterschiedlichen Maßstabs diskutiert, wobei vor allem die Konflikte betrachtet werden, die auftauchen, wenn die Geometrien mehrfach repräsentierter Objekte nur partiell übereinstimmen. Des Weiteren werden hier die Implikationen einer simultanen Analyse von Geodaten in unterschiedlichen Maßstäben angesprochen. [Egenhofer et al. 94] untersuchen ebenfalls Mehrfachrepräsentationen in unterschiedlichen Maßstäben und schlagen vor, topologische Relationen zum Vergleich bzw. zur Bewertung der Konsistenz von Objekten unterschiedlicher Detailebenen zu verwenden. Sie gehen dabei von der Monotonie der topologischen Relationen im Zuge eines Generalisierungsprozesses aus, d.h. sowohl die eigenen topologischen Beziehungen eines Objektes sowie die topologischen Relationen zwischen mehreren Objekten müssen bei der Generalisierung entweder erhalten bleiben oder im Hinblick auf ihre Komplexität abnehmen.

Zuordnungsalgorithmen kann man auch anhand des geometrischen Objekttyps, der zugeordnet werden soll, unterscheiden. Dementsprechend gibt es punktbasierte, linienbasierte und flächenbasierte Verfahren sowie gemischte Ansätze. Punkt-basierte Methoden berücksichtigen generell die Koordinaten der zuzuordnenden Punkte (bzw. Knoten) und werten auch die Eigenschaften der inzidierenden Features aus, sofern diese vorhanden sind. Der Ansatz von [Pandazis 99] ist zu dieser Kategorie zu zählen (s.o.). Die Vorgehensweise von [Walter 97, 99] gehört hingegen zu den linienbasierten Methoden.

Eine flächenbasierte Methode wurde von [Kraft 95] entwickelt. Alle Flächen, deren Schwerpunkte einen geringeren Abstand aufweisen als durch einen definierten Schwellwert vorgegeben, werden als potentielle Zuordnungskandidaten gehandelt. Mittels einer Kostenfunktion, die verschiedene Parameter wie Fläche, Schwerpunktkoordinaten, Winkel und Anzahl der konstituierenden Linien berücksichtigt, können die möglichen Zuordnungen weiter eingeschränkt werden. Die Fläche, die die geringsten Kosten aufweist, kann schließlich als Ergebnisfläche des Zuordnungsprozesses angesehen werden.

Ein gemischter Ansatz wurde von [Kraut 03] realisiert. Er identifiziert zunächst anhand strikter geometrischer, attributiver und topologischer Parameter einander mit hoher Wahrscheinlichkeit entsprechende Punkte in den zuzuordnenden Datenbeständen. Diese werden als Startpunkte definiert und als Ausgangsbasis für eine Koordinatentransformation zur gegenseitigen Anpassung der Datensätze verwendet. Von jedem Startpunkt aus werden dann die angrenzenden Linien verfolgt und verglichen sowie deren Endpunkte und wiederum deren angrenzende Linien, usw., bis schließlich das ganze Netzwerk abgearbeitet ist.

Neben der Differenzierung von Zuordnungsverfahren anhand der Prozesskomplexität und des geometrischen Objekttyps kann man auch danach unterscheiden, welche Objektklassen zugeordnet werden sollen. Generell sind Verfahren zur Zuordnung von Straßenobjekten [z.B. Walter 97, Bishr et al. 98, Xiong & Sperling 03]

am weitesten verbreitet, da hier aufgrund der Interkonnektivität des Straßennetzes sehr komplexe Fälle auftauchen können und dieses Gebiet folglich ein hohes Potential für die Forschung enthält. Es gibt aber auch Forschungsarbeiten, die andere Objektklassen untersucht haben. Bei [Van Wijngarden et al. 97] wurden beispielsweise Gebäude zugeordnet. Korrespondenzen wurden gefunden, indem zwei Datensätze überlagert wurden und man anschließend die prozentuale Überdeckung von Gebäudeflächen berechnete. Zusätzlich wurde noch der Gebäudeschwerpunkt berücksichtigt.

4.3.3.2 Conflation-Techniken

Die Zuordnung von Objekten stellt den ersten Schritt beim Verschmelzen (in [Saalfeld 88] als „Conflation“ bezeichnet) unterschiedlicher Datensätze dar. [Cobb et al. 98] verstehen unter Conflation die Vorgehensweise zur Vereinigung zugeordneter Objekte in ein einzelnes, das qualitativ übergeordnet ist. Dabei müssen sämtliche Konflikte, die bei der Zuordnung von Objektinstanzen auftauchen, aufgelöst werden, d.h. es muss entschieden werden, welche geometrischen, attributiven und topologischen Eigenschaften bei einem verschmolzenen Objekt beibehalten werden. Der präsentierte Ansatz berücksichtigt zur Auflösung der Konflikte vor allem die Datenqualität der Ausgangsdaten. Insbesondere wird für Punktdaten des Vector-Product-Formats (VPF) gezeigt, wie die Unsicherheit der Positionsangaben beim geometrischen Conflation-Prozess einfließt.

[Laurini 98] spricht die Problematik beim Zusammenführen aneinandergrenzender Datensätze an. Er schlägt vor, heterogene Ausgangsdaten über elastische Transformationen in den Übergangsbereichen topologisch kohärent zu machen. Dazu wird für den Übergangsbereich zwischen zwei Datensätzen ein so genanntes elastisches Band definiert, innerhalb dessen homologe Punkte zusammengeführt werden. [Yuan & Tao 99] gehen davon aus, dass die Problematik im Bereich des Conflation zu komplex ist, um von einzelnen Algorithmen gelöst werden zu können. Vielmehr schlagen sie vor, durch die Entwicklung komponentenbasierter Software eine flexiblere Lösung zu finden.

Auch [Edwards & Simpson 02] haben sich mit dem Thema der Datenfusion auseinander gesetzt. In einem ersten, als „Linking“ bezeichneten Schritt, stellen sie Beziehungen zwischen homologen Objekten her. Hierbei werden lediglich korrespondierende Instanzen identifiziert, ohne dass bewertet wird, welche Instanz die „bessere“ ist. Der darauf folgende Schritt, „Best Map“ genannt, vereinigt dann die miteinander verbundenen Datensätze zu einem ganzen, d.h. die pluralistischen Auffassungen, die in verschiedenen Datensätzen über Realweltobjekte herrschen, werden in einzelnen resultierenden Repräsentationen vereinigt. Auf der Grundlage dieser verschmolzenen Daten können dann raumbezogene Abfragen und Analysen durchgeführt werden. Die Autoren unterstreichen die positiven Aspekte ihrer Vorgehensweise:

- Die Möglichkeit, alle verfügbaren Daten zu nutzen: Man kann davon ausgehen, dass eine einzelne Quelle niemals die perfekte Repräsentation eines Realweltobjektes liefern kann. Im Rahmen der Integration von Mehrfachrepräsentationen können theoretisch sämtliche Vorteile der einzelnen Repräsentationen kombiniert werden.
- Automatisierte Nutzung neu erzeugter Daten: Wenn neue Repräsentationen erzeugt werden, können sie mit den bestehenden verglichen werden. Im Falle größerer Abweichungen werden Hinweise geliefert, dass Nachführungen der existierenden Instanzen eingeleitet werden müssen.
- Umgang mit einer unsicheren Realität: Mehrfach repräsentierte Daten helfen dabei, das Bewusstsein hinsichtlich der Erkenntnis zu schärfen, dass die Wahrheit bzw. der Zustand in der Realität aus den Daten heraus nicht automatisch erschließbar ist.
- Schaffung neuen Wissens: Während des „Linkings“ entstehen neue Informationen, da bestimmte Objekteigenschaften bestätigt werden (nämlich jene, die in allen Repräsentationen identisch sind), während sich andere Eigenschaften als kontradiktorisch herausstellen können. Je mehr Repräsentationen identisch sind, desto mehr kann man sich auf die Korrektheit der Daten verlassen.

Im Rahmen des „Best Map“-Algorithmus wurden exemplarisch Straßendaten unterschiedlicher Auflösung fusioniert. Hierzu wurden zum einen raumbezogene, zum anderen thematische Regeln aufgestellt:

1. Wo verfügbar, wurden die hochaufgelösten Daten verwendet.

2. Außerhalb dieser Gebiete wurden die weniger detaillierten Daten verwendet.
3. In den Übergangsbereichen wurden so genannte Konnektivitätsgeometrien erzeugt, die die beiden Datensätze miteinander verbinden.
4. In den Gebieten, in denen hochaufgelöste und weniger detaillierte Daten verfügbar waren und zusätzlich keine Verbindung zwischen den beiden Ebenen bestand, wurden die größeren Daten ebenfalls erhalten und in die resultierende Karte aufgenommen.

Für die Thematik wurde ein gemeinsames Schema angelegt, innerhalb dessen sämtliche Attribute der Ausgangsdaten erhalten bleiben. Zusätzlich werden neue Attribute eingeführt, die u.a. angeben, ob Links zu mehreren Repräsentationen existieren etc.

[Devogele 02] entwickelte eine neuartige Methodik zum Verschmelzen homologer Geometrien auf der Basis der so genannten Fréchet-Distanz. Die Fréchet-Distanz wird als Maximaldistanz zweier gerichteter Linien definiert. Der Autor sieht sein Verfahren als Alternative zu den Fuzzy-Toleranz-Ansätzen, die nur eine lokale Verschmelzung von Geometrien erlauben und aufgrund der Definition von Schwellwerten häufig zu Ungenauigkeiten und fehlerhaften Vereinigungen führen. Das Verfahren erlaubt es, homologe Punkte von Linien und Polygonen auf der Basis der Fréchet-Distanz zu finden, und es kann dabei auch Zuordnungen der Kardinalität n:m erkennen. Anhand verschiedener Faktoren wie der Punktgenauigkeit oder der Kardinalität der gefundenen Zuordnung kann schließlich eine Gewichtung der Matches erfolgen, auf deren Basis die Verschmelzung erfolgt.

4.3.4 Anwendungen: Fortführung und Analyse mehrfach repräsentierter Objekte

Die Ansätze zur Integration von raumbezogenen Daten sollen die Grundlage dafür liefern, sowohl die Fortführung als auch die Analysen in mehrfach repräsentierten Datenbeständen realisieren zu können. Dabei wurde der Anwendungsbereich „Fortführung“ bisher nur vereinzelt betrachtet, der Anwendungsbereich „Analyse“ nahezu überhaupt nicht. Einige der vorhandenen Arbeiten werden in den folgenden Abschnitten besprochen.

4.3.4.1 Fortführung

Sowohl in [Uitermark et al. 98] als auch in [Uitermark et al. 99a] wird versucht, mit Hilfe der Integration raumbezogener Datenbanken den Bedarf an Eingriffen durch menschliche Operateure im Zuge der Datenfortführung zu minimieren. Letztere Arbeit widmet sich dabei besonders dem automatischen Update von Daten für Straßenverkehrsnetze. Die Fortführung wird anhand zweier Datensätze unterschiedlichen Maßstabs untersucht und in insgesamt 6 Schritte unterteilt:

1. Synchronisation der beiden Datensätze auf den gleichen zeitlichen Zustand: Aus dem aktuelleren Datensatz werden jene Objekte bzw. Aktualisierungszustände nicht betrachtet, die zeitlich nach dem Aktualisierungsdatum des älteren Datensatzes hinzugefügt wurden;
2. Erzeugung von flächenhaften Straßensegmenten (Road Elements) und Kreuzungen (Junctions) über eine Constrained Delauney Triangulation;
3. Auffinden korrespondierender Straßenobjekte über die Berechnung von Überlappungen zwischen den flächenhaften Straßensegmenten. Dieser Algorithmus arbeitet analog zu dem von [Van Wijngarden et al. 97] entwickelten Verfahren zur Detektion von korrespondierenden Gebäudeobjekten (vgl. Kapitel 4.3.3.1);
4. Entscheidung, ob eine Objektänderung wichtig genug ist, um ein Update einzuleiten;
5. Transformierung der Updates des einen Datensatzes in jene des anderen Datensatzes;
6. Propagieren der Updates in den jeweiligen Datensatz.

Die Autoren bemerken auch, dass Update-Prozesse von Straßendaten generell anders verlaufen als die für Gebäudeobjekte, da Gebäude häufig nicht direkt mit anderen Gebäuden verbunden sind, Straßen hingegen einer starken Vernetzung unterliegen. Änderungen einer Straße pflanzen sich somit an die Umgebung fort, was letztlich zu einer erhöhten Komplexität führt.

Die Fortführung von Gebäudestrukturen wird in [Van Wijngarden et al. 97] untersucht. Werden korrespondierende Instanzen in zwei Datensätzen gefunden, so erfolgt ein Update nur dann, wenn sich die Ausdehnung eines Gebäudes um einen bestimmten Schwellwert verändert hat. Kann für ein Objekt im aktuelleren Datensatz keine Korrespondenz gefunden werden, so muss im älteren Datensatz eine neue Instanz erzeugt werden. 1:n, n:1 oder n:m-Beziehungen versuchen die Autoren dadurch zu umgehen, dass sie im Vorfeld der Zuordnung korrespondierender Gebäude eine Aggregation sehr nahe beieinander liegender Strukturen vornehmen.

[Von Goesseln & Sester 03] nennen 3 Schritte, um Updates zwischen korrespondierenden Instanzen propagieren zu können. Zunächst muss der Zuordnungsschritt erfolgen, daraufhin werden die möglichen Änderungen klassifiziert und schließlich müssen diese in den betreffenden Repräsentationen realisiert werden. Die Untersuchung geschieht anhand von geologischen Kartendaten und ATKIS-Daten am Beispiel der Objektklasse „Wasserflächen“, die in beiden Beständen vorhanden ist. Um die Zuordnung zu ermöglichen, werden zunächst sämtliche Objekte in denselben Objekttyp – im Beispiel in den Typ Polygon – transformiert. Da korrespondierende Repräsentationen eine unterschiedliche Anzahl von Teilflächen bzw. Segmenten aufweisen, müssen die zueinander gehörenden polygonalen Segmente jeder Repräsentation vereinigt werden. Dabei sind benachbarte Segmente anhand eines Distanz-Schwellwertes zu detektieren und dann zusammenzuführen. Anschließend können die Zuordnungspaare der Repräsentationen gebildet und die Unterschiede zwischen ihnen über eine Verschneidung der korrespondierenden Polygone aufgedeckt werden. Zur Detektion der Objektunterschiede ist es jedoch notwendig, eine bessere geometrische Korrespondenz der zugeordneten Mehrfachrepräsentationen zu erreichen. Zu diesem Zweck wird eine lokale Transformation mit dem Iterative Closest Point Algorithmus (ICP, siehe [Besl & McKay 92]) angesetzt. Anhand der Transformationsparameter, die aus dem ICP resultieren, wird die Qualität eines Matching-Prozesses ausgedrückt. Anschließend können die Abweichungen zweier Repräsentationen korrekt festgestellt werden. Prinzipiell werden bei Unterschieden von Repräsentationen die Objektausprägungen von ATKIS aufgrund der höheren Aktualität als Referenz genommen und in die geologischen Karten transferiert. Sind beispielsweise Wasserflächen in ATKIS verzeichnet, im geologischen Bestand jedoch nicht, so wird ein neues Objekt in den Geologiedaten erzeugt. Ist ein Objekt hingegen in den geologischen Daten ausgewiesen, das in ATKIS nicht mehr vorhanden ist, so wird dieses gelöscht. Nicht eindeutige Fälle können in der Kartendarstellung markiert und der Behandlung durch einen Operateur überlassen werden.

[Mantel & Lipeck 04] versuchen, eine Zuordnung von Multiskalen-Repräsentationen innerhalb eines föderierten Systems zum Zwecke der Datenfortführung zu implementieren. Dabei soll nur die Repräsentation mit der höchsten Auflösung manuell aktualisiert werden, während alle zugehörigen Repräsentationen eine automatische Nachführung erfahren. Hierzu bedarf es Links, die im Zuge des Zuordnungsprozesses generiert und im Sinne einer Multi-Resolutions-Datenbank abgespeichert werden. Für die Implementierung der Zuordnung greifen die Autoren den „Buffer Growing“-Ansatz von [Walter 97] auf und erweitern diesen. Über Schwellwerte von Qualitätsmaßen der Zuordnung finden sie die optimalen Zuordnungen. Bei unklaren Zuordnungsergebnissen kann auch ein manueller Eingriff erfolgen.

4.3.4.2 Analyse

Analysen in mehrfach repräsentierten Datenbeständen wurden bisher kaum durchgeführt. Dies liegt in erster Linie daran, dass es noch keine funktionsfähigen Gesamtsysteme gibt, die Mehrfachrepräsentationen integrieren und mit den integrierten Ergebnissen arbeiten können. Einen ersten Ansatz hierzu lieferte das MurMur-Projekt [MurMur 02], in dessen Rahmen ein konzeptionelles Modell entwickelt wurde, um Mehrfachrepräsentationen in Datenbanken abspeichern zu können. Für dieses Modell wurde auch eine Abfragesprache mit eigener Algebra formuliert, die Mehrfachrepräsentationen berücksichtigt. Der Nutzer kann die Anfrage über einen visuellen Abfrageeditor ausführen. Diese wird dann auf die Algebra und letztlich auf die Syntax der Zielsprache wie z.B. SQL umgesetzt. Bei der Entwicklung der Abfragesprache wurden zwei Konzepte verfolgt: das „Intelligente Zoomen“ erlaubt es, alle Repräsentationen verschiedener Maßstäbe auszuwerten und nur jene Repräsentation in der Kartendarstellung anzuzeigen, die dem Betrachtungsmaßstab des Benutzers am ehesten entspricht. Die „Multidimensionalen Views“ ermöglichen es, die Evolution eines Objektes über die Betrachtung seiner zeitlich verschiedenen Repräsentationen nachzuverfolgen.

In [Wiederhold 99] wird ein Framework für die Analyse verteilter Daten vorgestellt, innerhalb dessen die Resultate von Anfragen an einzelne Server in einer Mediatorschicht (bzw. einer Föderation), die zwischen Anwendung und Datenquellen vermittelt, kombiniert werden können. Das Problem der Integration heteroge-

ner Quellen auf Mediatorebene wird hier zwar benannt, konkrete Ansätze zu einer gemeinsamen Datenanalyse werden aber nicht formuliert.

5 Abbildung bestehender Geodaten-Schemas auf das AWS

Die Qualität einer Informationsplattform für orts- und kontextbezogene Dienste hängt in entscheidendem Maße davon ab, wie umfangreich, vielfältig bzw. variabel verwendbar, detailliert und hochwertig das raumbezogene Datenmaterial ist, auf das sie einen Zugriff bietet. Analog zum World Wide Web muss eine solche Plattform also in erster Linie darauf ausgerichtet sein, möglichst vielen Datenanbietern die Möglichkeit zu geben, Informationen für sie zur Verfügung stellen zu können.

In Bezug auf Nexus bedeutet dies, dass ein globales Klassenschema entwickelt werden muss, auf das die bereits existierenden Geodatenmodelle über entsprechende Regeln abgebildet werden können, um eine Nutzung vorhandener Daten aus externen Quellen zu gewährleisten. Dies geschieht in dieser Arbeit nur für einen Teilbereich raumbezogener Daten, und zwar für den Bereich der Straßenverkehrsdaten, da eine weiterreichende Betrachtung den Rahmen der Untersuchung sprengen würde.

Als Grundlage für die Entwicklung entsprechender Klassen für Straßendaten innerhalb des Augmented World Schemas müssen einerseits die bestehenden Datenschemas GDF, ATKIS und ALK und deren Modellierungskonzepte untersucht werden. Andererseits werden die Anforderungen an das Datenmodell von Nexus aus der Sicht kontextbezogener Applikationen herausgearbeitet. Auf dieser Basis kann dann die Definition der Objektklassen für Straßendaten innerhalb des AWS erfolgen. Schließlich werden Abbildungsvorschriften (bzw. Schema-Mappings) für die existierenden Datenschemas auf das generische Nexus-Schema aufgestellt, die eine Übertragung vom Ursprungsformat ins Nexus-Austauschformat AWML zulassen (siehe Abbildung 5-1).

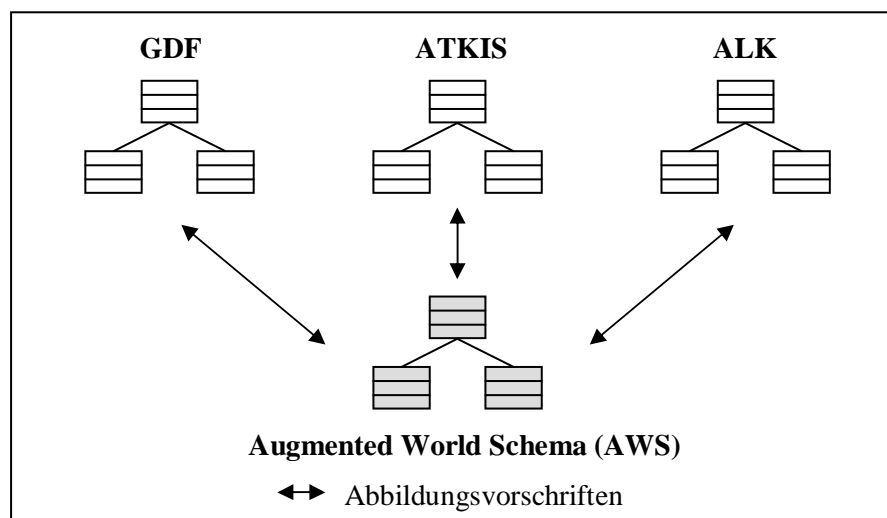


Abbildung 5-1: Um Geodaten in Nexus integrieren zu können, müssen Abbildungsvorschriften von bestehenden Geodaten-Schemas auf das Augmented World Schema existieren.

5.1 Untersuchung existierender Schemas für Geodaten

Im Folgenden soll eine kurze Übersicht über die in Deutschland verfügbaren Schemas bzw. Datenmodelle gegeben werden, die eine Repräsentation von Straßendaten erlauben. Es ist zu betonen, dass an dieser Stelle nur wichtige Grundlagen existierender Datenmodelle erläutert werden können. Umfassendere Informationen sind den jeweiligen Standards zu entnehmen.

Zunächst wird auf das Geographic Data Files (GDF)-Format eingegangen, das speziell für die Bedürfnisse der Fahrzeugnavigation entworfen wurde. Anschließend werden die Geodaten des amtlichen Vermessungswesens bzw. die Geobasisdaten betrachtet, die nicht auf spezielle Applikationen ausgerichtet sind und vielmehr von den verschiedensten Anwendungsbereichen als Ausgangsdaten herangezogen und entsprechend der jeweiligen Applikationsbedürfnisse erweitert werden können. Das Amtliche Topographisch-

Kartographische Informationssystem ATKIS wird dabei als erstes präsentiert. Danach folgt die Automatisierte Liegenschaftskarte ALK. Die ursprünglichen Konzepte von ALK und ATKIS stammen aus den 70er bzw. aus den 80er Jahren und werden daher in einem aktuellen Verfahren an das gegenwärtige Umfeld angepasst. Dieses Verfahren ist mit dem Ziel angelegt worden, ein auf den Standards von ISO und OGC basierendes, so genanntes AFIS-ALKIS-ATKIS-Referenzmodell aufzubauen (vgl. Kapitel 2.5). Dabei entstehen auch neue Objektartenkataloge für ALKIS (Nachfolgesystem der ALK) und ATKIS. Auf deren Basis wurden jedoch noch keine Daten erfasst bzw. in dieses Modell wurden noch keine bestehenden Daten migriert. Diese Arbeit beschränkt sich folglich auf die Beschreibung der herkömmlichen Datenmodelle von ALK und ATKIS, für die Daten verfügbar sind.

Nach der Vorstellung der Schemas erfolgt eine Gegenüberstellung der Modellierungskonzepte und abschließend wird kurz auf die gewonnenen Erkenntnisse eingegangen.

5.1.1 Das Datenschema von GDF

Das GDF-Datenmodell wurde für Anwendungen aus dem Bereich der Fahrzeugnavigation entwickelt und liegt derzeit in der Version 4.0, der ersten weltweiten Veröffentlichung des Standards, vor [ISO 04]. Allerdings standen für die vorliegende Untersuchung ausschließlich Daten der Vorgängerversion 3.0 zur Verfügung, weshalb an dieser Stelle nur auf diese ältere Version eingegangen wird. Für eine ausführliche Beschreibung des Standards sei auf dessen Spezifikation [CEN 95] verwiesen.

Im Zentrum des konzeptionellen Modells von GDF steht das Feature, das analog zur Auffassung des OGC als Phänomen der Realwelt (wie z.B. Straßen, Gebäude oder Eisenbahnen) zu interpretieren ist. Features existieren auf verschiedenen Levels bzw. Ebenen.

Level 0: Geometrie

Beschreibt die Geometrie von Kartenobjekten über die Grundprimitive Punkt, Linie oder Fläche.

Level 1: Einfache Features

Beschreibt Grundeinheiten einer Straßenkarte wie Straßenabschnitte (Road Elements) als Linien-Features oder einfache Kreuzungen (Junctions) als Punkt-Features.

Level 2: Komplexe Features

Beschreibt eine Aggregation von Features; So gibt es komplexe Straßen (Roads) oder komplexe Kreuzungen (Intersections), die aus den konstituierenden, einfachen Features aufgebaut sind.

Abbildung 5-2 soll die Idee der einfachen und komplexen Features illustrieren:

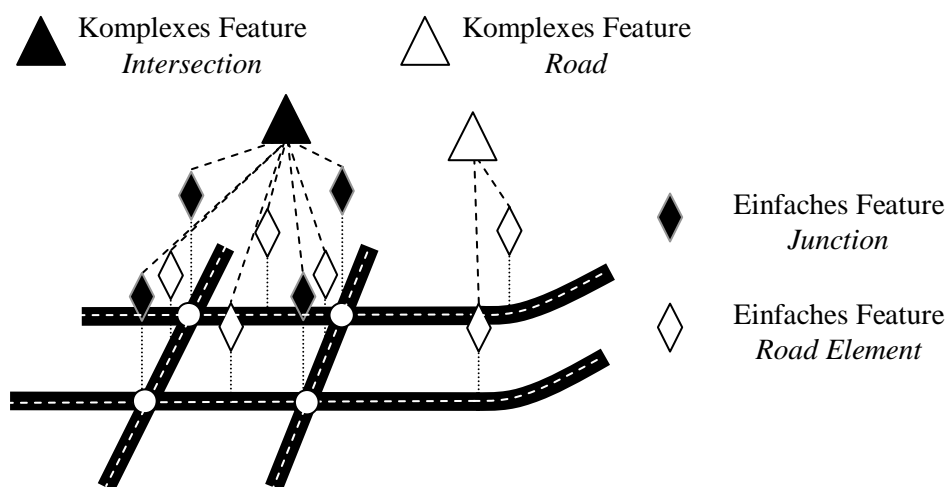


Abbildung 5-2: Einfache und komplexe Straßenverkehrs-Features (verändert nach [Walter 97]); ein Kreuzungsbereich (Intersection) auf Level 2 besteht aus den Level-1-Features Road Elements und Junctions, eine Straße (Road) auf Level 2 besteht aus den Level-1-Features Road Elements.

Jedes Feature gehört exakt einer bestimmten Feature-Klasse an. Für den Straßenverkehr existieren verschiedene dieser Feature-Klassen; neben den Junctions (Kreuzungspunkte) und den Road Elements (Straßensegmente) z.B. die Brunnels (Kunstwort aus Brücke und Tunnel), die Brücken und Tunnels, welche länger als 200 m sind, kennzeichnen, etc. Ähnliche Feature-Klassen werden wiederum in so genannten „Feature Themes“ zusammengefasst. Die Feature-Klassen des Straßenverkehrs gehören den Feature Themes „Roads and Ferries“ und „Brunnels“ an. Daneben gibt es u.a. Feature Themes für Verwaltungseinheiten, bebaute und bezeichnete Gebiete, Eisenbahnen, usw.

Ein Feature ist also eine konkrete Instanz einer Feature-Klasse und somit eine Repräsentation eines Realweltobjektes. Es kann durch Attribute näher beschrieben werden und steht mit den Features seiner Umgebung über Relationen in Beziehung. Relationen verfügen ebenfalls über Attribute.

• **Attribute**

Attribute können entweder einfach oder zusammengesetzt und zusätzlich zeitabhängig oder segmentiert sein. Bei der Segmentierung wird dem Umstand Rechnung getragen, dass ein Attribut wie z.B. die Straßenbreite auf einem Straßenabschnitt uneinheitlich sein kann, obwohl keine topologische Änderung auftritt. Unterschiedliche Teilabschnitte können durch diese Modellierungsvariante mit unterschiedlichen Werten belegt werden (siehe Abbildung 5-3).

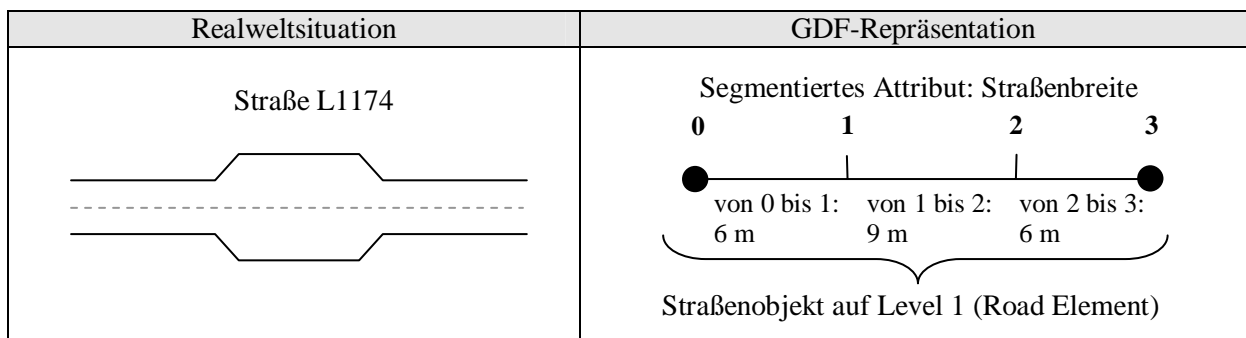
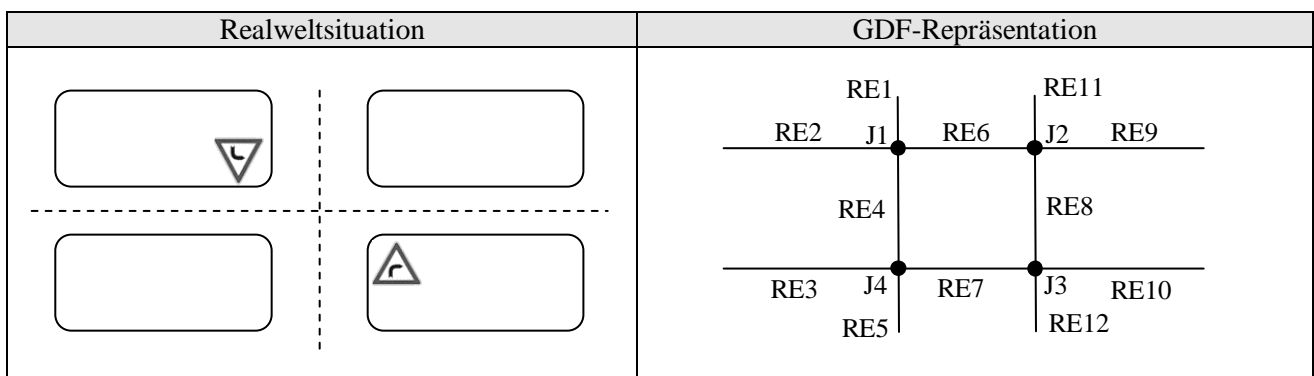


Abbildung 5-3: Darstellung der Modellierungsform „segmentiertes Attribut“. Innerhalb *eines* Attributes werden verschiedene Abschnitte angegeben, für die jeweils unterschiedliche Attributwerte gültig sind.

• **Relationen**

Eine Relation besteht aus zwei oder mehreren Features und beschreibt eine Assoziation zwischen ihnen.



Relationship-Code	Road Element	Junction	Road Element	Junction	Road Element	Junction	Road Element
53077	RE1	J1	RE4	J4	RE7	J3	RE10
93461	RE12	J3	RE10	-	-	-	-

Abbildung 5-4: Modellierung von Relationen in GDF am Beispiel von Abbiegeverboten (in Anlehnung an [NavTech 00])

So kann eine Relation z.B. ein Abbiegeverbot von Straße A nach Straße B darstellen, die als Sequenz jener Objekte repräsentiert wird, die nicht der Reihenfolge nach durchfahren werden dürfen (siehe Abbildung 5-4). Eine Relation kann eigene Attribute besitzen.

5.1.2 Das Datenschema von ATKIS

Das ATKIS-Datenmodell ist die Grundlage für eine möglichst weitreichende digitale Abbildung der Realwelt und soll so genannte Geobasisdaten liefern. Für eine nähere Betrachtung des Standards sei auf [AdV 88] und [AdV 02] verwiesen.

ATKIS erfasst Daten zu insgesamt sieben verschiedenen Objektbereichen wie z.B. Siedlung, Vegetation, Gewässer und auch Verkehr. Die Objektbereiche sind wiederum in so genannte Objektgruppen gegliedert. Beim Verkehr handelt es sich dabei u.a. um die Gruppen Straßenverkehr, Schienenverkehr, Flugverkehr usw. Eine weitere Ebene tiefer sind die Objektarten definiert, die den geometrischen Typ sowie die Attribute von Objekttypen der Realwelt im Schema festlegen.

Konkrete Instanzen der Objektarten sind die ATKIS-Objekte, die ihrerseits wiederum aus Objektteilen aufgebaut sind. Die Objektteile enthalten die eigentliche Geometrie in Form von Vektorelementen. Daher beinhaltet jedes Objekt mindestens einen Objektteil, wobei sowohl Objekten als auch Objektteilen Attribute zugeordnet werden können. Abbildung 5-5 illustriert die Bildung von Objekten und Objektteilen in ATKIS. Es sind vier Straßenobjekte dargestellt, die über ihr Attribut „Kurzbezeichnung“ (z.B. „L 1714“) zu identifizieren sind. Überall dort, wo topologische (Straßenkreuzungen in den Punkten A, B und D) oder attributive (Wechsel der Straßenbreite in Punkt C) Änderungen stattfinden, wird ein Straßenobjekt in Objektteile (also {AB}, {BC}, etc.) zerlegt. In ATKIS können sowohl Objekte als auch Objektteile Attribute haben. Im Gegensatz zu GDF können Attribute in ATKIS hingegen nicht segmentiert werden.

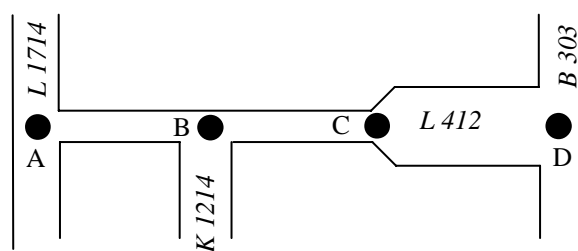


Abbildung 5-5: Objekte und Objektteile in ATKIS (aus [Walter 97]). Objektteile können eigene Attribute besitzen.

Allerdings ist es in ATKIS ebenfalls möglich, komplexe bzw. zusammengesetzte Objekte zu bilden. Im Falle von Straßen wird diese Modellierungsform u.a. benötigt, wenn eine Straße aus mehreren Fahrbahnen besteht. Ein komplexes Straßenobjekt setzt sich dann aus den Objekten Fahrbahn und Straßenkörper zusammen (siehe Abbildung 5-6).

Realweltsituation	ATKIS-Repräsentation

Abbildung 5-6: Bildung von komplexen Straßenobjekten in ATKIS mittels Straßenkörper- und Fahrbahn-Objekten.

ATKIS sieht die Bildung einfacher Relationen vor. Im Falle von Straßen gibt es die Über- und Unterführungsreferenzen. Mittels solcher Referenzen bzw. Relationen kann beispielsweise angegeben werden, dass eine Straße über eine Brücke eine andere Straße überführt.

Im so genannten Digitalen Landschaftsmodell im Maßstab 1:25 000 (DLM 25) erfasst ATKIS Straßenverkehrsdaten mit einer Genauigkeit von ca. 3 Metern. Allerdings können viele wichtige Informationen wie Abbiegeverbote in ATKIS nicht abgespeichert werden. Es muss auch darauf hingewiesen werden, dass die in ATKIS realisierte Objekthierarchie nicht dem objektorientierten Prinzip folgt, d.h. es gibt beispielsweise keine Vererbungsbeziehungen zwischen den Objektklassen.

5.1.3 Das Datenschema der ALK

Liegenschaftskataster und Grundbücher dienen als Nachweise der Eigentumsverhältnisse an Grund und Boden. Seit der Einführung des Automatisierten Liegenschaftskatasters Anfang der 80er Jahre werden die Katasterinformationen in Deutschland digital erfasst. Das Automatisierte Liegenschaftskataster setzt sich aus der Verfahrenslösung Automatisierte Liegenschaftskarte (ALK) und dem Automatisierten Liegenschaftsbuch (ALB) zusammen, wobei die ALK als darstellender Bestandteil entwickelt wurde und das ALB den beschreibenden Teil enthält. Die ALK wird in Deutschland von den Vermessungsverwaltungen der Bundesländer geführt, das ALB bei den Amtsgerichten [Bill 99b].

Die ALK ersetzt die früher in analoger Form geführten Flur- und Stadtkarten, in denen alle Informationen auf Zeichenfolien festgehalten wurden. Diese enthielten bestimmte Zeichenobjekte wie Objektgeometrien, Beschriftungen, Schraffuren, etc., deren Bedeutung über eine Legende ersichtlich wurde. Im Grunde besteht die ALK aus diesen Zeichenobjekten in digitaler Form, namentlich der ALK-Punktdatei, die die nummerierten Punkte des Liegenschaftskatasters enthält, und der so genannten Grundrissdatei, in der Objektgeometrien als Punkte, Linien oder Flächen sowie Texte vorgehalten werden. Die Informationen der Grundrissdatei werden aus unterschiedlichen fachlichen Perspektiven zu logischen Einheiten, so genannten Grundrissobjekten, zusammengefasst. Die Regeln, denen diese Objektbildung unterliegt, sind im Objektabbildungskatalog (OBAK) der ALK festgehalten [OBAK NRW 03].

Innerhalb der ALK werden Objekte auf separaten Folien bzw. in verschiedenen fachlichen Gruppen gespeichert. Die Bedeutung der Legende analoger Karten übernimmt ein so genannter Objektschlüsselkatalog (OSKA). In ihm sind alle Objekte und Objektbestandteile, die in der Automatisierten Liegenschaftskarte auftreten können, aufgeführt. Jedem Objekt werden im OSKA ein dreistelliger Folienschlüssel und ein vierstelliger Objektschlüssel zugewiesen, wodurch eine eindeutige Objektidentifikation gewährleistet werden kann. Diese Kodierung ist notwendig, um eine Verknüpfung von kartographischen Informationen aus der ALK und fachlichen Daten (z.B. aus dem ALB) eines Objektes zu ermöglichen.

Die ALK weist eine Vielzahl von Folien bzw. fachlichen Gruppierungen auf, z.B. für Flurstücke, Gebäude oder politische Grenzen. Tabelle 5-1 zeigt einen Auszug aus dem Objektschlüsselkatalog des Landes Nordrhein-Westfalen [OSKA NRW 03].

Folie	Folienbezeichnung	Objektschlüssel	Bezeichnung	Kennung für Objekt/Objektteil
021	Tatsächliche Nutzung	5100	Straße	O
		5110	Straße mehrbahnig	O
		5120	Straße einbahnig	O
	
		5140	Gehweg an Straße	O

Tabelle 5-1: Auszug aus dem Objektschlüsselkatalog Nordrhein-Westfalens [OSKA NRW 03].

ALK-Objekte erhalten ihre Semantik folglich über die Zugehörigkeit zu einer Folie, über den Objektschlüssel und die Objektbezeichnung. Die Differenzierung der Objektarten in verschiedene Folien ist teilweise länderspezifisch. In Baden-Württemberg werden Straßenobjekte in der Regel nicht weiter unterschieden und zumeist in Form von Flurstücken repräsentiert, die eine flächenhafte Geometrie aufweisen. Auch für die zur Verfügung stehenden Testdaten des Stadtgebiets Stuttgart wurden die Daten auf diese Weise erfasst.

Die Erfassung der ALK erfolgt auf der Grundlage analoger Flurkarten (so genannter Rahmen- oder Inselkarten) in unterschiedlichen Maßstäben (1:500, 1:1000, 1:1750, 1:2000, usw.). Die geometrische Genauigkeit

liegt im dm bis cm-Bereich. Es wird geschätzt, dass eine flächendeckende Realisierung der ALK in Deutschland mindestens bis zum Jahr 2005 dauern wird [Bill 99b].

5.1.4 Gegenüberstellung der Konzepte zur Modellierung von Straßendaten

Die unterschiedlichen Ansätze zur Modellierung von Straßenverkehrsdaten in GDF, ATKIS und ALK sollen hier noch einmal kurz gegenübergestellt werden (siehe Tabelle 5-2).

Konzeptionelles Datenmodell	GDF	ATKIS	ALK
Modellierung der Welt	Features auf unterschiedlichen Komplexitätsebenen	Einteilung in Objekte und konstituierende Objektteile. Komplexe Features sind ebenfalls möglich.	Einteilung in Objekte und Objektteile; Keine komplexen Objekte möglich.
Modellstruktur	Hierarchisch ohne Vererbung	Hierarchisch ohne Vererbung	Keine Objekthierarchien, sondern Ebenenstruktur
Attributkonzept	GDF	ATKIS	ALK
Segmentierte Attribute	Möglich	Nicht möglich	Nicht möglich
Zusammengesetzte Attribute	Möglich	Nicht möglich	Nicht möglich
Zeitabhängige Attribute	Möglich	Nicht möglich	Nicht möglich
Semantisches Relationenkonzept	GDF	ATKIS	ALK
Semantische Relationen	Möglich (z.B. Abbiegere-lationen)	Nur Über-/Unterführungs-Relationen	Nicht möglich
Attribute für Relationen	Möglich	Nicht möglich	Nicht möglich

Tabelle 5-2: Zusammenschau wichtiger Charakteristika bei der Modellierung von Straßendaten in GDF, ATKIS und ALK (in Anlehnung an [Kraut 03]).

Wie aus Tabelle 5-2 ersichtlich ist, stellt GDF das umfangreichste und differenzierteste Modellierungskonzept bereit. Dies ist sicherlich darauf zurückzuführen, dass GDF in erster Linie für die Abbildung von Straßenverkehrsdaten gedacht ist. Auf der Ebene des konzeptionellen Datenmodells sind noch die geringsten Unterschiede zwischen den Datenschemas auszumachen. Das Attributkonzept ist in ATKIS nur sehr einfach ausgeprägt und in der ALK aufgrund der Trennung zwischen Geometrie und Fachdateien überhaupt nicht vorhanden. Auch bei den Relationen existiert im Bereich der Straßenverkehrsdaten in ATKIS nur die Möglichkeit, Über- und Unterführungsreferenzen anzugeben. In der ALK fehlt der Mechanismus der Verbindung von Objekten mittels Relationen komplett.

In Abbildung 5-7 soll noch einmal aufgezeigt werden, wie sich die Datenmodelle von GDF, ATKIS und ALK konkret bei der Objektbildung von Straßen unterscheiden. In (I) ist der Kreuzungsbereich „Kurt-Georg-Kiesinger-Platz“ in der Innenstadt von Stuttgart aus der Deutschen Stadtgrundkarte (DSGK) zu erkennen. Die Darstellung in (II) zeigt die Repräsentation der Straßenobjekte aus der ALK in Form von Flurstücken. Mehrbahnige Straßen werden in ATKIS (III) als komplexe Straßenobjekte, die aus den Objekten Fahrbahn und Straßenkörper bestehen, abgebildet. Einbahnige Straßen sind als einfache Straßenobjekte repräsentiert. Ausschnitt (IV) zeigt eine Level 1-Repräsentation der GDF-Daten, die aus Junctions und Road Elements besteht.

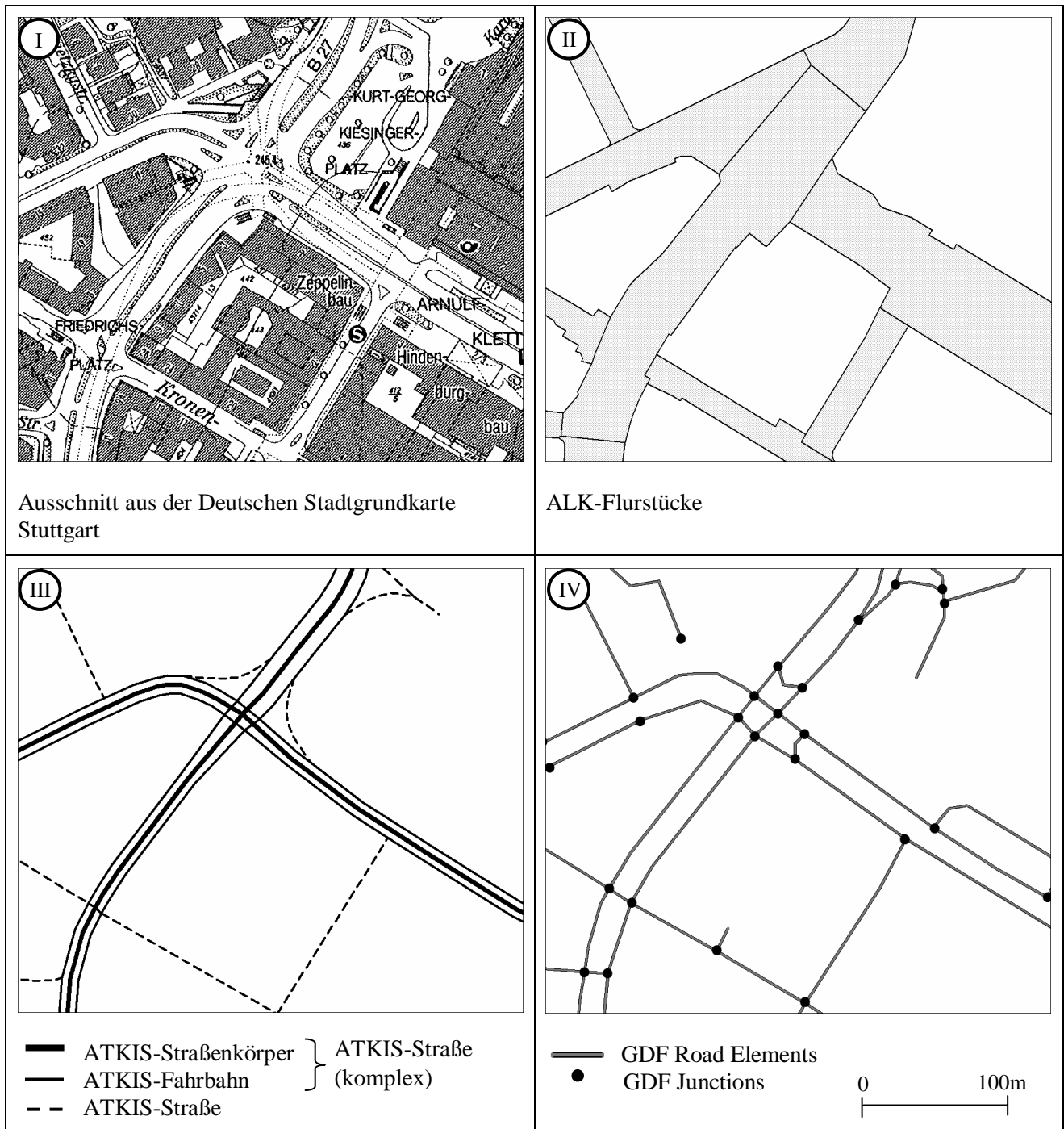


Abbildung 5-7: Objektbildung in GDF, ATKIS und ALK in den verfügbaren Datenbeständen

5.1.5 Erkenntnisse aus der Schema-Analyse

Die Analyse von bestehenden Schemas bzw. von Mehrfachkonzeptionen ist zum einen notwendig, um Informationen über die unterschiedlichen Konzepte zur Repräsentation von Straßendaten zu erhalten, welche bei der Gestaltung des globalen Schemas berücksichtigt werden können. Zum anderen möchte man Erkenntnisse darüber gewinnen, welche Korrespondenzen zwischen den Objektklassen der unterschiedlichen Ausgangsschemas vorhanden sind, um diese im übergeordneten Schema zusammenführen bzw. integrieren zu können. Die Identifikation semantischer Korrelationen ist ein wissensintensiver und somit zeitaufwendiger Prozess für den menschlichen Operateur, den es über geeignete Methoden zu optimieren gilt. Die vorliegende Arbeit bietet hierzu einen Ansatz (siehe Kapitel 8.1).

5.2 Anforderungen an das Straßen-Datenschema aus der Sicht von Nexus

Im Folgenden werden zunächst die allgemeinen Anforderungen an das Datenschema von Nexus in Bezug auf Straßendaten dargestellt, woraus auch die Begründung für die Ableitung eines eigenen, Nexus-spezifischen Schemas abzuleiten ist. Daraufhin werden die speziellen Anwendungsanforderungen bzw. Use Cases, die für Straßendaten in Nexus existieren, aufgelistet.

5.2.1 Allgemeine Anforderungen

Bei der Entwicklung eines neuen Datenschemas ist zunächst die Frage zu stellen, weshalb man dieses überhaupt benötigt und ob es nicht besser wäre, ein bereits bestehendes, bewährtes bzw. standardisiertes Schema zu übernehmen. Betrachtet man den GDF-Standard, so könnte man davon ausgehen, dass dieser eigentlich sämtliche Anforderungen erfüllen müsste. Die folgenden Gründe sprechen aber dennoch dafür, eine Nexus-spezifische Abstraktion und damit ein neues Schema zu entwerfen:

1. *Reduzierter Umfang und reduzierte Komplexität:* Das Nexus-Schema für Straßenverkehrsdaten soll deutlich weniger umfangreich und komplex sein als GDF, was die Transparenz der Daten für den Anwender erhöht.
2. *Reduzierte Information möglich:* Das Nexus-Schema soll – im Sinne des Anspruchs einer offenen Plattform – für den Datenanbieter nicht zu viele Vorgaben enthalten; jedem soll es möglich sein, bei Beachtung einfacher Vorschriften seine Daten in die Plattform einspeisen zu können. So sind beispielsweise die meisten Attribute optional angelegt. Vom GDF-Schema hingegen wird die Angabe vieler Informationen verlangt.
3. *Konformität zu (OpenGIS-)Standards bei der Abbildung der Geometrie:* Innerhalb der Augmented World Modeling Language von Nexus wird GML zur Abspeicherung der Geometrie von raumbezogenen Objekten verwendet. Dadurch wird eine Standard-Konformität erreicht, die in GDF bislang noch nicht gegeben ist.
4. *Verschiedene Geometrien für ein Objekt möglich:* In GDF haben Feature-Klassen einen bestimmten Geometrietyp. Diese Einschränkung kann im AWS nicht akzeptiert werden. Hier müssen Objekte in verschiedenen Detailstufen und damit auch in verschiedenen Geometrietypen vorgehalten werden können.
5. *Vernetzung von Navigationsdaten zum Zwecke der intermodalen Navigation:* Neben den Straßendaten müssen im Sinne einer intermodalen Navigation auch Daten von Fußgänger- und Fahrradwegen, Bus- und Bahnlinien und Luftverkehrswegen zusammengeführt werden. Dies sollte in einem einheitlichen Schema geschehen.

5.2.2 Use Cases für Straßenverkehrsdaten

Die Anforderungen an ein Datenmodell ergeben sich aus der Fragestellung, wofür die Anwendungen die Daten verwenden wollen. In Bezug auf Straßendaten sind die folgenden Anwendungstypen identifizierbar, die im Sinne der objektorientierten Modellierung in so genannten Use Cases formuliert werden:

- *Straßentypen in der Karte darstellen:* In Karten sollen die verschiedenen Arten von Straßen unterschiedlich dargestellt werden können, so z.B. Tunnels und Brücken, Straßen verschiedener Ordnung, etc. Das bedeutet, dass diese Differenzierung aus dem Datenschema ableitbar sein muss.
- *Straßennamen in der Karte darstellen:* Die Straßennamen oder -bezeichner sollen in der Karte als Labels erscheinen. Das bedeutet, dass es möglich sein muss, Straßensegmente mit dem gleichen Straßennamen zu einem Objekt zusammenfassen zu können, um eine anschauliche kartographische Darstellung zu erreichen. Folglich muss das Datenschema Möglichkeiten zur Zusammenfassung von Basisobjekten und damit zur Bildung von Komplexobjekten unterstützen.
- *Geonachrichten an mobile Objekte versenden:* In Nexus existiert die Möglichkeit, Nachrichten in bestimmte geographische Gebiete zu senden. Beispielsweise könnten an alle Fahrzeuge, die in eine bestimmte Richtung unterwegs sind, Warnungen vor Unfallgefahren (Staus, aktuelle Unfälle, verschmutzte Fahrbahn, etc.) weitergegeben werden. Das bedeutet, dass Straßen möglichst in die unter-

schiedlichen Fahrbahnen entsprechend der Richtung, in der sie befahren werden dürfen, differenziert werden müssen. Dies muss im Schema ausgedrückt werden können.

- *Navigation:* Die Navigation ist eine zentrale Komponente ortsbezogener Dienste. Daher müssen die Daten für diese Anwendung geeignet sein, d.h. es müssen möglichst Angaben wie Hausnummern, aber auch Abbiegeverbote usw. im Datenschema vorgesehen werden. Darüber hinaus sind Geschwindigkeitsangaben für Straßen notwendig, um beispielsweise den schnellsten Weg herausfinden zu können. Das Datenschema muss die Möglichkeit bieten, die Straßendaten in Form eines Graphen zu repräsentieren.
- *Zonengenerierung um Straßen- und Kreuzungsobjekte:* Um Straßen- und Kreuzungsobjekte müssen Puffer gelegt werden können, um z.B. alle Gebäude entlang einer bestimmten Straße oder an einer bestimmten Kreuzung effizient finden zu können. Bezüglich der Straßen ist es ähnlich wie beim Labeling effizienter, wenn man aus mehreren, zur gleichen Straße gehörenden Segmenten ein Objekt bilden kann. Ansonsten müsste man für jedes Segment einen Puffer bilden oder vorher alle gleichnamigen Abschnitte aggregieren. Dasselbe Argument gilt auch für das Versenden von Geonachrichten oder von Ereignissen in bestimmte Straßen. Aus diesem Use-Case resultiert die Anforderung an das Zielschema, Straßen gemäß dem Straßennamen zusammenfassen und Kreuzungsobjekte explizit modellieren zu können.
- *Entwicklung von Mobilitätsmodellen:* Die Erstellung von Bewegungsmodellen mobiler Nutzer des Nexus-Systems erfordert u.a. eine detaillierte Bereitstellung der Straßengeometrie und der Navigationsmöglichkeiten innerhalb von Straßennetzen. Dementsprechend ergeben sich dieselben Anforderungen an das Schema wie aus dem Use-Case „Navigation“ heraus.

Neben den genannten, eher anwendungsbezogenen Use Cases existieren weitere, die sich u.a. auch aus der Anforderung, Mehrfachrepräsentationen innerhalb von Nexus optimal zuordnen zu können, ergeben. So ist es beispielsweise notwendig, nicht nur kantenbasierte, sondern auch knotenbasierte Zuordnungen durchführen zu können. Daraus resultiert im Bezug auf Straßendaten die Bedingung, die Verbindungen von Straßenabschnitten explizit modellieren zu können.

5.3 Entwicklung eines globalen Schemas

In den voran gegangenen Kapiteln wurden die in Deutschland existierenden Schemas zur Beschreibung von Straßendaten sowie die sich aus Nexus heraus ergebenden Anforderungen im Hinblick auf die Modellierung von Straßenobjekten vorgestellt. Diese beiden Schritte liefern die Grundlage dafür, um ein globales Schema zur Repräsentation von Straßenverkehrsdaten in Nexus ableiten zu können. Dabei sind (in Anlehnung an [Conrad 02b]) die folgenden generellen Anforderungen an ein globales Datenmodell und damit auch für ein globales Schema zu berücksichtigen. Die beiden erstgenannten Anforderungen besitzen höchste Priorität:

1. Es soll für die Anwendungen bestmöglich geeignet sein.
2. Lokale Schemas sollen möglichst einfach, automatisiert und komplett (d.h. mit möglichst geringem Semantikverlust) in das globale Schema transformiert werden können.
3. Es soll möglichst viele und ausdrucksfähige Modellierungskonzepte besitzen, um eine optimierte Nachbildung der lokalen Schemas zu erlauben, damit die Eigenschaften der Quellobjekte aus den Ausgangsdaten so gut wie möglich erhalten bleiben können.

5.3.1 Anforderungskriterien für die Klassen

Aus den eben genannten Anforderungen an das resultierende Schema können folgende Anforderungen an die resultierenden Klassen abgeleitet werden:

Aus der Anforderung 1 folgt, dass alle Anforderungen, die in den Use Cases gestellt wurden, im Zielschema umsetzbar sein müssen. Das bedeutet beispielsweise, dass Straßensegmente mit dem gleichen Namen zu einem komplexen Objekt zusammengefügt werden können, etc. (vgl. Kapitel 5.2.2). Aus Anforderung 2 ergibt sich, dass es möglich sein muss, alle Objektklassen der Ausgangsschemas im AWS abbilden zu können (siehe Tabelle 5-3).

GDF Feature-Klassen	ATKIS-Objektarten	ALK-Objektarten
Road Element	Straße	Flurstück
Junction	Weg	
Road	Straßenkörper	
Intersection	Fahrbahn	
Brunnel	Tunnel	
	Brücke, Überführung, Unterführung	

Tabelle 5-3: Objektklassen aus den einzelnen Datenschemas, wie sie in den zur Verfügung stehenden Daten vorkommen

Aus der Anforderung 3 resultiert schließlich, dass Konzepte aus GDF wie die Bildung von Komplexobjekten, die Segmentierung von Features oder die Darstellung von Relationen zwischen Features möglichst erhalten bleiben sollen.

5.3.2 Realisierung der Klassen

Die Realisierung der Klassen war in das Nexus-Umfeld eingebettet. Beim Aufbau des Schemas musste daher den Vorgaben des Projektes Folge geleistet werden. Insbesondere war zu beachten, dass in Nexus das Konzept verfolgt wird, statische raumbezogene Objekte (vgl. Kapitel 3.2.2) in topologische und geographische Objekte zu unterteilen (siehe Abbildung 5-8).

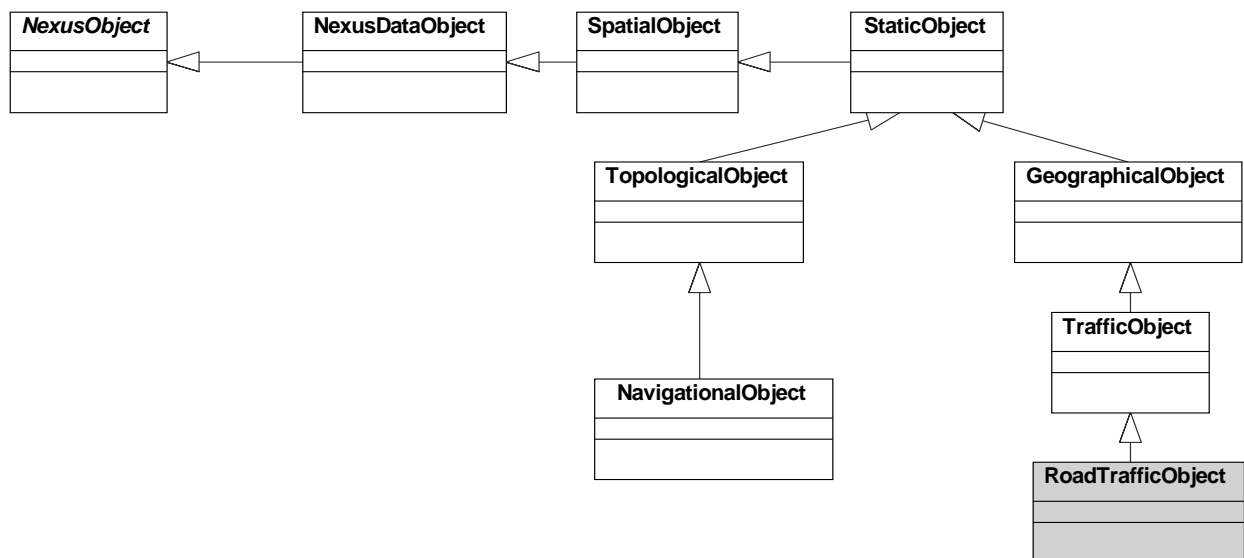


Abbildung 5-8: Differenzierung von topologischen und geographischen Objekten in Nexus

Im Rahmen dieser Arbeit wurden lediglich die geographischen Objekte aus dem Bereich des Straßenverkehrswesens im Augmented World Schema realisiert, die in der Klassenhierarchie alle von der Oberklasse *RoadTrafficObject* abgeleitet sind. Alle Anforderungen, die aus dem Use Case „Navigation“ in Bezug auf Straßendaten erwachsen, wurden in Nexus dagegen in Form von *NavigationalObjects* umgesetzt. Dabei wurde eine Klasse *NavigationalObject* geschaffen, die zu den Subklassen von *TopologicalObject* gehört. Von dieser Klasse können Typen zum Aufbau von Graphen (Navigationsknoten bzw. *NavigationNodes* und Navigationskanten bzw. *NavigationEdges*) und zur Unterstützung der Navigation innerhalb dieser Graphen wie z.B. Restriktionen für das Traversieren von Kanten, etc. vererbt werden. Im Fall einer Navigation in Straßennetzen enthalten die Navigationsobjekte Referenzen auf ihre geographischen Entsprechungen, um über diese auf Straßenattribute wie Straßennamen, usw. oder geometrische Eigenschaften wie Streckenlängen zugreifen zu können. Das Navigationskonzept von Nexus soll an dieser Stelle jedoch nicht weiter vertieft werden. Für eine nähere Betrachtung sei auf [Drosdol et al. 06] oder [Volz et al. 02] verwiesen.

Aus dem Vergleich der existierenden Geodatenschemas sowie aus den gegebenen Anforderungen heraus wurden schließlich die folgenden instanziierten Objektklassen zur Repräsentation von Straßendaten in Nexus aus der Oberklasse *RoadTrafficObject* (siehe Abbildung 5-9) abgeleitet:

Junction Element:

Objekt zur Begrenzung von *RoadElements*. *JunctionElements* kennzeichnen entweder topologische, attributive oder relationsabhängige Änderungen zwischen angrenzenden Straßenabschnitten in den Ausgangsdaten oder die Grenze eines Straßennetzes (Sackgassen).

RoadElement:

Straßensegment, das durch zwei *JunctionElements* begrenzt wird. *RoadElements* tragen die wichtigsten Attribute, z.B. den Straßennamen, Referenzen auf Hausnummern, Bedeutung der Straße, etc. Die meisten Attribute sind optional angelegt.

UndirectedRoad:

Komplexes Objekt, das sich aus einem bis n *RoadElements*, die den gleichen Straßennamen haben und in beide Richtungen befahrbar sind, zusammensetzt.

DirectedRoad:

Komplexes Objekt, das sich aus einem bis n *RoadElements*, die den gleichen Straßennamen und die gleiche Fahrtrichtung haben, zusammensetzt.

ComplexRoad:

Komplexes Objekt, das sich aus mindestens 2 *DirectedRoads* und 0 bis n *UndirectedRoads* oder mindestens einer *DirectedRoad* und 1 bis n *UndirectedRoads*, die alle den gleichen Straßennamen tragen, zusammensetzt. Die Klasse wird hauptsächlich für die Repräsentation von Straßen mit mehreren Fahrbahnen verwendet.

ComplexJunction:

Ggf. kann auch ein komplexes Objekt (nicht in der Abbildung 5-9 dargestellt), das sich aus Instanzen der genannten Objektklassen zusammensetzen kann, eingerichtet werden, um Elemente, die zu einem größeren, zusammenhängenden Kreuzungsbereich gehören, zu gruppieren.

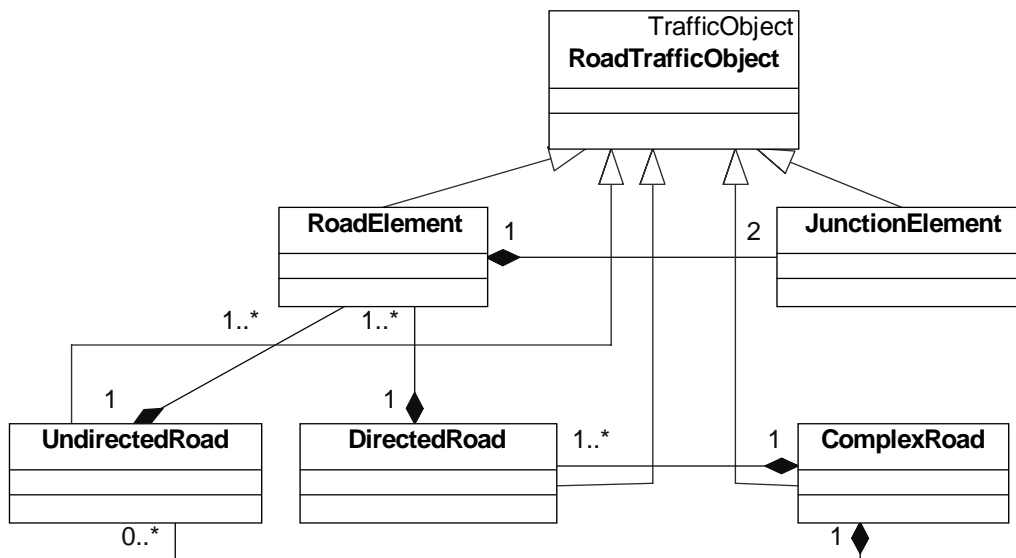


Abbildung 5-9: Einordnung der Klassen für Straßendaten in die Nexus-Hierarchie.

Die Angaben über die Art der Straße werden mittels der Attributierung von *RoadElements* gelöst, d.h. es werden im AWS keine separaten Klassen für Tunnel oder Brücken realisiert. Auch die Bedeutung von Straßen (Bundesstraße/Landstraße/etc.) wird den *RoadElements* in Form von Attributen mitgegeben. Ob ein *JunctionElement* ein Kreuzungsobjekt oder eine Sackgasse oder nur eine Verbindung zwischen zwei *RoadElements* repräsentiert, wird aus dessen Attribut „Valency“ abgeleitet, das die Anzahl der inzidierenden Kanten angibt.

Aus GDF und ATKIS wurde das Konzept der Bildung von zusammengesetzten Objekten übernommen. Allerdings wurden hier eigene Regeln für die Bildung von Komplexobjekten definiert. Der Grund dafür ist, dass Komplexobjekte in verschiedenen Schemas auf unterschiedliche Weise gebildet werden (im hier vorliegenden Fall existieren in GDF und ATKIS unterschiedliche Konzepte) und die Vereinbarkeit dieser Vorgehensweisen nicht immer gewährleistet ist bzw. bei der Umsetzung ins AWS Konflikte auftreten können. Außerdem bestehen von Nexus aus nur einfache Anforderungen an die Erzeugung von zusammengesetzten Objekten. Für den Fall, dass man aber noch Zugriff auf die Informationen zur Bildung von Komplexobjekten im Ursprungsformat haben möchte, wird eine Relation *BelongsToFormerComplex* auf ein Objekt *FormerComplex* eingeführt, das eine Datenstruktur zur Abspeicherung aller im Ursprungsformat zu einem Komplexobjekt zusammengefassten Instanzen anbietet. Auch das Prinzip der segmentierten Attribute kann für die *RoadElements* in Nexus angewandt werden. Hierzu werden die Anfangs- und Endpunkte eines Segments und der zugehörige Attributwert in einer speziellen Datenstruktur *SegmentedAttribute* aufgelistet.

5.4 Abbildungsvorschriften bestehender Schemas auf das AWS

Dieses Kapitel zeigt auf, welche Relationen zwischen den Ausgangs-Schemas GDF, ATKIS und ALK und dem Augmented World Schema bestehen und spezifiziert Abbildungsvorschriften für die Ausgangsschemas auf das AWS. Dabei gilt als Prämisse, eine möglichst einfache Übertragung der Ausgangsschemas in das Zielschema zu realisieren.

5.4.1 Abbildung von GDF auf das AWS

Die Abbildung der Level 1-Features von GDF auf das AWS ist eindeutig: Jedem Junction Feature entspricht im AWS genau ein *JunctionElement*, genauso entspricht jedem GDF Road Element Feature ein AWS *RoadElement* (wie in Abbildung 5-7/IV dargestellt). Die komplexen Features von GDF werden nicht im AWS repräsentiert. Die Abbildungsrelationen von GDF auf AWS zeigt Tabelle 5-4:

GDF	Abbildungsrelation	AWS
Junction	1:1	<i>JunctionElement</i>
Road Element	1:1	<i>RoadElement</i>

Tabelle 5-4: Abbildungsrelationen von GDF auf das AWS

5.4.2 Abbildung von ATKIS auf das AWS

Die Abbildung von ATKIS auf das AWS wird folgendermaßen realisiert, um eine Eindeutigkeit der Abbildung zu erreichen (siehe Tabelle 5-5).

ATKIS	Abbildungsrelation	AWS
Anfangspunkt Objektteil	1:1	<i>JunctionElement</i>
Endpunkt Objektteil	1:1	<i>JunctionElement</i>
Objektteil	1:1	<i>RoadElement</i>

Tabelle 5-5: Abbildungsrelationen von ATKIS auf das AWS

Es ist vorzuschicken, dass Straßenkörper-Objekte für die Abbildung im AWS keine Rolle spielen, da die reale Ausprägung von komplexen Straßen nur durch Fahrbahnen repräsentiert wird. Aus Straßenkörper-Objekten werden darum lediglich die Attribute extrahiert und den zugehörigen Fahrbahn-Instanzen zugeordnet. Die Aussagen im Hinblick auf die Abbildung von ATKIS-Objektteilen auf das AWS beziehen sich also ausschließlich auf die Objekte Straße, Weg und Fahrbahn.

In ATKIS existieren keine Elemente, die Anfang und Ende von Straßenabschnitten kennzeichnen. Sie müssen folglich aus den ATKIS-Daten abgeleitet werden, um eine Konformität mit dem Nexus-Schema zu realisieren. Somit werden sowohl der Anfangs- als auch der Endpunkt von Objektteilen zu AWS *JunctionElements*. Bei aneinandergrenzenden Objektteilen wird nur jeweils ein *JunctionElement* gebildet, um eine topologisch konsistente Datenstruktur zu gewährleisten. Objektteile werden im AWS 1:1 als *RoadElements*

abgebildet. Die Objektteile übernehmen dabei sämtliche Attribute des Objekts, um einen Informationsverlust zu vermeiden. In Datensätzen, in denen das Attribut „Straßenname“ belegt ist, kann auch die Bildung von komplexen AWS-Objekten erfolgen. Die Vorgehensweise zur Bildung von AWS-Objekten aus ATKIS wird in Abbildung 5-10 näher gebracht.

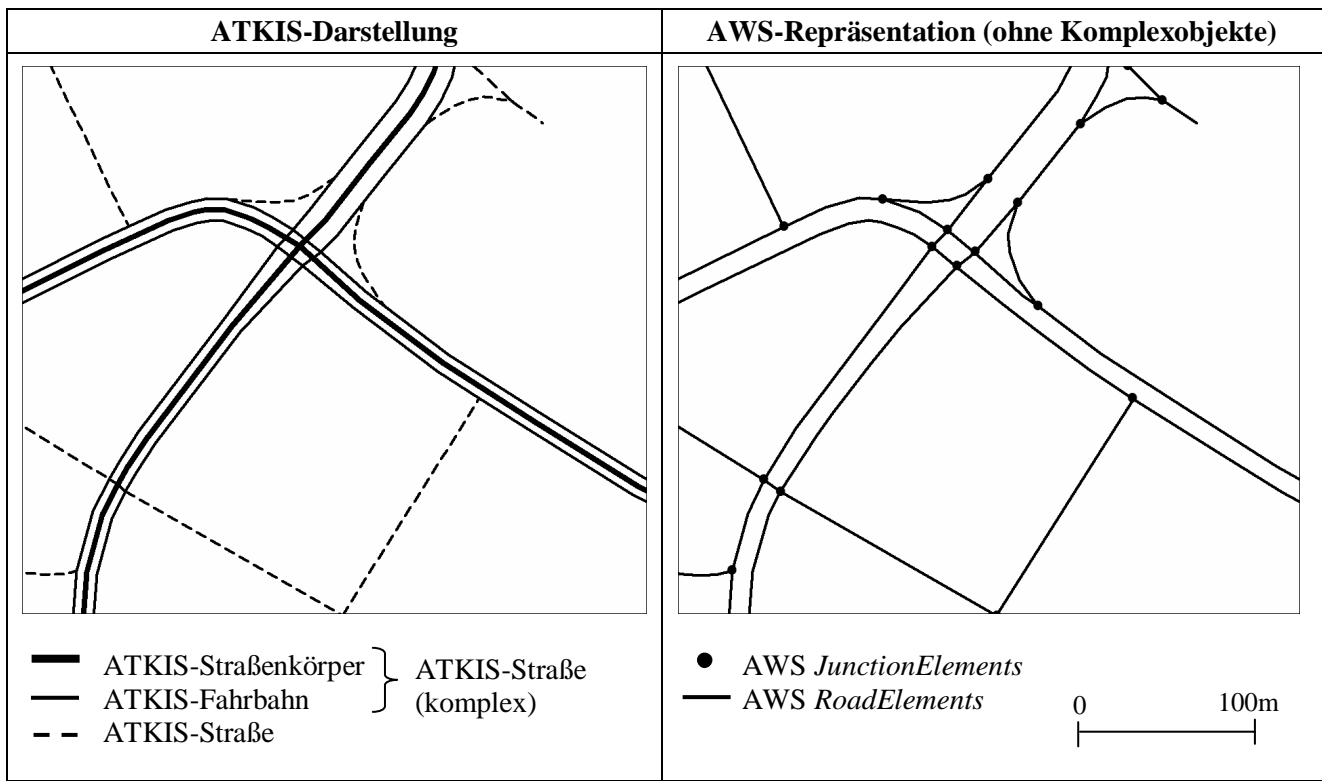


Abbildung 5-10: Repräsentation einer Szene in ATKIS und AWS

In [Volz & Bofinger 02] erfolgte eine Umsetzung von Straßendaten aus GDF und ATKIS nach AWS bzw. AWML anhand verschiedener Testdatensätze. Über XSL Transformationen [W3C 99] konnten die AWML-Daten dann in das Scalable Vector Graphics Format (SVG) konvertiert [W3C 02] werden, um deren Visualisierung über Standard-Browser zu ermöglichen.

5.4.3 Abbildung von ALK auf das AWS

Straßen existieren in den verfügbaren ALK-Daten als flächenhafte Objekte, während sie in ATKIS und GDF als Linien erfasst werden. Um eine gemeinsame Datenverarbeitung zu ermöglichen, müssen die Straßendaten der unterschiedlichen Schemas jedoch denselben geometrischen Typ aufweisen. Ansonsten würde man gegen die Anforderungen verschiedenster Use-Cases verstoßen. Beispielsweise könnten Straßen in Karten einmal als Linien und einmal als Flächen gezeichnet werden, was für deren Interpretation wenig förderlich wäre.

Es ist also notwendig, die Straßengeometrien zu vereinheitlichen. Einerseits könnte man die linienhaften Daten aus GDF und ATKIS über die Auswertung des Attributes zur Angabe der Straßenbreite in Flächen konvertieren. Andererseits ist es möglich, die flächenhafte Repräsentation der ALK-Straßen in eine linienhafte zu transformieren. Letztere Alternative ist dabei im Kontext von Nexus in der Regel angezeigt, da die meisten Use Cases eine Darstellung von Straßen in Form von Linien erfordern. Ohnehin können AWS *JunctionElements* und AWS *RoadElements* nicht eindeutig aus den verfügbaren flächenhaften ALK-Daten abgeleitet werden, da Kreuzungsbereiche aufgrund der völlig anderen, an Eigentumsverhältnissen orientierten Erfassungsvorgabe bei Flurstücken sehr heterogen in Erscheinung treten können (siehe Abbildung 5-11).

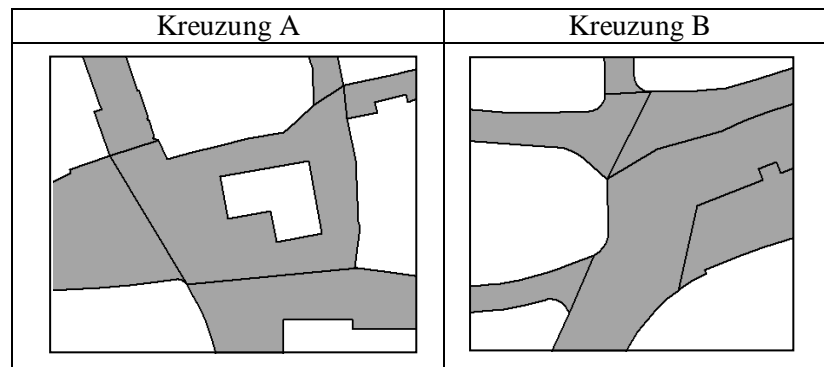


Abbildung 5-11: Verschiedene Kreuzungsbereiche aus der ALK (Ausschnitt aus dem Bereich Stuttgart Vaihingen)

Die flächenhaften ALK-Straßen müssen also in linienhafte Elemente transformiert werden. Eine Möglichkeit hierzu besteht darin, die Flurstücke zunächst als binäres Rasterbild zu repräsentieren (siehe Abbildung 5-12.1) und anschließend durch eine Skelettierungsoperation zu linearisieren (siehe Abbildung 5-12.2). Schließlich muss die erhaltene Rasterstruktur über eine Raster-Vektor-Konvertierung wieder in eine Vektordarstellung (siehe Abbildung 5-12.3) überführt werden.

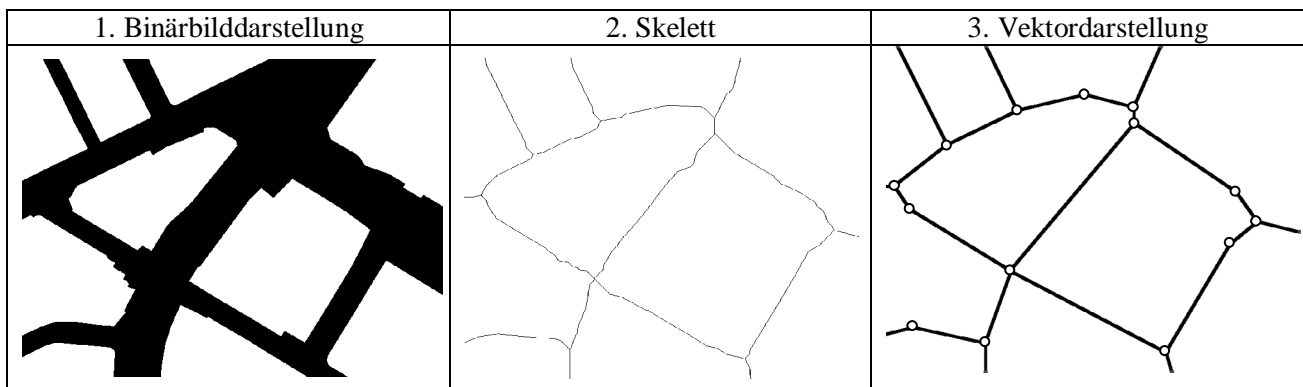


Abbildung 5-12: Linearisierung von ALK-Straßenobjekten: Mittels Skelettierung auf Rasterebene und anschließender Konvertierung in Vektorobjekte können anfangs flächenhaft vorliegende Objekte in Linienobjekte umgewandelt werden.

In Abbildung 5-13 werden die Ergebnisse der Linearisierung für die komplexen Kreuzungsbereiche aus Abbildung 5-11 den Darstellungen aus ATKIS gegenübergestellt. Obwohl es sich um eine einfache, nicht auf die spezielle Ausprägung von Straßenobjekten ausgerichtete Skelettierung handelt, sind Zuordnungen auf der Basis der Geometrie erkennbar.

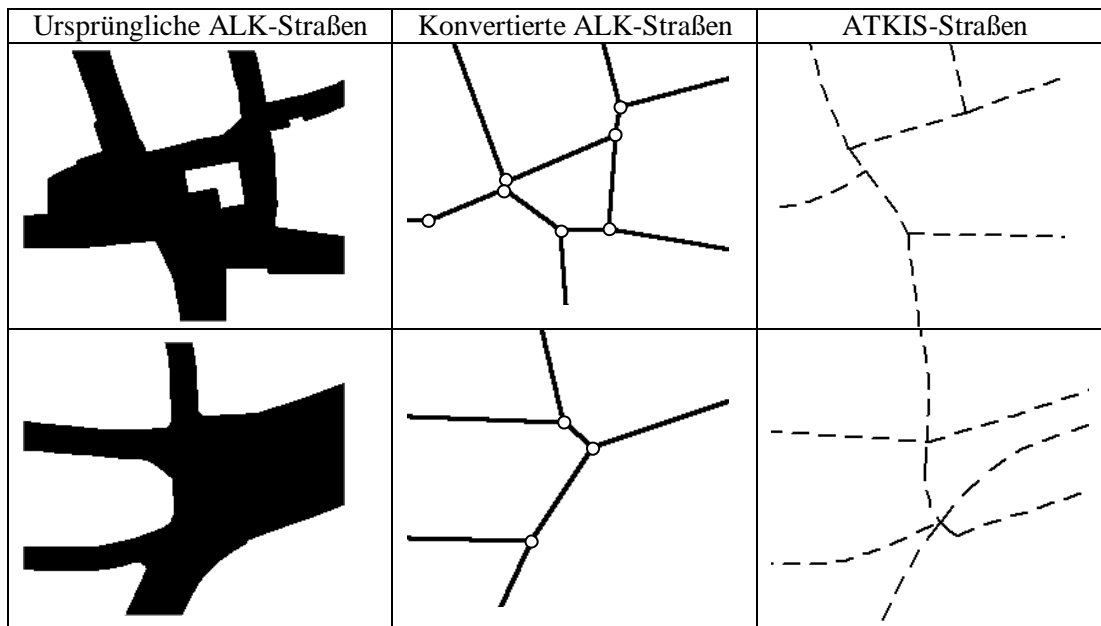


Abbildung 5-13: Gegenüberstellung der ursprünglichen und der abgeleiteten ALK-Straßenobjekte mit der ATKIS-Repräsentation

Somit können eindeutige Abbildungsrelationen der ALK-Straßenobjekte analog zu GDF und ATKIS wie folgt erreicht werden (siehe Tabelle 5-6).

ALK	Abbildungsrelation	AWS
Angrenzungspunkte der Segmente, die sich aus der Raster-Vektor-Konversion ergaben	1:1	<i>JunctionElement</i>
Segmente aus der Raster-Vektor-Konversion	1:1	<i>RoadElement</i>

Tabelle 5-6: Abbildungsrelationen von ALK auf AWS.

Wie aus Abbildung 5-12 und Abbildung 5-13 zu erkennen ist, werden in manchen Angrenzungsbereichen von Flurstücken allerdings mehrere *JunctionElements* gebildet, da der Prozess der Raster-Vektor-Konvertierung keine semantischen Gesichtspunkte berücksichtigt. Insgesamt kann das vorgestellte Verfahren daher, vor allem in komplexen Bereichen, zu Problemen führen. Demzufolge müssen weitere Ansätze untersucht werden. Denkbar wäre es z.B., die topologischen Beziehungen, die zwischen den Flurstücken bestehen, auf eine Graph- bzw. eine Knoten-/Kantenstruktur abzubilden und die Knoten als *JunctionElements* sowie die Kanten als *RoadElements* zu interpretieren. Im Hinblick auf eine optimierte Vorgehensweise zur Lösung der Problematik muss an dieser Stelle ebenfalls auf Verfahren der kartographischen Generalisierung verwiesen werden, wie sie z.B. in [Sester 00] untersucht wurden.

6 Modellierung von Relationen zwischen Mehrfachrepräsentationen

Modellierung in Geo-Informationssystemen umfasste bisher ausschließlich die formale Beschreibung der Geometrie, der Thematik und gegebenenfalls der Topologie eines Objektes. In Anbetracht der Tatsache, dass es durch die Anhäufung raumbezogenen Datenmaterials aber immer häufiger zu Mehrfachrepräsentationen kommt, die im Zuge des Datenaustauschs gemeinsam verarbeitet oder fortgeführt werden müssen, wird in dieser Arbeit ein Ansatz für die formale Beschreibung von Relationen zwischen Mehrfachrepräsentationen in Form so genannter *MRep-Relationen* vorgeschlagen (siehe Abbildung 6-1). Durch die explizite Modellierung der Beziehungen zwischen Mehrfachrepräsentationen soll zum einen eine Optimierung bei der Analyse sowie bei der Verschmelzung und der Fortführung von Mehrfachrepräsentationen erreicht werden. Zum anderen soll dadurch die Grundlage dafür geschaffen werden, über die Auswertung der Relationen auf Instanzebene in automatisierter Weise Relationen auf Schemaebene abzuleiten. Die Untersuchung geschieht dabei anhand von Straßendaten.

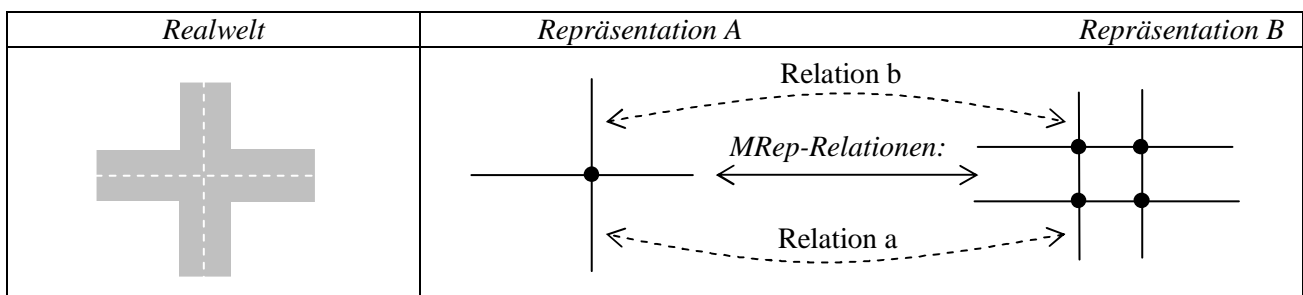


Abbildung 6-1: Das Grundkonzept: Relationen zwischen Mehrfachrepräsentationen sollen explizit in Form von *MRep-Relationen* ausgedrückt werden. Die Untersuchung wird dabei anhand von mehrfach repräsentierten Straßendaten durchgeführt.

Im Folgenden wird die Erzeugung von *MRep-Relationen* kurz motiviert. Anschließend wird der Begriff „Relation“ beleuchtet und es wird aufgezeigt, welche zusätzlichen Rahmenbedingungen beachtet werden müssen, wenn *MRep-Relationen* innerhalb von Nexus bzw. innerhalb des Augmented World Schemas repräsentiert werden. Darüber hinaus werden Anforderungen an das Modell für *MRep-Relationen* aus der Sicht der Anwendungen formuliert. Es werden Ähnlichkeitsmaße zwischen raumbezogenen Objekten vorgestellt und die Struktur von *MRep-Relationen* wird formalisiert. Schließlich wird eine Vorgehensweise zur Typifizierung von Relationen zwischen Mehrfachrepräsentationen andiskutiert, bevor letztlich das auf der Basis der *MRep-Relationen* modifizierte Konzept zum Umgang mit Mehrfachrepräsentationen in Nexus vorgestellt wird.

6.1 Motivation für die Speicherung von *MRep-Relationen*

Geht man davon aus, dass innerhalb einer Geodateninfrastruktur wie Nexus mehrfach repräsentierte Objekte vorliegen, ohne dass diese Kenntnis voneinander besitzen, so ist jedes Mal, wenn zwei Datensätze miteinander zu verarbeiten sind, ein zeitaufwendiger Zuordnungsprozess durchzuführen. Dies ist in der Praxis nur teilweise möglich, da hierdurch in der Regel zu lange Antwortzeiten entstehen würden. Deshalb sollten zumindest die Zuordnungsbeziehungen über Objektreferenzen oder explizite Zuordnungslisten vorgehalten werden. Will man nun jedoch Mehrfachrepräsentationen miteinander verschmelzen, gemeinsam analysieren oder gemeinsam fortführen (vgl. Kapitel 4.2.2), so benötigt man weitere Informationen über die Art der Beziehungen von Mehrfachrepräsentationen. Diese können entweder bei jedem dieser Vorgänge von neuem abgeleitet oder sie können persistent gespeichert werden. Die im Rahmen der vorliegenden Arbeit verfolgte Speicherung der Objektrelationen hat den Vorteil, dass ohne weitere Prozessierung direkt auf die notwendigen Informationen zugegriffen werden kann, was eine zeitliche Einsparung bringt. Man kann somit sehr schnell erkennen, ob eine gemeinsame Weiterverarbeitung von Mehrfachrepräsentationen möglich bzw. sinnvoll ist. Schließlich erlauben die *MRep-Relationen* auch eine standardisierte Vorgehensweise bei der Analyse, der Verschmelzung und der Fortführung von Mehrfachrepräsentationen.

6.2 Zum Begriff „Relation“

Der Begriff der Relation kann ganz allgemein als Beziehung zwischen zwei oder mehreren Objekten definiert werden. Im objektorientierten Sinne sind Relationen ebenfalls Objekte, d.h. es kann potentiell auch Relationen zwischen Relationen geben. So könnte man sich beispielsweise vorstellen, dass zwei mittels Relationen repräsentierte Abbiegeverbote in den Straßendatensätzen A und B über eine Relation verknüpft sind. Theoretisch können auch mehrfach repräsentierte Relationen auftreten.

In der Mathematik werden Relationen als Beziehungen zwischen den Elementen einer Menge oder den Elementen verschiedener Mengen aufgefasst. Sie werden über das kartesische Produkt von Mengen definiert [Reinhardt & Soeder 98]:

Definition des Kartesischen Produktes:

$M_1 \times \dots \times M_n := \{(x_1, \dots, x_n) \mid x_i \in M_i\}$ heißt kartesisches Produkt der Mengen M_1, \dots, M_n .

Definition n-stellige Relation:

Jede Teilmenge $R \subseteq M_1 \times \dots \times M_n$ heißt n-stellige Relation.

Zweistellige Relationen $R \subseteq M_1 \times M_2$ können als Zuordnungen von Elementen einer Menge 1 zu Elementen einer Menge 2 gemäß einer definierten Vorschrift interpretiert werden, wobei sich diese als Graph darstellen lassen (siehe Abbildung 6-2).

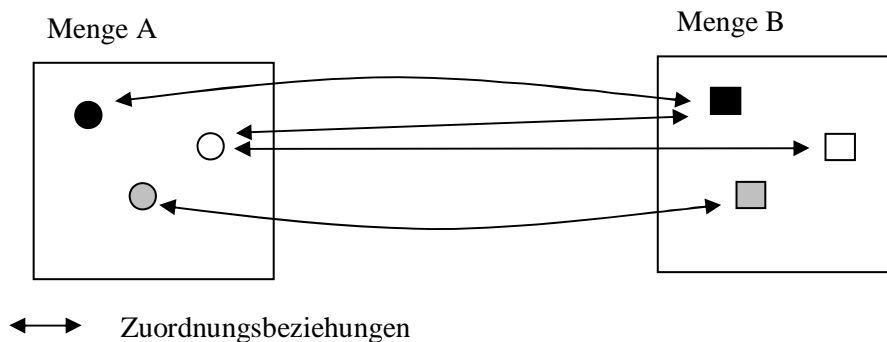


Abbildung 6-2: Zuordnungsbeziehungen zwischen den Elementen zweier Mengen können als Graph repräsentiert werden (nach [Reinhardt & Soeder 98])

Relationen existieren in verschiedenen Domänen. Zeitliche Relationen wurden u.a. von [Allen 83] betrachtet. Er entwarf Operatoren, um Relationen zwischen Zeiträumen wie z.B. *before*, *meet*, *equal* ausdrücken zu können. [Parent et al. 99] verstehen unter raumbezogenen Relationen diejenigen, die auf der Topologie, der Ausrichtung oder der Metrik von Objekten beruhen. Interessant ist auch eine weitere Kategorie von Relationen, die die Autoren betrachteten: jene der dynamischen Relationen. Hiermit können beispielsweise Objektwandlungen von anfänglichen zu resultierenden Objektklassen ausgedrückt werden. So entwickelt sich beispielsweise eine Instanz der Klasse „Student“ über die Relation „hatDiplomErhalten“ zur Instanz der Objektklasse „Graduierter“. Zu den dynamischen Relationen gehören auch die Erzeugungs-Relationen, die festhalten, durch welche Prozesse neue Objekte entstanden. Somit kann z.B. nachvollzogen werden, dass ein Flurstück durch die Aufteilung eines anderen oder durch die Verschmelzung mehrerer Einzelobjekte erzeugt wurde. [Kent 00] bemerkt zu Relationen: „*Relationships are the stuff of which information is made. Just about everything in the information system looks like a relationship.*“ Er weist ferner darauf hin, dass es einen Grund geben muss, weshalb zwei oder mehrere Dinge in Beziehung stehen. Im Hinblick auf Mehrfachrepräsentationen besteht der Grund für deren Beziehung in der Tatsache, dass sie dasselbe Realweltobjekt repräsentieren.

Im Datenbankbereich werden Relationen über bestimmte Begriffe wie Grad bzw. Stelligkeit, Rollen, etc. charakterisiert. Diese Begriffe wurden bereits in Kapitel 2.4.1.1 erklärt, als es um die Beschreibung des Relationskonzeptes des OGC ging. In relationalen Datenbanksystemen sind Relationen gleichzusetzen mit Tabellen, deren Zeilen (bzw. Tupel) den Datenobjekten (Instanzen) entsprechen, während die Spalten die

Eigenschaften der Datenobjekte (bzw. das Schema) repräsentieren. Beim konzeptuellen Entwurf von Datenbanken ist das Entity-Relationship-Modell weit verbreitet. Dabei werden „Entities“ bzw. Gegenstände über „Relationships“ bzw. Beziehungen miteinander verbunden. Für weitere Ausführungen zu Relationen auf dem Gebiet der Datenbanksysteme sei auf [Kemper & Eickler 99] verwiesen.

Im Kontext der vorliegenden Arbeit soll unter dem Begriff der Relation in erster Linie eine Beziehung zwischen zwei raumbezogenen Mehrfachrepräsentationen verstanden werden. Die Beziehungen von raumbezogenen Objekten haben besondere Charakteristika. Diesen soll bei der Modellierung von MRep-Relationen Rechnung getragen werden. Es gilt dabei formal auszudrücken, auf welche Weise Mehrfachrepräsentationen in GIS miteinander in Beziehung stehen.

6.3 Rahmenbedingungen für die Modellierung von MRep-Relationen

Prinzipiell hat der Ansatz zur Modellierung von Relationen zwischen Mehrfachrepräsentationen einen generischen Charakter, d.h. MRep-Relationen in der präsentierten Form sollen zwischen beliebigen raumbezogenen Mehrfachrepräsentationen gespeichert werden können. Allerdings ist dieses Konzept im besonderen auch für die Nexus-Plattform vorgesehen. Sollen MRep-Relationen innerhalb von Nexus abgespeichert werden können, so ergeben sich zusätzliche Rahmenbedingungen. Diese werden hier kurz skizziert. Darüber hinaus stellen auch die Anforderungen der Anwendungen, die MRep-Relationen auswerten, gewisse Vorgaben für die Modellierung der Relationen zwischen Mehrfachrepräsentationen dar. Sie sollen an dieser Stelle ebenfalls Beachtung finden.

6.3.1 Anforderungen durch Nexus

Sollen MRep-Relationen in Nexus möglich sein, so müssen sie auch innerhalb des Augmented World Schemas repräsentiert werden. Deshalb unterliegt die Modellierung von MRep-Relationen einerseits den allgemeinen Regeln des AWS. Andererseits wurde für die Darstellung von Relationen in Nexus ein explizites Relationenmodell aufgesetzt [Drosdol et al. 04], das auch die Richtlinie für die Abbildung von MRep-Relationen im AWS darstellt.

Das zu entwickelnde Schema für MRep-Relationen muss sämtliche Relationen zwischen Mehrfachrepräsentationen innerhalb des Augmented World Modells abspeichern können. In der Folge wird aufgezeigt, in welchen Punkten sich die Objekte bedingt durch die Regeln des Augmented World Schemas unterscheiden können.

6.3.1.1 Allgemeine Vorgaben durch das AWS

Um die Konformität mit dem Augmented World Schema zu gewährleisten, muss bei der Modellierung von MRep-Relationen beachtet werden, dass die über eine MRep-Relation verbundenen Objekte die folgenden unterschiedlichen Eigenschaften haben können:

Komplexitätsgrad

Als erstes muss berücksichtigt werden, dass raumbezogene Objekte innerhalb des AWS entweder einfach oder komplex sein können.

Einfache raumbezogene Objekte: Einfache raumbezogene Objekte oder auch Basisobjekte sind zu definieren als Objekte, die nicht weiter unterteilt werden können. Basisobjekte enthalten die Objektgeometrien.

Komplexe raumbezogene Objekte: Komplexe raumbezogene Objekte oder auch Komplexobjekte sind zu definieren als Objekte, die sich aus weiteren Komplexobjekten und/oder Basisobjekten zusammensetzen können. Ein Komplexobjekt besteht aus mindestens zwei Unterobjekten. In Analogie zum Datenmodell von GDF, in dem Komplexobjekte wie Roads oder Intersections existieren, gibt es in Nexus ebenfalls zusammengesetzte Objekte wie z.B. *ComplexRoads*, die auf die entsprechenden Basis-Objekte, also die konstituierenden *RoadElements*, über Referenzen verweisen.

Klassenzugehörigkeit

Mehrfach repräsentierte Objekte können unterschiedlichen Objektklassen des AWS angehören. Dies kann beispielsweise der Fall sein, wenn ein mehrfach repräsentiertes Objekt einer Superklasse (z.B. „Restaurant“) und eine korrespondierende Repräsentation einer Subklasse (z.B. „italienisches Restaurant“) angehört. Außerdem ist dieser Fall möglich, falls ein erweitertes Klassenschema eine Klasse des Standardschemas spezialisiert und eine Instanz dieser Klasse das gleiche Objekt repräsentiert wie eine Instanz der Standardklasse.

Attribute

Da die meisten Attribute innerhalb des Augmented World Schemas optional sind, kann man bei Mehrfachrepräsentationen nicht davon ausgehen, dass sie bezüglich ihrer Attribute übereinstimmen müssen. Es können daher auch nur partielle Übereinstimmungen oder im Extremfall überhaupt keine Übereinstimmungen existieren.

Geometriertyp

Aus der Struktur des AWS ergibt sich darüber hinaus, dass Mehrfachrepräsentationen potentiell unterschiedliche Geometriertypen haben können, da das AWS seinen Objektklassen keine vordefinierten Geometrien zuweist. Dies geschieht deshalb, weil pro Objekt grundsätzlich verschiedene Detailstufen bzw. Levels of Detail (LOD) möglich sein müssen. Innerhalb einer Detailstufe bzw. für eine Anwendung müssen die Geometriertypen jedoch einheitlich sein.

Zugehörigkeit zu einer Augmented Area

Des Weiteren gilt die Vorgabe, dass pro Augmented Area nur eine Repräsentation eines Realweltobjektes existieren darf, d.h. innerhalb einer Augmented Area kann es keine Mehrfachrepräsentationen geben. Gleichzeitig bedeutet dies auch, dass ein Objekt einer Augmented Area A nur eine MRep-Relation zu einem Objekt einer Augmented Area B haben kann.

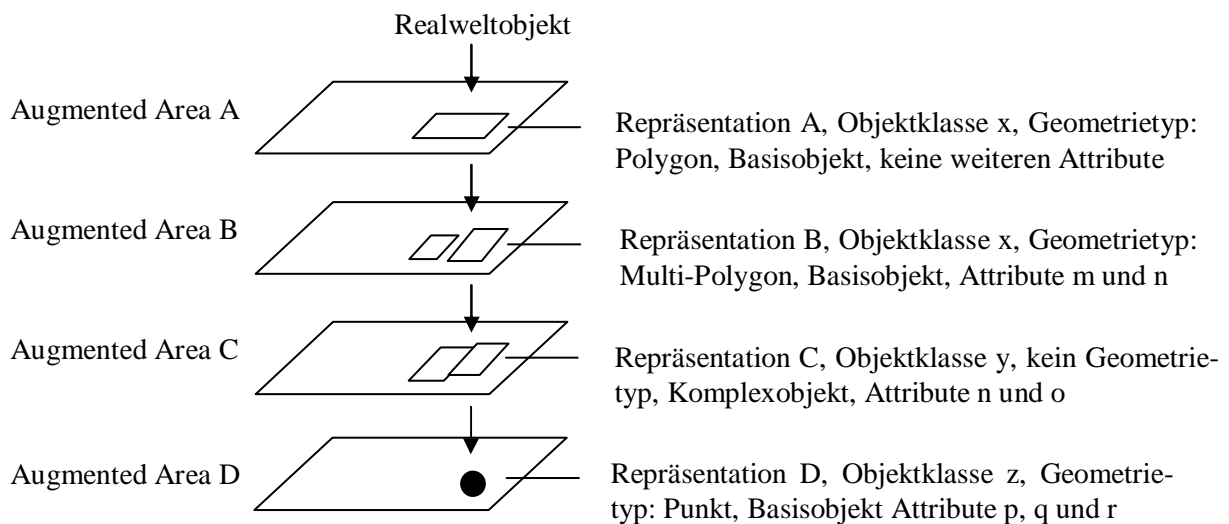


Abbildung 6-3: AWS-bedingte Unterschiede von Mehrfachrepräsentationen

Abbildung 6-3 zeigt in der Zusammenschau, welche Unterschiede zwischen Mehrfachrepräsentationen aus dem Ansatz des Augmented World Schemas heraus möglich sein müssen: Sie müssen unterschiedlichen Objektklassen angehören und unterschiedliche Komplexitätsstufen aufweisen sowie theoretisch unterschiedliche geometrische Typen haben können. Pro Augmented Area darf es zudem nur eine Repräsentation eines Realweltobjektes geben.

6.3.1.2 Anforderungen durch das Relationenmodell von Nexus

Innerhalb des Nexus-Projektes wurde auch ein Ansatz zur Realisierung von Relationen zwischen Objekten entwickelt [Drosdol et al. 04] und im AWS verankert. Er baut auf der Abstrakten Spezifikation des OGC zum Thema „Relationen zwischen Features“ [OpenGIS 99b] auf, die bereits in Kapitel 2.4.1.1 vorgestellt wurde. Sollen MRep-Relationen in Nexus eingeführt werden, so muss der Nexus-spezifischen Richtlinie zur Bildung von Relationen Folge geleistet werden.

Relationen als eigenständige Objekte

Für die Modellierung von Relationen zwischen Objekten sind grundsätzlich verschiedene Ansätze denkbar. Gemäß dem OGC wird hier zwischen so genannten leicht- und schwergewichtigen (light- und heavyweight) Relationen unterschieden. Leichtgewichtige Relationen sind ausschließlich binär und existieren nicht als eigenständiges Objekt, sondern lediglich als Referenz innerhalb eines Objektes. Daher können sie auch keine eigenen Attribute tragen. In Nexus kommen zwar Relationen vor, für die die genannten Limitierungen belanglos sind, doch die Mehrzahl der Objektbeziehungen muss durch eigenständige, attributierbare Relationsobjekte realisiert werden, die auch eine höhere Stelligkeit erlauben. In Nexus wird daher das Konzept der schwergewichtigen Relationen unterstützt. Diese Vorgehensweise bedeutet keine Einschränkung bei der Erstellung von Relationen, da alle leichtgewichtigen Relationen auch über schwergewichtige ausgedrückt werden können. Für die vorliegende Arbeit gilt also, dass alle MRep-Relationen ebenfalls schwergewichtig sein müssen. Der Unterschied zwischen schwer- und leichtgewichtigen Relationen wird in der folgenden Abbildung 6-4 noch einmal verdeutlicht.

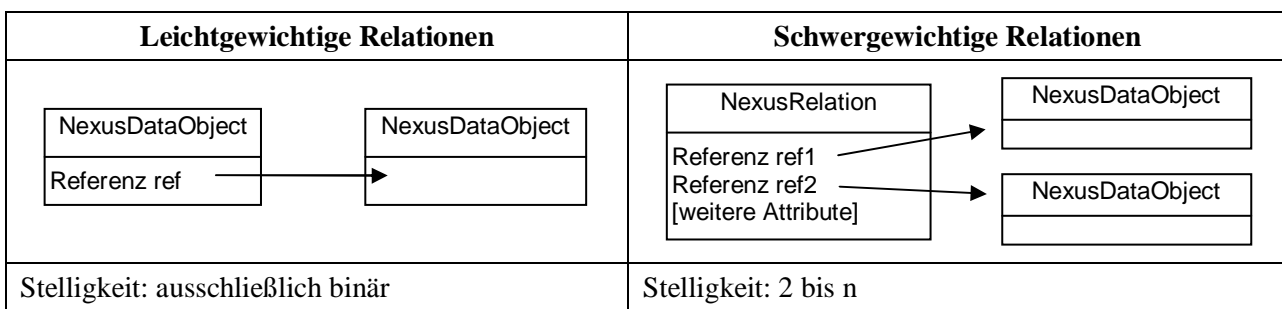


Abbildung 6-4: Der Unterschied zwischen leicht- und schwergewichtigen Relationen (nach [Drosdol et al. 04])

Die Stelligkeit einer Relation wird durch die Zahl der referenzierten Objekte bestimmt. Heavyweight-Relationen erlauben zudem die Möglichkeit, nur bestimmte Objektklassen für einen Relationstyp zuzulassen. So könnte beispielsweise eine Relation „*istBefreundetMit*“ nur Objekte der Objektklasse „*Person*“ in Beziehung setzen. Mit Hilfe dieser Maßnahme können unlogische Relationen und damit Inkonsistenzen vermieden werden.

Symmetrie von Relationen

Eine binäre Relation gilt als symmetrisch, wenn gilt: $(a,b) \in R \Leftrightarrow (b,a) \in R$. Ein Beispiel für eine binäre symmetrische Relation wäre etwa eine Partnerstadt-Relation, da a Partnerstadt von b und b Partnerstadt von a ist, d.h. die an der Relation teilnehmenden Objekte haben dieselben Rollen (siehe Abbildung 6-5.1). Daneben gibt es die unsymmetrischen Relationen, bei denen die Rollen der teilnehmenden Objekte unterschiedlich sind. Ein Beispiel hierfür ist die topologische Beziehung des Enthaltenseins, da hier unterschieden werden muss, welches Objekt ein anderes enthält bzw. welches in einem anderen enthalten ist (siehe Abbildung 6-5.2).

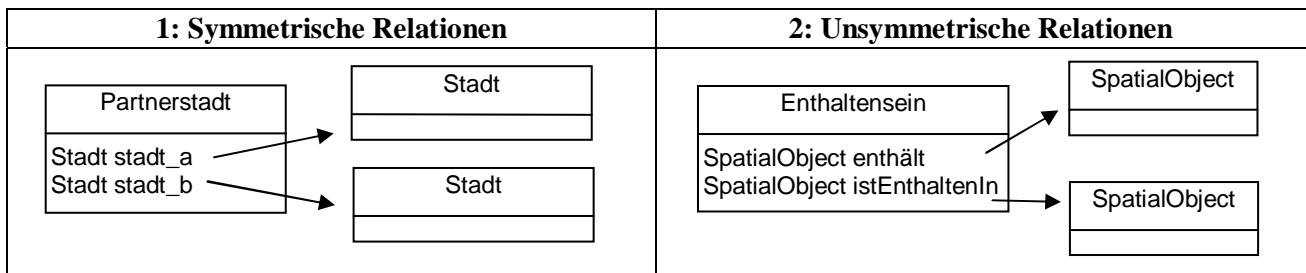


Abbildung 6-5: Symmetrische und unsymmetrische Relationen (nach [Drosdol et al. 04])

Bei nicht-symmetrischen Relationen ist die Zuordnung eindeutig, da die einzelnen Rollen der referenzierten Objekte anhand der referenzierenden Attribute unterschieden werden können, wie dies in Abbildung 6-5.2 bei der Enthaltensein-Relation ersichtlich ist. Symmetrische Relationen müssen jedoch über andere Konzepte abgebildet werden können. Dies kann ebenfalls am Beispiel der Partnerstadt-Relation aus Abbildung 6-5.1 erklärt werden. Wird nämlich ohne weitere Vorkehrungen eine solche Relation instanziiert und dem Attribut stadt_a="Stuttgart" und dem Attribut stadt_b="Brünn" zugewiesen, so wird eine Anfrage nach Partnerstadt (stadt_a="Brünn", stadt_b="Stuttgart") nicht beantwortet werden können, da "Stuttgart"!="Brünn" gilt (vgl. [Drosdol et al. 04]). Entweder müssen daher alle symmetrischen Relationen explizit als solche markiert werden oder die Attribute einer symmetrischen Relation werden als Mengen modelliert. In Nexus entschied man sich für letztere Variante, da hierdurch eine höhere Effizienz bei den Anfragen an die Föderationskomponente gewährleistet werden kann. Diese Vorgehensweise hat auch dann Vorteile, wenn es sich nicht nur um binäre symmetrische Relationen handelt, sondern wenn eine höhere Stelligkeit vorliegt und nur ein Teil der Referenzen einer Relation symmetrisch ist (siehe Abbildung 6-6). Da MRep-Relationen grundsätzlich symmetrisch sind, müssen diese ebenfalls als Mengen gespeichert werden.

Verwaltung von Relationen

Im Rahmen der Nexus-Plattform ist es unerlässlich, Relationsobjekte effizient auffinden zu können. Eine Möglichkeit, dies zu erreichen, bestünde theoretisch darin, Relationen ebenfalls eine räumliche Ausdehnung zu geben, die sich aus den Geometrien der an ihr beteiligten Objekte zusammensetzt, um Relationsobjekte über den Ortsbezug suchen zu können. Da manche Objekte aber eine große räumliche Distanz haben können (z.B. Partnerstädte) und somit die Ausdehnung der entsprechenden Relationsobjekte ebenfalls sehr groß wäre, sah man hier aber für Relationen im Allgemeinen gewisse Effizienzprobleme. Optional ist es aber dennoch möglich, ein Relationsobjekt auch mit einem Raumbezug zu versehen. Somit können Relationsobjekte potentiell auch über raumbezogene Abfragen gefunden werden. Dies ist insbesondere für MRep-Relationen relevant, da sie ja nur eine sehr begrenzte räumliche Ausdehnung haben. Sie können folglich gegenüber normalen Nexus-Relationen effizienter verwaltet werden.

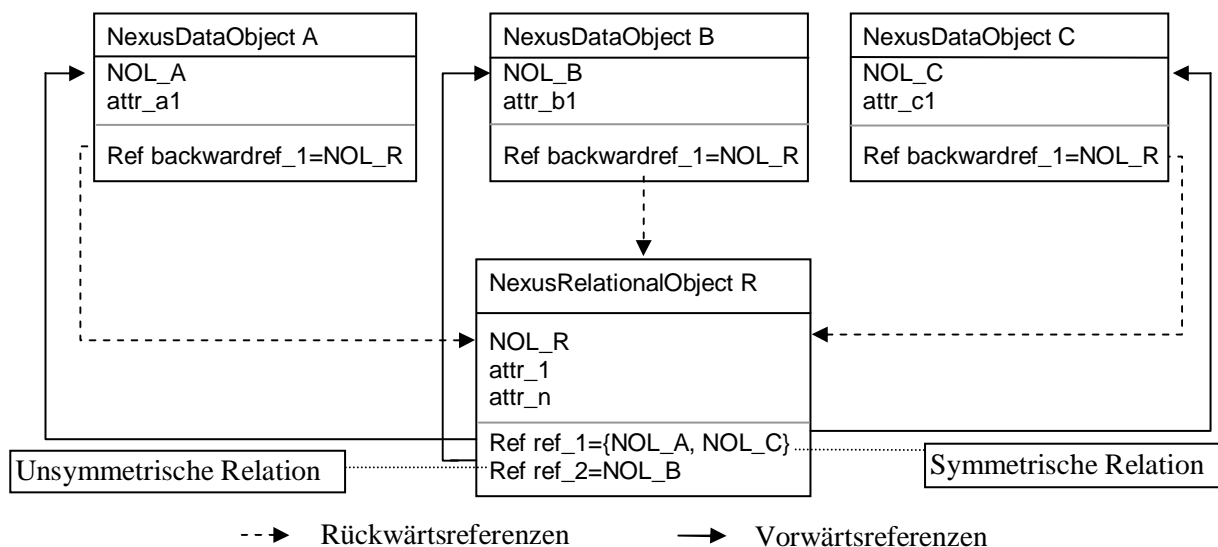


Abbildung 6-6: Das Relationskonzept von Nexus in der Übersicht

Als Suchstrategie für Nexus-Relationen wurde die Lösung favorisiert, von den an einer Relation beteiligten Objekten Rückwärtsreferenzen auf das zugehörige Relationsobjekt anzulegen. In Abbildung 6-6 wird dieses Konzept ersichtlich. Jedes *NexusDataObject* (also A, B und C), das eine Relation (also R) eingeht, bekommt ein Attribut („Ref“), das die Nexus-Identifikatoren, die so genannten *NexusObjectLocators* (NOLs), all jener Relationsobjekte enthält, die selbst auf es verweisen. Zusätzlich zur NOL kann noch angegeben werden, welche Rollen das Objekt innerhalb der Relation einnimmt und um welchen Relationstyp es sich handelt.

Die Verwaltung von Relationsobjekten bzw. von Objekten, die über Relationen miteinander in Beziehung gesetzt werden, schließt auch deren Aktualisierung mit ein. Das bedeutet, wenn Relationen oder partizipierende Objekte modifiziert oder gelöscht werden, sind alle Objekte, die von dieser Änderung betroffen sind, nachzuführen. Abbildung 6-7 zeigt ein Beispiel dafür. In Stadium I ist Objekt B in Objekt A enthalten. Durch eine Änderung der Geometrie von Objekt A ändert sich in Stadium II die Relation zwischen A und B: Die Enthaltensein-Beziehung wird aufgelöst und durch eine Überlappungs-Relation ersetzt. Durch die Löschung von Objekt B in Stadium 3 wird schließlich auch die Beziehung zwischen den Objekten A und B obsolet.

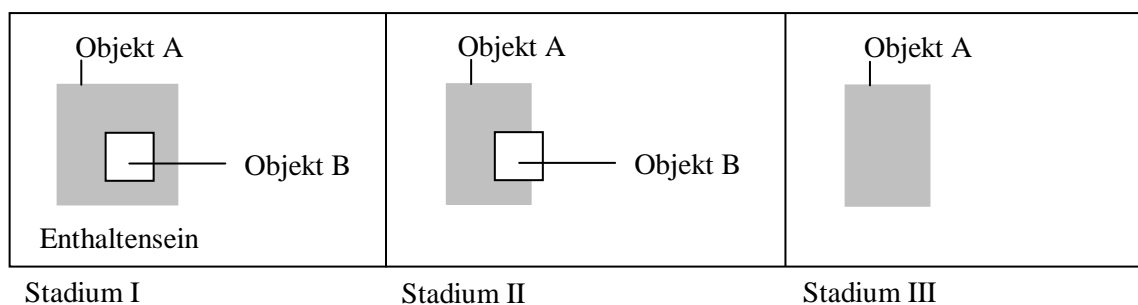


Abbildung 6-7: Ein Beispiel für die Nachführung von Relationen und Objekten, die an Relationen beteiligt sind

Ändert sich in Nexus ein Objekt bzw. eine Relation, so können über die Vorwärts- und Rückwärtsreferenzen jene Objekte identifiziert werden, die von der Änderung in Kenntnis gesetzt und gegebenenfalls selbst modifiziert werden müssen. Die jeweiligen Context Server der Plattform aktualisieren die Objekte auf diese Weise untereinander, ohne dass die Föderation einbezogen werden muss. Dieser Mechanismus gilt analog auch für Relationen zwischen Mehrfachrepräsentationen.

Die Klassenhierarchie der Nexus-Relationen

Die Superklasse der Relationen in Nexus ist *NexusRelation*. Sie ist eine direkte Unterklasse von *NexusDataObject*. Somit können alle instanziierten Objektklassen auch über Relationen verbunden werden. Bislang wurden 4 Subklassen von *NexusRelation* gebildet:

- *Spatial Relation*: Relationen, die durch die raumbezogenen Eigenschaften von Objekten zustande kommen können. Hier gibt es drei weitere allgemeine Unterklassen:
 - *TopologicalRelation*: beschreibt die topologischen Beziehungen;
 - *DistanceRelation*: beschreibt die Beziehungen, die als Distanzen zwischen geographischen Objekten ausgedrückt werden können, z.B. die direkte Distanz, die Distanz innerhalb eines Netzwerkes, etc.;
 - *DirectionalRelation*: beschreibt die Beziehungen der gegenseitigen Lage von Objekten, z.B. „liegt östlich von“, etc. Sie werden mittels einer Richtung ausgehend von einem Bezugsobjekt angegeben;
- *TemporalRelation*: Relationen, die die zeitlichen Zusammenhänge von Objekten ausdrücken können, um z.B. Aussagen darüber zu treffen, ob sich zwei Ereignisse zeitlich überlappt haben, ob ein Ereignis vor einem anderen eingetreten ist, etc.

- *ThematicRelation*: Relationen, die sich auf die Sachdaten von Objekten beziehen, also auf alle Attribute, die weder raum- noch zeitbezogen sind. Die erwähnte Relation „istPartnerstadtVon“ gehört zu dieser Kategorie.
- *MultiRepresentationalRelation*: Die Klasse zur Modellierung von Relationen zwischen Mehrfachrepräsentationen. Sie wird untenstehend näher erläutert.

Abbildung 6-8 zeigt die Basisklassen für die Realisierung von Relationen in Nexus in der Übersicht. Die aufgeführten Relationen stellen einen Grundbestand dar, der im weiteren Verlauf des Projektes erweitert und differenziert werden wird. So steht derzeit beispielsweise die Formulierung von Relationen im dreidimensionalen Raum an, um z.B. ausdrücken zu können, ob ein Objekt auf/neben/unter einem anderen steht usw.

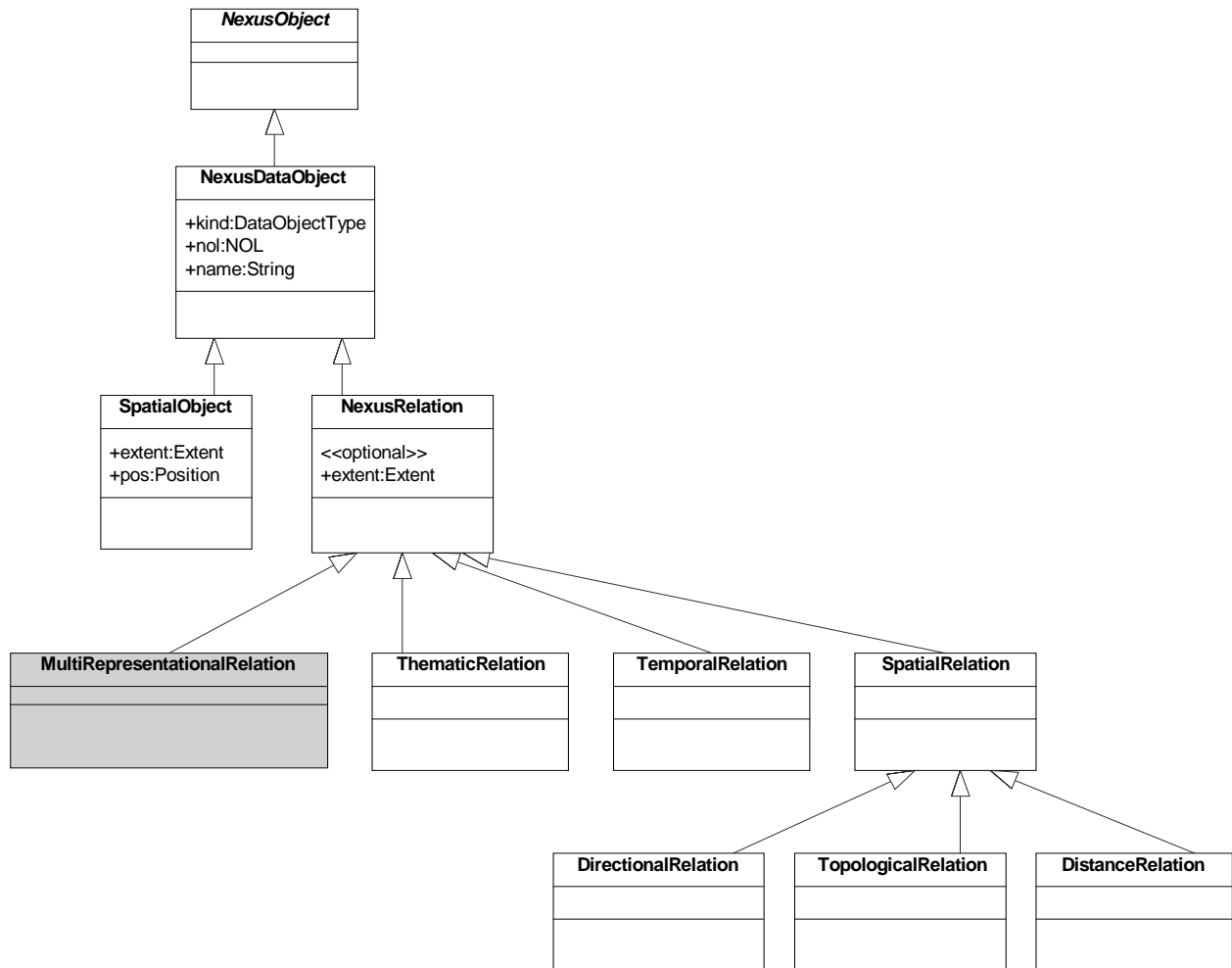


Abbildung 6-8: Die Hierarchie von Nexus-Relationen innerhalb des Augmented World Schemas (nach [Drosdol et al. 04])

6.3.2 Anforderungen durch die Anwendungen

Die Modellierung eines bestimmten Sachverhaltes wird stets vor dem Hintergrund durchgeführt, dass gewissen Anwendungsanforderungen Genüge zu leisten ist. MRep-Relationen sollen zum einen dazu dienen, die Analyse, Verschmelzung und Fortführung von Mehrfachrepräsentationen zu verbessern und zum anderen sollen mittels der Auswertung der MRep-Relationen auch Relationen zwischen korrespondierenden Objektklassen unterschiedlicher Schemas abgeleitet werden können. Im Anschluss wird daher darauf eingegangen, welche Anforderungen zur Ausgestaltung von MRep-Relationen sich aus den genannten Zielanwendungen ergeben.

6.3.2.1 Anforderungen durch Analyse, Verschmelzung und Fortführung

Um eine Analyse, Verschmelzung oder Fortführung von Mehrfachrepräsentationen überhaupt durchführen zu können, müssen korrespondierende Objekte aus verschiedenen Datenbeständen bekannt sein. Sie werden mit Hilfe von Zuordnungsprozessen identifiziert. Im Zuge entsprechender Algorithmen geschieht ein Vergleich wichtiger raumbezogener und sonstiger Eigenschaften von Objekten (siehe Abbildung 6-9). Aus diesem Objektvergleich heraus wird dann ersichtlich, wie ähnlich sich zwei Repräsentationen sind, d.h. ob sie entweder identisch, partiell übereinstimmend oder disjunkt sind. Die Informationen, die im Rahmen der Zuordnung ausgewertet werden, lassen sich im Grunde als MRep-Relationen ausdrücken. Es sind aber bisher keine Algorithmen bekannt, welche diese Relationen in Folgeprozesse wie Analyse, Verschmelzung und Fortführung auswerten. Diese Intention liegt dem Ansatz dieser Arbeit zugrunde.

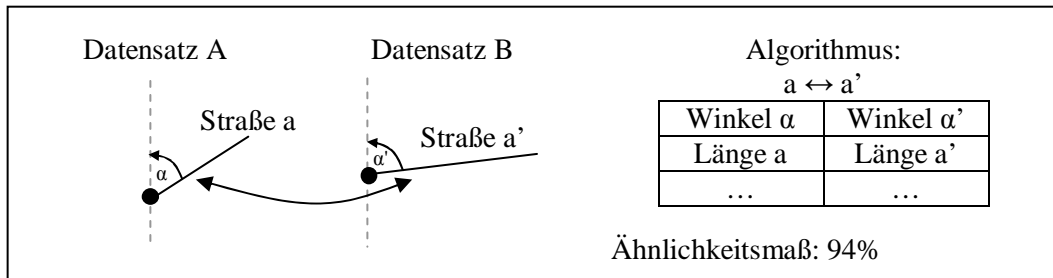


Abbildung 6-9: Auswertung von Objekteigenschaften im Zuge der Zuordnung. Mittels eines Ähnlichkeitsmaßes wird festgelegt, wie stark Repräsentationen übereinstimmen.

Um die Analyse, Verschmelzung und Fortführung von Mehrfachrepräsentationen eindeutiger, einfacher und verlässlicher machen zu können, ist es von großer Notwendigkeit, *den Grad der Ähnlichkeit* korrespondierender Repräsentationen möglichst detailliert beschreiben zu können. Diese Anforderung ist allen genannten Anwendungen gemeinsam. Es bestehen nur Unterschiede darin, welche Ähnlichkeitsmaße für die jeweiligen Prozesse von besonderer Bedeutung sind. Soll beispielsweise eine reine Verschmelzung der Geometrien durchgeführt werden, so sind hierfür ausschließlich topologische und geometrische Ähnlichkeitsmaße relevant. Auch Ähnlichkeitsmaße auf Metadatenebene, wie z.B. zur Datenqualität oder zum Erfassungszeitpunkt, liefern wertvolle Hinweise. Einen Überblick über mögliche Ähnlichkeitsmaße für Geodaten gibt Kapitel 6.4. An dieser Stelle ist anzumerken, dass ein wichtiges Ergebnis bei der Auswertung der in MRep-Relationen gespeicherten Informationen auch darin liegt, zeigen zu können, welche Mehrfachrepräsentationen *nicht* fusioniert oder gemeinsam analysiert und fortgeführt werden können.

6.3.2.2 Anforderungen aus der Sicht des automatischen Schema-Matchings

Sollen aus der Auswertung von MRep-Relationen zwischen Instanzen automatisch Beziehungen zwischen Schemas abgeleitet werden, so sind zunächst nur Informationen über die Klassen und die Attribute sowie die Attributwerte der zugeordneten Objekte bedeutsam. Es sind jedoch auch weiterführende Ansätze denkbar, die z.B. nur jene MRep-Relationen für die Ableitung von Schema-Korrespondenzen heranziehen, deren referenzierte Repräsentationen eine sehr hohe Ähnlichkeit aufweisen, d.h. auch hier ist die Bestimmung der Ähnlichkeit von Mehrfachrepräsentationen von Bedeutung. Dies wird in Kapitel 8.1 näher erklärt.

6.4 Ähnlichkeitsmaße für raumbezogene Objekte

Merkmale von GIS-Objekten können unterschiedlich ausgeprägt sein. Um herauszufinden, wie stark die Objektcharakteristika voneinander abweichen bzw. wie ähnlich sie einander sind, werden so genannte Ähnlichkeitsmaße eingesetzt. Prinzipiell handelt es sich dabei naturgemäß um geometrische, topologische oder attributive Ähnlichkeitsmaße. An dieser Stelle sollen einige der Maße Beachtung finden, wobei aufgrund der Fokussierung dieser Untersuchung auf Straßendaten hauptsächlich jene für linienhafte Geo-Objekte erwähnt werden, ohne dass dabei jedoch ein Anspruch auf Vollständigkeit erhoben wird. Abschließend wird darauf eingegangen, wie aus den partiellen Ähnlichkeitsmaßen ein Gesamtmaß abgeleitet werden kann, das es erlaubt, den Grad der Ähnlichkeit zweier Mehrfachrepräsentationen insgesamt auszudrücken.

6.4.1 Geometrische Ähnlichkeitsbestimmung

Geometrische Ähnlichkeitsmaße vergleichen die geometrischen Eigenschaften bzw. die Form, Charakteristik und Distanz von Datenobjekten. Einige Ähnlichkeitsmaße ergeben sich dadurch, dass Objektmerkmale pro Repräsentation separat berechnet und dann die erhaltenen Werte miteinander in Beziehung gesetzt bzw. klassifiziert werden. So ist es z.B. bei folgenden Ähnlichkeitsmaßen:

1. Längenvergleich,
2. Winkelvergleich (Parallelitätsgrad),
3. Vergleich der Anzahl der konstituierenden Punkte.

So wird z.B. beim Längenvergleich die Länge aller Instanzen einer Repräsentation errechnet und mit der Länge aller Instanzen einer anderen Repräsentation verglichen. Diese Werte müssen dann klassifiziert werden, um ein Maß für die Ähnlichkeit zu erhalten. Es gibt jedoch auch Ähnlichkeitsmaße, die als Distanzmaße zwischen den zu vergleichenden Geometrien zweier Repräsentationen errechnet werden. So z.B.

4. Durchschnittliche Liniendistanz,
5. Hausdorff-Distanz,
6. Fréchet-Distanz (vgl. [Devoegele 02]).

Eine Übersicht über Methoden zum Messen der Ähnlichkeit oder der Distanz von geometrischen Formen geben [Alt & Guibas 96]. Sie erwähnen, dass die Hausdorff-Distanz eines der am meisten verwendeten Maße ist. Die einseitige Hausdorff-Distanz zweier Geometrien A und B im d-dimensionalen Raum \mathfrak{R}^d berechnet sich zu

$$\delta_{eH}(A,B) = \max \min \|a-b\|,$$

wobei $\| \cdot \|$ die euklidische Distanz in \mathfrak{R}^d darstellt. Bei der Bestimmung der einseitigen Hausdorff-Distanz wird jedem Punkt einer Geometrie der am nächsten gelegene Punkt einer anderen Geometrie zugeordnet und die Distanz zwischen diesen Punkten bestimmt. Die Hausdorff-Distanz entspricht dann der maximalen aller gefundenen Distanzen. Die bidirektionale Hausdorff-Distanz macht dasselbe in beide Richtungen. Sie definiert sich wie folgt:

$$\delta_{bH}(A,B) = \max (\delta_{eH}(A,B), \delta_{eH}(B,A))$$

Da die Hausdorff-Distanz nicht den kompletten Linienverlauf betrachtet, sondern lediglich Punktmengen, führt sie in manchen Fällen zu fehlerhaften Zuordnungen [Alt & Guibas 96]. Bei der Zuordnung von Geodaten treten solche Fälle aber nur selten auf, deshalb betrachten [Davis & Aquino 03] das Maß für diese Anwendung als äußerst nützlich.

Es ist wichtig zu bemerken, dass manche Ähnlichkeitsmaße auf verschiedene Art und Weise bestimmt werden können. Natürlich ergeben sich entsprechend unterschiedliche Werte für die Ähnlichkeit, wenn Merkmale unterschiedlich errechnet wurden. Auch bei der Klassifikation von numerischen Werten wie z.B. der Längendifferenz können unterschiedliche Bewertungen zustande kommen. Ein Vorzug einer formalen Beschreibung von MRep-Relationen liegt darin begründet, dass man Berechnungsvorschriften einführen kann, die zu vergleichbaren Ergebnissen führen. Das Gesagte gilt selbstverständlich nicht nur für geometrische Ähnlichkeitsmaße, sondern für Ähnlichkeitsmaße im Allgemeinen.

6.4.2 Topologische Ähnlichkeitsbestimmung

Topologische Ähnlichkeitsmaße können einerseits Adjazenz-/Inzidenzbeziehungen von raumbezogenen Objekten vergleichen oder es werden graphbasierte Merkmale gegenübergestellt. Adjazenz- und Inzidenzangaben:

1. Anzahl der benachbarten Knoten von Anfangs- und Endpunkten einer Linie (Knotengrade);

2. Anzahl der benachbarten Knoten von Zwischenpunkten einer Linie;

Graphbasierte Merkmale:

3. Richtung einer Kante;
4. Erreichbarkeit eines Knoten innerhalb eines Graphen;
5. Exzentrizität eines Knoten innerhalb eines Graphen.

6.4.3 Ähnlichkeitsbestimmung von Attributen

Für die Bestimmung der Ähnlichkeit von alphanumerischen Attributen und Attributwerten werden weitere Ähnlichkeitsmaße eingesetzt. Im Datenbankbereich wird hier im Zuge der Identifizierung von Datenduplikaten häufig die so genannte „Edit Distance“ verwendet, die als die minimale Anzahl von Einfüge-, Lösch- und Ersetzungsoperationen, um einen String s in einen String s' zu transformieren, definiert werden kann [Weis & Naumann 04]. Für numerische Werte kann die absolute Differenz als Ähnlichkeitsmaß herangezogen werden. Wichtige Attribute, die beim Matching von Straßendaten untersucht werden, sind z.B.:

- Bezeichnung einer Straße; Straßename oder Kurzbezeichnung
- Bedeutung einer Straße; z.B. Hauptstraße oder Nebenstraße, etc.
- Anzahl der Fahrbahnen

Gerade bei Straßennamen müssen beim String-Vergleich im Deutschen Sprachraum auch Präfixe (z.B. „Straße des“) oder Umlaute beachtet werden (vgl. [Bofinger 01]).

Zum Vergleich der Eigenschaften von Objekten gehört es auch, die Ähnlichkeit der Metadaten zu bestimmen. Dabei handelt es sich vorwiegend um Angaben zur Datenqualität und -aktualität sowie um allgemeine Angaben zur Erfassung:

- Angaben zur Datenqualität und -aktualität:
 - Positionsgenauigkeit der Koordinaten
 - Verlässlichkeit der erfassenden Institution (Trust-Werte)
 - Erfassungszeitpunkt
 - Updaterate
- Angaben zur Datenerfassung:
 - Art des Koordinatensystems
 - Primär- oder Sekundärdatenerfassung
 - Erfassungsmethode (z.B. über GPS, über photogrammetrische Auswertung, etc.)
 - Erfassungsmaßstab
 - Objektbildungsregeln

Die Bestimmung der semantischen Ähnlichkeit wird traditionell auch auf dem Forschungsgebiet des „Information Retrieval“ untersucht [Baeza-Yates & Ribeiro-Neto 99]. So ist es beispielsweise für Suchmaschinen im Internet wesentlich, nur die Informationen an den Benutzer weiterzuleiten, an denen er auch tatsächlich interessiert ist, d.h. es muss hier ein Ranking auf der Basis semantischer Ähnlichkeitsmaße erfolgen

6.4.4 Gemischte Ähnlichkeitsmaße

Neben den bereits genannten Ähnlichkeitsmaßen existieren auch Kombinationen von Maßen aus verschiedenen Bereichen. Beispielsweise kann man nicht nur die Anzahl inzidierender Kanten eines Knoten (also ein topologisches Maß), sondern zusätzlich deren geometrische Eigenschaften wie Abgangswinkel oder auch deren nicht-geometrische Merkmale wie – im Falle von Straßen – Straßennamen miteinander vergleichen.

6.4.5 Bestimmung eines Gesamtmaßes für die Ähnlichkeit von Repräsentationen

Ein Gesamtmaß (GM) für die Ähnlichkeit von Repräsentationen kann aus den verschiedenen individuellen Ähnlichkeitsmaßen (IM) berechnet werden. Um ausdrücken zu können, ob ein IM gegenüber einem anderen IM einen größeren Einfluss bei der Berechnung des GM hat, kann man die individuellen Maße einzeln gewichten und so eine gewichtete Summe bilden:

$$GM = \sum_{i=0}^k IM_i * \text{Gewicht}_i$$

In [Walter 97] und [Walter & Fritsch 99] werden die Gewichtungsfaktoren über die statistische Betrachtung der gegenseitigen Information von Merkmalen wie Länge oder Winkel von Linienobjekten aus Trainingsgebieten abgeleitet, d.h. das Verfahren ist frei von explizit vergebenen Gewichten.

[Bilenko & Mooney 03] versuchen, ein Gesamtmaß für die Ähnlichkeit von Datenbank-Tupeln zu finden. Die Datenbank-Tupel besitzen dabei mehrere Felder bzw. Attribute, für die jeweils die Ähnlichkeit ihrer Attributwerte bestimmt werden kann. Jedes beliebige Tupel-Paar einer Tabelle kann daher über einen Vektor repräsentiert werden, wobei jede Komponente dieses Vektors die Ähnlichkeit zweier Attributwerte repräsentiert. Um aus den individuellen Ähnlichkeitsmaßen für die einzelnen Attribute ein Gesamtmaß für die Ähnlichkeit zweier Datenbank-Tupel abzuleiten, werden Lernverfahren aus dem Bereich der Künstlichen Intelligenz eingesetzt. Zunächst werden dabei Trainingsdaten generiert. Diese bestehen aus den Tupel-Vektoren, die, je nachdem, ob es sich um Mehrfachrepräsentationen handelt oder nicht, einen positiven oder einen negativen Wert bekommen. Auf der Basis dieser Trainingsdaten wird schließlich anhand eines binären Klassifikators, der die Ähnlichkeitsmaße der einzelnen Attribute in einer Zuordnungsfunktion kombiniert, ein Gesamtmaß für die Ähnlichkeit zweier Tupel erzeugt.

6.5 Modellierung von MRep-Relationen

Zur Repräsentation von Relationen zwischen mehrfach repräsentierten Objekten innerhalb des Augmented World Schemas dient die Klasse *MultiRepresentationalRelation*, die direkt von der Oberklasse aller Nexus-Relationen, der Klasse *NexusRelation*, abgeleitet wird. Durch Instanziierung dieser Klasse werden MRep-Relationen erzeugt. In der Folge wird die Struktur von *MultiRepresentationalRelation* vorgestellt. Abbildung 6-10 zeigt die UML-Repräsentation der Klasse.

Die *MultiRepresentationalRelation*-Klasse hat keine Unterklassen. Sie erbt die folgenden Attribute von ihren Superklassen. Alle optionalen Attribute sind durch die Signatur <<optional>> gekennzeichnet.

Von *NexusDataObject*:

- kind:*DataObjectType* – der Typ eines instanziierten Nexus-Objektes;
- nol:*NexusObjectLocator* – die global eindeutige ID;
- name:String – der Name des Objektes;

Von *NexusRelation*:

- <<optional>> extent:*Extent* – die geometrische Ausdehnung des Objektes;

Die *MultiRepresentationalRelation*-Klasse hat die folgenden eigenen Attribute:

Allgemeine Attribute:

- staticObjectsA:*SetOfStaticObjectNOL* – die global eindeutigen IDs jener Objekte, die eine Repräsentation A aufbauen;
- staticObjectsB:*SetOfStaticObjectNOL* – die global eindeutigen IDs jener Objekte, die die korrespondierende Repräsentation B aufbauen;
- cardinality:*Cardinality* – die Kardinalität der Relation;

- `matchType:MatchType` – die Art der Zuordnung der geometrischen Elemente (vgl. Kapitel 7.1.1.3);
- `totalSimilarityMeasure:TotalSimilarityMeasure` – das Gesamtmaß der Ähnlichkeit zweier Repräsentationen;
- `<<optional>> complexityA:SetOfComplexity` – die Komplexität der Instanzen von Repräsentation A;
- `<<optional>> complexityB:SetOfComplexity` – die Komplexität der Instanzen von Repräsentation B;
- `<<optional>> humanEstimation:HumanEstimation` – die Einschätzung des menschlichen Operateurs bezüglich der Sicherheit der Zuordnung;

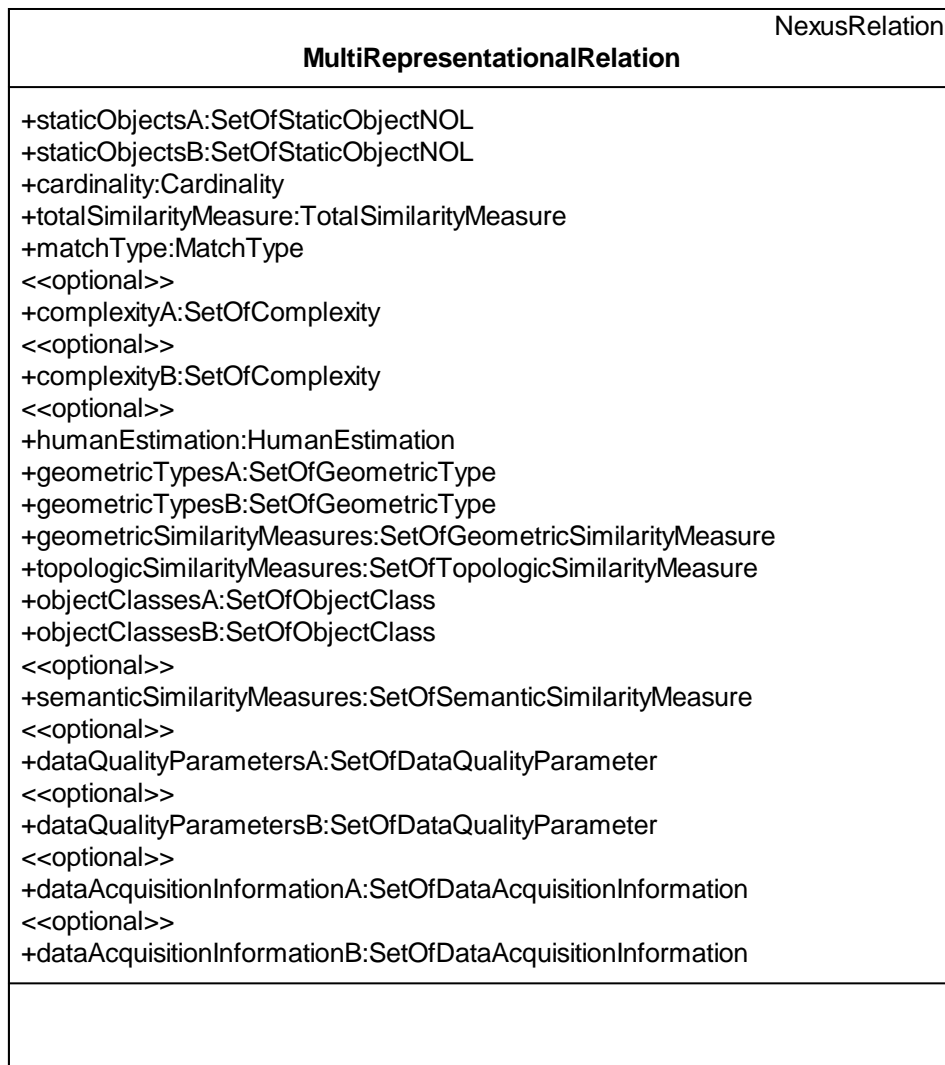


Abbildung 6-10: Darstellung der *MultiRepresentationalRelation*-Klasse in UML

Geometrische Attribute:

- `geometricTypesA:SetOfGeometricType` – die geometrischen Typen der Instanzen von Repräsentation A;
- `geometricTypesB:SetOfGeometricType` – die geometrischen Typen der Instanzen von Repräsentation B;
- `geometricSimilarityMeasures:SetOfGeometricSimilarityMeasure` – ein Set verschiedener geometrischer Ähnlichkeitsmaße;

Topologische Attribute:

- `topologicSimilarityMeasures:SetOfTopologicSimilarityMeasure` – ein Set verschiedener topologischer Ähnlichkeitsmaße;

Semantische und Meta-Attribute:

- `objectClassesA:SetOfObjectClass` – die Objektklassen der Instanzen von Repräsentation A;
- `objectClassesB:SetOfObjectClass` – die Objektklassen der Instanzen von Repräsentation B;
- `<<optional>> semanticSimilarityMeasures:SetOfSemanticSimilarityMeasure` – ein Set verschiedener semantischer Ähnlichkeitsmaße;
- `<<optional>> dataQualityParametersA:SetOfDataQualityParameter` – die Datenqualitätsparameter der Instanzen von Repräsentation A;
- `<<optional>> dataQualityParametersB:SetOfDataQualityParameter` – die Datenqualitätsparameter der Instanzen von Repräsentation B;
- `<<optional>> dataAcquisitionInformationA:SetOfDataAcquisitionInformation` – Angaben zur Datenerfassung der Instanzen von Repräsentation A;
- `<<optional>> dataAcquisitionInformationB:SetOfDataAcquisitionInformation` – Angaben zur Datenerfassung der Instanzen von Repräsentation B.

6.6 Typifizierung des Identitätsgrades von Mehrfachrepräsentationen

Mehrfachrepräsentationen entstehen dann, wenn sich zwei oder mehr Computerrepräsentationen auf das gleiche Realweltobjekt beziehen, d.h. es muss eine Überdeckung der Identität der digitalen Objekte existieren. An dieser Stelle soll kurz diskutiert werden, inwiefern eine Typifizierung des Identitätsgrades mehrfach repräsentierter Objekte möglich ist.

In diesem Zusammenhang stellt sich die Frage, ob sich MRep-Relationen in Kategorien einteilen lassen, wie das beispielsweise innerhalb des 9-Intersection-Modells von [Egenhofer 91] für topologische Relationen geschehen ist. Im 9-Intersection-Modell wurde dabei die Ausprägung der räumlichen Überlappung von Geobjekten als Kriterium für die Bildung von Klassen innerhalb eines formalen Modells gewählt.

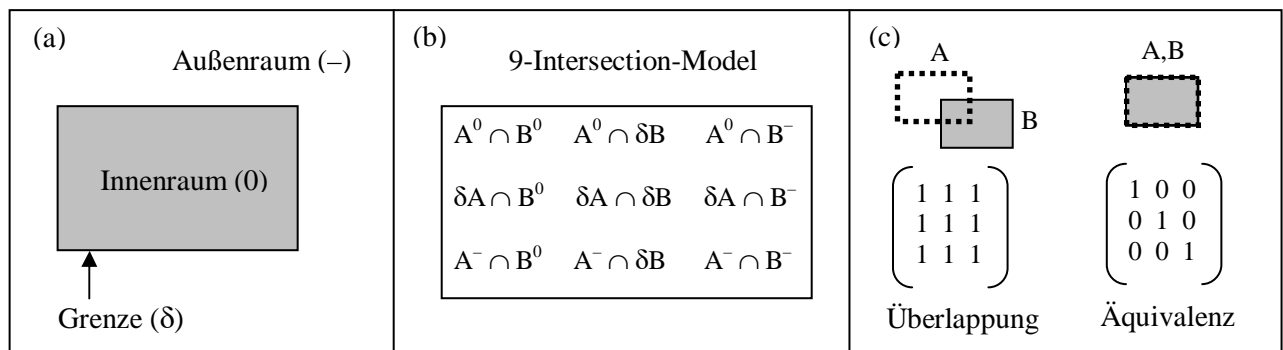


Abbildung 6-11: Das 9-Intersection-Modell von [Egenhofer 91] erlaubt es, die topologischen Beziehungen zwischen zwei Flächen in Form einer Matrix zu beschreiben.

Jedes flächenhafte Objekt wird im 9-Intersection-Modell als geschlossene, homogene zweidimensionale Teilmenge des \mathbb{R}^2 aufgefasst, das aus einem Außenraum (-), einer Grenze (δ) und einem Innenraum (0) besteht (siehe Abbildung 6-11 (a)). Alle topologischen Relationen zwischen zwei Objekten A und B lassen sich klassifizieren, indem man alle Schnittmengen zwischen dem Innenraum (0), der Grenze (δ) und dem Außenraum (-) von A und B betrachtet. Daraus ergibt sich eine 3 x 3 Matrix: die 9-Intersection-Matrix (siehe Abbildung 6-11 (b)). Jeder Eintrag in der Matrix kann entweder leer (0) oder nicht-leer (1) sein: es kann daher zwischen $2^9 = 512$ verschiedenen topologischen Relationen unterschieden werden. Für zwei einfache Flächen sind davon jedoch nur 8 realisierbar. In Abbildung 6-11 (c) sind beispielhaft die Matrizen zur Abbil-

derung der Überlappung und der Identität dargestellt. Eine Erweiterung des 9-Intersection-Modells stellt das Dimensionally Extended 9-Intersection-Modell dar [Clementini & Di Felice 94]. Es erlaubt eine weitere Differenzierung der Ergebnismenge bei paarweisen Verschneidungen von Objekten in die jeweiligen geometrischen Objektdimensionen. Bei der Verschneidung von Objektgrenzen können beispielsweise Linien oder Punkte oder beide Geometrietypen entstehen.

Will man die Beziehungen zwischen Mehrfachrepräsentationen typifizieren, so darf nicht nur die räumliche Überlappung zweier Repräsentationen betrachtet werden, sondern es ist zu untersuchen, wie bzw. in welchem Maße die Identität zweier Repräsentationen überlappt. Doch wie kann dies festgestellt werden? Geodaten verfügen nicht über identitätsbildende Merkmale wie z.B. die DNA beim Menschen, die sozusagen jede Instanz eindeutig definiert. Die Identität von Geodaten lässt sich nur anhand des Vergleiches der Eigenschaften verschiedener Repräsentation ausmachen bzw. sie kann nur näherungsweise über die entsprechenden Ähnlichkeitsmaße (vgl. Kapitel 6.4) bestimmt werden. Prinzipiell gibt es – analog zum 9-Intersection-Modell – dreierlei Zustände, wie Repräsentationen in Beziehung stehen können:

1. Repräsentationen sind komplett identisch (Kongruenz),
2. Repräsentationen sind partiell identisch (Partielle Identität),
3. Repräsentationen sind disjunkt bzw. es existieren keine Mehrfachrepräsentationen (Disjunktheit).

Besteht eine Überlappung der Identität von Repräsentationen, so kann deren MRep-Relation anhand der verfügbaren Ähnlichkeits- bzw. Qualitätsmaße weiter differenziert werden. So kann man z.B. den Grad der geometrischen Übereinstimmung ausdrücken und auf die Relationen des 9-Intersection-Modells abbilden. Per Definition kann etwa festgelegt werden, dass eine geometrische Identität dann gegeben ist, wenn das geometrische Ähnlichkeitsmaß 95% oder mehr beträgt. Ähnliche Schwellwerte könnten für die Überlappung, die Berührung etc. vergeben werden. Analog kann dies auch für die topologischen und semantischen Ähnlichkeitsmaße oder auch insgesamt für das Gesamtmaß der Ähnlichkeit durchgeführt werden, so dass man einen natürlichsprachlichen Ausdruck für die Beziehungen zwischen Mehrfachrepräsentationen erhält. Beispielsweise könnte eine schwache geometrische Ähnlichkeit zweier Repräsentationen über die Relation „Berührung“, eine deutliche semantische Ähnlichkeit als „Überlappung“ und eine totale topologische Ähnlichkeit als „Äquivalenz“ kategorisiert werden. Die Gesamrelation könnte dann entweder als „Berührung“ oder als „Überlappung“ repräsentiert werden, je nachdem, wie stark die jeweiligen Ähnlichkeitsmaße untereinander gewichtet werden. Entsprechend des Dimensionally Extended 9-Intersections-Modells kann auch eine weitere Differenzierung der Beziehungen von Mehrfachrepräsentationen erfolgen. Generell beruht eine solche Typifizierung jedoch im Gegensatz zu den formalisierten Ansätzen im Rahmen des 9-Intersection-Modells auf Schwellwerten. Sie kann als Forschungsgegenstand betrachtet werden. An dieser Stelle erscheint eine weitere Vertiefung allerdings nicht sinnvoll, da die vorliegende Arbeit einen anderen Fokus hat.

6.7 Umsetzung des MRep-Relationskonzeptes in Nexus

Durch die Modellierung von Relationen zwischen Mehrfachrepräsentationen ist die Grundlage dafür geschaffen, mehrfach repräsentierte Objekte im Sinne einer Multirepräsentations-Datenbank (MRDB) gemeinsam zu verwalten. Die Realisierung von MRep-Relationen ist dabei nicht an Nexus gebunden. MRep-Relationen können genauso zwischen den Instanzen verschiedener Dateien oder zwischen den Tupeln innerhalb eines raumbezogenen Datenbanksystems erzeugt werden. In Nexus sollen MRep-Relationen dazu verwendet werden, korrespondierende Objekte innerhalb einer verteilten, föderierten Dateninfrastruktur zu verwalten und gemeinsam zu verarbeiten.

In Kapitel 3 wurde das Konzept von Nexus und der Datenverwaltung innerhalb der Plattform beschrieben. Es wurde erwähnt, dass alle Daten des Nexus-Systems auf so genannten Context-Servern zur Verfügung gestellt werden. Wenn ein Datenanbieter seine Informationen innerhalb von Nexus bereitstellen möchte, muss er sich bei einem Verzeichnis, dem Area Service Register, registrieren. Der gegenüber der Darstellung in Kapitel 3.4 modifizierte Ansatz zur Verwaltung von Mehrfachrepräsentationen in Nexus sieht vor, dass immer dann, wenn ein neues lokales Umgebungsmodell beim Nexus-System angemeldet wird, ein spezieller Dienst des Area Service Registers nach anderen Augmented Areas sucht, die mit dem neu eingebrachten Datenbestand räumlich überlappen und die gleichen Objektklassen des Augmented World Schemas enthalten. Die gefundenen lokalen Umgebungsmodelle können dann auf Mehrfachrepräsentationen mit der neu zur

Verfügung gestellten Augmented Area hin untersucht werden. Dies kann auf automatisierte Weise oder in einem manuellen Schritt durch den Datenanbieter erfolgen. Relationen zwischen Mehrfachrepräsentationen können dann analog zu anderen Relationen von den Context-Servern verwaltet werden, so dass z.B. beim Löschen von Objekten auch alle Relationen auf das Objekt gelöscht werden. Durch diesen Ansatz kann auch ein gemeinsames Update von Mehrfachrepräsentationen erfolgen. Allerdings ist es hierzu notwendig, die Möglichkeit der Durchführung einer Aktualisierung genau zu prüfen. Findet beispielsweise eine geometrische Fortführung einer Repräsentation statt, deren Geometrie eine hohe Ähnlichkeit zu jener der korrespondierenden Repräsentation aufweist, kann diese Änderung in der Regel auch an die zugehörige Repräsentation weitergeleitet werden. Bei einer geringeren Eindeutigkeit der Zuordnung zweier Repräsentationen muss entschieden werden, ob ein Update überhaupt legitim ist und nicht vielmehr die MRep-Relation geändert oder sogar neu gebildet werden muss. Mit dem Konzept der MRep-Relationen steht aber zumindest ein Mechanismus zur Verfügung, um das Problem der Aktualisierung von Mehrfachrepräsentationen und damit der Konsistenz des Datenbestandes in Nexus angehen zu können. Allerdings ist zu beachten, dass dieses Konzept voraussetzt, dass Schreibrechte auf die Context-Server der Plattform eingeräumt werden müssen, was wiederum den Aspekt der Datensicherheit betrifft. Das MRep-Relationskonzept benötigt folglich geeignete Sicherheitskonzepte, die eine mutwillige Manipulation der Daten im Zuge der Aktualisierung verhindern können.

Sollen nun räumlich überlappende Mehrfachrepräsentationen gemeinsam analysiert werden, so kann mittels des MRep-Relationskonzeptes der aufwendige Zuordnungsprozess auf Föderationsebene umgangen werden, indem einfach die MRep-Relationen aller beteiligten Mehrfachrepräsentationen ebenfalls von der Föderation geladen werden und diese dann entweder einer Verschmelzung durchführt oder aber eine Analyse unter Auswertung der verfügbaren MRep-Relationen. Ein Ansatz dieser Arbeit zielt darauf ab, nachzuweisen, dass eine Analyse auf der Basis der Betrachtung von MRep-Relationen tatsächlich möglich und folglich eine Datenverschmelzung nicht immer notwendig ist (vgl. Kapitel 8.2). Dies bedeutet eine weitere Optimierung hinsichtlich der Effizienz der Plattform.

Prinzipiell können Augmented Areas auch weiterhin in Nexus eingebracht werden, ohne dass korrespondierende Repräsentationen gesucht werden. In diesem Falle ist es, wie in Kapitel 3.4 erwähnt, notwendig, den Prozess der Erzeugung von MRep-Relationen auf der Ebene der Föderation anzusiedeln. Dies bedeutet allerdings eine Einschränkung des Antwortverhaltens der Plattform, sofern umfangreiche Zuordnungsprozesse für die Erzeugung eines konsistenten Datenbestandes, auf dem die entsprechende Abfrage ausgeführt werden kann, notwendig sind. Da die Föderation in Nexus auch kein „Gedächtnis“ hat, gingen die MRep-Relationen nach der Beantwortung einer Anfrage prinzipiell wieder verloren oder könnten nur temporär im Cache aufbewahrt werden, falls keine geeigneten Maßnahmen zur Speicherung von MRep-Relationen auf Context-Servern oder in einem separaten Dienst getroffen würden. Insgesamt ist diese Vorgehensweise daher weniger praktikabel. Allerdings würde man sich auf diese Weise den Overhead der Verwaltung von MRep-Relationen sparen. Der hier zu betreibende Aufwand ist in einer so umfangreichen, global ausgerichteten Infrastruktur wie Nexus nicht zu unterschätzen. Dadurch bedingte Einschränkungen müssen in Zukunft evaluiert werden.

7 Erzeugung von MRep-Relationen

In diesem Kapitel wird die Realisierung des Konzeptes von MRep-Relationen beschrieben. Zunächst wird die Software vorgestellt, mittels derer MRep-Relationen generiert und visualisiert werden können. Anschließend wird der Prozess der Generierung von MRep-Relationen anhand von zwei Testgebieten aus dem Bereich der Stuttgarter Innenstadt, für die Straßendaten aus ATKIS und GDF mehrfach repräsentiert vorlagen, beschrieben. Schließlich wird auf Probleme und Ergebnisse der Vorgehensweise eingegangen.

7.1 Die Software-Umgebung zur Erzeugung und Visualisierung von MRep-Relationen

Um Mehrfachrepräsentationen über MRep-Relationen miteinander verknüpfen zu können, wurde im Rahmen dieser Arbeit ein semiautomatisches Software-Tool entwickelt. Es basiert auf dem in Java implementierten Open Source GIS JUMP (Java Unified Mapping Platform) und seinen Basismodulen Java Topology Suite (JTS) und Java Conflation Suite (JCS) [Ramsey 04]. Die Java Topology Suite stellt die Basisalgorithmen zur Verarbeitung von Geodaten wie z.B. Verschneidung, Vereinigung, Pufferung, etc. bereit, während JUMP die graphische Benutzeroberfläche (GUI) und die Standardfunktionalitäten von GIS wie Zoomen, Anzeige von Objektinformation, Distanzmessung, Dateneditierung, usw. liefert. Mit der Java Conflation Suite bietet die JUMP-Umgebung eine Möglichkeit, Mehrfachrepräsentationen zuzuordnen und zusammenzuführen zu können. Auf der Basis der JCS wurde im Rahmen des JUMP-Projektes auch ein spezielles Modul für die Zuordnung und Verschmelzung von Straßendaten, die so genannte RoadMatcher Applikation, zur Verfügung gestellt. Bei der Implementierung orientiert sich das JUMP-Framework an den Vorgaben des OpenGIS Konsortiums. JUMP wird kontinuierlich weiter entwickelt. Für einen kompletten Überblick über das JUMP-Projekt sei auf [JUMP 04] verwiesen.

Die Funktionalität von JUMP kann beliebig durch eigene, in Java implementierte Module, so genannte Plug-Ins, erweitert werden. Im Rahmen der vorliegenden Arbeit sind die folgenden Plug-Ins entstanden (siehe Tabelle 7-1):

Nr.	Modul	Beschreibung
1	<i>RelationBuilderToolboxPlugIn</i>	Ein Tool zur Erzeugung von MRep-Relationen
2	<i>RelationViewerToolboxPlugIn</i>	Ein Tool zur Visualisierung von MRep-Relationen
3	<i>RelationAnalyzerToolboxPlugIn</i>	Ein Tool zur Auswertung von MRep-Relationen zum Zwecke der Ableitung von Korrespondenzen auf Schema-Ebene
4	<i>MRepNetworkAnalysisPlugIn</i>	Ein Tool zur Netzwerkanalyse auf mehrfach repräsentierten Daten unter Nutzung von MRep-Relationen

Tabelle 7-1: JUMP-PlugIns, die im Rahmen dieser Arbeit entstanden

An dieser Stelle sollen die beiden ersten Module vorgestellt werden. Die Module 3 und 4 werden später bei der Auswertung von MRep-Relationen in Kapitel 8 erläutert.

7.1.1 Die RelationBuilder-Toolbox

Mittels des RelationBuilderToolboxPlugIns ist es möglich, MRep-Relationen zwischen linienhaften Straßendaten zu erzeugen. An dieser Stelle werden die Vorgänge erläutert, die bei der Erzeugung einer MRep-Relation ablaufen. Außerdem wird darauf eingegangen, wie MRep-Relationen gespeichert werden und welche Richtlinien bei der Zuordnung von Objekten zu beachten sind.

7.1.1.1 Erzeugung von MRep-Relationen

Im Vorfeld der Erzeugung von Relationen müssen die zu bearbeitenden Datenbestände über eine Rubber-sheeting-Transformation angepasst werden, um die globalen geometrischen Unterschiede zu minimieren (vgl. [Walter 97]). Anschließend können die Datensätze in die RelationBuilder-Toolbox geladen werden. Die zugehörigen Schematypen sind aus einer Liste zu selektieren, wobei bislang nur Daten aus GDF und ATKIS zugelassen werden. In zwei separaten Feldern werden die Attribute der geladenen Datensätze angezeigt. Diejenigen Attribute, die für die Bestimmung der Schemarelationen herangezogen werden sollen, müssen hier spezifiziert werden. Sind die beiden Datenlayer, deren Schematypen und die gewünschten Attribute spezifiziert, kann mit der Erzeugung von MRep-Relationen begonnen werden (vgl. Abbildung 7-1).

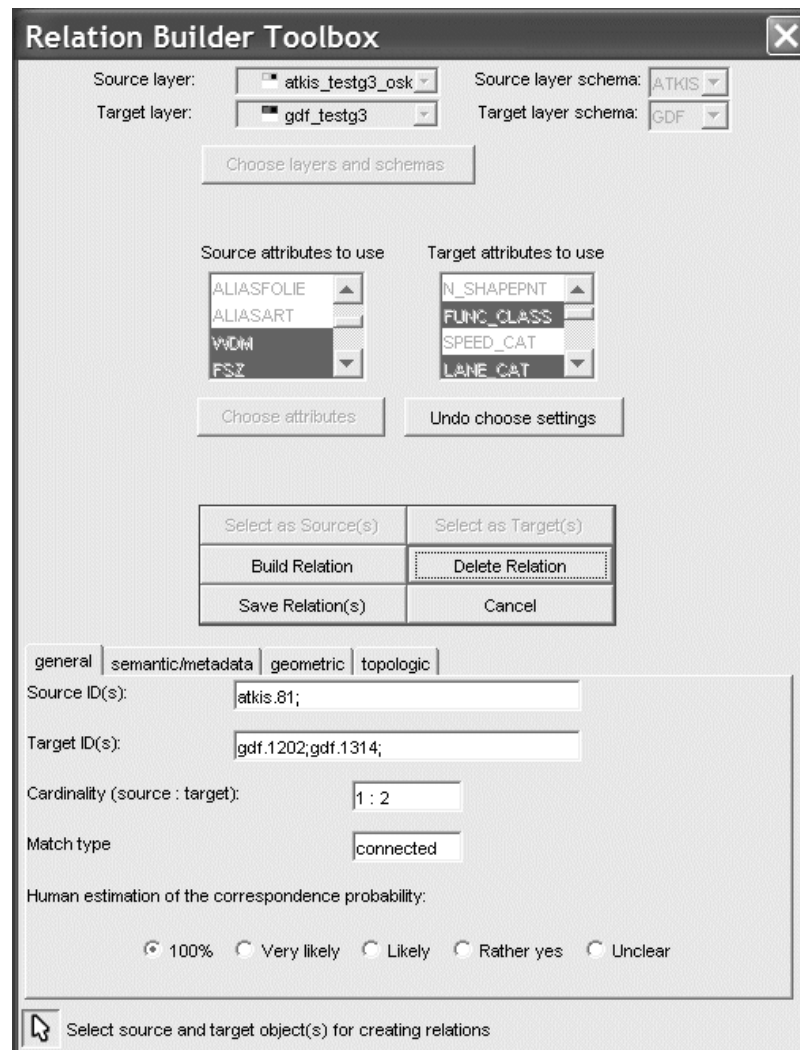


Abbildung 7-1: Mittels der RelationBuilder-Toolbox können MRep-Relationen erzeugt und in einer MRRL-Datei gespeichert werden.

Mit Hilfe des Cursors werden die Instanzen, die zu einer Quell- und einer Ziel-Repräsentation gehören, in der Kartendarstellung selektiert. Die Identifikatoren der zu einer Repräsentation gehörenden Instanzen werden daraufhin ebenso wie die Kardinalität der Relation sowie der Zuordnungstyp (siehe unter 7.1.1.3) angezeigt. Nach der Auswahl der Repräsentationen ist vom Operateur eine individuelle Einschätzung hinsichtlich der Übereinstimmung der zuzuordnenden Repräsentationen vorzunehmen. Durch Drücken des „Build Relation“ Knopfes wird dann eine MRep-Relation erzeugt. Während dieses Prozesses wird zunächst aus den potentiellen Teilgeometrien bzw. aus den einzelnen Liniensegmenten der zu einer Repräsentation gehörenden Instanzen eine einzige resultierende Geometrie pro Repräsentation in Form einer Gesamtlinie abgeleitet. Daraufhin werden die folgenden geometrischen und topologischen Ähnlichkeitsmaße automatisch berechnet:

- **Längendifferenz:** Die Differenz der Gesamtlängen der beiden Repräsentations-Geometrien.
- **Winkeldifferenz:** Über die Anfangs- und Endknoten einer Repräsentation wird eine Kante gebildet, für die der Winkel in Grad gemessen wird. Die Winkeldifferenz ergibt sich folglich aus der Differenz der Winkel beider Kanten.
- **Durchschnittliche Liniendistanz:** Die durchschnittliche Liniendistanz wird aus den Teildistanzen aller Punkte einer Repräsentation zu den Punkten einer anderen Repräsentation, die die gleiche Distanz vom jeweiligen Startpunkt aufweisen, errechnet.
- **Vertex-Hausdorff-Distanz:** Die Vertex-Hausdorff-Distanz ist eine vereinfachte Version der Hausdorff-Distanz (vgl. Kapitel 6.4.1). Die Berechnung der normalen Hausdorff-Distanz ist für Linien mit mehreren Segmenten sehr komplex und aufwendig. Mit der Vertex-Hausdorff-Distanz erreicht man eine gute Approximation der Hausdorff-Distanz für die meisten Fälle, die bei der Zuordnung von Geodaten auftreten [Davis & Aquino 03].
- **Differenzen der Adjazenzen der Start- und Endknoten:** Es werden die Abweichungen in der Anzahl der Adjazenzen der Start- und Endknoten einer Repräsentation berechnet.
- **Anzahl der topologischen Unterbrechungen einer Repräsentation:** Falls im Verlauf einer zugeordneten Repräsentation topologische Unterbrechungen auftreten, bei der korrespondierenden jedoch nicht, werden diese aufsummiert und verringern die topologische Ähnlichkeit entsprechend der Anzahl der Unterbrechungen (vgl. Abbildung 7-4.B).

Weitere Ähnlichkeitsmaße, wie sie für MRep-Relationen vorgesehen sind – z.B. zum Attributvergleich – wurden nicht berücksichtigt, da hierfür keine adäquaten Daten zu Verfügung standen, d.h. die Ähnlichkeitsmaße, die im Rahmen dieser Arbeit umgesetzt wurden, basieren auf der Auswertung topologischer und geometrischer Eigenschaften von Mehrfachrepräsentationen.

Aus einer Begutachtung der Testdaten heraus wurden die jeweiligen Absolutwerte der erfassten individuellen Ähnlichkeitsmaße (IM) in verschiedene Klassen unterteilt, so dass pro IM ein so genannter Evaluationswert (EVW) festgelegt werden kann. Er liegt pro IM zwischen einem Maximalwert von 6 (beste Kategorie) und einen Minimalwert von 0 (schlechteste Kategorie). Die Evaluationswerte der einzelnen Ähnlichkeitsmaße werden je nach Relevanz für die Gesamtbewertung der Ähnlichkeit zweier Repräsentationen mit einem Faktor gewichtet. Da die Hausdorff-Distanz beim Matching von Geodaten als sehr nützlich betrachtet wird [Davis & Aquino 03], erhielt diese den höchsten Faktor. Auch die topologischen Beziehungen wurden mit dem Maximalwert versehen, wobei diese insgesamt als ein Evaluationswert repräsentiert werden. Längen- und Winkelunterschiede fließen mit einem mittleren Gewichtungsfaktor ein, während die durchschnittliche Liniendistanz am geringsten gewertet wird. So lässt sich schließlich über eine gewichtete Summe der Evaluationswerte das Gesamtmaß (GM) der Ähnlichkeit zweier Repräsentationen berechnen. Es gilt:

$$\begin{aligned} \text{GM} = & 3 \times \text{EVW_Länge} + 3 \times \text{EVW_Winkel} + \\ & 2 \times \text{EVW_Durchschnittsdistanz} + 4 \times \text{EVW_Hausdorffdistanz} + \\ & 4 \times \text{EVW_Adjazenzunterschiede} \end{aligned}$$

Das GM kann somit einen Maximalwert von 96 und einen Minimalwert von 0 erhalten. Um einen besseren Vergleich zu ermöglichen, wird es auf hundert normiert, d.h. das Gesamtmaß der Ähnlichkeit kann einen Wert zwischen 0 und 100 annehmen.

Nachdem eine MRep-Relation gebildet wurde, werden die bearbeiteten Repräsentationen in der Kartendarstellung kenntlich gemacht (siehe Abbildung 7-2) und in der Datenstruktur als bearbeitet markiert, so dass die Zuordnung einer Instanz zu mehreren Repräsentationen ausgeschlossen werden kann.

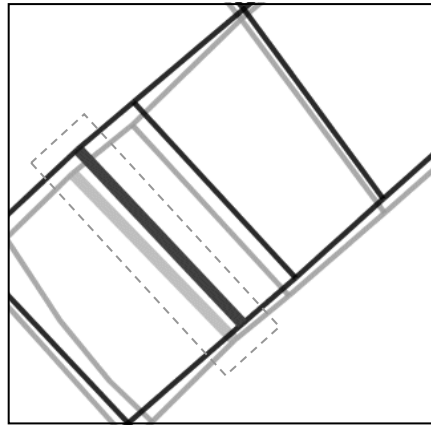


Abbildung 7-2: Bearbeitete Objekte werden markiert, indem sie verbreitert werden.

Eine MRep-Relation kann auch wieder gelöscht werden, wenn der Bearbeiter seine Meinung ändert und eine andere Zuordnung durchführen möchte. Sind alle gewünschten MRep-Relationen vorhanden, so können diese in einem XML-Format persistent gespeichert werden. Hierzu wurde eigens die so genannte Multi-Representational Relation Language (MRRL, siehe folgender Abschnitt) spezifiziert.

7.1.1.2 Speicherung von MRep-Relationen

Die Speicherung von MRep-Relationen geschieht mittels eines XML-Formats [W3C 04], das im Rahmen der vorliegenden Arbeit definiert wurde. Es wird als MultiRepresentational Relation Language (MRRL) bezeichnet. Die Speicherung von MRep-Relationen wird deshalb in XML durchgeführt, weil auch innerhalb der Nexus-Plattform sämtliche Daten über XML-Formate ausgetauscht werden [Bauer et al. 04b]. Die Implementierung von MRRL erfolgte mittels der JDOM-Bibliothek [JDOM 04] zur Verarbeitung von XML unter Java.

MRRL besteht stets aus drei verschiedenen Dateisektionen. In einem Headerabschnitt werden allgemeine Informationen über die beiden einander zugeordneten Datenlayer, d.h. deren Namen und deren Speicherort, deren Schematyp sowie die jeweilige Anzahl der enthaltenen Objekte und die räumliche Ausdehnung in Form eines minimal umschließenden Rechtecks (MUR) in GML-Koordinaten, festgehalten. Außerdem ist hier der ungefähre Überlappungsgrad der Datensätze in Prozent eingetragen. Er wird aus einer Verschneidung der beiden MURs berechnet.

Die zweite Sektion enthält die gespeicherten MRep-Relationen. Wie aus dem Ausschnitt einer MRRL-Instanzdatei in Abbildung 7-3 zu erkennen ist, verfügt eine MRep-Relation über eine eindeutige ID sowie verschiedene Attribute. Die allgemeinen Attribute (*general_atts*) enthalten die IDs (bzw. im Nexus-Kontext die NOLs) der zu einer Repräsentation gehörenden Instanzen (*source_ids*, *target_ids*), die Kardinalität der Relation (*cardinality*), den Zuordnungstyp im Element *match_type* (siehe nächster Abschnitt), die Einschätzung des Operators bezgl. der Wahrscheinlichkeit der Zuordnung (*human_estimation*) und das Gesamtmaß der Ähnlichkeit (*total_similarity*). Es folgen daraufhin die semantischen Attribute (*semantic_atts*) mit den Angaben zur Klassenzugehörigkeit zugeordneter Instanzen (*source_classes*, *target_classes*) sowie diejenigen Attribute der Instanzen, die für die Speicherung innerhalb der MRep-Relationen ausgewählt wurden. Das *geometric_atts* Element gibt an, welchen geometrischen Typen die zugeordneten Instanzen angehören und listet außerdem die erfassten geometrischen Ähnlichkeitsmaße (z.B. *length_difference*) auf. Analog hierzu stehen in den topologischen Attributen einer MRep-Relation die topologischen Ähnlichkeitsmaße (z.B. *startnode_deg_diff*).

In der dritten Sektion von MRRL werden schließlich alle Instanzen beider Datensätze aufgeführt, die nicht zugeordnet werden konnten und somit nicht an einer MRep-Relation beteiligt sind.

```

<relations>
...
  <mreprelation>
    <mreprelation_id>http://nexus.uni-stuttgart.de/nol/mreprelations/181</mreprelation_id>
    <attributes>
      <general_atts>
        <source_ids>
          <id>atkis.1003</id>
        </source_ids>
        <target_ids>
          <id>gdf.1670</id>
          <id>gdf.1664</id>
        </target_ids>
        <cardinality>1 : 2</cardinality>
        <match_type>separate</match_type>
        <human_estimation>100%</human_estimation>
        <total_similarity>91.66666666666667</total_similarity>
      </general_atts>
      <semantic_atts>
        <source_classes>
          <class>
            <class_name>Strasse</class_name>
            <class_attributes>
              <WDM>1305</WDM>
              <FSZ>9998</FSZ>
              <FKT_STR>2301</FKT_STR>
              <FKT_WEG>none</FKT_WEG>
            </class_attributes>
          </class>
        </source_classes>
        <target_classes>
          <class>
            <class_name>ROAD</class_name>
            <class_attributes>
              <func_class>3</func_class>
              <speed_cat>6</speed_cat>
              <lane_cat>1</lane_cat>
            </class_attributes>
          </class>
          <class>
            <class_name>ROAD</class_name>
            <class_attributes>
              <func_class>3</func_class>
              <speed_cat>6</speed_cat>
              <lane_cat>1</lane_cat>
            </class_attributes>
          </class>
        </target_classes>
      </semantic_atts>
      <geometric_atts>
        <source_geotypes>
          <geo_type>LineString</geo_type>
        </source_geotypes>
        <target_geotypes>
          <geo_type>LineString</geo_type>
          <geo_type>LineString</geo_type>
        </target_geotypes>
        <length_difference>2.017465451252221</length_difference>
        <angle_difference>1.0172069034131148</angle_difference>
        <vertex_hausdorff_distance>5.331939466003357</vertex_hausdorff_distance>
        <avg_line_distance>4.028095256363919</avg_line_distance>
      </geometric_atts>
      <topologic_atts>
        <startnode_deg_diff>1</startnode_deg_diff>
        <endnode_deg_diff>1</endnode_deg_diff>
        <topologic_interrupts>0</topologic_interrupts>
      </topologic_atts>
    </attributes>
  </mreprelation>
...
</relations>

```

Abbildung 7-3: Aufbau einer MRep-Relation gemäß der Format-Spezifikation von MRRL

7.1.1.3 Richtlinien bei der Zuordnung

Prinzipiell wäre es wünschenswert, MRep-Relationen vollautomatisch zu generieren. Ein automatisiertes Matching ist jedoch nicht das Ziel dieser Arbeit. Es soll lediglich untersucht werden, für welche Anwendungen MRep-Relationen genutzt werden können. Daher wurde ein semiautomatisches Verfahren zur Erzeugung von MRep-Relationen entwickelt. Durch die Einbeziehung eines menschlichen Operateurs wird allerdings eine subjektive Komponente bei der Zuordnung von Mehrfachrepräsentationen eingeführt. Zwei Operateure mögen eine räumliche Situation unterschiedlich interpretieren, so dass sie auch unterschiedliche

Zuordnungsergebnisse bzw. unterschiedliche Ergebnisse bei der Generierung von MRep-Relationen liefern. Um eine möglichst einheitliche Vorgehensweise zu gewährleisten, muss den Operateuren daher ein entsprechender Verfahrenskatalog für die Erfassung von MRep-Relationen an die Hand gegeben werden. Dieser enthält Richtlinien dafür, wie MRep-Relationen in bestimmten Situationen zu bilden sind.

Eine Vorgabe besteht beispielsweise darin, generell immer dann neue MRep-Relationen zu bilden, wenn eine 1:1-Zuordnung möglich ist, auch wenn mehrere Repräsentationen eindeutig zusammengehören (siehe Abbildung 7-4.A). Sind keine klaren Entsprechungen zu erkennen, müssen korrespondierende Repräsentationen solange um weitere Instanzen ergänzt werden, bis eine eindeutige Abbildung möglich ist, auch wenn *eine* der beiden Repräsentationen topologisch unterbrochen wird (siehe Abbildung 7-4.B). Treten jedoch topologische Wechsel beider Repräsentationen an ähnlicher Stelle auf, ist deren Bedeutung zu priorisieren, d.h. die Repräsentationen sind an den Stellen topologischer Wechsel „aufzubrechen“ (siehe Abbildung 7-4.C).

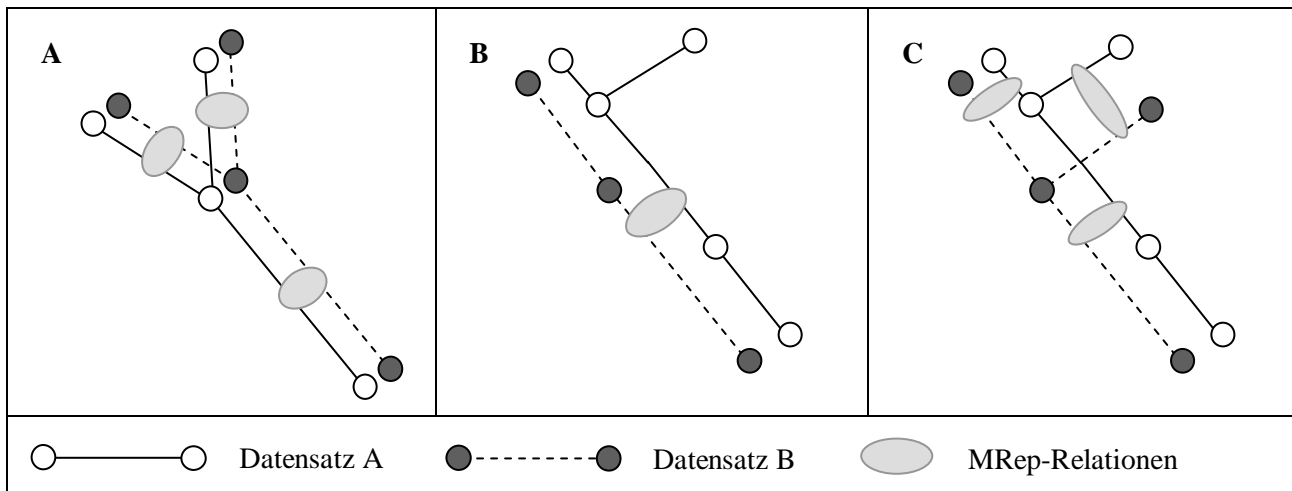


Abbildung 7-4: Vorgaben zur Bildung von MRep-Relationen

Bei der Erzeugung von MRep-Relationen haben die Operateure die Wahl zwischen exakt drei unterschiedlichen Zuordnungsvarianten bzw. Zuordnungstypen (siehe Abbildung 7-5). Diese decken nahezu alle korrekten Zuordnungsmöglichkeiten, die in den Testdaten auftreten, ab.

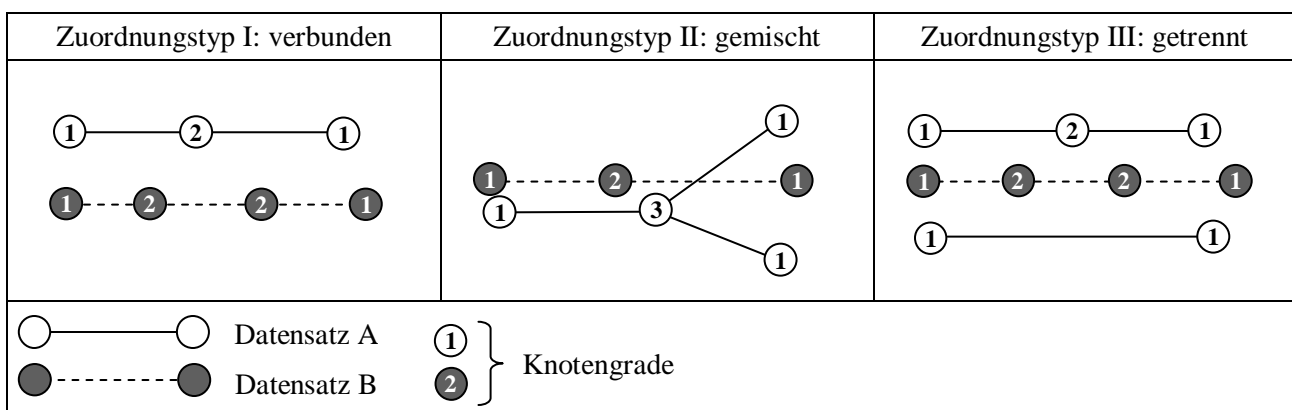


Abbildung 7-5: Zugelassene Zuordnungstypen im RelationBuilder. Zwischen den korrespondierenden Repräsentationen besteht jeweils *eine* MRep-Relation.

Potentiell fehlerhafte Zuordnungen, so z.B. eine Zuordnung von einer zusammenhängenden Repräsentation des einen Datensatzes zu einer Repräsentation eines anderen Datensatzes, die mehr als zwei topologisch getrennte Liniensegmente aufweist, werden bei der Erzeugung von MRep-Relationen nicht zugelassen. So unterstützt die RelationBuilder-Software den Bearbeiter also durch eigene Konsistenzprüfungen und bewahrt

ihn vor Fehlern. Zu dieser Art von Hilfestellung gehört auch die bereits erwähnte Funktionalität der Software, die Teilnahme einer Instanz an mehreren MRep-Relationen zu verhindern.

Aus Abbildung 7-4 und Abbildung 7-5 wird auch ersichtlich, dass generell Zuordnungen der Kardinalität $n:m$ zugelassen sind. Im Normalfall des Zuordnungstyps I wird eine MRep-Relation zwischen n zusammenhängenden Liniensegmenten des Datensatzes A und m zusammenhängenden Liniensegmenten des Datensatzes B erzeugt, wobei n entweder $= 1$ oder > 1 sein kann. In den untersuchten Datensätzen treten jedoch auch Fälle auf, in denen sich eine Repräsentation verzweigt und wo dennoch eine Zuordnung möglich sein muss. Diese Fälle werden als Zuordnungstyp II ausgewiesen. Sie werden vom Normalfall anhand des Knotengrades der Anfangs- und Endpunkte der beteiligten Kanten unterschieden. So existieren beim Zuordnungstyp I pro Repräsentation zwei Außenknoten, die nur mit einem Nachbarknoten adjazieren. Dahingegen tritt beim Zuordnungstyp II ein Zentrumsknoten auf, der mehr als zwei inzidierende Kanten aufweist. Folglich gibt es auch mehr als 2 Außenknoten mit nur jeweils einer Adjazenz.

Beim Zuordnungstyp III wird eine Repräsentation aus topologisch voneinander getrennten Instanzen eines Datensatzes A einer zusammenhängenden Repräsentation des Datensatzes B zugeordnet. Diese Konstellation tritt in realen Datensätzen dann in Erscheinung, wenn im einen Datensatz die verschiedenen Fahrbahnen einer Straße einzeln erfasst wurden, im anderen jedoch nicht. Zuordnungstyp III kann von Zuordnungstyp I und von Zuordnungstyp II dadurch unterschieden werden, dass es hier exakt vier Außenknoten mit einer Adjazenz von 1 und *keinen* Knoten mit einem Knotengrad größer als 2 gibt.

Bei den Zuordnungstypen II und III ergibt sich nun die Problematik, dass die Ähnlichkeitsmaße aus der gegebenen Geometrie nicht korrekt abgeleitet werden können. So ergäben sich beispielsweise erhebliche Differenzen beim Längenvergleich von Mehrfachrepräsentationen der beiden Zuordnungstypen. Um dennoch eine Vergleichsmöglichkeit zu erhalten, berechnet der RelationBuilder bei gemischten oder topologisch getrennten Elementen eine mittlere Geometrie, die dann für die Berechnung der geometrischen Ähnlichkeitsmaße herangezogen wird. Abbildung 7-6 illustriert den Vorgang für die beiden Zuordnungstypen. Die Vorgehensweise ist einfach, da die Zuordnungstypen zumeist dann vorkommen, wenn die Liniengeometrien gerade verlaufen. Beim Zuordnungstyp II werden zuerst die äußeren Knoten {A, B, C} betrachtet und die Distanzen d_1 , d_2 , und d_3 zwischen ihnen berechnet. Aus den beiden Knoten, die die geringste Distanz zueinander aufweisen (also B und C), wird der Mittelpunkt einer virtuellen Linie zwischen ihnen bestimmt und als Endpunkt der mittleren Geometrie genommen. Der Anfangspunkt der mittleren Geometrie ist identisch mit jenem Außenknoten, der die größte Distanz zu den anderen Außenknoten aufweist. Beim Zuordnungstyp III werden zunächst die voneinander getrennten Liniensegmente in zwei Teilmengen einsortiert. Anschließend wird jeweils zwischen den Anfangs- und Endpunkten der Liniensegmente (also A und B sowie C und D) ein mittlerer Punkt bestimmt. Die beiden erzeugten Punkte werden über eine neue Linie miteinander verbunden, die die mittlere Geometrie repräsentiert (siehe Abbildung 7-6).

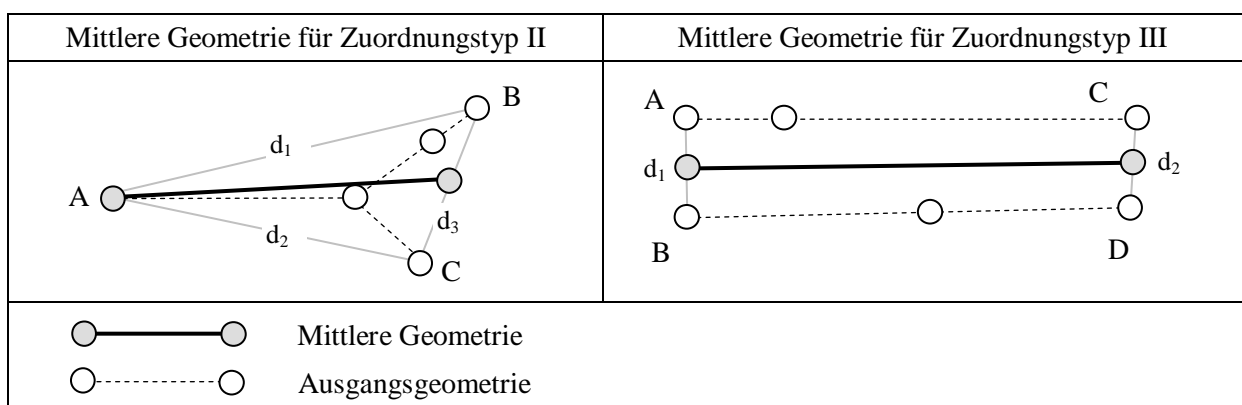


Abbildung 7-6: Bildung von mittleren Geometrien für die Zuordnungstypen II und III

Ein optimierter Ansatz zur Bestimmung der mittleren Geometrie könnte darin bestehen, alle Zwischenpunkte einer Linie (originale Linienpunkte) auf die zugehörige Linie zu übertragen (die im Normalfall weitgehend parallel verläuft) und jeweils aus dem übertragenen und dem ursprünglichen Zwischenpunkt den mittleren

Punkt zu bestimmen. Die Verbindung aller mittleren Punkte lieferte hier eine bessere Approximation, insbesondere bei kurvigen Linienverläufen. Abbildung 7-7 illustriert diese Vorgehensweise für Zuordnungstyp III.

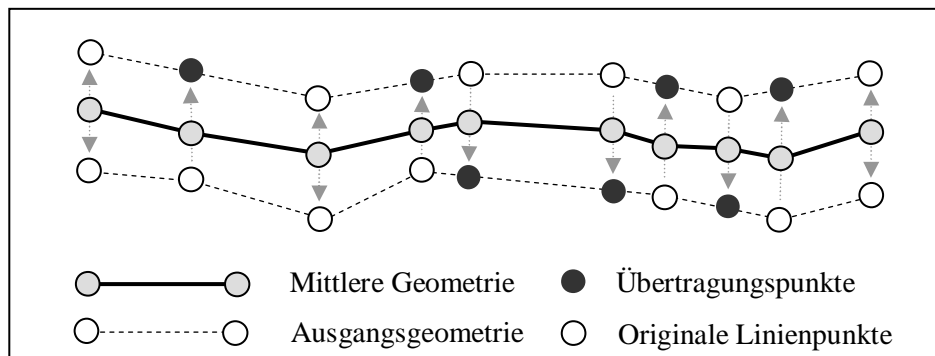


Abbildung 7-7: Bildung von mittleren Geometrien durch Punktübertragung

Außerdem wäre es denkbar, eine Mittelachstransformation über eine Triangulation (z.B. eine Delaunay-Triangulation) zu erreichen, wobei die Mittelpunkte aller Dreiecksseiten der Reihe nach zu verbinden sind, um die mittlere Geometrie zu erhalten (siehe Abbildung 7-8). Die beiden Ansätze wurden jedoch aufgrund des geringen Auftretens solcher Fälle in den Testdaten nicht realisiert.

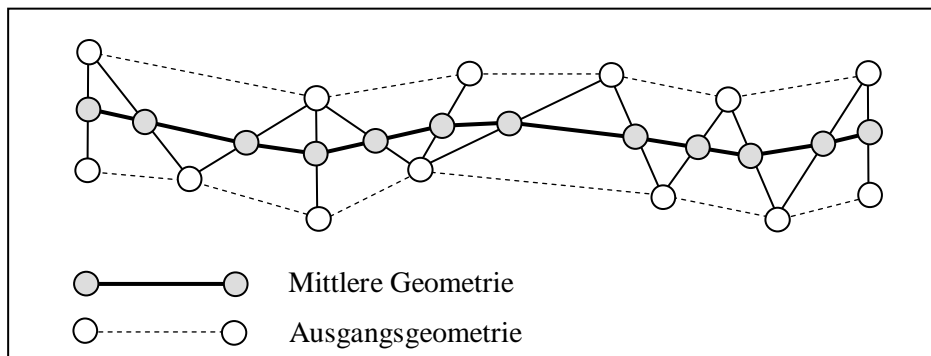


Abbildung 7-8: Bildung von mittleren Geometrien über eine triangulationsbasierte Mittelachstransformation

Neben den Geometrien müssen für die Zuordnungstypen II und III auch die topologischen Beziehungen gesondert berechnet werden. Dabei werden Adjazenzen, die zwischen den Außenknoten existieren und Adjazenzen der Außenknoten zum selben Nachbarknoten eliminiert.

Wie in Kapitel 7.2.2 ersichtlich wird, ist die Problematik der Zuordnung von mehrfach repräsentierten Straßendaten sehr vielfältig und komplex, und es ist nicht möglich, für alle in den Daten auftretenden Fälle adäquate Zuordnungsrichtlinien aufzustellen. Dies kann an manchen Stellen, insbesondere in unübersichtlichen Szenen, zu abweichenden Zuordnungen durch menschliche Operateure führen. Solche Fälle treten in den untersuchten Daten aber nicht signifikant in Erscheinung und sind daher für die Gesamtuntersuchung unproblematisch.

7.1.2 Die RelationViewer-Toolbox

Die RelationViewer-Toolbox erlaubt es, bereits generierte MRep-Relationen zu visualisieren und auf ihre Korrektheit zu überprüfen. Dazu müssen die beiden Datensätze, für welche die Relationen erzeugt wurden, sowie die zugehörige MRRL-Datei geladen werden. Anschließend kann man eine beliebige Instanz aus einem der beiden Datensätze auswählen und bekommt alle zur selben Repräsentation gehörenden Instanzen des gleichen Datensatzes sowie alle korrespondierenden Instanzen des anderen Datensatzes in der Kartendarstellung farblich hervorgehoben dargestellt. Alle Informationen, die in der jeweiligen MRep-Relation gespeichert sind, werden ebenfalls aufbereitet. Abbildung 7-9 zeigt die RelationViewer-Toolbox.

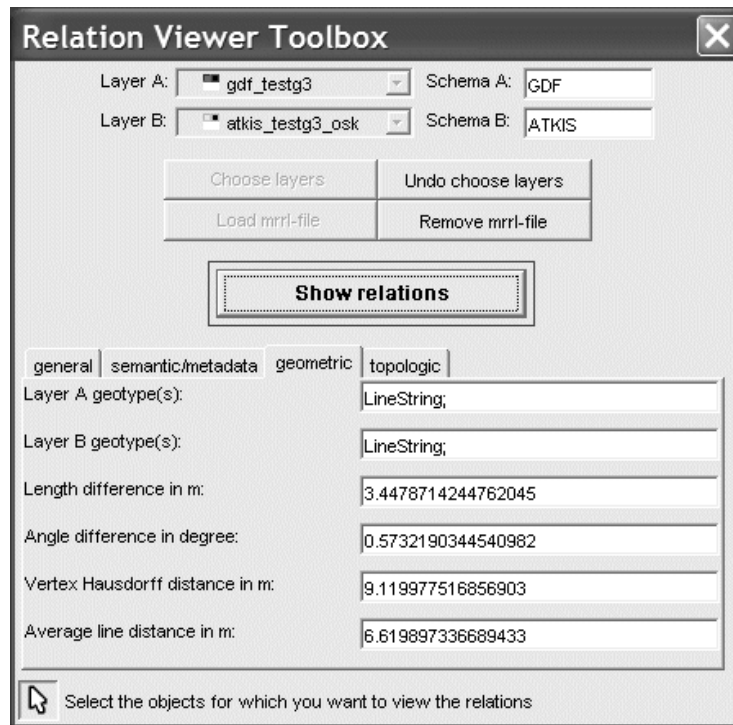


Abbildung 7-9: Die RelationViewer-Toolbox erlaubt die Anzeige von bereits generierten MRep-Relationen und deren Eigenschaften.

7.2 Ableitung von MRep-Relationen anhand verschiedener Beispieldatensätze

Die Vorgehensweise zur Erstellung von MRep-Relationen wurde anhand konkreter Daten untersucht. Die Charakteristika der Daten werden im Anschluss aufgeführt. Daraufhin werden Probleme, die bei der Bildung von MRep-Relationen auftraten, skizziert und die erhaltenen Zuordnungsergebnisse ausgewertet und erläutert.

7.2.1 Beschreibung der verwendeten Daten

Im Stadtgebiet von Stuttgart wurden zwei Testgebiete von jeweils ca. 4 Quadratkilometern Größe ausgewählt, für die Straßendaten aus ATKIS und GDF vorlagen (siehe Anhang II). Bei der Auswahl der Testgebiete wurde darauf geachtet, dass diejenigen Klassen, die auf eine Korrespondenz untersucht werden sollten, in genügender Anzahl vorhanden sind, um möglichst repräsentative Ergebnisse zu erhalten. Außerdem sollten die beiden Testgebiete eine möglichst ähnliche Struktur aufweisen. Schließlich können „*nur Elemente zugeordnet werden, welche dieselben Objekte der Landschaft beschreiben*“ [Walter 97], d.h. nur wenn zwei Datensätze korrespondierende Objektklassen enthalten, kann man sie miteinander vergleichen. Darum befinden sich beide Testgebiete im inneren Stadtbereich von Stuttgart und weisen eine ähnliche Objektausprägung und -dichte auf. In Tabelle 7-2 und Tabelle 7-3 werden die jeweiligen Objektklassen der verschiedenen Datenmodelle und deren Auftreten in den Datensätzen aufgezeigt:

		ATKIS		
	Gesamt	Straßen	Wege	Fahrbahnen
Testgebiet 1	998	672 (67,33%)	180 (18,04%)	146 (14,63%)
Testgebiet 2	1065	633 (59,44%)	320 (30,04%)	112 (10,52%)

Tabelle 7-2: Verteilung der Instanzen auf die verschiedenen Objektklassen von ATKIS in den Testdatensätzen

	GDF			
	Gesamt	Road Elements	Road Elements von Roads	Road Elements von Intersections
<i>Testgebiet 1</i>	1070	906 (84,67%)	123 (11,49%)	41 (3,84%)
<i>Testgebiet 2</i>	1039	872 (83,93%)	128 (12,32%)	39 (3,75%)

Tabelle 7-3: Verteilung der Instanzen auf die verschiedenen Objektklassen von GDF in den Testdatensätzen

Bei ATKIS wurden die Objektarten Straße, Weg und Fahrbahn betrachtet. Die GDF-Daten bestanden aus normalen Road Elements und Road Elements, die einem komplexen Objekt – also entweder einem Road- oder einem Intersection-Objekt – angehören. Der Einfachheit halber werden Road Elements, die zu einer komplexen Klasse gehören, im weiteren Verlauf mit dem Namen der komplexen Klasse (also Road oder Intersection) bezeichnet.

Die Zuordnung der Mehrfachrepräsentationen fand weitestgehend auf der Basis der Betrachtung von Geometrie und Topologie der jeweiligen Datenobjekte statt. In komplexen Bereichen konnte jedoch auch auf die Semantik der Objekte zurückgegriffen werden. Allerdings lagen für die ATKIS-Daten keine Straßennamen vor. Somit konnte dieses bei Straßenobjekten wichtigste Attribut für die Objektbildung nicht für den Zuordnungsprozess herangezogen werden. Die folgenden Sachdaten standen für die betrachteten Datenbestände aus ATKIS (siehe Tabelle 7-4) und GDF (siehe Tabelle 7-5) zur Verfügung.

Attribut	Mögliche Attributwerte
Widmung	1303 (Bundesstraße), 1305 (Landesstraße, Staatsstraße), 1306 (Kreisstraße), 1307 (Gemeindestraße), kein Eintrag
Fahrstreifenzahl	1, 2, 3, 4, kein Eintrag
Funktion Straße	1808 (Fußgängerzone), 2301 (Straßenverkehr), kein Eintrag
Funktion Weg	1701 (Hauptwirtschaftsweg, Verbindungsweg (Fahrweg)), 1702 (Wirtschaftsweg (Feld-, Waldweg)), 1703 (Fußweg), 1704 (Park-, Friedhofsweg), kein Eintrag

Tabelle 7-4: Verfügbare Attribute der ATKIS-Testdaten

Die Bedeutung der ATKIS-Attribute ist selbsterklärend. Es ist lediglich anzumerken, dass bei der Aufbereitung der Daten darauf geachtet wurde, dass Fahrbahnen das Widmungsattribut des zugehörigen Straßenkörper-Objektes übertragen bekamen, um einen Informationsverlust zu vermeiden.

Bei den GDF-Attributen gibt „Functional Class“ den Ordnungsgrad einer Straße an. Der höchste Ordnungsgrad ist 1, der geringste 5. Je geringer der Wert des Attributes ist, desto bedeutsamer ist die Straße innerhalb des Gesamtnetzwerkes. In den GDF-Daten waren außerdem Straßenobjekte unterschiedlicher Geschwindigkeitskategorien („Speed Categories“) und unterschiedlicher Fahrstreifenanzahl („Lane Category“) vorhanden.

Attribut	Mögliche Attributwerte
Functional Class	2 First Class Road, 3 Second Class Road, 4 Third Class Road, 5 Fourth Class Road
Speed Category	5 (51-70 km/h), 6 (31-50 km/h), 7 (11-30 km/h), 8 (<11 km/h)
Lane Category	1 (1 Fahrstreifen), 2 (2-3 Fahrstreifen)

Tabelle 7-5: Verfügbare Attribute der GDF-Testdaten.

Um Missverständnissen im weiteren Verlauf vorzubeugen, wird noch darauf hingewiesen, dass die Speicherung der Daten in einer nicht-objektorientierten layerbasierten Datenstruktur geschah, in der alle Instanzen eines Layers unabhängig von ihrer Klassenzugehörigkeit dieselben Attribute aufweisen. Trifft ein Attribut für eine Instanz nicht zu, so ist dessen Wert nicht belegt. So ist z.B. bei Fahrbahnen oder Straßen das Attribut „Funktion Weg“ ohne Eintrag, während die Zahl der Fahrstreifen bei Wegen nicht vergeben ist.

7.2.2 Probleme bei der Erzeugung von MRep-Relationen

Bei der Bildung von MRep-Relationen zwischen den ATKIS- und GDF-Objekten der untersuchten Testdatensätze trat an einigen Stellen das Problem auf, dass eine eindeutige Zuordnung nicht zu leisten war. In Bestätigung der Aussagen von [Walter 97] war dies genau dann der Fall, wenn im einen Datensatz ein Kreuzungsbereich mit weniger Kanten repräsentiert wurde als im anderen (vgl. Abbildung 7-10).

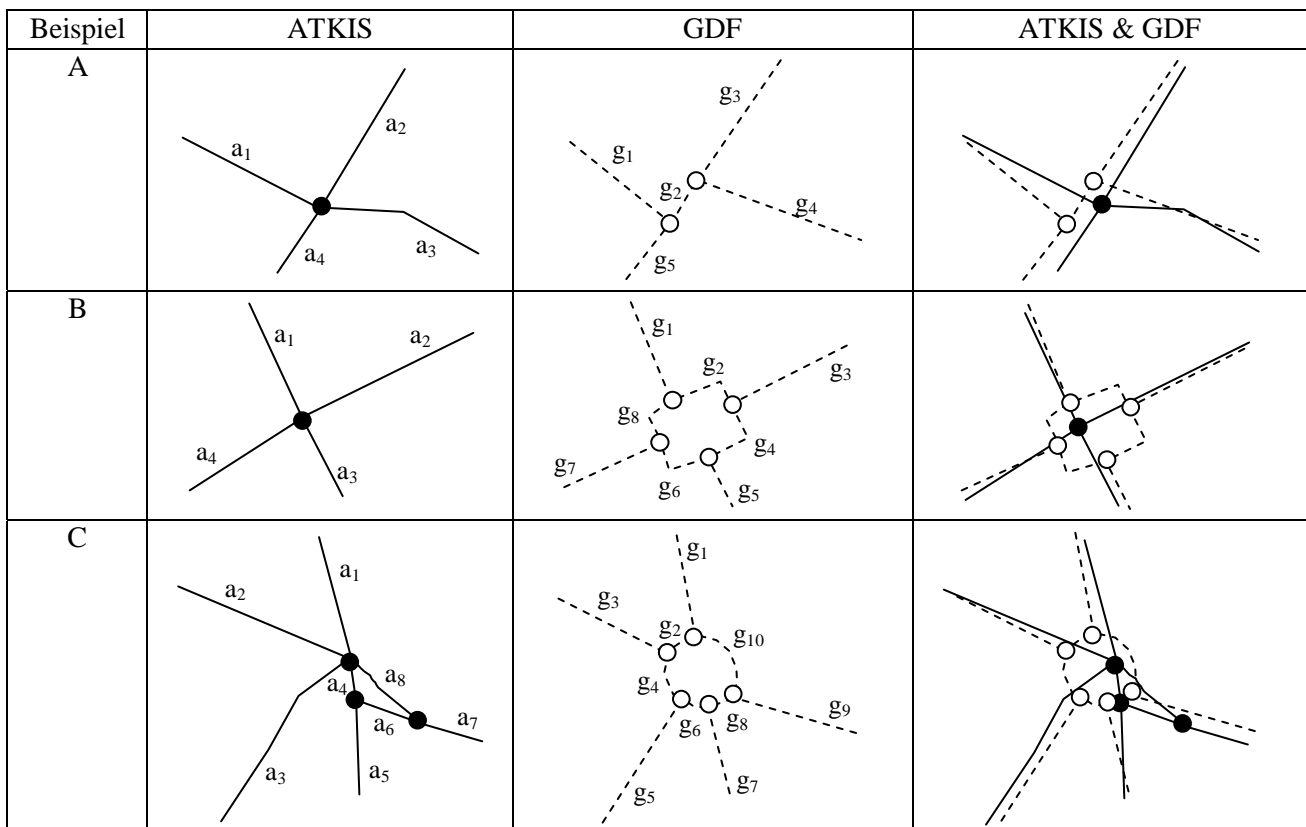


Abbildung 7-10: Darstellung von Szenen, in denen eine eindeutige Zuordnung nicht durchführbar ist (in Anlehnung an [Walter 97])

In den Beispielen A und B werden Kreuzungsbereiche von ATKIS jeweils nur durch einen, in Beispiel C durch zwei Knoten repräsentiert. Unabhängig davon, für welche Zuordnung sich der Operator in Beispiel A entscheidet, können die topologischen Beziehungen in keinem Fall korrekt abgebildet werden. Eine Eindeutigkeit liegt in Fall A deshalb nicht vor, weil man einerseits g_5 und g_2 zu a_4 zuordnen könnte, genauso aber auch g_3 und g_2 zu a_2 . Element g_2 könnte auch ganz außer Acht gelassen werden, so dass also a_1 auf g_1 , a_2 auf g_3 , a_3 auf g_4 und a_4 auf g_5 abgebildet würde, wobei diese Variante gemäß der Vorgaben für die Operateure zu vermeiden ist.

In Beispiel B entsprechen einem Knoten in ATKIS vier Knoten bzw. Kanten in GDF. Auch hier ist eine Eindeutigkeit nicht gegeben. Potentielle Zuordnungen würden beispielsweise je eine längere und eine kürzere Kante in GDF, also z.B. g_1 und g_2 bzw. g_1 und g_8 , g_3 und g_4 bzw. g_3 und g_2 , usw. zusammenfassen und dann der Entsprechung in ATKIS (für die erwähnten Fälle also a_1 oder a_2) zuordnen. Mittels des RelationBuilder-Tools konnte die Zuordnung so durchgeführt werden, dass jeweils 2 mal 3 Kanten von GDF, also z.B. g_6 , g_7 und g_8 sowie g_2 , g_3 und g_4 mit jeweils einer Kante von ATKIS, also entsprechend a_4 sowie a_2 , über eine MRep-Relation des Zuordnungstyps II miteinander verbunden wurden (vgl. Kapitel 7.1.1.3). Die übrig gebliebenen Kanten g_1 und g_5 konnten dann über eine 1:1-Relation auf a_1 und a_3 abgebildet werden.

Beispiel C zeigt, dass auch Fälle größerer Komplexität möglich sind. So ist hier zwar klar erkennbar, dass es sich um denselben Kreuzungsbereich handelt, eine eindeutige oder gar standardisierte Vorgehensweise für die Zuordnung kann daraus allerdings nicht abgeleitet werden.

Wie die folgenden Beispiele (siehe Abbildung 7-11) zeigen, treten außerdem Fälle auf, in denen eine partielle Zuordnung von Teilabschnitten zweier Elemente zwar wahrscheinlich ist, die Objekte als Ganzes jedoch nur mit einer gewissen Unsicherheit zuzuordnen sind. So ist in Beispiel D a_3 vermutlich teilweise identisch mit g_3 . Es liegt im Ermessen des Operators, ob er die Zuordnung akzeptiert oder verwirft.

Beispiel	ATKIS	GDF	ATKIS & GDF
D			
E			

Abbildung 7-11: Die Geometrien von Mehrfachrepräsentationen sind oft nur partiell identisch. Eine Zuordnung ist in diesen Fällen unsicher.

Beispiel E zeigt einen weiteren Konflikt, der im Zuge der Zuordnung in Erscheinung treten kann. So sind hier im GDF-Datensatz mehr Objekte repräsentiert als in ATKIS. Die „zusätzlichen“ Objekte sind aber mit jenen verknüpft, die auch eine Korrespondenz in ATKIS haben. So entspricht bei Beispiel E a_3 den GDF-Elementen g_2 und g_4 . Das Objekt a_2 in ATKIS ist hingegen gewissermaßen eine Mischung der Objekte g_5 und g_6 aus GDF. Zuordnungen sind daher nicht klar durchführbar.

Die dargestellten Beispiele sollen einen Eindruck von der Vielfältigkeit und Komplexität der Zuordnungsproblematik und der dabei auftretenden Konflikte geben. Sie zeigen lediglich einige charakteristische Situationen, doch gibt es weitere Schwierigkeiten. An dieser Stelle können allerdings nicht alle Probleme zur

Sprache kommen. Abschließend ist zu bemerken, dass die vorgestellten Fälle zwar in Einzelsituationen zu Problemen führen können, in der Gesamtbetrachtung aber keine signifikante Ausprägung erreichen und somit das Zuordnungsergebnis *insgesamt* auch nur in vernachlässigbarem Maße beeinflussen. Außerdem ist in diesen Fällen – falls der Operateur sich für die Zuordnung entschließt – in der Regel über das automatisch erzeugte Ähnlichkeitsmaß oder über die menschliche Bewertung zu erkennen, dass die Zuordnung nicht sicher ist (vgl. folgendes Kapitel).

7.2.3 Zuordnungsergebnisse für die Testgebiete

Bei der Auswertung der MRep-Relationen für die beiden Testgebiete konnten die folgenden Ergebnisse gewonnen werden, die untenstehend in tabellarischer Form veranschaulicht sind (siehe Tabelle 7-6). Zu beachten ist, dass die Berechnung der prozentualen Anteile von 1:1-, 1:n- und n:m- Zuordnungskardinalitäten auf der Gesamtzahl der MRep-Relationen (im Testgebiet 1 also 739) beruht. Im Falle von nicht zugeordneten Objekten bzw. 1:0 Zuordnungen konnten keine MRep-Relationen gebildet, weshalb hier auch die Prozentangaben in Tabelle 7-6 fehlen.

Art der Zuordnung	Testgebiet 1		Testgebiet 2	
	ATKIS	GDF	ATKIS	GDF
1:0	111	110	209	179
1:1	512 (69,28%)		427 (65,09%)	
1:n	111 (15,02%)	62 (8,39%)	92 (14,02%)	72 (10,98%)
n:m	54 (7,31%)		65 (9,91%)	
Gesamtzahl MRep-Relationen	739		656	

Tabelle 7-6: Zuordnungsergebnisse für die beiden untersuchten Testgebiete

Wie der Tabelle 7-6 zu entnehmen ist, ist das Auftreten der verschiedenen Zuordnungskardinalitäten innerhalb der MRep-Relationen für beide Testgebiete vergleichbar. Demnach machen die 1:1-Zuordnungen ca. 2/3 aller Zuordnungen aus, während die 1:n-Zuordnungen bei jeweils ungefähr einem Viertel liegen. N:m-Zuordnungen treten zu knapp 10% in Erscheinung.

Allerdings unterscheiden sich die prozentualen Anteile der nicht zugeordneten Objekte in den beiden Testgebieten. So konnten in Testgebiet 1 insgesamt lediglich 11,12% aller ATKIS- und 10,28% aller GDF-Objekte nicht zugeordnet werden, während es in Testgebiet 2 19,62% aller ATKIS- und 17,23% aller GDF-Objekte waren. Eine Betrachtung der Objektklassenzugehörigkeit nicht zugeordneter Instanzen kann hier Aufschluss geben (siehe Tabelle 7-7).

	ATKIS				GDF			
	Gesamt	Straße	Fahrbahn	Weg	Gesamt	Road Element	Road	Intersection
Testgebiet 1	111	31	3	77	110	100	3	7
Testgebiet 2	209	52	7	150	179	170	4	5

Tabelle 7-7: Verteilung der nicht zugeordneten Objekte auf die verschiedenen Objektklassen

Bei den nicht zugeordneten Objekten aus beiden Datenbeständen zeigen sich deutliche Verteilungen hinsichtlich der Klassenzugehörigkeit von Instanzen, die nicht zugeordnet werden konnten. So waren bei ATKIS in erster Linie Objekte der Klasse „Weg“ nur schwer zuzuordnen: für annähernd 43% aller Wege aus Testgebiet 1 und für nahezu 47% aller Wege aus Testgebiet 2 konnten keine Entsprechungen gefunden werden. Bei GDF wiesen fast ausschließlich normale Road Elements keinen Partner im ATKIS-Datensatz auf. Die nicht gefundenen Zuordnungen sind darauf zurückzuführen, dass in vielen Bereichen ATKIS Wege und GDF Road Elements nicht als ähnlich genug erkannt wurden, um eine Zuordnung zu rechtfertigen oder dass überhaupt keine möglichen Zuordnungskandidaten existieren. Abbildung 7-12 zeigt zwei Beispiele, die

das Gesagte belegen. Sowohl in Szene 1 als auch in Szene 2 sind sämtliche Objekte ohne klar ersichtliche Entsprechung entweder Wege oder Road Elements.

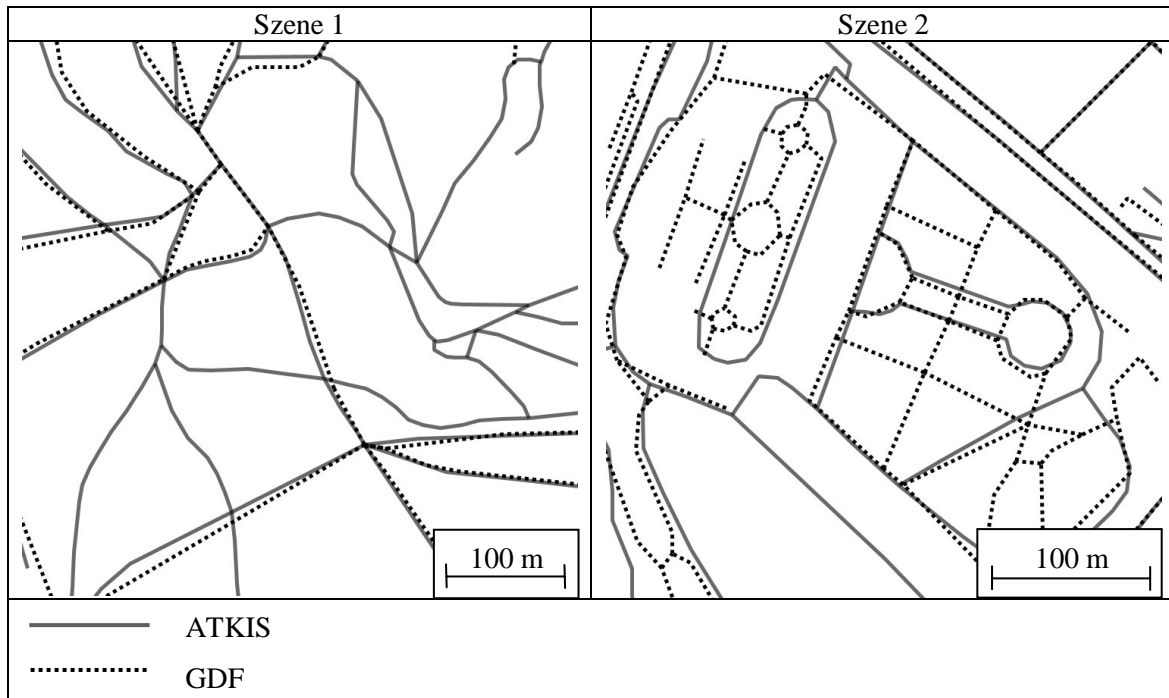


Abbildung 7-12: Zwei Beispiele, die die eingeschränkte Zuordenbarkeit von Road Elements und Wegen in manchen Bereichen illustrieren

Das gegenüber Testgebiet 1 deutlich erhöhte Vorkommen von nicht zugeordneten Instanzen in Testgebiet 2 resultiert den obigen Ausführungen zufolge einerseits daraus, dass in Testgebiet 2 mehr Weg-Objekte vorkommen als in Testgebiet 1 (vgl. Tabelle 7-7). Andererseits gibt es gegenüber Testgebiet 1 mehr Gebiete, in denen Road Elements ohne Zuordnungspartner bleiben.

Für alle Objekte aus den Testgebieten, die an einer MRep-Relation partizipieren, ist die folgende Verteilung der verschiedenen Zuordnungstypen (vgl. Kapitel 7.1.1.3) festzuhalten (siehe Tabelle 7-8). Wie man erkennen kann, wurden die Zuordnungstypen „gemischt“ und „getrennt“ nur selten benötigt.

Testgebiet	Zuordnungstyp: verbunden	Zuordnungstyp: gemischt	Zuordnungstyp: getrennt
1	721	5	13
2	645	3	8

Tabelle 7-8: Verteilung der verschiedenen Zuordnungstypen bei den MRep-Relationen

Für die über MRep-Relationen verbundenen Mehrfachrepräsentationen existieren weitere Charakteristika. An dieser Stelle sollen die Häufigkeitsverteilungen des Gesamt-Ähnlichkeitsmaßes von Repräsentationen beider Testgebiete präsentiert werden (siehe Abbildung 7-13). Sie werden in Prozent ausgedrückt.

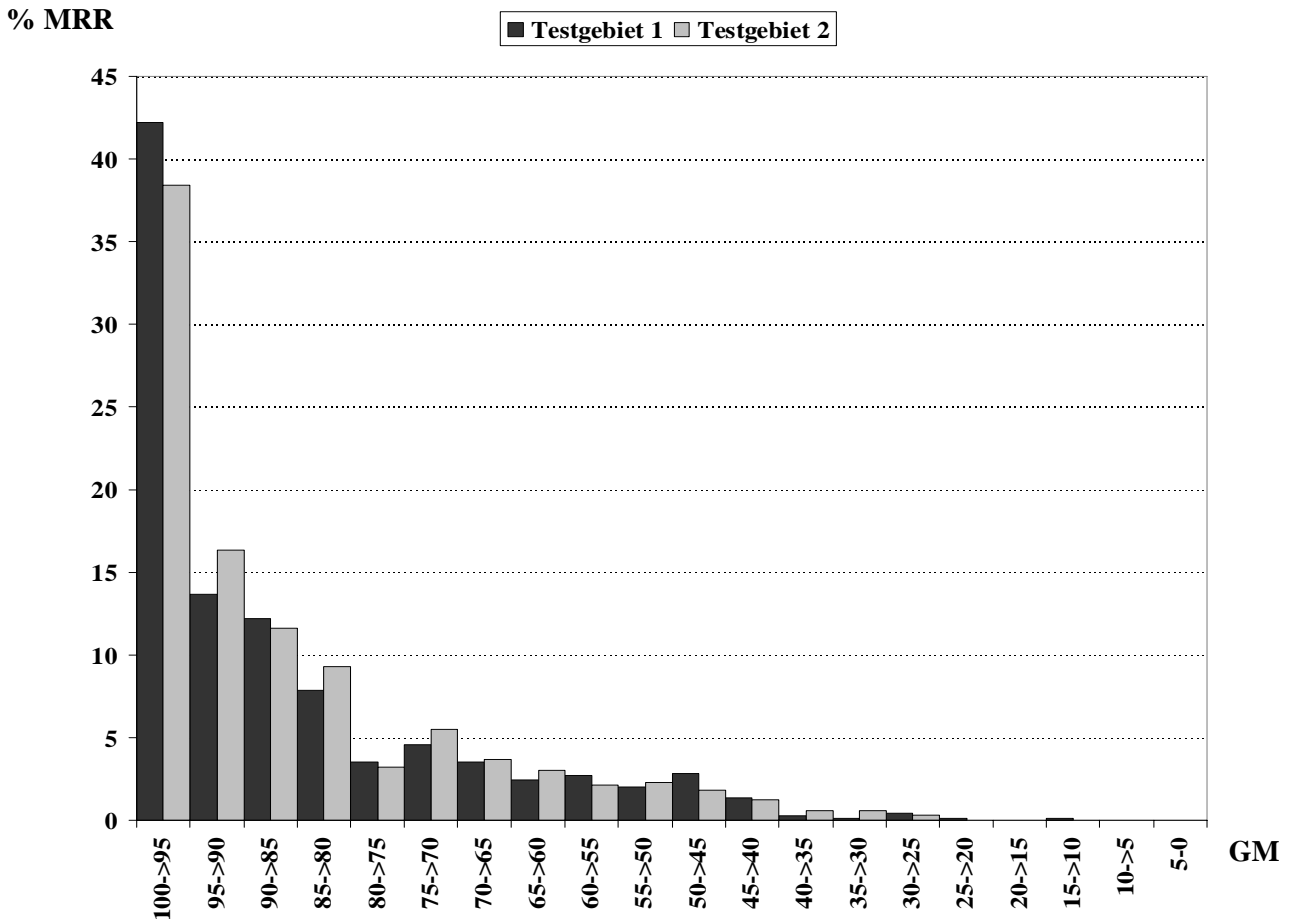


Abbildung 7-13: Prozentuale Häufigkeitsverteilungen des Gesamt-Ähnlichkeitsmaßes (GM) für die MRep-Relationen (MRR) der beiden Testgebiete

Wie zu erkennen ist, zeichnen sich in beiden Testgebieten über drei Viertel aller MRep-Relationen durch ein Gesamtmaß der Ähnlichkeit zugeordneter Repräsentationen von > 80 aus. Weitere knapp 15% MRep-Relationen enthalten ein GM zwischen 80 und 60. Die restlichen ca. 10% der MRep-Relationen haben weitestgehend einen Ähnlichkeitsindex von 60 bis 40, bei nur ca. 1% liegt das GM unter 40.

Wie Abbildung 7-14 zeigt, korrespondiert die jeweilige Anzahl der MRep-Relationen mit schlechtem Ähnlichkeitsmaß allerdings deutlich mit der Zahl der MRep-Relationen, bei denen sich der Operator ebenfalls nicht sicher war und entweder die Bewertungen „undeutlich“ oder „eher ja“ für die Einschätzung der Wahrscheinlichkeit einer Entsprechung vergab. Bei der Betrachtung der subjektiven Einschätzungen des Operateurs bei der Zuordnung fällt darüber hinaus eine nahezu identische Verteilung der Werte für die beiden Testgebiete ins Auge. Dies liefert ein weiteres Indiz dafür, dass sich die Struktur der beiden Testgebiete stark ähnelt.

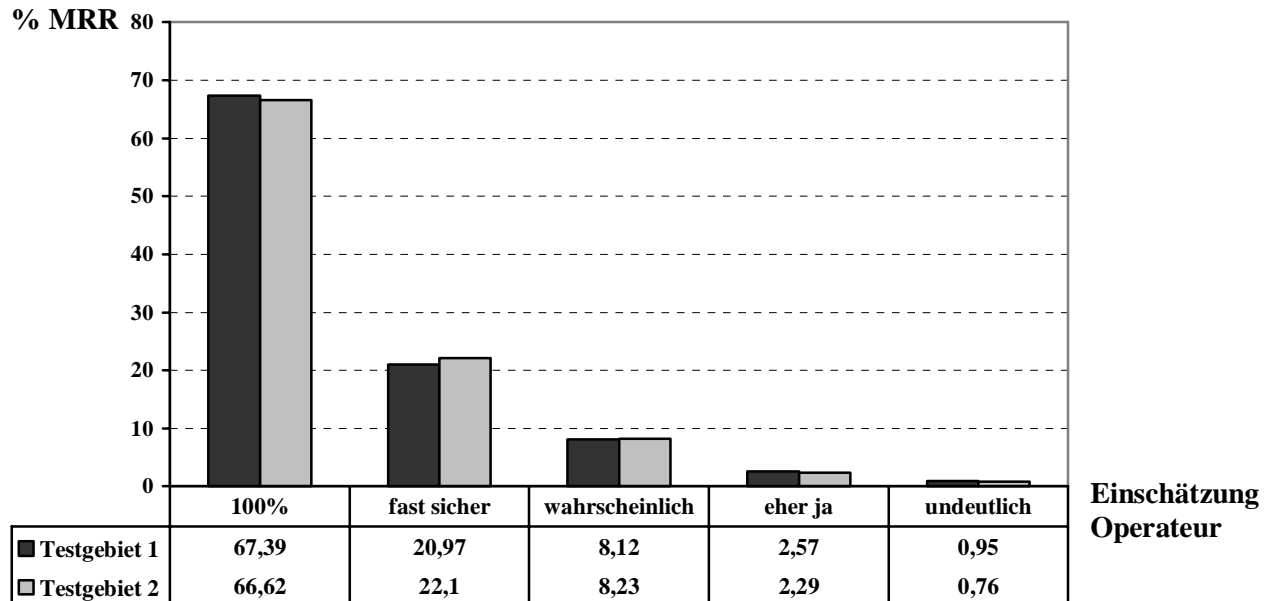


Abbildung 7-14: Prozentuale Häufigkeitsverteilungen der Einschätzungen des Operateurs bezüglich der Sicherheit der Zuordnung für die MRep-Relationen (MRR) beider Testgebiete.

In einem weiteren Schritt kann das Gesamtmaß der Ähnlichkeit zu den jeweiligen Kategorien für die Einschätzung des menschlichen Operateurs in Beziehung gesetzt werden. In den untenstehenden Diagrammen (siehe Abbildung 7-15) wurde jeweils die Gesamtanzahl aller in einer Kategorie auftretenden Zuordnungen als 100% betrachtet. Daraufhin wurde der jeweilige prozentuale Anteil jeder Stufe des GM berechnet. Beispielsweise waren in Testgebiet 1 498 Zuordnungen vom Operateur der Kategorie „100%“ zugeteilt worden. Von diesen Zuordnungen wiesen 364, also ca. 73%, ein GM zwischen 100 und > 90 auf, etc.

Insgesamt ist zu erkennen, dass das GM signifikant mit der menschlichen Bewertung korreliert. So entsprechen z.B. die meisten Zuordnungsfälle (im Mittel der beiden Testgebiet ca. 96%), in denen vom Operateur eine hundertprozentige Übereinstimmung festgestellt wurde, einem Gesamt-Ähnlichkeitsmaß von > 70, wobei mit ca. 73% dieser Fälle der überwiegende Teil sogar ein GM von > 90 aufweist. In den weiteren Diagrammen ist eine Verlagerung der Verteilungskurve nach rechts ersichtlich, d.h. je weniger sicher die Zuordnung war, desto geringer fiel prinzipiell auch das GM aus. Zu den Diagrammen aus Abbildung 7-15 ist allerdings anzumerken, dass für die Kategorien „eher ja“ und „undeutlich“ nur kleine Stichproben verfügbar waren

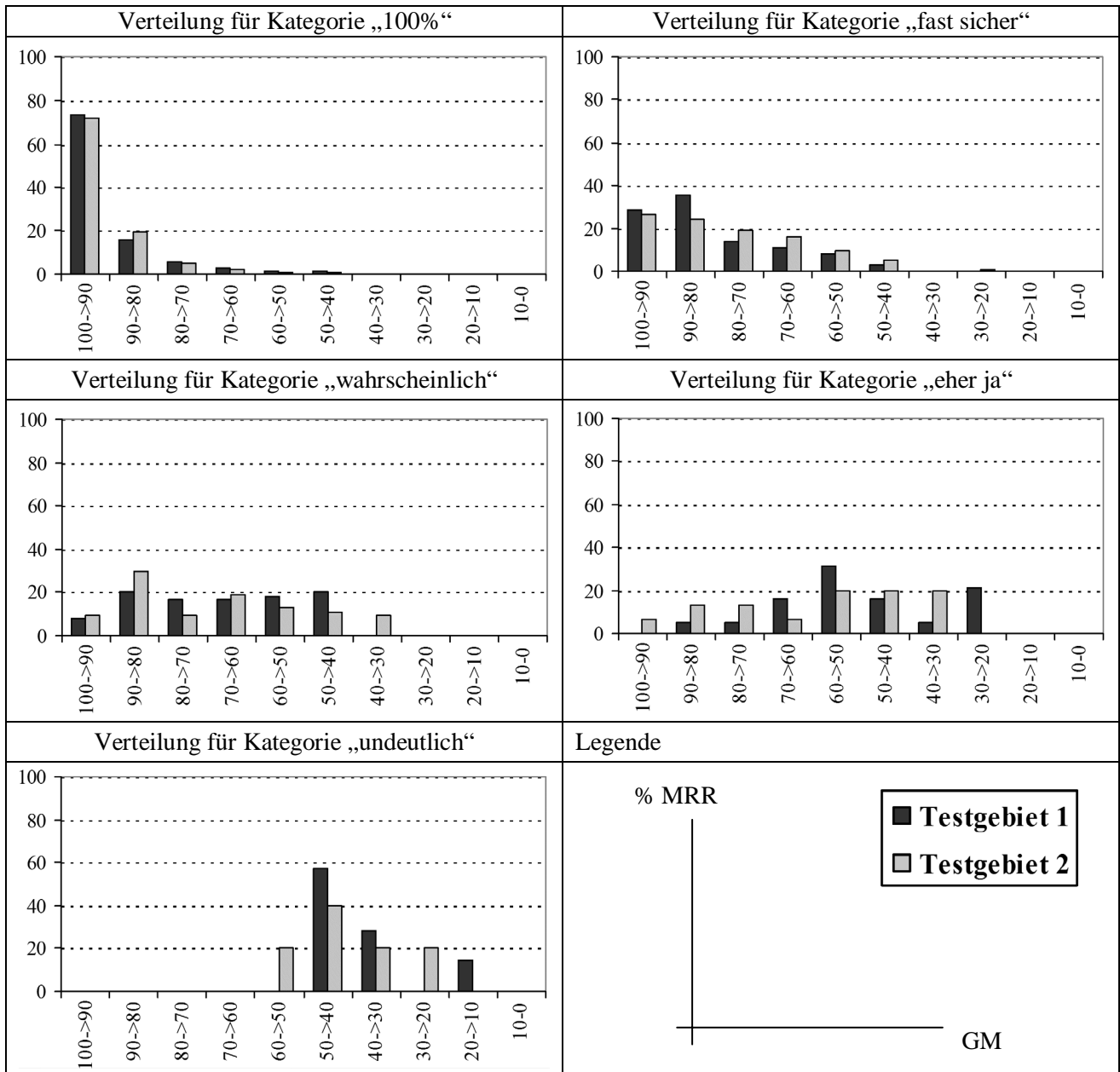


Abbildung 7-15: Häufigkeitsverteilung des GM in den jeweiligen Zuordnungskategorien.

Aus dem Gesagten ist abzuleiten, dass das GM ein nützlicher Indikator ist, um die Ähnlichkeit von über MRep-Relationen verknüpften Mehrfachrepräsentationen ausdrücken zu können. Allerdings existieren offensichtlich einige Ausreißer, bei denen das GM die Eindeutigkeit der Bewertung durch den menschlichen Operator nicht zu bestätigen vermag. Andererseits treten – wenn auch äußerst selten – Fälle auf, in denen das GM verhältnismäßig hoch ist, aber die Einschätzung durch den Operator Zweifel bei der Zuordnung vermuten lässt. In den folgenden Darstellungen (siehe Abbildung 7-16) werden einige charakteristische Beispiele für diese Fälle abgebildet.

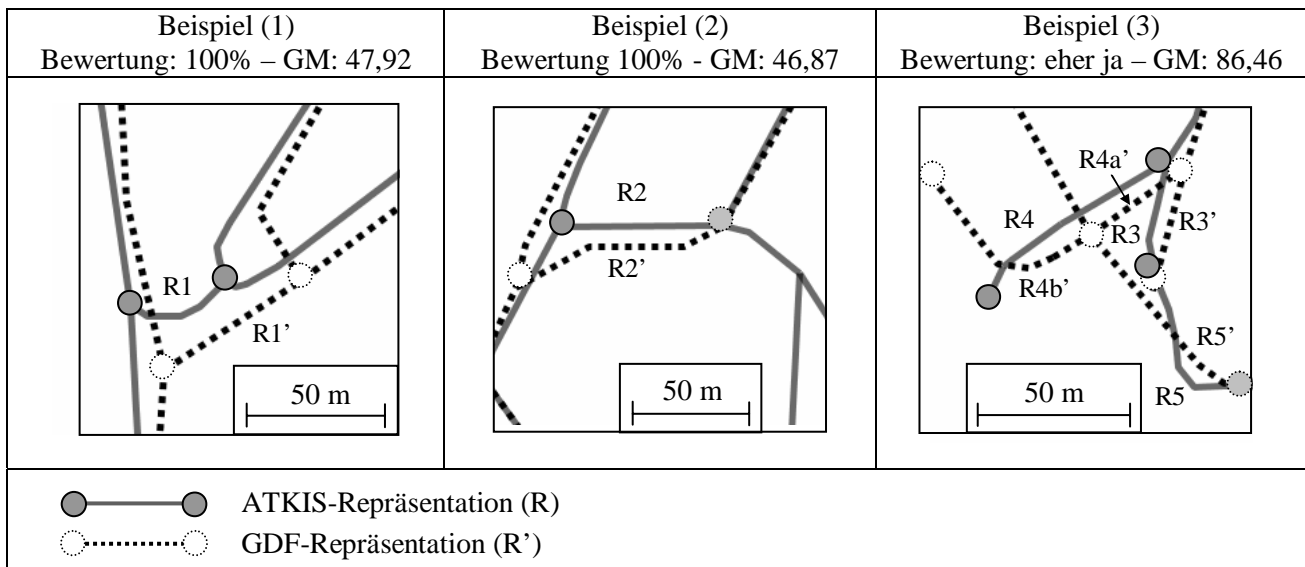


Abbildung 7-16: Szenen, in denen die menschliche Bewertung und das errechnete Gesamtmaß deutlich voneinander abwichen.

Eine eindeutige menschliche Einschätzung gegenüber einem niedrigen Wert des GMs ist in erster Linie in jenen Situationen zu beobachten, in denen zwar die „Gesamtsituation“ für den Operator klar war, sich die Geometrie der Objekte aber stark unterschied und sie auch eine größere Distanz zueinander aufwiesen. In vielen dieser Fälle spielten die topologischen Beziehungen der zugeordneten Repräsentationen eine wichtige Rolle bei der Bewertung durch den Operator, was wiederum nahe legt, der Topologie bei der Berechnung des GM einen noch stärkeren Einfluss zu geben (siehe Abbildung 7-16 (1)). Im Gegensatz dazu gibt es aber durchaus auch Fälle, in denen für die *angrenzenden* Objekte einer Repräsentation im einen Datensatz keine Entsprechungen im anderen Datensatz gefunden werden können und so die Topologie ggf. unterschiedlich ist, obwohl die Repräsentationen für den Operator eindeutig identisch sind (siehe Abbildung 7-16 (2)). Hieraus ist der Konflikt bei der automatischen Zuordnung abzuleiten: an jede Situation werden die gleichen Maßstäbe angelegt. Es ist jedoch vielmehr wichtig, aus der Gesamtsicht auf eine Situation diejenigen Ähnlichkeitsmaße anzuwenden bzw. diejenigen Objektmerkmale zu betrachten, die gerade in der speziellen Szene von Bedeutung sind. Der Mensch tut dies intuitiv. Um ein automatisches System ähnlich „intelligent“ zu machen, könnten hier lernende Verfahren herangezogen werden (vgl. Kapitel 6.4.5 oder z.B. [Sester 99]). Dieser Bereich birgt ein großes Potential für die weitere Forschung.

Sehr seltene Unterschiede zwischen einem niedrigen GM und einer eindeutigen menschlichen Bewertung kamen ferner dadurch zustande, dass die Berechnung der mittleren Geometrie für den Zuordnungstyp III eine zu einfache Approximation war und es so bei kurvigen Linien hohe geometrische und distanzbedingte Abweichungen gab. Bei der Untersuchung der Zuordnungsergebnisse traten auch einige wenige Erfassungsfehler auf. Z.B. gab es vereinzelt Fälle, in denen ein Segment bei der Bildung einer 1:n bzw. n:1 oder einer n:m MRep-Relation fälschlicherweise nicht selektiert wurde. Solche Fälle können durch eine verbesserte Benutzeroberfläche eliminiert werden, indem der Operateur darauf hingewiesen wird, wenn die menschliche Bewertung und das erhaltene GM starke Diskrepanzen aufweisen.

Beispiel 3 in Abbildung 7-16 zeigt einen Fall, in dem das GM sehr hoch war, die Zuordnung aus Sicht des Operateurs aber als „eher ja“ eingestuft wurde und somit als unsicher zu charakterisieren ist. Zwar ist die Geometrie in diesem Fall sehr ähnlich, allerdings repräsentiert R3' eine Sackgasse, während die korrespondierende ATKIS-Repräsentation R3 eine unmittelbar anschließende Fortsetzung in R5 hat und nur aufgrund einer nicht-topologischen Änderung, die aus den vorhandenen Objekteigenschaften heraus nicht zu erkennen war, als separates Segment erfasst wurde. Daher wäre auch eine Alternativzuordnung der GDF-Repräsentation {R4a' plus R5'} zur ATKIS-Repräsentation {R3 plus R5} denkbar, zumal R4 und {R4a' plus R4b'} nicht mit Sicherheit identisch sind. Dieses Beispiel zeigt eine weitere Problematik der Zuordnung: oft reicht es nicht aus, nur die potentiell zueinander gehörenden Objekte zu betrachten, sondern es müssen auch ihre Nachbarobjekte und wiederum deren Nachbarn betrachtet werden, um ein klares Bild zu erhalten. Dies

ist aber für ein automatisches System äußerst schwer zu realisieren. Auch hier sind Ansatzpunkte für die weitere Forschung zu sehen.

8 Auswertung von MRep-Relationen

Das Ziel dieses Kapitels besteht darin, aufzuzeigen, für welche Zwecke MRep-Relationen genutzt werden können, um die Verarbeitung von Mehrfachrepräsentationen in offenen Dateninfrastrukturen wie Nexus zu erleichtern. Dabei werden zwei konkrete Anwendungsfälle untersucht. Zum einen wurde ein System entwickelt, mittels dessen Korrespondenzen zwischen Objektklassen unterschiedlicher Schemas bzw. zwischen Mehrfachkonzeptionen aufgedeckt sowie deren Entsprechungsgrade ausgedrückt werden können. Zum anderen wurde eine Netzwerkanalyse durchgeführt, die unter Nutzung von MRep-Relationen eine Suche von kürzesten Wegen in mehrfach repräsentierten, aus Straßendaten abgeleiteten Graphen realisiert.

8.1 Automatische Erzeugung von Schemarelationen

Im vorangegangenen Kapitel wurde gezeigt, welche Relationen zwischen Mehrfachrepräsentationen auf Instanzebene existieren und wie diese beschrieben werden können. In diesem Teil der Arbeit sollen nun durch eine quantitative Analyse der MRep-Relationen von Straßenverkehrsdaten aus ATKIS und GDF Relationen zwischen den beiden Schemas abgeleitet werden. Im Sinne von [Do und Rahm 02] wird damit versucht, ein Verfahren zum Zwecke des Schema-Matchings zu realisieren (vgl. Kapitel 4.3.1.1).

Im Folgenden wird zunächst der theoretische Ansatz für diese Vorgehensweise erläutert. Dann wird die Software vorgestellt, die durch die automatische Analyse der Instanzrelationen eine Ableitung von Relationen auf Objektklassenebene erlaubt.

8.1.1 Theoretischer Ansatz

Wenn in zwei verschiedenen Datenschemas Entitäten der Realwelt auf unterschiedliche Weise in Form von Objektklassen und zugehörigen Attributen typisiert werden, so wird im Rahmen dieser Arbeit von Mehrfachkonzeptionen gesprochen. Betrachtet man ausschließlich die Schemas, muss man mit Hilfe von Ontologien oder über Verfahren des strukturellen Schema-Matchings Übereinstimmungen zwischen Mehrfachkonzeptionen herausfinden (vgl. Kapitel 4.3.1). Auf diese Weise können Ähnlichkeitsmaße, welche die semantische Nähe von Objektklassen (siehe z.B. [Rodriguez & Egenhofer 04]) ausdrücken, abgeleitet werden. Dadurch kann gezeigt werden, welche Objektklassen der verschiedenen Schemas teilweise oder komplett denselben Sachverhalt der Realwelt modellieren.

Der Ansatz, der dieser Arbeit zugrunde liegt, hat prinzipiell dieselbe Zielsetzung, geht im Gegensatz zu den bisherigen Verfahren jedoch davon aus, dass die unterschiedlichen Schemas auch mit Daten gefüllt sind. Sofern diese Voraussetzung erfüllt ist, kann man die Instanzen der verschiedenen Schemas einander zuordnen und die Charakteristik dieser Zuordnungen analysieren. Dabei gilt es, festzustellen, ob es signifikante Zusammenhänge in der Hinsicht gibt, dass Instanzen bestimmter Objektklassen des einen Schemas stets mit Instanzen bestimmter anderer Objektklassen des zweiten Schemas korrelieren. In der folgenden Darstellung (siehe Abbildung 8-1) soll die Idee des Ansatzes kurz skizziert werden.

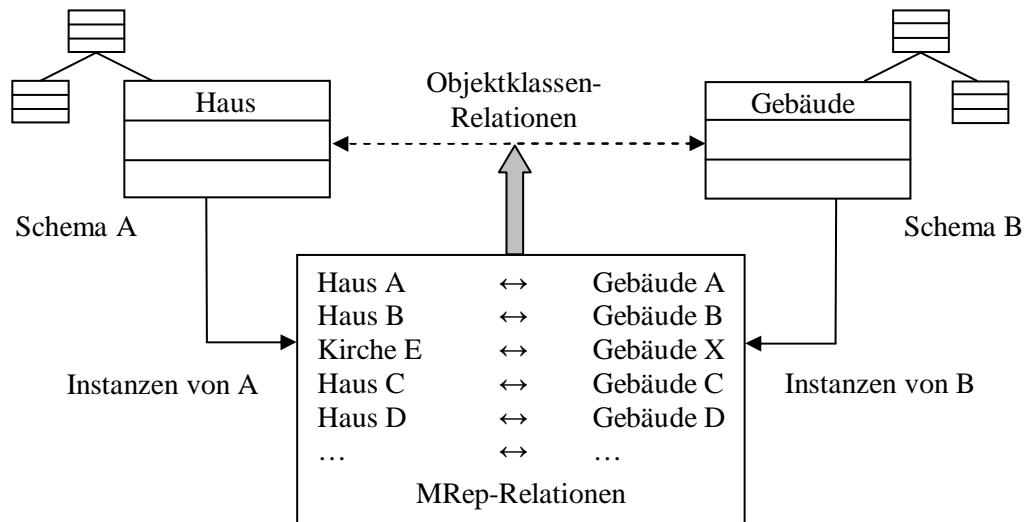


Abbildung 8-1: Durch die Auswertung von MRep-Instanzrelationen kann auf Korrespondenzen zwischen Schemas geschlossen werden.

In Schema A ist eine Objektklasse „Haus“ vorhanden, in Schema B eine Objektklasse „Gebäude“. Würde man den Sinn der Begriffe nicht verstehen, hätte man keine Möglichkeit, herauszufinden, ob diese beiden Objektklassen in irgendeiner Weise semantisch miteinander korrelieren. Betrachtet man aber die Instanzen, so ist sehr leicht zu erkennen, dass Objekte vom Typ „Haus“ aus Schema A meist den Objekten der Klasse „Gebäude“ aus Schema B zugeordnet werden. Wenn das Auftreten dieser Zuordnungscharakteristik nun signifikant bzw. statistisch nachweisbar ist, so ist es möglich, mittels der Betrachtung der Instanzen auf eine Korrelation von Objektklassen zu schließen. Andererseits kann auf diese Weise auch aufgedeckt werden, welche Objektklassen semantisch unähnlich bzw. disjunkt sind.

8.1.2 Ein System zur automatischen Ableitung von Schemarelationen

Im Rahmen des vorgestellten Ansatzes wird in der Folge versucht, durch die Auswertung der MRep-Instanzrelationen von Straßenobjekten der beiden untersuchten Testgebiete aus Kapitel 7 die semantischen Zusammenhänge der Objektklassen aus den Datenschemas ATKIS und GDF näher zu beschreiben. Innerhalb der betrachteten Daten gibt es dabei folgende potentielle Zuordnungsmöglichkeiten von Objektklassen (siehe Abbildung 8-2).

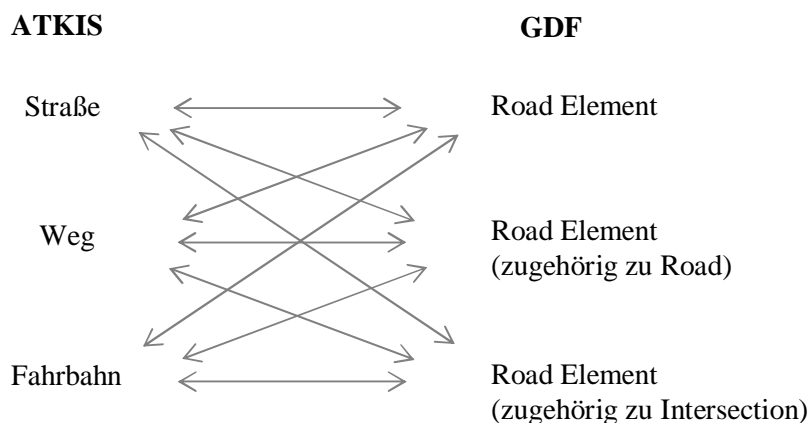


Abbildung 8-2: Potentielle Zuordnungsmöglichkeiten von Objektklassen der untersuchten Testdaten

Jeder Objektklasse aus ATKIS kann theoretisch jede Objektklasse aus GDF zugeordnet werden und umgekehrt, d.h. bei der Betrachtung der Testdaten existieren pro Objektklasse drei verschiedene Korrespondenzmöglichkeiten. Es ist jedoch zu beachten, dass diese 1:1-Abbildungen von Objektklassen nicht die einzig möglichen Fälle darstellen. Bei der Erzeugung von MRep-Relationen der Kardinalitäten 1:n oder n:m kön-

nen auch Instanzen zugeordnet werden, die unterschiedlichen Klassen angehören. Dies bedeutet, dass potentiell beliebige Klassenkombinationen auftreten können. In Abbildung 8-3 wird dieser Sachverhalt verdeutlicht. An dieser Stelle soll noch einmal darauf hingewiesen werden, dass die für die Untersuchung zur Verfügung stehenden GDF-Daten aus Road Elements bestanden, die entweder einem komplexen Road-Objekt oder einem komplexen Intersection-Objekt oder keinem komplexen Objekt angehörten. Der Einfachheit halber werden Road Elements, die zu einer komplexen Klasse gehören, in den nachfolgenden Darstellungen mit dem Namen der komplexen Klasse bezeichnet.

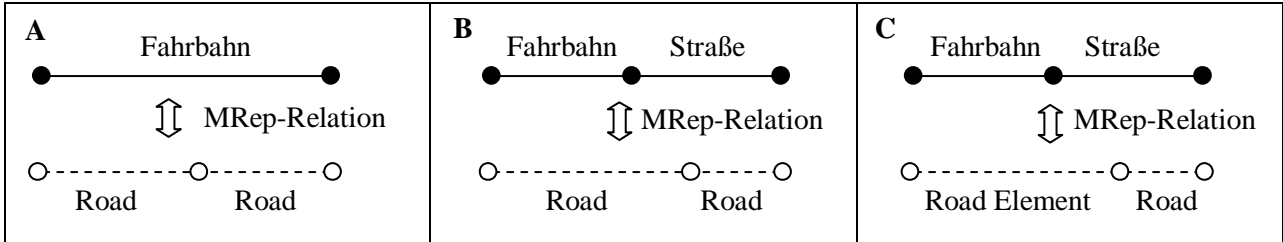


Abbildung 8-3: Eindeutige sowie einseitig und doppelt gemischte Klassenzuordnungen

Beispiel A aus Abbildung 8-3 zeigt eine MRep-Relation der Kardinalität 1:2 zwischen zwei Repräsentationen aus ATKIS und GDF. Die ATKIS-Repräsentation besteht dabei lediglich aus einem einzigen Objekt, das der Klasse Fahrbahn angehört. Dagegen enthält die GDF-Repräsentation zwei Instanzen, die beide der gleichen Objektklasse Road entstammen. Die Klassenzuordnung ist in diesem Falle eindeutig, da nur Instanzen *einer* Objektklasse auf Instanzen *einer* anderen Objektklasse abgebildet werden. Anders ist dies in Beispiel B. Hier besteht die ATKIS-Repräsentation aus zweierlei Instanzen und diese weisen eine unterschiedliche Klassenzugehörigkeit auf: ein Objekt ist vom Typ Fahrbahn, das andere vom Typ Straße. Dies wird als einseitig gemischte Klassenzuordnung bezeichnet, da hier Instanzen *einer* Objektklasse (Road) des einen Schemas auf Instanzen *verschiedener* Objektklassen (Fahrbahn, Straße) des anderen Schemas abgebildet werden. In Beispiel C enthalten beide Repräsentationen Instanzen, die zu verschiedenen Klassen gehören. In diesem Fall spricht man von doppelt gemischten Klassenzuordnungen.

Der vorgestellte Ansatz kann analog auch auf die Analyse von Korrespondenzen zwischen Attributen übertragen werden, wobei die Attributzuordnungen stets im Kontext der zugehörigen Klasse untersucht werden müssen, sofern man keinerlei Vorwissen über die zu vergleichenden Schemas voraussetzt. Entsprechend existieren dann ebenfalls eindeutige sowie gemischte Attributzuordnungen pro Klassenkorrespondenz. Abbildung 8-4.A zeigt ein Beispiel einer eindeutigen Klassenkorrespondenz der Klassen Straße und Road Element einschließlich einer eindeutigen Korrespondenz der Attribute *Widmung: 1307* von Objekt Straße und *Functional_Class: 5* von Objekt Road Element. Dahingegen ist die Attributzuordnung in Abbildung 8-4.B nicht mehr eindeutig, da sich die Attributwerte für das Attribut *Functional_Class* unterscheiden. Somit unterscheiden sich die beiden Fälle A und B und werden deshalb bei der Auswertung (siehe Kapitel 8.1.2.2) separaten Kategorien zugewiesen.

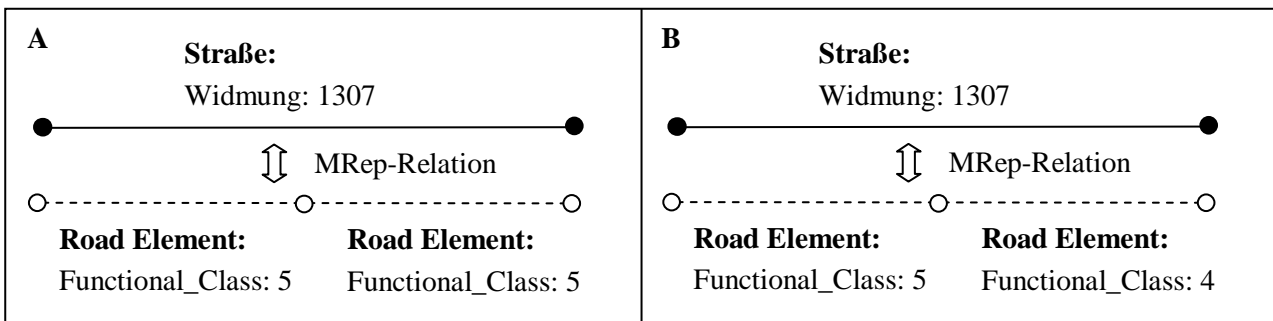


Abbildung 8-4: Zwei Beispiele für Attributzuordnungen von Straßenobjekten aus ATKIS und Road Element-Objekten aus GDF

Prinzipiell können bei der Erzeugung von MRep-Relationen beliebig viele Attribute zugelassen werden. Die Anzahl der potentiellen Zuordnungsmöglichkeiten steigt demzufolge umso stärker, je mehr Objektklassen

und Attribute von verschiedenen Datenschemas betrachtet werden. Werden sämtliche Objektklassen und Attribute der beiden Testgebiet einbezogen, so ist beispielsweise die nachfolgende Zuordnung möglich, die auch in den Testdaten nachzuweisen war (siehe Abbildung 8-5).

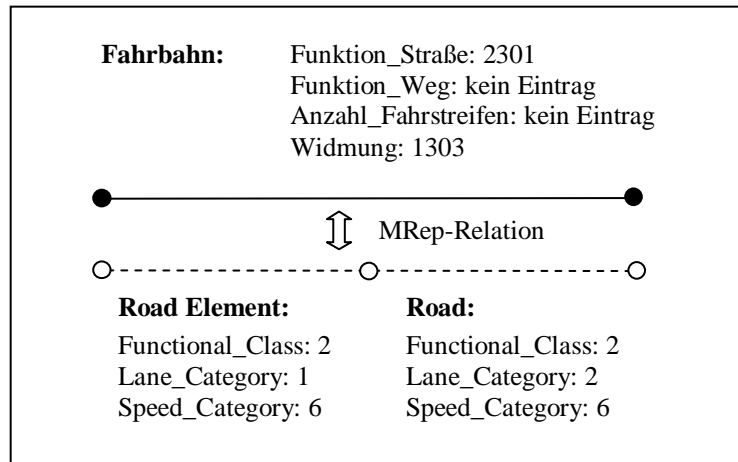


Abbildung 8-5: Beispiel einer möglichen Klassen- und Attributzuordnung

In den anschließenden Abschnitten wird zunächst die entwickelte Software vorgestellt, die zur Ableitung der Schemarelationen dient. Anschließend werden die Ergebnisse der Untersuchungen analysiert und diskutiert.

8.1.2.1 Das RelationAnalyzer-Tool

Wie bereits in Kapitel 7 erwähnt, ist das RelationAnalyzer-Tool ein Programm zur Analyse von MRRL-Dateien. Nach dem Laden eines zu untersuchenden MRRL-Files erlaubt die Software eine Darstellung allgemeiner Informationen über die verknüpften Datensätze wie z.B. die Namen der Eingabelayer oder deren Schemas (siehe Abbildung 8-6). Außerdem können auf einem zweiten Reiter Informationen über die Häufigkeitsverteilungen der verschiedenen Zuordnungskardinalitäten sowie des Gesamt-Ähnlichkeitsmaßes abgelesen werden.

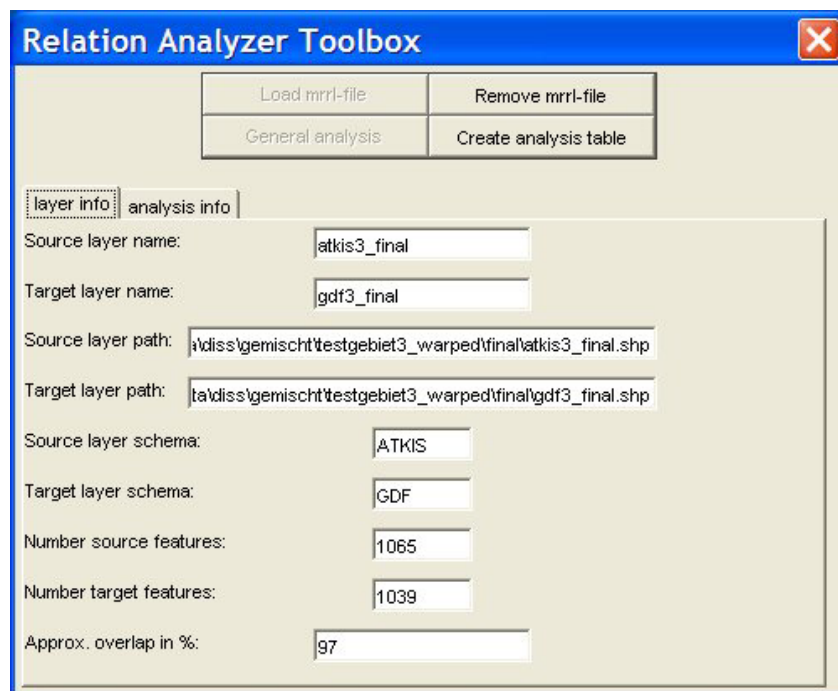


Abbildung 8-6: Darstellung der Relation Analyzer-Toolbox

Nachdem die allgemeine Analyse abgeschlossen ist, können die Analyse-Tabellen generiert werden, die den Zusammenhang von Objektklassen und Attributen verschiedener Schemas deutlich machen. Abbildung 8-7 zeigt die Auswertungstabelle für die Objektklassenzuordnungen des zweiten Testgebiets.

Abbildung 8-7: Analyse-Tabelle für die Klassenzuordnungen des zweiten Testgebietes

Große Tabellen, die auch noch die Attributzuordnungen enthalten, sind innerhalb der entwickelten Oberfläche nur schwer darstellbar. Daher wurde eine Möglichkeit zum Export der Tabellendaten ins ASCII-Format geschaffen. Somit können die Informationen in Standard-Kalkulationsprogramme wie z.B. Microsoft Excel eingeladen und dort untersucht werden.

8.1.2.2 Ergebnisse der Analyse

Die Auswertung der MRep-Relationen zur Ableitung von Hinweisen über die Schemakorrespondenz von ATKIS und GDF erfolgte in zwei Schritten. Zunächst wurden ausschließlich die Zusammenhänge zwischen den Objektklassen Fahrbahn, Straße und Weg aus ATKIS und Road Element, Road und Intersection aus GDF analysiert. In einem zweiten Abschnitt wurden die Untersuchungen auch auf die Betrachtung von Attributkorrespondenzen ausgedehnt.

Korrelationen zwischen Objektklassen

Auf der Basis der mittels des RelationAnalyzer-Tools automatisch erzeugten Tabellen zur Objektklassenkorrespondenz (siehe Abbildung 8-7) erfolgte eine individuelle statistische Auswertung für alle Objektklassen beider Schemas. Die Berechnung der Klassenkorrelationen einer bestimmten Objektklasse wurde dabei folgendermaßen durchgeführt: In einem ersten Schritt wurden alle MRep-Relationen betrachtet, an denen Instanzen der Klasse beteiligt waren. Die Gesamtzahl dieser Relationen (für die Klasse Road Element des Testgebiets 1 also 636) entsprach für die betrachtete Klasse 100%. Danach wurden die Anteile der Korrespondenzen einer Klasse sowie aller Klassenkombinationen (mit Klassen des eigenen Schemas, also z.B. Road Element/Intersection), an denen sie beteiligt war, zu allen Klassen und Klassenkombinationen des anderen Schemas als prozentuale Anzahl an den Gesamrelationen der Klasse errechnet. Ein Beispiel für die GDF-Klasse Road Element soll die Vorgehensweise verdeutlichen (siehe Tabelle 8-1). Die Tabellen für die anderen Klassen sind im Anhang II angefügt. Zum besseren Verständnis der Tabellen sei angemerkt, dass es sich bei Darstellungen mit einem Schrägstrich (/) um Klassenkombinationen innerhalb desselben Schemas handelt. Die Pfeile (→) deuten die Zuordnungen zu einer Objektklasse bzw. einer Klassenkombination des jeweils anderen Schemas an.

Korrespondenzen Road Element	Testgebiet 1	Testgebiet 2	Gesamt
Road Element→Straße	526 (82,7%)	429 (75,26%)	955 (79,19%)
Road Element→Weg	67 (10,53%)	107 (18,77%)	174 (14,43%)
Road Element→Fahrbahn	15 (2,36%)	5 (0,88%)	20 (1,66%)
Road Element→Fahrbahn/Straße	1 (0,16%)	-	1 (0,08%)
Road Element→Straße/Weg	18 (2,83%)	13 (2,28%)	31 (2,57%)
Road Element/Intersection→Straße	4 (0,63)	5 (0,88%)	9 (0,75%)
Road Element/Intersection→Fahrbahn	1 (0,16%)	-	1 (0,08%)
Road Element/Road→Straße	-	4 (0,7%)	4 (0,33%)
Road Element/Road→Fahrbahn	3 (0,47%)	4 (0,7%)	7 (0,58%)
Road Element/Road→Fahrbahn/Straße	1 (0,16%)	3 (0,53%)	4 (0,33%)
Gesamtzahl	636	570	1206

Tabelle 8-1: Erster Schritt bei der Errechnung der Korrelationswerte am Beispiel der Klasse Road Element

Der zweite Schritt umfasste eine Neuauswertung aller einseitig und doppelt gemischten Klassenzuordnungen. Sie sollten in die Werte für eindeutige Zuordnungen einfließen. In einem einfachen Ansatz wurden hierbei die gemischten Klassenzuordnungen unter Annahme einer Gleichverteilung der jeweiligen Klassenzuordnungen auf die eindeutigen Zuordnungen verteilt. Dadurch wurden beispielsweise die Zuordnungen {Road Element→Straße/Weg}, deren Anteil insgesamt 31 betrug, zu jeweils gleichen Teilen (also jeweils 15,5) den Zuordnungen {Road Element→Straße} und {Road Element→Weg} zugeschrieben. Ebenso wurde die Zuordnung {Road Element/Road→Straße}, die insgesamt 4 Mal auftrat, nur 2 Mal für die Zuordnung {Road Element→Straße} berücksichtigt usw. Das Verfahren führte zu den in Tabelle 8-2 dargestellten Korrelationswerten.

Korrelationen Road Element	Werte gesamt
Road Element→Straße	978,5 (81,98%)
Road Element→Weg	189,5 (15,88%)
Road Element→Fahrbahn	25,5 (2,14%)
Gesamtzahl	1193,5

Tabelle 8-2: Korrelationswerte für die Klasse Road Element zu allen ATKIS-Klassen

Führt man die Berechnungen analog für alle betrachteten Klassen durch, so gelangt man zu der folgenden Darstellung (siehe Abbildung 8-8). Da Wege ausschließlich auf Road Elements abgebildet wurden, sind deren gegenseitige Verbindungen zu Roads und Intersections ausgespart worden.

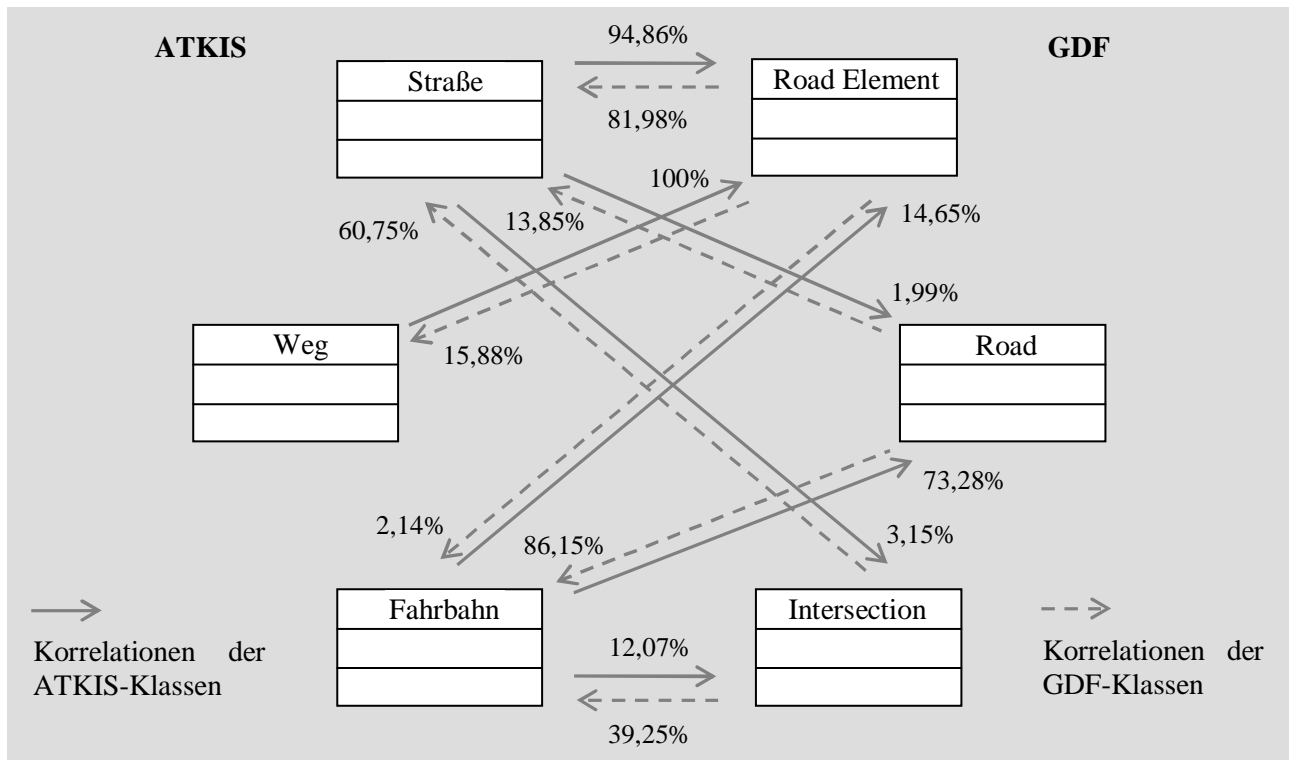


Abbildung 8-8: Aus der Untersuchung der Testgebiete abgeleitete Klassenkorrelationen (alle MRep-Relationen berücksichtigt).

Wie aus Abbildung 8-8 ersichtlich wird, gibt es einige signifikante Korrelationen zwischen den untersuchten Klassen. So werden Wege ausschließlich zu Road Elements zugeordnet. Darüber hinaus entsprechen nahezu 95% der Straßen ebenfalls Road Elements, wogegen die dritte untersuchte ATKIS-Objektklasse Fahrbahn zu fast drei Vierteln auf die Klasse Road abgebildet wurde. Die GDF-Klasse Road Element zeigt mit 81,98% ebenfalls deutliche Korrespondenzen zur Klasse Straße aus ATKIS. Mit 86,15% ist das Verhältnis von Road zu Fahrbahn noch augenscheinlicher. Bei der Klasse Intersection zeigt sich allerdings keine klare Verteilung. Sie entspricht zu knapp 40% der Klasse Fahrbahn und zu gut 60% der Klasse Straße. Wichtige Erkenntnisse liefert auch die Tatsache, dass in allen MRep-Relationen niemals eine Zuordnung von Wege-Objekten aus ATKIS und Intersection- oder Road-Objekten aus GDF stattgefunden hat.

Die vorgestellte Untersuchung von Korrelationen zwischen Objektklassen der Schemata GDF und ATKIS wurde in analoger Weise noch einmal durchgeführt, doch wurden bei der zweiten Analyse lediglich diejenigen MRep-Relationen einbezogen, bei denen das Gesamtmaß der Ähnlichkeit über 80 lag. Hiermit sollte festgestellt werden, ob man deutlichere Korrespondenzen auf Objektklassenebene ableiten kann, wenn auch die betrachteten Instanzen eine sehr hohe Zuordnungswahrscheinlichkeit aufweisen. Abbildung 8-9 zeigt die Ergebnisse für diesen Fall.

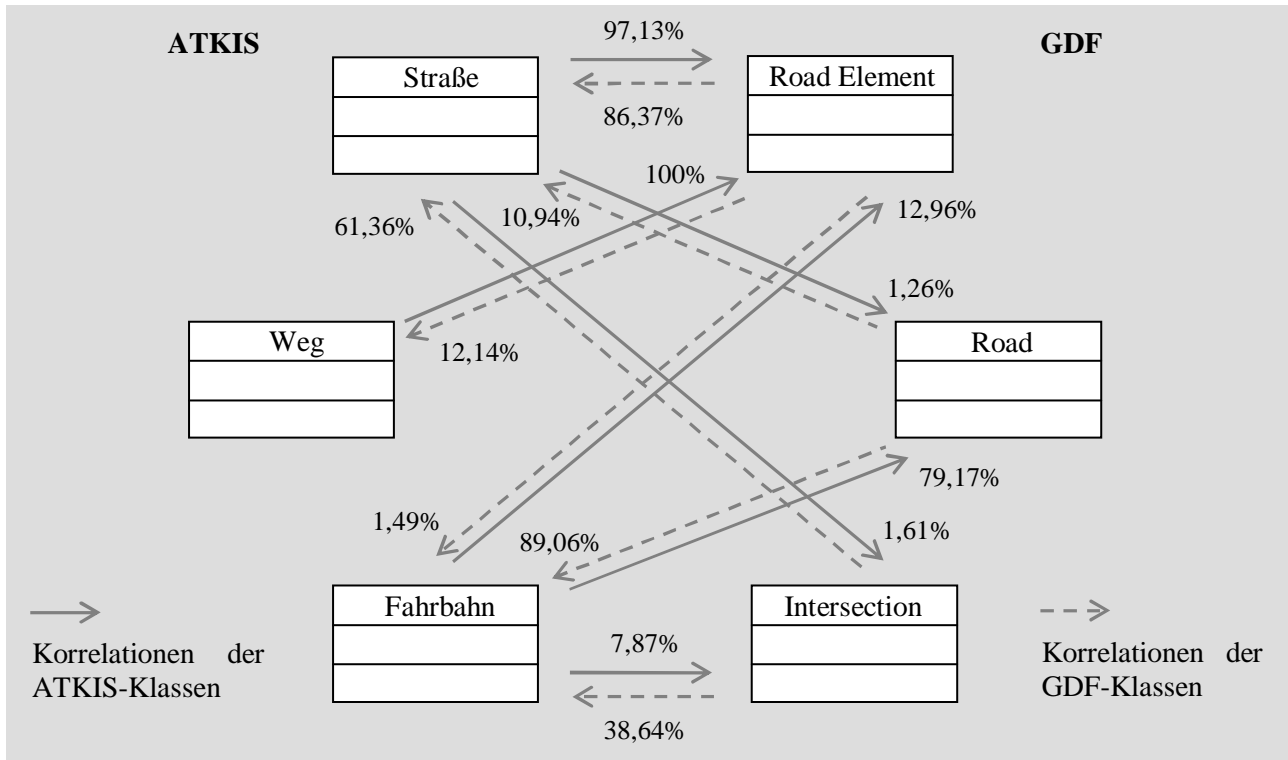


Abbildung 8-9: Aus der Untersuchung der Testgebiete abgeleitete Klassenkorrelationen (nur MRep-Relationen mit GM > 80 berücksichtigt)

Aus Abbildung 8-9 ist zu erkennen, dass die Korrelationsmaße für die Objektklassen in nahezu allen Fällen tatsächlich deutlicher ausfallen. So wird die Zuordnung von Straßen-Objekten aus ATKIS auf Road Element-Objekte aus GDF nun zu gut 97% bestätigt, während es in der ersten Untersuchung knapp 95% waren. Die Werte für den Zusammenhang zwischen der Klasse Fahrbahn und der Klasse Road konnten um knapp 6% auf 79,19% gesteigert werden. Auch die GDF-Klassen Road und Road Element zeigen nun noch klarere Korrespondenzen zu den ATKIS-Klassen Fahrbahn und Straße. Lediglich bei der Klasse Weg aus ATKIS hat sich nichts verändert, da die Abbildung auf die Klasse RoadElement ja schon zuvor vollständig war. Nur geringfügige Veränderungen der Korrelationsmaße gibt es auch bei der Klasse Intersection, die weiterhin zu gut 60% mit der Klasse Straße und zu knapp 40% mit der Klasse Fahrbahn korreliert.

Korrelationen unter Einbeziehung von Attributen

Bezieht man nun zusätzlich Attribute in die Untersuchungen mit ein, so erhält man ein stärker differenziertes Bild. In Abbildung 8-10 ist ein Auszug einer mittels des RelationAnalyzer-Tools erzeugten Analysetabelle für Testgebiet 1 wiedergegeben, die auch Attributzuordnungen enthält. Insgesamt besteht die Tabelle aus 22 Spalten und 33 Zeilen.

---	Strasse FKT_STR-2301;FKT_WEG-none;FSZ-2;WDM-1307;
RDEL func_class-5;lane_cat-1;speed_cat-7;	249 34 10 6 -- sum: (299)
RDEL func_class-5;lane_cat-1;speed_cat-8;	15 5 2 0 -- sum: (22)
ROAD func_class-2;lane_cat-2;speed_cat-6;	0 1 0 0 -- sum: (1)
ISEC func_class-2;lane_cat-2;speed_cat-6;	0 0 0 0 -- sum: (0)
RDEL func_class-2;lane_cat-1;speed_cat-6;	0 0 0 0 -- sum: (0)
RDEL func_class-4;lane_cat-1;speed_cat-6;	25 1 3 0 -- sum: (29)
RDEL func_class-2;lane_cat-1;speed_cat-6;-ROAD func_class-2;lane_cat-2;speed_cat-6;	0 0 0 0 -- sum: (0)
RDEL func_class-3;lane_cat-1;speed_cat-6;	24 3 4 1 -- sum: (32)
RDEL func_class-5;lane_cat-1;speed_cat-7;-RDEL func_class-5;lane_cat-1;speed_cat-8;	0 10 0 1 -- sum: (11)
ISEC func_class-3;lane_cat-1;speed_cat-6;	1 0 2 0 -- sum: (3)

Abbildung 8-10: Ausschnitt einer Analysetabelle, die auch Attributzuordnungen enthält

Die Darstellung sämtlicher Klassen- und Attributkonstellationen würde den Rahmen dieser Arbeit sprengen. Daher wird die Vorgehensweise exemplarisch für die Beziehungen der Klasse Road Element (RDEL) aus GDF und der Klassenkombinationen von Road Element mit den Klassen Road (ROAD) und Intersection (ISEC) mit dem Attribut „Functional_Class“ (FC) und der Klasse Straße (STR) aus ATKIS sowie der Klassenkombinationen von Straße mit Weg (WEG) und Fahrbahn (FRB) mit dem Attribut „Widmung“ (WDM) aufgezeigt. Aus den MRep-Relationen beider Testgebiete konnten hierfür die folgenden Zuordnungen abgeleitet werden (siehe Tabelle 8-3).

	STR; WDM 1307	STR; WDM 1305	STR; WDM 1303	STR; WDM ohne	STR; WDM 1307/ STR; WDM ohne	STR; WDM 1307/ WEG; WDM ohne	STR; WDM 1303/ FRB; WDM 1303	STR; WDM 1307/ STR; WDM 1303	STR; WDM 1307/ FRB; WDM 1305
RDEL; FC 2	-	-	27 (64,29%)	-	-	-	1 (25%)	-	-
RDEL; FC 3	64 (7,23%)	21 (61,76%)	13 (30,95%)	-	-	4 (12,9%)	-	-	-
RDEL; FC 4	93 (10,51%)	7 (20,59%)	-	-	-	-	-	-	-
RDEL; FC 5	714 (80,68%)	2 (5,88%)	1 (2,38%)	1 (100%)	3 (100%)	27 (87,1%)	-	-	-
RDEL; FC 2/ RDEL; FC 3	-	-	1 (2,38%)	-	-	-	-	-	-
RDEL; FC 4/ RDEL; FC 5	2 (0,23%)	-	-	-	-	-	-	-	-
RDEL; FC 3/ RDEL; FC 5	2 (0,23%)	-	-	-	-	-	-	1 (50%)	-
RDEL; FC 2/ RDEL; FC 5	1 (0,11%)	-	-	-	-	-	-	-	-
RDEL; FC 3/ RDEL; FC 4/ ISEC; FC 3	1 (0,11%)	-	-	-	-	-	-	-	-
RDEL; FC 3/ ISEC; FC 3	4 (0,45%)	-	-	-	-	-	-	-	-
RDEL; FC 4/ ISEC; FC 4	-	1 (2,94%)	-	-	-	-	-	1 (50%)	-
RDEL; FC 5/ ISEC; FC 5	3 (0,34%)	-	-	-	-	-	-	-	-
RDEL; FC 2/ ROAD; FC 2	-	-	-	-	-	-	3 (75%)	-	-
RDEL; FC 3/ ROAD; FC 3	-	3 (8,82%)	-	-	-	-	-	-	-
RDEL; FC 5/ ROAD; FC 5	1 (0,11%)	-	-	-	-	-	-	-	-
RDEL; FC 4/ ROAD; FC 3	-	-	-	-	-	-	-	-	1 (100%)

Tabelle 8-3: Zuordnung von Instanzen der Klasse Road Element (RDEL) mit Attribut Functional Class (FC) zu Straßen-Objekten (STR) mit Attribut Widmung (WDM).

Wie aus Tabelle 8-3 zu entnehmen ist, sind die Zuordnungsergebnisse mit einem gewissen Rauschen behaftet. Einige Zuordnungen (z.B. jene der 3 äußeren rechten Spalten) sind aufgrund der Seltenheit ihres Auftretens ohne jegliche Aussagekraft. Die signifikanten Werte (grau eingefärbt) lassen aber dennoch Schlüsse hinsichtlich der Attributkorrespondenz zu. So werden Straßen-Objekte mit der Widmung 1307 (Gemeindestraße) vornehmlich auf GDF-Road Elements niedriger Ordnung (also Functional Class: 5) abgebildet.

Landes- und Staatsstraßen aus ATKIS (Widmung 5) finden zumeist Entsprechungen bei Road Elements der Functional Class 3, während ATKIS-Bundesstraßen (Widmung 1303) mit ca. 65% am häufigsten zu Road Elements der Functional Class 2 zugeordnet werden. Die aus der automatischen Ableitung erhaltenen Informationen entsprechen dem, was man auch als Mensch bei Kenntnis beider Schemas erwarten würde. Somit kann auch für den Fall der Attributzuordnung die Funktionsfähigkeit des Verfahrens bestätigt werden.

8.1.3 Diskussion der Ergebnisse

Mit der Realisierung des vorgestellten Verfahrens (siehe auch [Volz 05a]) konnte gezeigt werden, dass eine „datengetriebene“ Ableitung von Korrespondenzen zwischen verschiedenen Schemas aus dem Bereich der Geodaten sowohl auf Objektklassen- als auch auf Attributebene möglich ist. Somit bietet das Verfahren eine Alternative zu den eher strukturell geprägten Ansätzen des Schema-Matchings, die z.B. von [Do und Rahm 02] präsentiert wurden, sowie zu den Verfahren, die die Ähnlichkeit von Objektklassen über Ontologien zu erfassen versuchen (vgl. Kapitel 4.3.1). Die entwickelte Methodik hilft dabei nicht nur, Korrespondenzen zwischen Mehrfachkonzeptionen aufzudecken, sondern sie liefert auch entsprechende Korrelationsmaße, die vergleichbar mit dem Gesamtmaß der Ähnlichkeit auf Instanzebene den *Grad* der Übereinstimmung von korrespondierenden Objektklassen auszudrücken vermögen. Dabei sind diese Maße nachweislich umso aussagekräftiger, je stärker sich die Instanzen, auf Grundlage derer sie abgeleitet wurden, in geometrischer und topologischer Hinsicht ähneln. Es muss allerdings darauf hingewiesen werden, dass die Korrelationsmaße auf der Auswertung von Testgebieten mit ähnlicher Struktur beruhen. In Zukunft sollten daher auch Auswertungen auf der Basis unterschiedlich strukturierter Testgebiete erfolgen, um ermitteln zu können, inwiefern die erhaltenen Korrelationen hier bestätigt werden.

Interessant wäre es auch, die Ergebnisse mit anderen Ansätzen zu vergleichen oder evtl. sogar mögliche Synergieeffekte durch die Kombination verschiedener Ansätze zu erreichen, wie es auch in [Do und Rahm 02] für das Schema-Matching propagiert wird. Dies bedeutet, dass das vorgestellte Verfahren eine Teilkomponente beim Aufdecken von Schema-Korrespondenzen darstellen könnte, die durch weitere „Zuordnungsfunktionen“ aus den Bereichen Schema-Matching oder Ontologien ergänzt werden kann.

Gemäß [Rahm & Bernstein 01] und [Conrad 02b] ist das Aufdecken von Korrespondenzen zwischen Elementen verschiedener Schemas bzw. zwischen Mehrfachkonzeptionen im Zuge des Schema-Matchings zwar ein zentraler, aber eben nur ein Schritt auf dem Weg zur Schemaintegration, d.h. das hier vorgestellte Verfahren führt noch nicht zu einer kompletten Schemaintegration, sondern ist eher als Teilabschnitt des Integrationsprozesses zu verstehen. Nach dem hier präsentierten Schema-Matching-Ansatz müssen daher weitere Schritte folgen, um eine wirkliche Schemaintegration in Form eines globalen Schemas zu erreichen.

So ist es möglich, im Sinne der von [Conrad 02b] erwähnten zusicherungs-basierten Schemaintegration (vgl. Kapitel 4.3.1.1) eine Klassifikation der Zusicherungen (bzw. der Inter-Schema-Korrespondenzen) vorzunehmen. An dieser Stelle soll dies für Objektklassen- bzw. Element-Korrespondenzen durchgeführt werden:

1. $X_1 \equiv X_2$ (Äquivalenz): X_1 und X_2 repräsentieren immer dieselbe Menge von Objekten der Realwelt. Sie sind semantisch äquivalent.
2. $X_1 \supseteq X_2$ (Einschluss/Inklusion): Die Objektmenge X_1 enthält die Objektmenge X_2 . X_2 ist semantisch immer eine Teilmenge von X_1 .
3. $X_1 \cap X_2$ (Überlappung): Zwischen den Mengen von Realweltobjekten X_1 und X_2 existiert eine nicht-leere Schnittmenge.
4. $X_1 \neq X_2$ (Disjunktheit): Die Schnittmenge von X_1 und X_2 ist stets leer, d.h. es gibt kein Objekt, das gleichzeitig zu X_1 und X_2 gehört.

Die im Rahmen des vorgestellten Verfahrens abgeleiteten Beschreibungen über die semantischen Zusammenhänge zwischen Mehrfachkonzeptionen können auf diese Form der Klassifikation angewendet werden. So konnte für die betrachteten Objektklassen aus GDF und ATKIS in keinem Falle eine Äquivalenzbeziehung festgestellt werden, die anderen Korrespondenztypen sind hingegen alle nachzuweisen.

- Einschluss: Die Objektmenge der Objektklasse Road Element aus GDF schließt die Objektmenge der Objektklasse Weg aus ATKIS ein.
- Überlappung: Es konnten z.B. die folgenden Überlappungen festgestellt werden:

- Die Objektmenge der Objektklasse Road aus GDF überlappt mit der Objektmenge der Objektklasse Fahrbahn aus ATKIS.
- Die Objektmenge der Objektklasse Straße aus ATKIS überlappt mit der Objektmenge der Objektklasse Road Element aus GDF.
- Etc.
- Disjunktheit: Die Objektmenge der Objektklasse Intersection aus GDF und die Objektmenge der Objektklasse Weg aus ATKIS sind disjunkt.

Im Falle der zusicherungs-basierten Integrationstechnik wird auf der Basis der festgestellten Zusicherungen bzw. Korrespondenzen über den Einsatz von Integrationsregeln der letzte Schritt der Schemaintegration durchgeführt. „Eine Integrationsregel legt dabei für eine bestimmte Art der Korrespondenz fest, wie die betroffenen Schemaelemente, für die diese Korrespondenz gilt, ins integrierte Schema abgebildet werden“ [Conrad 02b]. Ein mögliches Verfahren zur Aufstellung von Integrationsregeln unter Ausnutzung der abgeleiteten Zusicherungen ist beispielsweise der Ansatz der „Upward Inheritance“ (siehe Abbildung 8-11), welcher ebenfalls in [Conrad 02b] erläutert wird. Stellt man eine Äquivalenz fest, werden die Klassen der Ausgangsschemas im integrierten Schema vereinigt. Bei Inklusionen wird die einschließende Klasse als Oberklasse der eingeschlossenen Klasse festgelegt. Treten Überlappungen oder disjunkte Klassen auf, so wird im integrierten Schema eine neue Klasse als Oberklasse der Ausgangsklassen eingeführt.

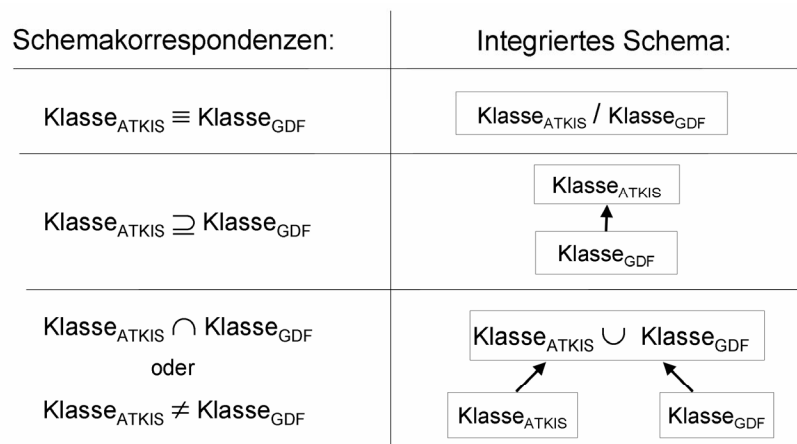


Abbildung 8-11: Schemaintegration gemäß des Upward Inheritance-Prinzips (nach [Conrad 02b])

Integrationsregeln stellen im Grunde Konfliktlösungsstrategien dar, wie sie z.B. von [Spaccapietra et al. 96] vorgeschlagen wurden (vgl. Kapitel 4.3.1.1). Im Rahmen dieser Arbeit wurden keine Integrationsregeln spezifiziert. Hier müssen daher zukünftige Arbeiten ansetzen. Dabei kann man sich den Vorteil des präsentierten Verfahrens zu Nutze machen, dass nicht nur der Typ der Korrespondenz zweier Klassen festgestellt, sondern im häufigsten Falle der Überlappung auch der Grad der Überlappung mittels der Korrelationsmaße angeben werden kann. Im Hinblick auf Nexus kann die Schemaintegration auf der Basis des vorgestellten Verfahrens dabei helfen, globale AWS-Objektklassen für andere Objekttypen aus dem Bereich der Geodaten (z.B. Gebäude) zu entwerfen. Außerdem können auf diese Weise Mehrfachkonzeptionen, die aus unterschiedlichen Erweiterungen des AWS für gleiche Entitäten der Realwelt entstehen, ggf. im Standard-Klassenschema zusammengeführt werden.

Es ist aber auch möglich, Schemas nicht komplett zu integrieren, sondern analog zu den MRep-Relationen auf Instanzebene die Korrespondenzen von Objektklassen explizit zu formulieren und so eine eher „lose“ Integration der Ausgangsschemas zu realisieren. [Balley et al. 04] verfolgen diesen Ansatz in einer aktuellen Arbeit und übertragen die Objektklassen der lokalen Schemas (also z.B. ATKIS und GDF) ohne Veränderungen in das globale integrierte Schema. Über so genannte Korrespondenz-Relationen werden einander ähnliche Objektklassen dann im integrierten Schema miteinander verknüpft. Diese Korrespondenz-Relationen drücken die potentiellen Beziehungen zwischen den Instanzen der jeweiligen Objektklassen aus; dies können Äquivalenzbeziehungen (1:1), Aggregationsbeziehungen (1:n bzw. n:1) oder so genannte „Set-to-Set“-Beziehungen (n:m), also Beziehungen zwischen Mengen, sein. Sind für zwei Klassen verschiedene Beziehungen möglich, so werden alle möglichen Beziehungen und die entsprechenden Kardinalitäten im

Schema angegeben. Diese Art der Modellierung könnte unter Nutzung des vorgestellten Verfahrens noch stärker differenziert werden. So wäre es beispielsweise möglich, den Korrespondenz-Relationen auf Objektklassenebene nicht nur die Korrelationsmaße als Attribute mitzugeben, sondern auch den Prozentsatz der jeweiligen Kardinalitäten von Zuordnungen auszudrücken, die aus der vom RelationAnalyzer-Modul erzeugten Analyse-Tabelle hervorgeht. Abbildung 8-12 zeigt einen Ansatz hierzu.

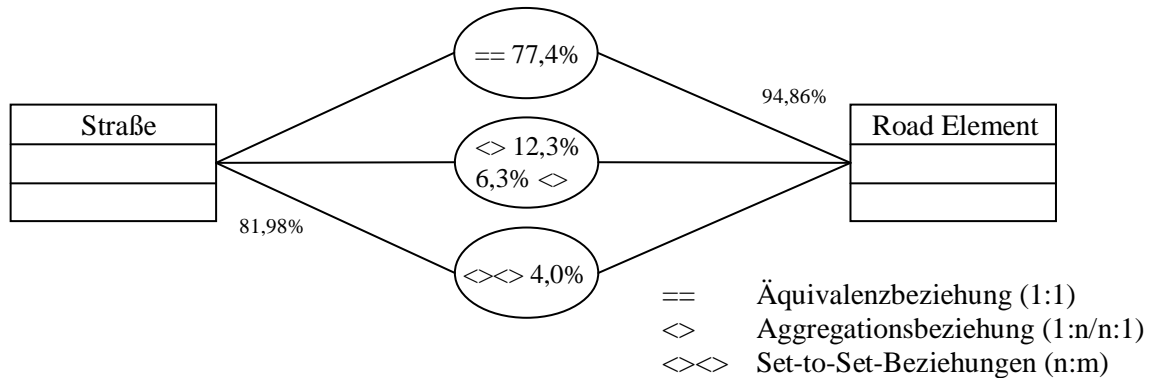


Abbildung 8-12: Modellierung von Korrespondenzen zwischen Objektklassen (in Anlehnung an die Notation von [Balley et al. 04])

Wie Abbildung 8-12 zu entnehmen ist, kann durch diese Vorgehensweise im integrierten Schema über explizite Symbolik der Zusammenhang zwischen Objektklassen verschiedener Ausgangsschemas deutlich gemacht werden. So kann man nicht nur erkennen, dass ca. 95% aller Straßen Road Elements und umgekehrt ca. 82% aller Road Elements Straßen zugeordnet wurden, sondern es ist auch die Verteilung der Zuordnungen auf die verschiedenen Kardinalitäten zu sehen. Beispielsweise haben 12,3% aller Road Elements eine Aggregationsbeziehung zu Straßen, während es umgekehrt 6,3% sind. Zumeist können bei Klassenzuordnungen zwischen Road Elements und Straßen aber Äquivalenzbeziehungen festgestellt werden.

Betrachtet man die mittels des entwickelten Verfahrens abgeleiteten Korrelationsmaße zwischen Objektklassen unterschiedlicher Schemas, so stellt sich die Frage, ob die hier gewonnenen Informationen auch für andere Anwendungen als die Schemaintegration zur Verfügung gestellt werden können. Dabei liegt ein Applikationsbereich unmittelbar auf der Hand: die Zuordnung von Mehrfachrepräsentationen. Bei der Zuordnung von korrespondierenden Instanzen können die Korrelationsmaße auf Objektklassenebene als Vorinformationen dienen. So ist es beispielsweise unter Berücksichtigung der abgeleiteten Korrelationsmaße zwischen den betrachteten Objektklassen von ATKIS und GDF ausgeschlossen, dass ein Weg-Objekt aus ATKIS einem Fahrbahn-Objekt aus GDF zugeordnet werden kann. Sollte eine solche Zuordnung also in einem automatischen oder einem semi-automatischen Verfahren zustande kommen, könnte man aufgrund der gegebenen Vorinformationen mit hoher Wahrscheinlichkeit auf eine Fehlzuordnung schließen. Solche Ansätze sind bislang nicht betrachtet worden und stellen eine Herausforderung für zukünftige Forschungsarbeiten dar.

Ferner ist es denkbar, die Korrelationsmaße, die zwischen Mehrfachkonzeptionen existieren, auch für die Abbildung von Daten des einen Schemas in ein anderes zu nutzen. Wollte beispielsweise ein Datenanbieter an Nexus teilnehmen und gäbe es für sein Datenschema noch keine Abbildungsvorschrift auf das AWS, so könnte man mit dem vorgestellten Verfahren ein Schema-Matching mit Daten des Anbieters und AWM-L-Daten durchführen und aus den erhaltenen Informationen evtl. entsprechende Abbildungsregeln ableiten. Diese Annahme stützt sich auf die Aussage von [Rahm & Bernstein 01], die bemerken, dass das Schema-Matching für Transformationen von einem Format ins andere nützliche Hinweise liefern kann. Auch für die von [Kuhn 03] geforderten Transformationen zum Austausch von Informationen zwischen verschiedenen semantischen Referenzsystemen (vgl. Kapitel 4.3.1.2) könnte das entwickelte Verfahren dienlich sein.

8.2 Auswertung von MRep-Relationen zur Navigation in mehrfach repräsentierten Graphen

Existieren Relationen zwischen Mehrfachrepräsentationen auf Instanzebene, so können diese für verschiedene GIS-Prozesse genutzt werden (vgl. Kapitel 4.2.2). Im Zuge dieser Prozesse spielen insbesondere die verschiedenen Ähnlichkeitsmaße sowie das Gesamtmaß der Ähnlichkeit zentrale Rollen. Beispielsweise können Daten, die ein hohes Maß an Übereinstimmung zeigen, gemeinsam fortgeführt werden, um auf diese Weise den Aufwand bei der Datenaktualisierung zu minimieren. Darüber hinaus wird es möglich, stark ähnliche Objekte miteinander zu verschmelzen, d.h. aus den bestehenden Mehrfachrepräsentationen verschiedener Datenbestände neue, lediglich einmalig existierende Repräsentationen oder auch Mono-Repräsentationen zu generieren und somit insgesamt *einen* resultierenden, konsistenten Datensatz zu schaffen, auf dem GIS-Anwendungen operieren können.

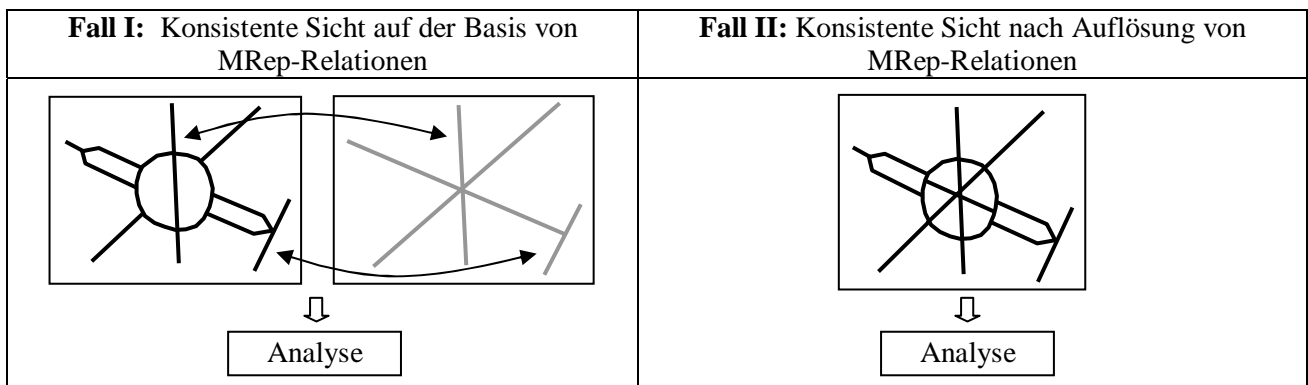


Abbildung 8-13: Zwei Varianten einer „konsistenten Sichtweise“: Mehrfachrepräsentationen bleiben als getrennte Instanzen erhalten oder werden zu einer resultierenden (Mono-)Repräsentation verschmolzen.

Wie in der Folge gezeigt werden kann, müssen Mehrfachrepräsentationen zum Zwecke der Analyse aber nicht miteinander verschmolzen werden, sondern sie können auch getrennt voneinander erhalten bleiben, wobei die Auswertung der in MRep-Relationen gespeicherten Informationen dann in den Analyseprozess verlagert wird. Vom Standpunkt dieser Arbeit und auch vom Nexus-Standpunkt aus betrachtet, kann eine konsistente Sicht auf Föderationsebene also auf zweierlei Arten realisiert werden (siehe Abbildung 8-13).

Für den Fall II aus Abbildung 8-13, bei dem MRep-Relationen aufgelöst und Mehrfachrepräsentationen verschmolzen werden, sind die bestehenden Mechanismen bei der Datenanalyse anwendbar. Will man jedoch Mehrfachrepräsentationen im Zuge der Analyse erhalten (Fall I), so sind neue Ansätze notwendig. An dieser Stelle wird ein solcher Ansatz für eine Netzwerkanalyse in mehrfach repräsentierten Graphen unter Auswertung von MRep-Relationen entwickelt. Im Folgenden wird zunächst die allgemeine Vorgehensweise zur Lösung des Problems erläutert, bevor dann die konkrete Implementierung beschrieben wird.

8.2.1 Mögliche Strategien zum Aufbau eines integrierten Navigationsgraphen

Die grundlegende Zielsetzung des zu entwickelnden Verfahrens besteht darin, eine Möglichkeit zu finden, von einem Startknoten in einem Netz bzw. in einem Graphen I zu einem Zielknoten in einem Netz II zu gelangen und dabei den kürzesten Weg zu finden. Dabei kann der Graph beliebig oft gewechselt werden. Die Problematik wird an einem einfachen Beispiel in Abbildung 8-14 illustriert: Es sind zwei kleine Graphen dargestellt, die jeweils gewichtet sind. Richtungen wurden jedoch nicht vergeben. Aus beiden Graphen sind jeweils drei Kanten und 4 Knoten mehrfach repräsentiert vorhanden. MRep-Relationen wurden allerdings nur für die mehrfach repräsentierten Kantenobjekte (MRep-Kanten), jedoch nicht für die zugehörigen Knotenobjekte, die so genannten MRep-Knoten, erzeugt, d.h. Ähnlichkeitsmaße liegen lediglich für die zugeordneten Kanten vor.

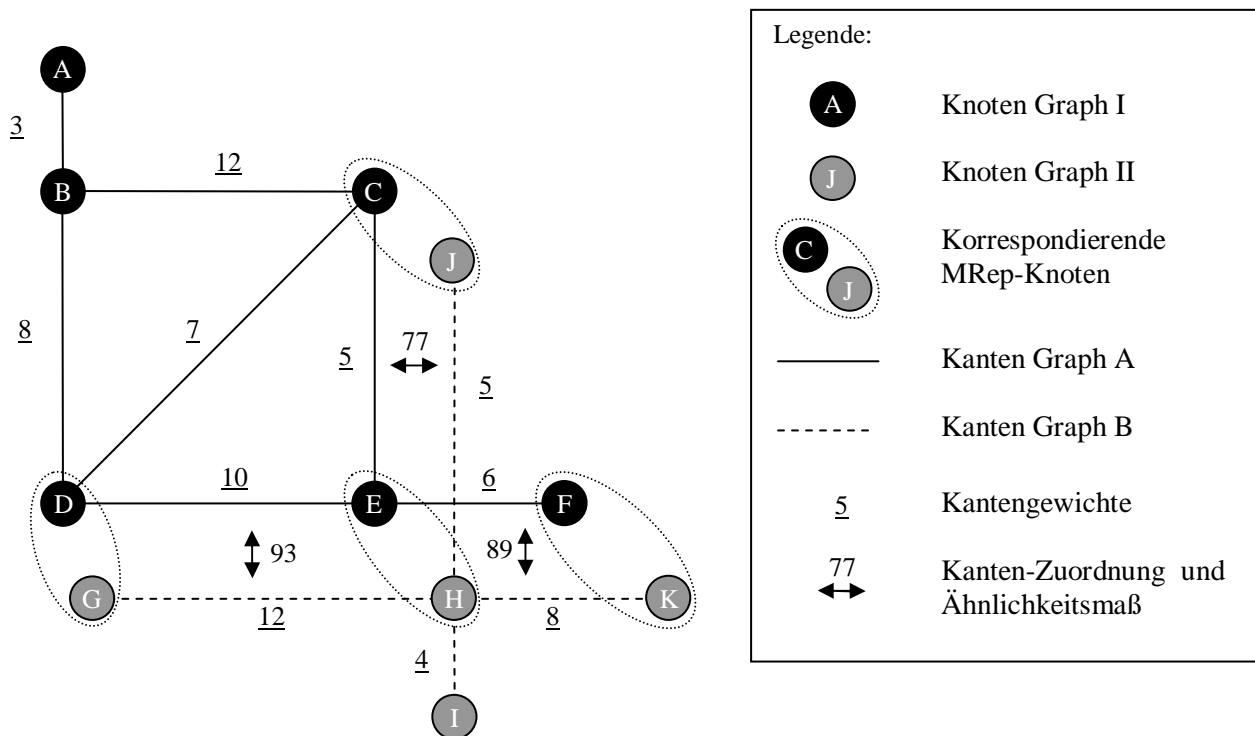


Abbildung 8-14: Mehrfach repräsentierte Graphen, die über MRep-Relationen miteinander verbunden sind

Um auf den dargestellten Daten eine Netzwerkanalyse ablaufen lassen zu können, sind verschiedene Strategien denkbar. Die Vorgehensweisen sollen anhand der Wegesuche von Startknoten A aus Graph I zum Zielknoten K aus Graph II analysiert werden.

Als erstes müssen die Adjazenzen aller Knoten bestimmt werden. Hier gibt es zwei denkbare Varianten:

1. Entweder können zwischen den MRep-Knoten, die die potentiellen Übergangsknoten von einem in den anderen Graphen darstellen, Übergangskanten (bzw. im Sinne von [Edwards & Simpson 02] (vgl. Kapitel 4.3.3.2) so genannte Konnektivitätsgeometrien) eingefügt werden, deren Länge entweder 0 ist oder sich aus der Auswertung des Ähnlichkeitsmaßes der Kantenzuordnung ergibt (vgl. Kapitel 8.2.4).
2. Oder die MRep-Knoten werden nur als ein einziger Knoten betrachtet und alle adjazenten Knoten beider MRep-Knoten werden addiert und gelten als deren jeweilige Nachbarknoten, wobei zu den adjazenten Knoten auch die korrespondierenden Knoten von adjazierenden MRep-Knoten zu zählen sind. So übernimmt im obigen Beispiel Knoten G neben der Adjazenz zu Knoten H alle adjazierenden Knoten von Knoten D. Da einer dieser adjazierenden Knoten, nämlich Knoten C, ein MRep-Knoten ist, ist G auch adjazent mit Knoten J aus dem eigenen Datensatz. Es ist zu bemerken, dass diese logische Zusammenfassung nicht einer Verschmelzung gleichkommt, da die Geometrien der Knoten erhalten bleiben.

Die beiden Varianten werden in der Folge dargestellt (siehe Abbildung 8-15):

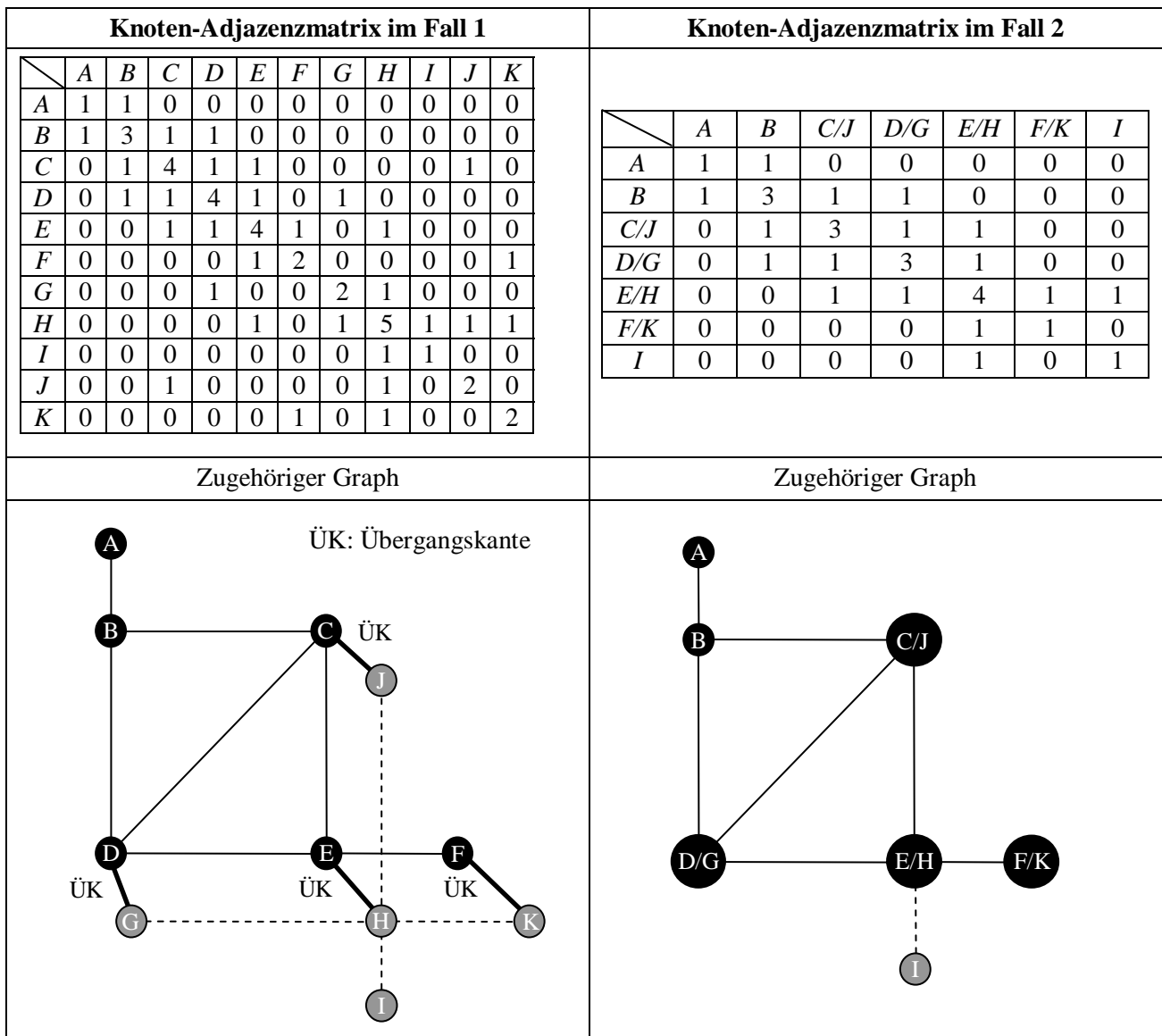


Abbildung 8-15: Die zwei möglichen Varianten zur Abbildung von Adjazenzen bei mehrfach repräsentierten Kanten

Im Rahmen dieser Arbeit wurde der erstgenannte Ansatz verfolgt, da hierbei die Übergänge von einem zum anderen Graphen explizit gewichtet werden können. Dies spielt bei der Bestimmung des kürzesten Weges eine wichtige Rolle, da nur so die Unsicherheiten, die beim Wechsel von einem Graphen in einen anderen entstehen, zu berücksichtigen sind.

8.2.2 Transformation des Navigationsgraphen in eine Baumstruktur

Algorithmen zur Suche von kürzesten Wegen können auf der Basis von Adjazenzmatrizen für gerichtete und ungerichtete Graphen ausgeführt werden. Beispiele hierfür sind so genannte „Single Source Shortest Path“ Algorithmen wie der Dijkstra-Algorithmus oder „All Pairs Shortest Path“ Algorithmen wie der Floyd-Algorithmus (siehe [Aho et al. 87]). In der Informatik werden die Lösungsräume jedoch häufig auch als Baumstrukturen repräsentiert. Ein Baum ist ein Sonderfall eines Graphen, der einen Wurzelknoten enthält, welcher den Eingangsgrad 0 hat, während alle anderen Knoten des Baums den Eingangsgrad 1 haben. Auf Baumstrukturen können verschiedene Suchstrategien operieren.

Im Rahmen des entwickelten Ansatzes werden Graphen, die aus mehrfach repräsentierten Daten gemäß der Variante 1 (aus dem vorigen Kapitel) abgeleitet wurden, in Bäume transformiert. Die grundsätzliche Vorgehensweise besteht darin, einen Wurzel- bzw. Startknoten zu definieren und von diesem aus den Baum aufzu-

bauen. Um Zyklen zu vermeiden, müssen für jeden Knoten im Baum seine Vorgänger bis zum Wurzelknoten gespeichert werden. Sollte ein Nachfolgeknoten identisch mit einem Vorgängerknoten sein, darf dieser Ast nicht erweitert werden. Um das Gesagte zu veranschaulichen, wird in Abbildung 8-16 der Aufbau eines Baums für den Graphen der Variante 1 aus dem vorhergehenden Kapitel illustriert. Startknoten sei Knoten A, Zielknoten sei Knoten K. Der Anschaulichkeit halber wird nur *ein* beispielhafter Weg dargestellt.

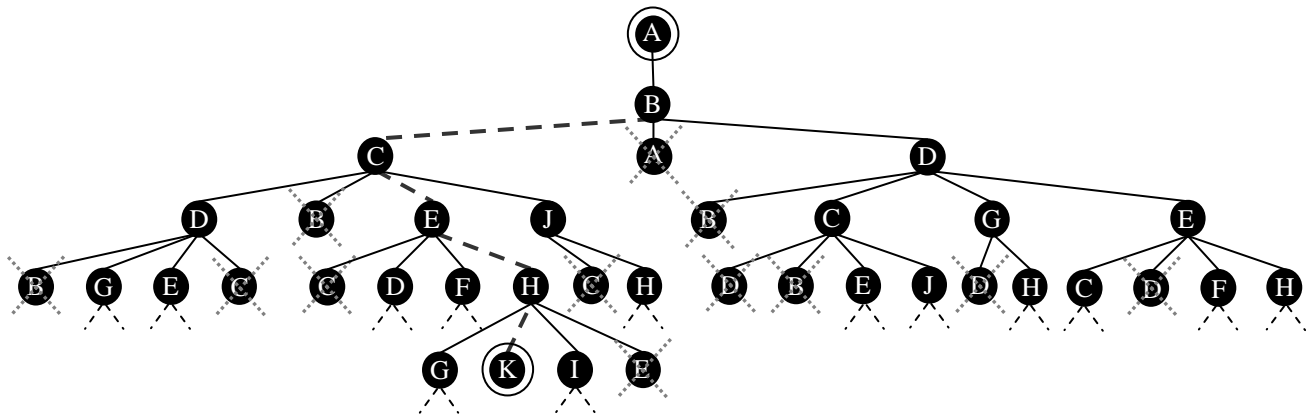


Abbildung 8-16: Teilausschnitt eines als Baum repräsentierten Navigationsgraphen mit Startknoten A und Zielknoten K.

8.2.3 Finden von kürzesten Wegen über eine Tiefensuche

Ist der als Adjazenzmatrix repräsentierte Navigationsgraph in eine Baumstruktur transformiert worden, so kann eine Baumsuche stattfinden. Hier existieren verschiedene Verfahren. Eine Übersicht zur Thematik geben [Aho et al. 87]. Im vorliegenden Ansatz wird eine Tiefensuche angewandt. Die Strategie gestaltet sich in diesem Fall wie folgt:

1. Bilde eine Liste L, die nur den Wurzelknoten enthält:
2. Wiederhole die folgenden Schritte, solange L nicht leer ist:
 - a. Untersuche, ob der erste Knoten in der Liste der Zielknoten ist:
 - i. Wenn ja, gib den Weg aus und entferne das erste Listenelement
 - ii. Wenn nicht, dann
 1. entferne das erste Listenelement von L;
 2. wenn das entfernte Listenelement Nachfolger hat, so füge jene Nachfolger in die Liste ein, die nicht bereits auf dem bisherigen Weg besucht wurden;
 3. speichere für alle Nachfolger den kompletten Weg vom Wurzelknoten zu sich selbst.

Für das obige Beispiel ergibt sich demgemäß die nachstehende Listenfolge:

1. {A (-)}
2. {B (A)}
3. {C (B, A), D (B, A)}
4. {E (C, B, A), D (C, B, A), J (C, B, A), D (B, A)}
5. {H (E, C, B, A), F (E, C, B, A), D (E, C, B, A), D (C, B, A), J (C, B, A), D (B, A)}
6. {**K (H, E, C, B, A)**, I (H, E, C, B, A), G (H, E, C, B, A), F (E, C, B, A), D (E, C, B, A), D (C, B, A), J (C, B, A), D (B, A)}
7. { I (H, E, C, B, A), G (H, E, C, B, A), ... } ...
- n. { - }

8.2.4 Berechnung der Wegekosten

Im Anschluss kann nun die Berechnung der Kosten K aller Wege W erfolgen. Sie kann sich bei der Bestimmung des kürzesten Pfades aus der Summe der Kosten K_r aller realen Kanten plus der Summe der Kosten K_{ii} aller Übergangskanten von einem Graphen in den anderen ergeben:

$$K(W) = \sum_{i=0}^j K_{ri} + \sum_{k=0}^l K_{ik}$$

Bei der Berechnung der Kosten der Übergangskanten muss beachtet werden, dass im hier verfolgten Ansatz lediglich die mehrfach repräsentierten Kanten über MRep-Relationen miteinander verknüpft sind, d.h. es existieren nur Ähnlichkeitsmaße für diese Kanten, jedoch nicht für deren Anfangs- und Endknoten (die MRep-Knoten). Bei der Bestimmung der Gewichtung (bzw. der Kosten) von Übergangskanten müssen aber genau diese Ähnlichkeitsmaße von mehrfach repräsentierten Knoten herangezogen werden. Auf diese Weise können Übergangskanten zwischen MRep-Knoten, die ein hohes Ähnlichkeitsmaß aufweisen, ein entsprechend niedriges Gewicht bekommen, während bei MRep-Knoten mit geringer Ähnlichkeit die zugehörige Übergangskante sehr hoch gewichtet wird, um damit die Unsicherheit des Übergangs von einem in den anderen Graphen auszudrücken.

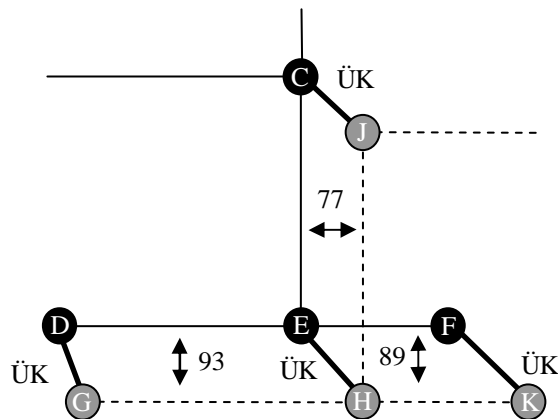


Abbildung 8-17: Übertragung des GMs von MRep-Kanten auf MRep-Knoten.

Es ist also notwendig, die Ähnlichkeitsmaße der MRep-Kanten an die MRep-Knoten weiterzugeben, um eine Bewertung der Übergangskanten zu ermöglichen. Dies ist für den Fall, dass MRep-Knoten (z.B. Knoten C und J aus Abbildung 8-17) lediglich jeweils eine inzidierende MRep-Kante haben, eindeutig: es wird einfach der Wert des Gesamtmaßes der Ähnlichkeit der zugehörigen Kante (also 77) als Maß für die Ähnlichkeit der jeweiligen MRep-Knoten verwendet. Hat aber ein MRep-Knoten mehrere inzidierende MRep-Kanten, so berechnet sich das Gesamtmaß der Ähnlichkeit zwischen zwei zueinander gehörenden MRep-Knoten aus dem arithmetischen Mittel der GMs *aller* an die MRep-Knoten angrenzenden, mehrfach repräsentierten Kanten (bei Übergangskante {EH} aus Abbildung 8-17 also aus dem arithmetischen Mittel von 77, 89 und 93 = 86,33). Diese Approximation erscheint als sinnvoll, da für die Ähnlichkeit mehrfach repräsentierter Knoten die Ähnlichkeit aller abgehenden MRep-Kanten ein vernünftiges Maß ist.

Aus dem Ähnlichkeitsmaß für die MRep-Knoten können nun über eine entsprechende Funktion die Kosten K_{ik} zur Traversierung der Übergangskante abgeleitet werden. Beispielsweise ist hierfür folgende Funktion denkbar:

$$K_{ik} = (GM_{\max} - GM_{MRepKn}) \cdot \omega$$

Dabei handelt es sich bei GM_{\max} um das maximale Gesamtmaß der Ähnlichkeit von MRep-Knoten (also 100) und bei GM_{MRepKn} um den tatsächlichen Ähnlichkeitswert der MRep-Knoten (im Beispiel von Übergangskante EH wäre $GM_{\max} - GM_{MRepKn}$ also 13,66). Über den Gewichtungsfaktor ω kann der Einfluß der Ähnlichkeit von MRep-Knoten bei der Kostenberechnung für die Übergangskante individuell angepasst werden.

Die Gewichtung der Übergangskanten erlaubt es, die Sicherheit des Wechsels von einem in den anderen Graphen und somit auch die Sicherheit des resultierenden Weges bei der Netzwerkanalyse berücksichtigen zu können. Somit werden einer Applikation zusätzliche Informationen zur Verfügung gestellt, die sie bei der Erzeugung einer Antwort auf Client-Anfragen einbeziehen kann. Ist beispielsweise eine hohe Sicherheit verlangt, so können Übergänge, deren Ähnlichkeitsindex unter einem gewissen Schwellwert liegt, potentiell

mit ∞ gewichtet werden, um diese Wege von vorn herein auszuschließen. Außerdem könnte auch die Anzahl der Übergänge von einem zum anderen Graphen bei der Wegesuche berücksichtigt werden, da mit jedem Wechsel die Unsicherheit des Weges zunimmt. Falls hingegen das Sicherheitskalkül ohne Belang ist, z.B. weil eine Anwendung die Möglichkeit vorsieht, sich zusätzliche Informationen über die Stimmigkeit des Weges (z.B. aus Mobilitätsmodellen) beschaffen zu können, können die Längen sämtlicher Übergangskanten auch 0 betragen. Die Bestimmung des kürzesten Weges folgt dann ausschließlich aus der Summe aller realen Kanten.

8.2.5 Optimierungsmöglichkeiten bei der Baumsuche

Die Berechnung von kürzesten Wegen über eine Tiefensuche kann zeitaufwendig sein. Daher ist es sinnvoll, eine Optimierung der Wegesuche zu realisieren. So können z.B. die Anforderungen einer Navigationsapplikation hinsichtlich der Verlässlichkeit des zu berechnenden kürzesten Weges als Eingabeinformationen bzw. Constraints verwendet werden, um den zeitlichen Aufwand bei der Tiefensuche zu verringern. U.a. könnte man das Zeitverhalten des Algorithmus steigern, indem Knoten von Übergangskanten, die ein gewisses Ähnlichkeitsmaß unterschreiten, bei der Baumsuche nicht mehr expandiert würden. Um eine Traversierung des gesamten Suchbaums bei der Tiefensuche zu vermeiden, sind auch heuristische Verfahren denkbar.

8.2.6 Realisierung des Ansatzes

Der vorgestellte Ansatz wurde in Form eines weiteren Moduls für das Java-GIS JUMP implementiert und anhand von ATKIS- und GDF-Straßendaten evaluiert [Volz 05b]. Das Datenmaterial hierfür war strukturell identisch mit jenem, das in Kapitel 7 vorgestellt wurde. Es wurde lediglich ein anderer geographischer Ausschnitt gewählt. Die Testdaten können dabei in drei Teilbereiche gegliedert werden:

1. Einen Teilbereich, in dem ausschließlich ATKIS-Daten vorliegen;
2. Einen Teilbereich, in dem ATKIS- und GDF-Daten räumlich überlappen und in dem folglich Mehrfachrepräsentationen vorhanden sind. In diesem Bereich werden MRep-Relationen gebildet;
3. Einen Teilbereich, in dem ausschließlich GDF-Daten vorliegen.

Aus den zur Verfügung stehenden Daten kann mittels der Software in einem ersten Schritt *eine* Datenstruktur gebildet werden, die die Adjazenzen sämtlicher Knoten, d.h. der Knoten der beiden Ausgangsdatenbestände und der Übergangsknoten, die aus den MRep-Relationen abzuleiten sind, enthält. Dabei laufen die nachstehend beschriebenen Prozesse ab.

Da die Ausgangsdaten aufgrund des verwendeten Datenformates nicht topologisch strukturiert vorliegen, müssen zunächst sowohl für den ATKIS- als auch für den GDF-Datensatz Adjazenzbeziehungen aufgebaut werden. Zusätzlich werden für alle aus den MRep-Relationen extrahierten MRep-Kanten die zugehörigen Übergangs- bzw. MRep-Knoten und deren Adjazenzbeziehungen zu korrespondierenden MRep-Knoten gebildet. Hier werden auch die Ähnlichkeitsmaße der Kanten an die Knoten übergeben. Pro Repräsentation existieren stets zwei Übergangsknoten, und zwar jeweils der Anfangs- und Endpunkt der Repräsentationsgeometrien. Der letzte Arbeitsabschnitt besteht darin, die drei verschiedenen Adjazenzrepräsentationen zu einer einzigen zu vereinen. Dabei werden die Adjazenzen der Übergangsknoten an jene Knoten in den Ausgangsbeständen übertragen, die identisch mit den Übergangsknoten sind (vgl. Abbildung 8-15). Die Zuordnung erfolgt jeweils über deren Position.

Sobald die Adjazenzstruktur erzeugt ist, kann der Benutzer über die Applikation Start- und Endpunkt der Wegesuche spezifizieren und auswählen, wie hoch die Verlässlichkeit des Weges sein soll. Dabei gibt es die drei Kategorien „Hohe Verlässlichkeit“, „Normale Verlässlichkeit“ und „Geringe Verlässlichkeit“. Anschließend wird durch Drücken des „Shortest Path“-Knopfes die Adjazenzstruktur in eine Baumstruktur transformiert, die Tiefensuche aktiviert und schließlich das Ergebnis in der Karte visualisiert. Fordert der Anwender eine hohe Verlässlichkeit des resultierenden Weges, so werden alle Übergänge (bzw. MRep-Knoten), deren GM unter 80 liegt, nicht berücksichtigt bzw. mit ∞ gewichtet. Ist das Verlässlichkeitsbedürfnis des Benutzers normal, so werden die Übergangskanten mit geringerer Ähnlichkeit als 50 mit dem Wert unendlich belegt. Spielt die Verlässlichkeit hingegen eine untergeordnete Rolle, werden alle Übergangskanten zugelassen und mit 0 gewichtet. Mittels der Applikation kann also aufgezeigt werden, inwiefern Wege

bei unterschiedlicher Bewertung der Übergangskanten differieren. Abbildung 8-18 zeigt eine Darstellung der Navigationsanwendung, des so genannten MRepNetworkAnalysis-Tools.

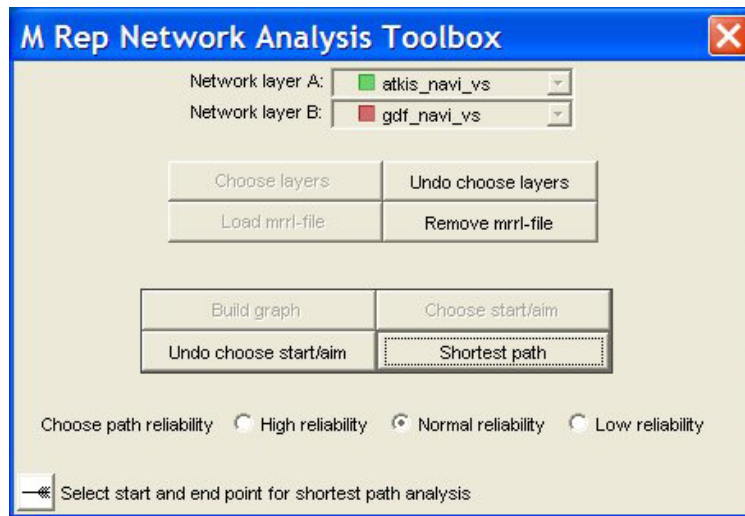


Abbildung 8-18: Das MRepNetworkAnalysis-Tool

8.2.7 Diskussion der Ergebnisse

An dieser Stelle sollen einige Probleme zur Sprache kommen, die bei dem entwickelten Verfahren zu beachten sind. Dies geschieht anhand der Darstellungen in Abbildung 8-19. Hier wird eine Netzwerkanalyse auf der Basis des entwickelten Verfahrens präsentiert. Darstellung (1) zeigt die Ausgangsdaten, (2) zeigt den kürzesten Weg unter der Bedingung einer hohen Verlässlichkeit, (3) zeigt schließlich den kürzesten Weg für eine Anfrage mit geringen Sicherheitsbedürfnissen.

Zunächst ist die Problematik der Kanten im Grenzbereich zu nennen. Alle Kanten am Rand eines Netzwerkes erscheinen als Sackgassen (siehe in (1)). Tatsächlich können sie aber doch Verbindungen haben, die jedoch nicht mehr im Datensatz bzw. in der Augmented Area vorhanden sind. Bei der Bildung von MRep-Relationen erhalten die Kanten, die fälschlicherweise als Sackgassen repräsentiert sind, aufgrund der unterschiedlichen topologischen Gegebenheiten ein niedrigeres Ähnlichkeitsmaß, als wenn die Kanten nicht am Rand lägen (z.B. die Kanten b und c in (1)), d.h. es treten Probleme auf, die bei der Wegesuche zu inkorrekten Ergebnissen führen können. Die Problematik lässt sich allerdings kaum umgehen.

Wurden zwei Kanten zugeordnet, von denen eine Kante eine (echte) Sackgasse darstellt, so kann dies bei der Navigation ebenfalls problematisch sein. So sind z.B. die beiden Kanten d zwischen den Knoten 1 und 6 in (1) zwar zugeordnet worden, doch stellt lediglich die GDF-Repräsentation eine Sackgasse dar, während die ATKIS-Repräsentation durchgängig erfasst wurde. Bei der Lösung der Navigationsanfrage, die in (3) dargestellt wird, wurde das betreffende Road Element aus GDF aufgrund der Zuordnung zur ATKIS-Straße als durchgängig interpretiert. Der Benutzer wird daher evtl. in eine Sackgasse gelotst, und zwar in Abhängigkeit davon, ob die durchgängig erfasste Straße oder die Sackgasse die Situation in der Realwelt korrekt widerspiegelt.

Der kürzeste Weg unter der Bedingung einer hohen Verlässlichkeit ist in (2) dargestellt. Der Übergang findet hier am MRep-Knoten 3 statt, da das arithmetische Mittel des Gesamt-Ähnlichkeitsmaßes der angrenzenden MRep-Kanten von MRep-Knoten 3 > 80 ist, wogegen es bei allen anderen MRep-Knoten, also 1, 2, 4, 5 und 6, unter 80 liegt und diese folglich der Anforderung einer hohen Verlässlichkeit nicht genügen. Wie zu erkennen ist, entsteht hier offensichtlich ein deutlicher Umweg, da zuerst zum möglichen Übergangsknoten navigiert werden muss, um den Wechsel realisieren zu können. Dadurch wird Kante a doppelt durchlaufen, und zwar einmal in der GDF- und einmal in der ATKIS-Repräsentation.

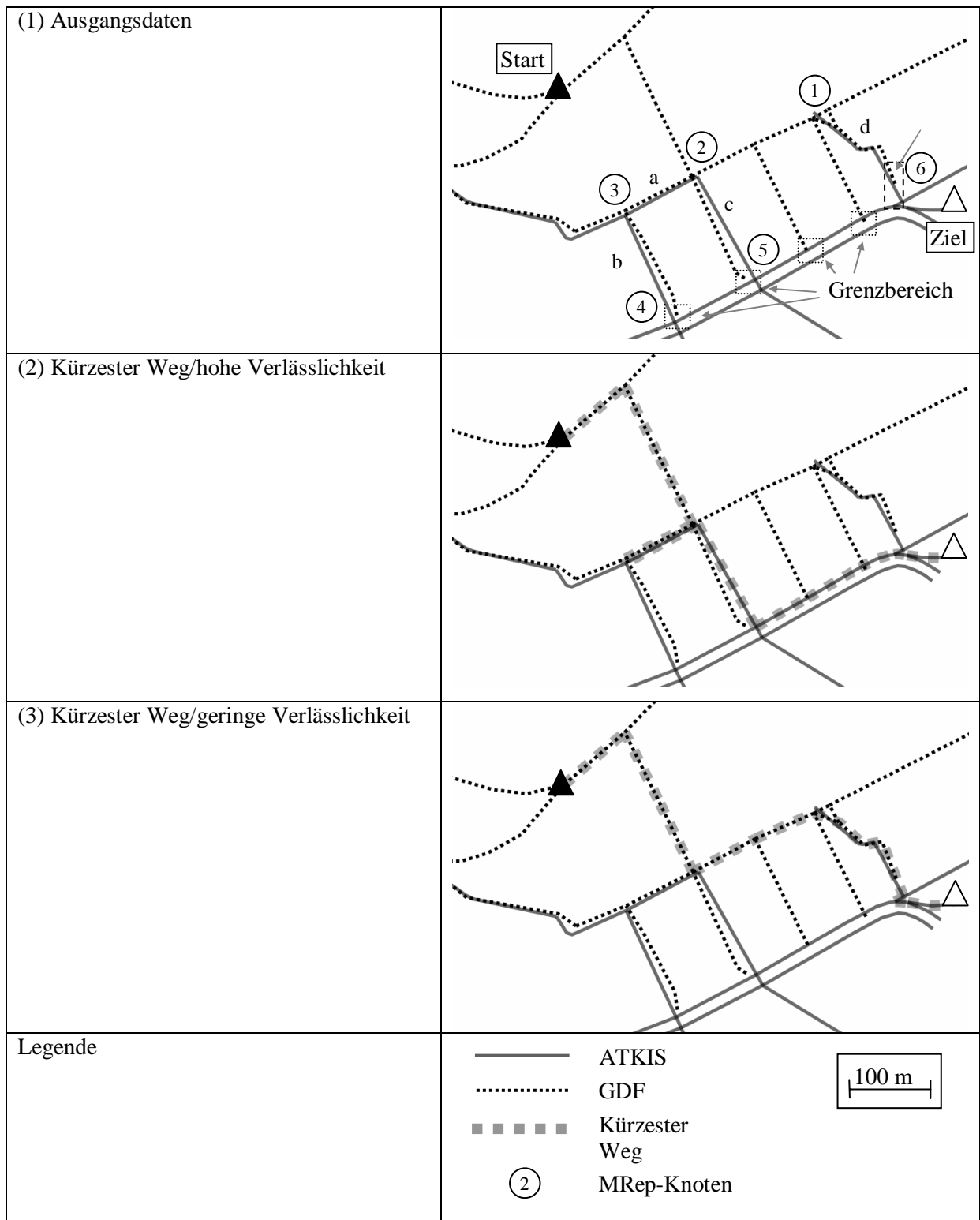


Abbildung 8-19: Darstellung von Ergebnissen einer Wegesuche mit unterschiedlichen Verlässlichkeitsstufen

Aus dem Vergleich von verlässlichstem (2) und kürzestem (3) Weg wird folglich eine deutliche Diskrepanz der Streckenlänge ersichtlich. Eine Navigationsanwendung könnte in solchen Fällen evtl. den kürzesten Weg anfordern und dessen Richtigkeit nochmals über zusätzliche Informationen überprüfen, anstatt den langen, sicheren Weg zu wählen. Solche zusätzlichen Informationen könnten im Fall von Nexus darin bestehen, existierende Mobilitätsmodelle heranzuziehen. Prinzipiell wäre es auch möglich, solche Informationen schon in die Ähnlichkeitsbetrachtungen von Mehrfachrepräsentationen einzubeziehen. Sind sicherster und unsi-

cherster Weg in Bezug auf das Optimierungskriterium wenig unterschiedlich, wird eine Anwendung natürlich immer den Weg wählen, der das höchste Maß an Sicherheit garantiert.

Je nachdem, wie gut die Ähnlichkeitsmaße bei der Zuordnung von zwei Datensätzen sind, mag es auch Bereiche geben, in denen keine Wege mit hoher Verlässlichkeit zur Verfügung gestellt werden können. Die Anwendung muss dann entscheiden, ob sie dennoch einen Weg an den Klienten weiterleitet und ihn darauf hinweist, dass dieser mit einer gewissen Unsicherheit behaftet ist, oder ob sie kein Ergebnis zurückliefert.

Gerade bei der Netzwerkanalyse ist die Richtigkeit bzw. Verlässlichkeit topologischer Informationen von großer Bedeutung. Als Lösungsansatz kann man sich daher vorstellen, für die Zwecke der Netzwerkanalyse alle MRep-Relationen, deren topologisches Ähnlichkeitsmaß unter einem bestimmten Schwellwert liegt, nicht zu berücksichtigen, d.h. die zugehörigen Repräsentationen würden quasi als nicht zugeordnet betrachtet. Aus dieser Vorgehensweise kann eine generelle Schlussfolgerung gezogen werden: die Ähnlichkeitsmaße müssen anwendungsabhängig interpretiert werden können, um eine bestmögliche Eignung für die jeweiligen Prozesse zu gewährleisten.

Insgesamt hat der vorgestellte Ansatz gezeigt, dass es möglich ist, eine Netzwerkanalyse auf mehrfach repräsentierten Daten unter Nutzung von MRep-Relationen durchzuführen. Daraus folgt die Erkenntnis, dass es nicht unbedingt notwendig ist, mehrfach repräsentierte Datenbestände zum Zwecke der Datenanalyse komplett miteinander zu verschmelzen. Dies ist insbesondere für offene Dateninfrastrukturen wie Nexus von Bedeutung. Dadurch kann auf Föderationsebene bei der Beantwortung von Anwendungsanfragen ein Schritt eingespart werden. Zusätzlich ergibt sich der Vorteil, dass man Anwendungen Informationen über die Verlässlichkeit eines Weges mitgeben kann und diese dann entscheiden können, wie sie mit diesen Informationen umgehen.

Was an dieser Stelle für die Netzwerkanalyse gezeigt wurde, ist in zukünftigen Arbeiten aber auch für andere Analysefunktionalitäten in GIS nachzuweisen, d.h. bestehende GIS-Operatoren sollen in der Zukunft nicht nur auf Daten arbeiten können, die sozusagen in einer Mono-Repräsentation vorliegen, sondern es muss auch eine Analyse auf mehrfach repräsentierten Daten möglich sein. Die Nutzung von MRep-Relationen ist ein möglicher Ansatz, wie eine solche Analyse realisiert werden kann.

Außerdem wäre es sinnvoll, Ergebnisse von Analysen auf verschmolzenen und auf mehrfach repräsentierten Daten miteinander zu vergleichen und die sich ergebenden Unterschiede zu untersuchen. Daraus können Hinweise bezüglich der Unsicherheit von Analyseergebnissen abgeleitet und es kann evtl. entschieden werden, welche Form der Datenanalyse auf Mehrfachrepräsentationen unter bestimmten applikationsseitigen Anforderungen vorzuziehen ist.

9 Zusammenfassung und Diskussion der Ergebnisse

Die Thematik der Mehrfachrepräsentationen spielt eine sehr wichtige Rolle auf dem Weg zur Verwirklichung von Interoperabilität in Geo-Informationssystemen. Wie gezeigt wurde, werden derzeit bereits viele Anstrengungen unternommen, um die technische Interoperabilität durch Definition von Schnittstellen, Funktionen, Austauschsprachen, etc. in den Griff zu bekommen. Auch das Problemfeld der Integration unterschiedlicher Semantik, zu dem die Thematik der Mehrfachrepräsentationen zu zählen ist, wurde bei Standardisierungsorganisationen erkannt, doch wurden diesbezüglich nur erste Ansätze verfolgt. Das liegt daran, dass es äußerst schwierig ist, die verschiedenen Auffassungen von GIS-Anwendern über dieselben Objekte der Realwelt in Einklang zu bringen, da die Sichtweisen applikationsbedingt stark differieren. Als Erkenntnis der dargelegten Betrachtungen folgt, dass Standardisierungen bei der Problematik der Mehrfachrepräsentationen – zumindest bislang – nur in bedingtem Maße wirksam sein können.

Die Problematik der Mehrfachrepräsentationen wird seit einigen Jahren intensiv aus Forschungssicht bearbeitet. Die Betrachtung geschieht dabei aus unterschiedlichen Fachdisziplinen heraus. Traditionell wird die Problematik der Informationsintegration aus Datenbanksicht untersucht, wobei vor allem Fragen der Schemaintegration und des Schema-Matchings sowie der Verwaltung von Mehrfachrepräsentationen in Datenbanken relevant sind. Aus dem GIS-Bereich gab es zahlreiche Vorschläge, das Problem der unterschiedlichen Semantik von Geodaten über die Einführung von Ontologien anzugehen. Wichtige Forschungsarbeit wurde auch auf dem Gebiet der Zuordnung und der Verschmelzung von korrespondierenden Instanzen geleistet, wogegen eine gemeinsame Fortführung und vor allen Dingen die Analyse mehrfach repräsentierter Daten bislang kaum untersucht wurden.

Als grundsätzliche Zielsetzung der vorliegenden Arbeit gilt es, einen Beitrag zur Forschung auf dem Gebiet der Mehrfachkonzeptionen und -repräsentationen zu leisten. Die Problematik wurde hierbei hauptsächlich im Kontext offener Dateninfrastrukturen und insbesondere aus dem Blickwinkel des Nexus-Projektes untersucht, das eine Plattform für orts- und kontextbezogene Dienste entwickelt.

Zunächst wurde analysiert, wie aus bestehenden Schemas für Geodaten (GDF, ALK und ATKIS) ein globales Schema für eine offene Dateninfrastruktur bzw. im engeren Sinne ein Schema für die Nexus-Plattform (AWS) abgeleitet werden kann. Die Untersuchung war dabei gezielt auf einen Teilausschnitt des Geodatenpektrums, den Bereich der Straßenverkehrsdaten, begrenzt worden. Es konnte gezeigt werden, dass zur Generierung eines globalen Schemas zwei vorbereitende Schritte nötig sind:

1. eine intensive Untersuchung der existierenden Schemas und der entsprechenden Modellierungskonzepte, die diesen Schemas zugrunde liegen, mit dem vorrangigen Ziel, Korrespondenzen zwischen den Schemas aufzudecken;
2. eine Analyse der Anforderungen, denen das globale Schema entsprechen muss.

Neben diesen Gesichtspunkten mussten bei der Schemaintegration auch die speziellen Bedingungen, denen das Zielschema, bzw. im vorliegenden Fall das Nexus-Schema, unterworfen ist, berücksichtigt werden. Schließlich konnte in einem rein manuellen Arbeitsschritt das Schema zur Repräsentation von Straßendaten in Nexus erzeugt werden.

Wie sich herausstellte, handelt es sich bei dieser Vorgehensweise zur Integration bestehender Schemas in ein übergeordnetes Schema um einen wissensintensiven und damit zeitaufwendigen Prozess. Potential zur Automatisierung weist dabei nur ein Teilschritt, die Ableitung von Schema-Korrespondenzen, auf. Es existieren bereits einige Ansätze zu diesem als Schema-Matching bezeichneten Vorgang (vgl. Kapitel 4.3.1.1). Im Rahmen der vorliegenden Arbeit wurde diese Thematik ebenfalls untersucht (siehe unten).

Ist ein globales Schema einmal abgeleitet, so können Abbildungsregeln definiert werden, die es erlauben, Datenbestände bestehender Schemas dem globalen Schema gemäß zu repräsentieren. In diesem Arbeitsschritt ist wiederum die Kenntnis der Ausgangsdatenbestände von Bedeutung, um möglichst wenig Informationsverlust bei der Übertragung hinnehmen zu müssen. Eine gemeinsame Verarbeitung von raumbezogenen

Daten innerhalb der Nexus-Plattform setzt voraus, dass die Geometrietypen der Ausgangsdaten angepasst werden. Im betrachteten Fall waren aus ATKIS und GDF linienförmige und aus der ALK flächenhafte Objekte vorhanden. Prinzipiell sind nun zwei Ansätze denkbar: entweder, die linienförmigen Objekte werden anhand eines Attributes, das die Straßenbreite angibt, in Flächen umgewandelt, oder die Fläche wird linearisiert. Für Nexus-Anwendungen müssen die Daten zumeist als Linien vorhanden sein, weshalb über eine Konvertierung ein rasterbasiertes Verfahren zur Erzeugung von Linien aus den ALK-Flächen angewendet wurde. Damit konnte ein mögliches Verfahren aufgezeigt werden, das jedoch in komplexen Situationen zu Problemen führen kann. Hier müssen daher in Zukunft Generalisierungsmethoden eingesetzt werden.

Die Grundidee der vorliegenden Arbeit besteht darin, Mehrfachrepräsentationen über explizite Relationen miteinander zu verbinden und somit eine neue Form der Modellierung von Geodaten einzuführen. Zu diesem Zweck wurde ein formales Schema in Form der Objektklasse *MultiRepresentationalRelation* erarbeitet. Mittels dieses Schemas konnte gezeigt werden, welche Informationen zur Beschreibung von Beziehungen zwischen mehrfach repräsentierten Objekten wichtig sind. Die Klasse besteht aus verschiedenen Attributen, die in erster Linie Angaben über die Ähnlichkeit von Mehrfachrepräsentationen zulassen. Es konnten hier verschiedene Maße vorgestellt werden, die insbesondere für die Feststellung der Ähnlichkeit raumbezogener Daten verwendet werden können. Instanzen von *MultiRepresentationalRelation* werden im Rahmen der Arbeit als MRep-Relationen bezeichnet. Die Klasse konnte in das Nexus Augmented World Schema eingegliedert werden, doch ist die Realisierung von MRep-Relationen nicht an Nexus gebunden und kann folglich auch für die Abspeicherung von Beziehungen zwischen Mehrfachrepräsentationen außerhalb der Nexus-Plattform verwendet werden.

Die Integration des Konzeptes der expliziten Modellierung von Relationen zwischen Mehrfachrepräsentationen in die Nexus-Plattform macht diese sozusagen „MRep-fähig“. Mehrfachrepräsentationen können von der Plattform bereits dann wahrgenommen werden, wenn Daten in die Infrastruktur eingebracht werden. Auf der Basis der MRep-Relationen kann eine Methodik zur gemeinsamen Aktualisierung mehrfach repräsentierter Objekte in Nexus aufgesetzt werden. Zum Zwecke der Datenanalyse können über MRep-Relationen verknüpfte Objekte zu einer konsistenten Sicht vereinigt werden. Entweder wird hier eine Verschmelzung durchgeführt oder die mehrfach repräsentierten Daten werden unter Auswertung der MRep-Relationen analysiert. Einschränkend muss jedoch gesagt werden, dass die Verwaltung von Relationen zwischen Mehrfachrepräsentationen in Nexus einen Mechanismus voraussetzt, um auch die MRep-Relationen konsistent zu halten. Sobald eine Repräsentation eine Änderung erfährt, die in der korrespondierenden Repräsentation nicht in exakt identischem Maße nachgeführt werden kann, ändert sich nämlich die MRep-Relation, die zwischen den beiden Repräsentationen existiert. Der Verwaltungsaufwand, der dabei anfällt, kann die Effizienz der Plattform beeinträchtigen.

Das Konzept der MRep-Relationen wurde anhand von Testdaten umgesetzt. Bei den Testdaten handelte es sich um Straßenverkehrsdaten aus ATKIS und GDF, d.h. die Daten wurden nicht ins Nexus-Schema transformiert, sondern blieben in den Ausgangsschemas erhalten. Für die verfügbaren Daten wurden MRep-Relationen erzeugt, wobei geometrische und topologische Ähnlichkeitsmaße abgeleitet wurden, aus denen ein Gesamtmaß für die Ähnlichkeit zweier Repräsentationen errechnet wurde. Für die Erzeugung von MRep-Relationen wurde ein semi-automatisches Zuordnungstool entwickelt, das es ermöglicht, während des Zuordnungsprozesses die persönliche Einschätzung des menschlichen Operateurs hinsichtlich der Wahrscheinlichkeit der Zuordnung angeben zu können. Ferner wurde im Rahmen dieser Arbeit ein XML-basiertes Format zum Austausch von MRep-Relationen, die so genannte *MultiRepresentational Relation Language* (MRRL) definiert. Um den manuellen Zuordnungsprozess so weit wie möglich zu vereinheitlichen, wurden dem menschlichen Operateur Vorgaben für die Zuordnung gemacht. Bei der der Bildung von MRep-Relationen für die Testdaten erkannte man jedoch, dass sehr vielfältige Zuordnungsfälle auftreten können, die nicht alle kategorisiert und in Form von Zuordnungsrichtlinien regulierbar sind. Dabei handelt es sich jedoch – zumindest in den untersuchten Testgebieten – um Einzelfälle, die die Gesamtuntersuchung nicht beeinflussen. Dennoch ist zu erkennen, dass bei der Zuordnung durch den Menschen ein gewisses Maß an Subjektivität vorhanden ist und somit eine Reproduzierbarkeit eines Zuordnungsergebnisses nicht komplett gewährleistet werden kann.

Bei einer Untersuchung der Verteilung des Gesamt-Ähnlichkeitsmaßes in den Testdaten konnte festgestellt werden, dass die meisten Mehrfachrepräsentationen einen sehr hohen Ähnlichkeitswert aufwiesen. Die

weitgehende Übereinstimmung der zugeordneten Repräsentationen wurde auch durch die menschliche Einschätzung bestätigt. Schließlich konnte durch einen Vergleich von automatisch generiertem Ähnlichkeitsmaß und menschlicher Einschätzung aufgezeigt werden, dass beide Bewertungen korrelieren. Betrachtet man die menschliche Zuordnung als optimal, so kann das in dieser Untersuchung verwendete Gesamt-Ähnlichkeitsmaß folglich als nützlicher Indikator für die Beschreibung der Ähnlichkeit von Mehrfachrepräsentationen angesehen werden.

Die erzeugten MRep-Relationen wurden in der Folge für zwei weitere, im Rahmen dieser Arbeit untersuchte Ansätze herangezogen. In einem ersten Ansatz wurde die Idee verfolgt, über eine Auswertung von Relationen zwischen Mehrfachrepräsentationen Rückschlüsse auf die Korrespondenzen zwischen verschiedenen Schemas ziehen zu können. Diese datengetriebene, automatisierte Form des Schema-Matchings wurde bislang auf dem Gebiet der Geoinformatik nicht verfolgt. Der Ausgangspunkt des Ansatzes besteht darin, dass die Klassenzugehörigkeiten der über MRep-Relationen zugeordneten Instanzen untersucht werden. Dabei ist zu klären, ob es zu einem signifikanten Zusammenhang in der Form kommt, dass Instanzen einer bestimmten Klasse des einen Schemas weitgehend auf die Instanzen der Klasse eines anderen Schemas abgebildet werden. Im Rahmen der Arbeit konnte gezeigt werden, dass anhand dieser Vorgehensweise tatsächlich deutliche Korrespondenzen zwischen Objektklassen zur Repräsentation von Straßendaten in ATKIS und GDF ableitbar sind. Der Ansatz erlaubt es nicht nur, eindeutige Klassenzuordnungen der Kardinalität 1:1 auszuwerten, sondern es können auch Klassenkombinationen der Kardinalität 1:n und n:m berücksichtigt werden. Aus der Untersuchung ergeben sich letztlich statistische Korrelationsmaße für die betrachteten Objektklassen, die als prozentuales Maß angegeben werden. So zeigte sich beispielsweise für die Objektklasse Weg aus ATKIS eine vollständige Abbildung auf die Objektklasse Road Element aus GDF. Ebenfalls wurden ATKIS-Straßen zu nahezu 95% auf Road Elements abgebildet. In umgekehrter Richtung war u.a. zu erkennen, dass 90% der Road-Objekte aus GDF Fahrbahn-Objekten aus ATKIS entsprachen.

In einer Erweiterung der Vorgehensweise wurden nur diejenigen MRep-Relationen betrachtet, deren Gesamt-Ähnlichkeitsmaß größer als 80 und damit sehr hoch war. Die Korrelationswerte für die Klassen konnten dadurch nochmals gesteigert werden, d.h. es konnte gezeigt werden, dass die Korrespondenzen auf Objektklassenebene umso deutlicher ausfallen, je deutlicher auch die Zuordnungen auf Objektebene erfolgen. Analog zu der Vorgehensweise auf Objektklassenebene wurde der Ansatz auch auf Attribute angewandt. Hier wurde die Untersuchung nur anhand jeweils eines Beispielattributes aus ATKIS und GDF durchgeführt, für die aus der Kenntnis der Schemas heraus Korrelationen als wahrscheinlich galten. Diese konnten mittels des automatischen Verfahrens ebenfalls nachgewiesen werden, auch wenn die Daten mit einem gewissen Rauschen behaftet waren. Somit konnte die grundsätzliche Anwendbarkeit der Methode auf Attributebene ebenfalls unter Beweis gestellt werden.

Das im Rahmen der Arbeit entwickelte Verfahren kann also verwendet werden, um Ähnlichkeiten von Schemas gänzlich unbekannter Semantik aufzudecken, d.h. es erfordert keinerlei Kenntnis über ein Schema und kann aus diesem Grund vollautomatisch ablaufen. Man könnte sich z.B. vorstellen, dass sämtliche Schemabezeichnungen ausschließlich in einer fremden Sprache verfügbar seien und das Verfahren dennoch die richtigen Zusammenhänge entdecken würde. Als Voraussetzung gilt jedoch, dass die betrachteten Schemas ausreichend mit Daten gefüllt sind, um eine repräsentative Untersuchung der Korrespondenzen zu erhalten.

Für die Nexus-Plattform ist dieser Ansatz praktikabel, wenn aus bestehenden Schemas verschiedener Anwendungsbereiche generische Objektklassen für das Augmented World Schema abgeleitet oder wenn verschiedene Erweiterte Klassenschemas integriert werden sollen. Allerdings ist mit dem bestehenden Verfahren lediglich ein erster Teilschritt getan, da eine komplette Integration der Formulierung von Integrationsregeln bedarf. Diese konnten bislang noch nicht realisiert werden.

Ein weiterer Aspekt der Arbeit liegt darin, erste Betrachtungen für die Analyse auf Mehrfachrepräsentationen anzustellen. Hierzu wurde ein Anwendungsfall, die Netzwerkanalyse in mehrfach repräsentierten Graphen, untersucht. Dabei sollte keine Verschmelzung der mehrfach repräsentierten Daten erfolgen. Vielmehr wurde angestrebt, die Mehrfachrepräsentationen getrennt zu belassen und bei der Analyse lediglich die MRep-Relationen zwischen ihnen auszuwerten. Aus der Untersuchung geht hervor, dass die Problematik zunächst darin besteht, eine Datenstruktur für einen Navigationsgraphen aufzubauen, in der auch die Über-

gänge von einem Graphen zum anderen gespeichert werden können. Hierzu wurden zwei Lösungsvarianten angeboten: eine logische Verschmelzung von mehrfach repräsentierten Knoten oder die Einführung von Adjazenzbeziehungen zwischen korrespondierenden Knoten verschiedener Graphen. Für die Implementierung wurde letztere Variante gewählt, da hier die Übergänge von einem Graphen in den anderen basierend auf dem Ähnlichkeitsgrad zweier mehrfach repräsentierter Knoten explizit gewichtet werden können. Zum Auffinden des kürzesten Weges aus der erzeugten Datenstruktur heraus wird diese bei Angabe von Start- und Zielpunkten in eine Baumstruktur transformiert, die über eine Tiefensuche durchlaufen werden kann.

Mittels des entwickelten Verfahrens wird es möglich, die Unsicherheiten bei der Analyse von Mehrfachrepräsentationen ausdrücken und an die Anwendungen weiterleiten zu können. Im Falle der Netzwerkanalyse strebt eine Applikation natürlich stets die verlässlichste Lösung an. Allerdings kann es sein, dass eine verlässliche Lösung für den Anwender inakzeptabel ist, weil beispielsweise der sichere Weg viel zu lang ist. Zudem ist es möglich, dass überhaupt kein Weg existiert, der den Sicherheitsanforderungen der Anwendung genügt. In diesen Fällen hat die Anwendung zumindest die Möglichkeit zu bewerten, ob sie Alternativen entwickeln und die Ergebnisse der unsicheren Analyse dennoch verwenden kann.

Das Konzept zur Navigation auf einer MRep-Datenstruktur hat gezeigt, dass eine Verschmelzung von mehrfach repräsentierten Daten bei Analysen in Nexus umgangen werden kann. Diese Erkenntnis bringt eine Effizienzsteigerung für die Plattform mit sich, da dadurch ein Schritt bei der Datenaufbereitung entfällt. Allerdings ist zu bemerken, dass sich durch die Auswertung von MRep-Relationen im Zuge eines Analyseprozesses auch ein erhöhter Aufwand bei der Datenanalyse ergibt. Dieser fällt aber im Vergleich zur Verschmelzung von Daten weniger schwer ins Gewicht.

10 Ausblick

Im Rahmen der vorliegenden Arbeit wurde ein automatisches Verfahren zur Erkennung von Schema-Korrespondenzen auf der Basis des Konzeptes der Modellierung von Relationen zwischen mehrfach repräsentierten Objekten entwickelt. Dabei handelt es sich aber lediglich um ein Schema-Matching, eine komplette, automatisierte Integration ähnlicher Schemas ist bislang noch nicht zu erreichen. Die Ergebnisse des Verfahrens müssen daher in weiteren Schritten bei der Formulierung von Integrationsregeln zur Bildung eines globalen Schemas herangezogen werden. Es ist zu prüfen, inwieweit hier ebenfalls Automatisierungen möglich sind, so dass in Zukunft die manuelle Komponente bei der Generierung von globalen Schemas zum Zwecke der Föderation von heterogenen Datenbanken weiter minimiert werden kann. In zukünftigen Forschungsansätzen ist auch zu evaluieren, inwiefern die erzeugten Korrelationsmaße auf Objektklassenebene für andere Verfahren genutzt werden können. Vielversprechend sind hier jene Konzepte einzuschätzen, die die Korrelationsmaße innerhalb von Zuordnungsprozessen als Vorinformationen auswerten, um Optimierungen bei der Verlässlichkeit der Zuordnungsergebnisse zu erzielen.

Ein weiterer Ansatz dieser Arbeit hat gezeigt, dass eine Analyse mehrfach repräsentierter Geodaten nicht unbedingt auf einem Datenbestand ausgeführt werden muss, in dem jedes Realweltobjekt nur einmal vertreten ist. Vielmehr kann eine Datenanalyse auch auf der Basis einer MRep-Datenstruktur erfolgen, indem die Relationen zwischen mehrfach repräsentierten Objekten ausgewertet werden. Bislang wurde die Vorgehensweise exemplarisch für die Netzwerkanalyse aufgezeigt. In weiteren Schritten ist es anzustreben, auch andere GIS-Operatoren an diese Art der Analyse anzupassen. Im Zuge dessen gilt es auch, die Anforderungen unterschiedlicher Analysefunktionalitäten an die Bestimmung des Ähnlichkeitsmaßes für MRep-Relationen zu untersuchen. Es kann ferner von Interesse sein, die Ergebnisse von Analysen auf Datenstrukturen, die durch Verschmelzung von Mehrfachrepräsentationen erzeugt wurden, mit den Resultaten auf MRep-strukturierten Daten zu vergleichen.

Das hier vorgeschlagene Konzept der expliziten Modellierung von Relationen zwischen korrespondierenden Repräsentationen kann darüber hinaus für weitere Anwendungsbereiche genutzt werden. So können auf der Basis von MRep-Relationen Mechanismen für die Datenfortführung und damit die Konsistenzwahrung von Daten angestoßen werden. Des Weiteren können die Relationen als Ausgangspunkt für Prozesse dienen, die eine Verschmelzung von Mehrfachrepräsentationen beabsichtigen.

Aus dem Gesagten heraus ist nachzuvollziehen, dass die Problematik der Behandlung von Mehrfachrepräsentationen und Mehrfachkonzeptionen auch in den kommenden Jahren im Fokus der Forschungsarbeit auf dem Gebiet der Geo-Informatik stehen wird. Eine Annäherung an die Gesamtlösung des Problems ist dabei nach Ansicht des Verfassers nur dann möglich, wenn die einzelnen Bausteine, die aus verschiedenen Fachrichtungen und Forschungsperspektiven beigetragen werden, zu einem Ganzen zusammengefügt werden können. Die vorliegende Arbeit versteht sich als ein solcher Baustein.

11 Literatur

- AdV [1988], Amtliches Topographisch-Kartographisches Informationssystem (ATKIS). Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), Bonn.
- AdV [2002], ATKIS-Objektartenkatalog, Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), Bonn.
http://www.atkis.de/dstinfo/dstinfo2.dst_gliederung
Zugriff: 25.05.2002
- AdV [2003], GeoInfoDok, Version 2.0, Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), Bonn.
http://www.adv-online.de/veroeffentlichungen/afis-alkis-atkis/geoinfodok_index.htm
Zugriff: 27.06.2003
- AdV [2004a], GeoInfoDok, Version 3.0, Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), Bonn.
<http://www.adv-online.de/extdeu/broker.jsp?uMen=bb2708b9-4c8a-3bff-de23-50376a112976>
Zugriff: 28.08.2004
- AdV [2004b], Homepage der Arbeitsgemeinschaft der Vermessungsverwaltungen der Bundesrepublik Deutschland.
<http://www.adv-online.de>
Zugriff: 17.09.2004
- Aho, A., Hopcroft, J. E., Ullman, J. D. [1987], Data Structures and Algorithms, Addison-Wesley Series in Computer Science and Information Processing, 427 p.
- Allen, J. F. [1983], Maintaining Knowledge about Temporal Intervals, Communications of the ACM, 26, No. 11, pp. 832-843.
- Alt, H., Guibas, L. J. [1996], Discrete Geometric Shapes: Matching, Interpolation, and Approximation – A Survey. Technical Report B 96-11, EVL-1996-142, Institute of Computer Science, Freie Universität Berlin. 34 p.
- Anders, K.-H. [2001], Data Mining for Automated GIS Data Collection, in: Fritsch, D., Spiller, R. (eds.), Photogrammetric Week '01, Heidelberg, pp. 263-272.
- Anders, K.-H., Sester, M. [1997], Methods of Data Base Interpretation – Applied to Model Generalization from Large to Medium Scale, in: Förstner, W., Plümer, L. (eds.), SMATI '97, Semantic Modelling for the Acquisition of Topographic Information from Images and Maps, Birkenhäuser, pp. 89-103.
- Baeza-Yates, R., Ribeiro-Neto, B. [1999], Modern Information Retrieval, Addison-Wesley.
<http://www.sims.berkeley.edu/~hearst/irbook/>
Zugriff: 19.06.2004
- Balley, S., Parent, C., Spaccapietra, S. [2004], Modelling Geographic Data with Multiple Representations, *Int. J. of Geographic Information Systems* 18 (4), pp. 327-352.
- Bartelme, N. [2000], Geoinformatik. Modelle, Strukturen, Funktionen, Springer-Verlag, Heidelberg, Berlin, 419 S.

- Bauer, M., Leonhardi, A. [2000], The VIT-System: Experiences with Developing a Location-Aware System for the Internet, *in: Workshop on Infrastructure for Smart Devices, How to Make Ubiquity an Actuality*, HUC2k, Bristol, Great Britain, 3 p.
- Bauer, M., Becker, C., Hähner, J., Schiele, G. [2003], ContextCube - Providing Context Information Ubiquitously, *in: Proceedings of the 23rd International Conference on Distributed Computing Systems Workshops (ICDCS 2003)*, pp. 308-313.
- Bauer, M., Rothermel, K. [2004], How to Observe Real-World Events through a Distributed World Model, *to appear in: Proceedings of the 10th International Conference on Parallel and Distributed Systems 2004 (ICPADS 2004)*, Newport Beach, California.
- Bauer, M., Jendoubi, L., Siemoneit, O. [2004a], Smart Factory – Mobile Computing in Production Environments, *in: Proceedings of the MobiSys 2004 Workshop on Applications of Mobile Embedded Systems (WAMES 04)*, 3 p.
- Bauer, M., Dürr, F., Geiger, J., Großmann, M., Hönle, N., Joswig, J., Nicklas, D., Schwarz, T. [2004b], Information Management and Exchange in the Nexus Platform, Fakultätsbericht der Fakultät Informatik der Universität Stuttgart Nr. TR-2004-04, 58 S.
- Bernard, L., Pundt, H. [1998], Semantikverlust in integrierten Systemen - Ein Fallbeispiel aus der Umweltplanung, *in: Haasis, H.-D. und K. C. Rance (Hrsg.), Umweltinformatik '98, 12. Intern. Symposium „Informatik für den Umweltschutz“ der Gesellschaft für Informatik (GI), Bremen, Band II, S. 528 - 539.*
- Besl, P., McKay, N. [1992], A Method for Registration of 3-D Shapes, *Trans. PAMI*, Vol. 14 (2), pp. 239-256.
- Bilenko, M., Mooney, R. J. [2003], Employing Trainable String Similarity Metrics for Information Integration, *in: Proceedings of the IJCAI-2003 Workshop on Information Integration on the Web, Mexico*, pp. 67-72.
- Bill, R. [1999a], Grundlagen der Geo-Informationssysteme, Band 1: Hardware, Software, Daten, Wichmann Verlag, Heidelberg, 454 S.
- Bill, R. [1999b], Grundlagen der Geo-Informationssysteme, Band 2: Analysen, Anwendungen und neue Entwicklungen, Wichmann Verlag, Heidelberg, 475 S.
- Bishr, Y. [1998], Overcoming the Semantic and other Barriers to GIS Interoperability, *Int. J. Geographical Information Science*, 12 (4), pp. 299-314.
- Bishr, Y., Pundt, H., Kuhn, W., Radwan, M. [1999a], Probing the Concept of Information Communities – A first Step towards a Road of Semantic Interoperability, *in: Goodchild, M. F., Egenhofer, M. J., Fegeas, R., Kottman, C. A. (eds.), Interoperating Geographic Information Systems*, Kluwer Academic Publishers, Boston, USA, pp. 55-71.
- Bishr, Y. A., Pundt, H., Rütger, C. [1999b], Proceeding on the Road of Semantic Interoperability - Design of a Semantic Mapper Based on a Case Study from Transportation, *in: Proceedings of the 2nd International Conference on Interoperating Geographic Information Systems, Zurich, Lecture Notes in Computer Science, Springer-Verlag, Heidelberg, Berlin*, pp. 203-215.
- Bofinger, J. [2001], Analyse und Implementierung eines Verfahrens zur Referenzierung geographischer Objekte, Diplomarbeit am Institut für Photogrammetrie, unveröffentlicht, 76 S.

- Böhlen, M., Güting, H., Erwig, M., Jensen, S., Lorentzos, A., Schneider, M., Vazirgiannis, M. [2000], A Foundation for Representing and Querying Moving Objects, *ACM Transactions on Database Systems* 25:1, pp. 1-42.
- Bruns, H., Egenhofer, M. [1996], Similarity of Spatial Scenes, *in: Proceedings of the 7th International Symposium on Spatial Data Handling (SDH) '96, Delft, The Netherlands*, pp. 4A.31-42.
- Bürklen, S., Marrón, P. J., Rothermel, K. [2004], An Enhanced Hoarding Approach Based on Graph Analysis, *in: Proceedings of the 5th IEEE International Conference on Mobile Data Management (MDM 2004), Berkeley, California, USA, January 19-22*, pp. 1-10.
- Burrough, P. A. [1996], Natural Objects with Indeterminate Boundaries, *in: Burrough, P. A., Frank, A. U. (eds.), ESF-GISDATA, Vol. 2*, pp. 3-28.
- CEN [1995], Geographic Data Files Version 3.0, TC 287, GDF for Road Traffic and Transport Telematics.
- Clementini, E., Di Felice, P. [1994], A Comparison of Methods for Representing Topological Relationships, *Information Sciences* 80, pp. 1-34.
- Cobb, M., Chung, M., Miller, V., Foley, H., Petry, F., Shaw, K. [1998], A Rule-based Approach for the Conflation of Attributed Vector Data, *GeoInformatica* 2 (1), pp. 7-35.
- Conrad, S. [2002a], Integration heterogener Datenbestände, *Jahrbuch der Heinrich-Heine-Universität Düsseldorf 2002*, S. 189-201.
- Conrad, S. [2002b], Schemaintegration – Integrationskonflikte, Lösungsansätze, aktuelle Herausforderungen, *Informatik – Forschung & Entwicklung*, 17, S. 101-111.
- Davis, M., Aquino, J. [2003], JCS Conflation Suite, Technical Report, 48 p.
Zugriff über: <http://www.jump-project.org/>
Zugriff: 17.02.2002
- Devogele, T., Parent, C., Spaccapietra, S. [1998], On Spatial Database Integration, *Int. J. Geographical Information Science*, 12 (4), pp. 335-352.
- Devogele, T. [2002], A new Merging Process for Data Integration Based on the Discrete Fréchet Distance, *in: Richardson, D., van Oosterom, P. (eds.), Proceedings of the 10th International Symposium on Spatial Data Handling (SDH) 02, Ottawa, Springer*, pp. 167-181.
- Do, H. H., Rahm, E. [2002], COMA – A System for Flexible Combination of Schema Matching Approaches, *in: Proceedings of the 28th Intl. Conference on Very Large Databases (VLDB), Hongkong*, 12 p.
- Drosdol, T., Dürr, F., Großmann, M., Hönle, N., Volz, S. [2004], Technischer Bericht der Taskforce Topologie, Universität Stuttgart, Nexus-Projekt, unveröffentlicht, 29 S.
- Drosdol, T., Dürr, F., Hönle, N., Volz, S. [2006], Navigation Component Report, Technischer Bericht des Sonderforschungsbereichs 627 (2006/10), 30 S.
- Dürr, F., Rothermel, K. [2003], On a Location Model for Fine-Grained Geocast, *in: Dey, A. K., Schmidt, A., McCarthy, J. F. (eds.), Proceedings of the 5th International Conference on Ubiquitous Computing, UbiComp '03, Seattle, Oct. 12-15*, pp. 18-35.

- Edwards, D., Simpson, J. [2002], Integration and Access of Multi-source Vector Data, *in*: Proceedings of the Joint International Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, on CD-ROM.
- Egenhofer, M. [1991], Reasoning about Binary Topological Relations, *in*: Günther, O. (edt.), Advances in Spatial Databases - Proceedings of the 2nd Symposium on Large Spatial Databases SSD '91, Lecture Notes in Computer Science 525, pp. 143-160.
- Egenhofer, M., Clementi, E., di Felice, P. [1994], Evaluating Inconsistencies among Multiple Representations, *in*: Proceedings of the 6th International Symposium on Spatial Data Handling (SDH) '94, Edinburgh, Scotland, pp. 901-920.
- Fonseca, F., Egenhofer, M., Davis, C. [2000], Ontology-Driven Information Integration, *in*: Proceedings of the AAAI 2000 Workshop on Spatial and Temporal Granularity, Austin, Texas, pp. 61-64.
- Fonseca, F., Egenhofer, M., Agouris, P., Câmara, G. [2002], Using Ontologies for Integrated Geographic Information Systems, *Transactions in GIS* 6 (3), pp. 231-257.
- Fonseca, F., Egenhofer, M., Davis, C., Câmara, G. [2003], Semantic Granularity in Ontology-Driven Geographic Information Systems, *Annals of Mathematics and Artificial Intelligence* 36 (1-2), pp. 121-151.
- Gabay, Y., Sester, M. [2002], Forming and Utilizing Communication between Two Spatial Representations at Different Scales – a Demonstration, submitted to *Geoinformatica*.
- Glemser, M. [2001], Zur Berücksichtigung der geometrischen Objektunsicherheit in der Geoinformatik, Dissertation, Fakultät für Bauingenieur- und Vermessungswesen, Universität Stuttgart, Deutsche Geodätische Kommission (DGK), Reihe C, Heft Nr. 539, 138 S.
- Gloss, B., Kühn, P. J. [2002], Modellierung und Entwurf von Kommunikations-Infrastrukturen für orts- und kontextabhängige Dienste, *in*: Tagungsmappe zum DFG-Workshop Modelle, Werkzeuge und Infrastrukturen zur Unterstützung von Entwicklungsprozessen, S. 1-4.
- Gösseln, G. v., Sester, M. [2003], Change Detection and Integration of Topographic Updates from ATKIS to Geoscientific Data Sets, *in*: Proceedings of the International Conference on Next Generation Geospatial Information, Boston, USA, 2003.
- Guarino, N. [1998], Formal Ontology and Information Systems, *in*: Guarino, N. (edt.), Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, FOIS '98, Trento, Italien, pp. 3-15.
- Hakimpour, F., Timpf, S., [2001], Using Ontologies for Resolution of Semantic Heterogeneity in GIS, *in*: AGILE (Editor), AGILE - GI in Europe: Integrative Interoperable Interactive. Masaryk University Brno, Brno, Czech Republic, pp. 385-395.
- Harvey, F. [1999], Designing for Interoperability: Overcoming Semantic Differences, *in*: Goodchild, M. F., Egenhofer, M. J., Fegeas, R., Kottman, C. A. (eds.), Interoperating Geographic Information Systems, Kluwer Academic Publishers, Boston, USA, pp. 85-97.
- Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., Riedemann, C. [1999], Semantic Interoperability: a Central Issue for Sharing Geographic Information, *The Annals of Regional Science*, Vol. 33 (2), pp. 213-232.
- Hauser, C., Leonhardi, A., Kühn, P. J. [2002], Sicherheitsaspekte in Nexus – einer Plattform für ortsbezogene Anwendungen, *it+ti, Zeitschrift für Informationstechnik und Technische Informatik*, Bd. 5/02, S. 268-277.

- Hohl, F., Kubach, U., Leonhardi, A., Rothermel, K., Schwehm, M. [1999], NEXUS – an Open Global Infrastructure for Spatial-Aware Applications, *in*: Proceedings of the 5th International Conference on Mobile Computing and Networking, Seattle, pp. 249-255.
- Holt, A. [1998], The Matching and Ranking of Surface and Deep Similarities of Spatial Data, *in*: Proceedings of the 10th Annual Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, pp. 133-143.
- Holt, A. [1999], Spatial Similarity and GIS: the Grouping of Spatial Kinds, *in*: Proceedings of the Annual Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand, pp. 241-250.
- Hubig, C. [2004], Künstliche intelligente Umwelten - über die Veränderungen unseres Wirklichkeitsverhältnisses, *in*: Liene, P., Kaufmann, S., Bruppacher, R. (Hrsg.), Kanalisation bis Handy - Verbreitungsbedingungen technischer Innovationen in der Gesellschaft (in Vorbereitung).
- ISO [2004], Intelligent Transport Systems - Geographic Data Files - Overall Data Specifications (GDF 4.0), ISO/TR 14825, TC 204, Working Group 3, (Intelligent TransportSystems), 2002.
- JDOM [2004], JDOM Project.
<http://www.jdom.org>
Zugriff: 15.02.2004
- JUMP [2004], Java Unified Mapping Platform.
<http://www.jump-project.org>
Zugriff: 17.02.2004.
- Kada, M. [2003], 3D Building Generalisation and Visualisation, *in*: Fritsch, D. (ed.), Proceedings of Photogrammetric Week '03, pp. 29-38.
- Kemper, A., Eickler, A. [1999], Datenbanksysteme – eine Einführung, München, 504 S.
- Kent, W. [2000], Data and Reality, 1st Books Library, 248 p.
- Klinec, D., Fritsch, D. [2001], Acquisition of Position Information for Location Aware Applications using Multi Sensors and Mobile Photogrammetry, *in*: Proceedings of ION GPS '01, Salt Lake City, USA, pp. 3113-3118.
- Klinec, D., Fritsch, D. [2003], Towards Pedestrian Navigation and Orientation, *in*: Proceedings of the 7th South East Asian Survey Congress (SEASC '03), Hong Kong, November 3-7, on CD-ROM.
- Klinger, R. [2000], Konzeption und Entwicklung eines Nexus-Demonstrators. Studienarbeit Nr. 1791, Fakultät für Informatik, Universität Stuttgart, 72 S.
- Kraft, W. [1995], Entwurf von Zuordnungsalgorithmen zur Fortführung und Überprüfung von raumbezogenen Datenbeständen. Diplomarbeit am Institut für Photogrammetrie, Universität Stuttgart, unveröffentlicht, 75 S.
- Kraut, M. [2003], Zuordnung und Conflation heterogener Straßendaten, Diplomarbeit am Institut für Photogrammetrie, Universität Stuttgart, unveröffentlicht, 109 S.
- Kubach, U., Rothermel, K. [2000], An Adaptive Location-Aware Hoarding Mechanism, *in*: Proceedings of the 5th IEEE Symposium on Computers and Communications (ISCC), pp. 615-620.
- Kuhn, W. [1994], Defining Semantics for Spatial Data Transfers, *in*: Proceedings of the 6th International Symposium on Spatial Data Handling (SDH) '94, Edinburgh.

- Kuhn, W. [2003], Semantic Reference Systems, *Int. J. Geographical Information Science*, 17 (5), pp. 405-409.
- Laurini, R. [1998], Spatial Multi-database Topological Continuity and Indexing: A Step Towards Seamless GIS Data Interoperability, *Int. J. Geographical Information Science*, 12 (4), pp. 373-402.
- Leonhardi, A., Kubach, U., Rothermel, K., Fritz, A. [1999], Virtual Information Towers - A Metaphor for Intuitive, Location-Aware Information Access in a Mobile Environment, *in: Proceedings of the 3rd International Symposium on Wearable Computers*, pp. 15-20.
- Leonhardi, A. [2003], Architektur eines verteilten skalierbaren Lokationsdienstes, Dissertation an der Fakultät Informatik der Universität Stuttgart, Institut für Parallele und Verteilte Höchstleistungssysteme, 186 S.
- Mackaness, W., Weibel, R., Buttenfield, B. [1997], Report of the 1997 ICA Workshop on Map Generalization, 19th until 21st of June 1997, Gävle, Sweden.
<http://www.geo.unizh.ch/ICA/docs/gaevle1997/report.html>
Zugriff: 28.08.2003
- Mantel, D., Lipeck, U. [2004], Matching Cartographic Objects in Spatial Databases, *in: Proceedings of the XXth ISPRS Congress, Comm. IV, Istanbul, Turkey*, pp. 172-176.
- McKee, L., Kottman, C. [1999], Inside the OpenGIS Specification, *Photogrammetric Engineering & Remote Sensing*, Vol. 65 (12), pp. 1345-1359.
- Messmer, J. [2001], Modellierung der Augmented World in Nexus, Diplomarbeit Nr. 1870 der Fakultät für Informatik, Universität Stuttgart, 120 S.
- Metaxides, A. [1975], "Information Bearing" and "Non-Information Bearing" Sets, *in: Douque, B. C. M., Nijssen, G. M. (eds.), Database Description, Proceedings of the IFIP TC-2 Special Working Conference, Wepion, Belgium*, pp. 363-368.
- Molenaar, M. [1998], An Introduction to the Theory of Spatial Object Modelling, London, 246 p.
- MurMur [2002], MurMur Project, Multi Representations – Multi Resolutions, Final Report,
<http://lbdwww.epfl.ch/e/MurMur/docs/RapportScient.doc>
Zugriff: 08.10.2003
- NavTech [2000], Navigation Technologies, GDF 3.0 (V3.20) User's Guide, 275 p.
- Neisser, U. [1996], Kognition und Wirklichkeit: Prinzipien und Implikationen der kognitiven Psychologie, zweite Ausgabe, Stuttgart.
- Nexus [2004], Homepage des Sonderforschungsbereiches 627,
<http://www.nexus.uni-stuttgart.de>
Zugriff: 22.07.2004
- Nexus C3 [2004], SFB 627, Teilprojekt „Interpretation multisensorieller Daten“,
<http://www.informatik.uni-stuttgart.de/ipvs/bv/projekte/nexus/tpc3.html>
Zugriff: 23.07.2004
- Nexus C5 [2004], SFB 627, Teilprojekt „Augmented Reality Anwendungen“,
<http://www.vis.uni-stuttgart.de/plain/research/proj/sfb627/tpc5.html>
Zugriff: 23.07.2004

- Nexus D2 [2004], SFB 627, Teilprojekt „Orientierungshilfen für Blinde“,
<http://www.vis.uni-stuttgart.de/plain/research/proj/sfb627/tpd2.html>
Zugriff: 23.07.2004
- Nicklas, D. [2001], Final Report of Design Workshop, Fakultätsbericht der Fakultät Informatik der Universität Stuttgart Nr. TR-2000-08, 59 S.
- Nicklas, D., Großmann, M., Schwarz, T., Volz, S., Mitschang, B. [2001], A Model-Based, Open Architecture for Mobile, Spatially Aware Applications, *in*: Jensen, Christian S., Schneider, M., Seeger, B., Tsotras, V. J. (eds), Proceedings of the 7th International Symposium on Spatial and Temporal Databases (SSTD) '01, Redondo Beach, California, USA, July 12-15, pp. 117-135.
- Nicklas, D., Mitschang, B. [2001], The Nexus Augmented World Model: An Extensible Approach for Mobile, Spatially-Aware Applications, *in*: Proceedings of the 7th International Conference on Object-Oriented Information Systems, Calgary, pp. 392-401.
- OBAK NRW [2003], Objektabbildungskatalog Liegenschaftskataster Nordrhein-Westfalen, 15 S.
http://www.lverma.nrw.de/produkte/druckschriften/verwaltungsvorschriften/images/OBAK/OBAK_Erlass_Titel_2003.pdf
Zugriff: 28.09.2003
- OpenGIS Consortium [1998], OpenGIS Guide, Datum: 1998-06-03.
<http://www.opengis.org/techno/guide/guide/Guide980629.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [1999a], OGC Abstract Specification, Topic 5: Features, Version 4, Project Document Number: 99-105r2, Datum: 1999-03-24.
<http://www.opengis.org/techno/abstract/99-105r2.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [1999b], OGC Abstract Specification, Topic 8: Relationships between Features, Version 4, Project Document Number: 99-108r2, Datum: 1999-03-26.
<http://www.opengis.org/techno/abstract/99-108r2.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [1999c], OGC Abstract Specification, Topic 9: Quality, Version 4, Project Document Number: 99-108r2, Datum: 1999-03-30.
<http://www.opengis.org/techno/abstract/99-109r1.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [1999d], OGC Abstract Specification, Topic 14: Semantics and Information Communities, Version 4, Project Document Number: 99-114, Datum: 1999-04-04.
<http://www.opengis.org/techno/abstract/99-114.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [2001a], OGC Abstract Specification, Topic 1: Feature Geometry, Project Document Number: 01-101; and ISO/DIS 19107, Geographic Information – Spatial Schema, Datum: 2001-04-01.
<http://www.opengis.org/techno/abstract/01-101.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [2001b], OGC History, Datum: 2001-11-27.
<http://www.opengis.org/pressrm/summaries/20011127.TS.OGChist.pdf>
Zugriff: 11.07.2003

- OpenGIS Consortium [2002], OGC Specification Program, Datum: 2002-08-13.
http://www.opengis.org/pressrm/summaries/2021021_TS_SpecProg.pdf
Zugriff: 11.07.2003
- OpenGIS Consortium [2003a], OpenGIS Geography Markup Language Implementation Specification, Reference Number OGC 02-023r4, Datum: 2003-01-29.
<http://www.opengis.org/techno/documents/02-023r4.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [2003b], Overview of OGC's Interoperability Program, Datum: 2003-04-11.
<http://www.opengis.org/pressrm/summaries/20020813.TS.IP.pdf>
Zugriff: 11.07.2003
- OpenGIS Consortium [2004a], OGC Specifications,
<http://www.opengis.org/specs/>
Zugriff: 17.08.2004
- OpenGIS Consortium [2004b], OGC Homepage,
<http://www.opengis.org/>
Zugriff: 17.08.2004
- OSKA NRW [2003], Objektschlüsselkatalog Liegenschaftskataster Nordrhein-Westfalen, 71 S.
http://www.lverma.nrw.de/produkte/druckschriften/verwaltungsvorschriften/images/Oska_gesamt_03.pdf
Zugriff: 28.09.2003
- Pandazis, J. [1999], TR 4011 EVIDENCE, Final Report, Brussels, July 1999.
- Parent, C., Spaccapietra, S., Zimányi, E. [1999], Spatio-Temporal Conceptual Models: Data Structures + Space + Time, *in: Proceedings of the 7th ACM International Symposium on Advances in Geographic Information Systems*, Kansas City, Missouri, USA, pp. 26-33.
- Parent, C., Spaccapietra, S., Zimányi, E. [2000], MurMur: Database Management of Multiple Representations, *in: Proceedings of the AAI-2000 Workshop on Spatial and Temporal Granularity*, Austin, Texas, USA, pp. 83-86.
- Rahm, E., Bernstein, P. A. [2001], A Survey of Approaches to Automatic Schema Matching, *VLDB Journal*, Vol. 10, No. 4, pp. 334-350.
- Ramsey, P. [2004], JUMP Project and Direction.
http://www.jump-project.org/assets/JUMP_Project_and_Direction.pdf,
Zugriff: 17.02.2004.
- Reinhardt, F., Soeder, H. [1998], dtv-Atlas Mathematik, Bd. 1: Grundlagen, Algebra und Geometrie, München, 270 S.
- Reinhardt, W., Joos, G., Kuhlmann, H., Müller-Hermes, R., Schmitt, B., Seeberger, S., Walter, V. [2004], Raumbezogene Informationssysteme, *in: Möser, M., Müller, G., Schlemmer, H., Werner, H. (Hrsg.), Handbuch Ingenieurgeodäsie*, 226 S.
- Ressel, W. [2004], Modelling and Simulation of Mobility, *in: Proceedings of the 1st International Workshop on Intelligent Transportation (WIT 2004)*, 5 S.
- Rodríguez, F. H., Aranda, G. B., Navarro, M. [1996], An Ontology-based Approach to Spatial Information Modelling, *in: Kraak, M. J., Molenaar, M. (eds.), Proceedings of the 7th International Symposium on Spatial Data Handling (SDH) '96, Vol. II, Delft*, pp. 12B.17-12B.29.

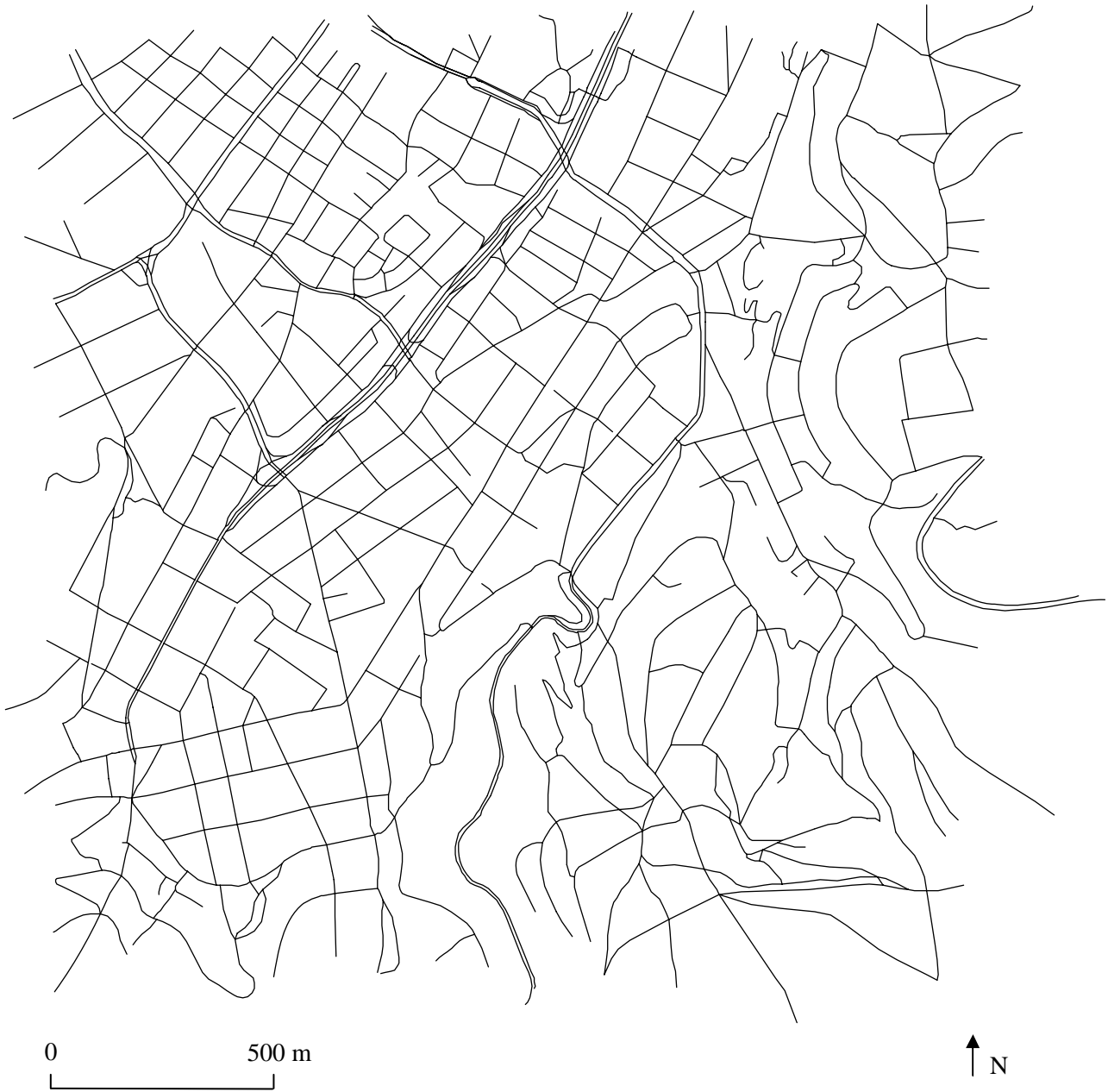
- Rodríguez, A., Egenhofer, M. [2004], Comparing Geospatial Entity Classes: an Asymmetric and Context-dependent Similarity Measure, *Int. J. of Geographic Information Systems* 18 (3), pp. 229-256.
- Saalfeld, A. [1988], Automated Map Compilation, *Int. J. of Geographic Information Systems* 2 (3), pp. 217-228.
- Sester, M., Anders, K.-H., Walter, V. [1998], Linking Objects of Different Spatial Data Sets by Integration and Aggregation, *GeoInformatica* (4), pp. 335-358.
- Sester, M. [1999], Acquiring Transition Rules between Multiple Representations in a GIS, *Computers, Environment and Urban Systems* (23), pp. 5-17.
- Sester, M. [2000], Maßstababhängige Darstellungen in digitalen räumlichen Datenbeständen, Habilitation, Fakultät für Bauingenieur- und Vermessungswesen, Universität Stuttgart, Deutsche Geodätische Kommission (DGK), Reihe C, Nr. 544, 108 S.
- Sester, M., Butenuth, M., Goesseln, G. v., Heipke, C., Klopp, S., Lipeck, U., Mantel, D. [2003], New Methods for Semantic and Geometric Integration of Geoscientific Data Sets with ATKIS, Applied to Geobjects from Geology and Soil Science, *in: Geotechnologien Science Report, Part 2, Koordinierungsbüro Geotechnologien, Potsdam*, pp. 51-62.
- Sheth, A. P., Larson, J. A. [1990], Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, *ACM Computing Surveys* 22 (3), pp. 183-236.
- Sheth, A. [1999], Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics, *in: Goodchild, M. F., Egenhofer, M. J., Fegeas, R., Kottman, C. A. (eds.), Interoperating Geographic Information Systems, Kluwer Academic Publishers, Boston, USA*, pp. 5-29.
- Siemoneit, O. [2004], Ubiquitous Computing. Neue Dimensionen technischer Kultur, *in: Institut zur Erforschung und Förderung österreichischer und internationaler Literaturprozesse (INST) (edt.), Das Verbindende der Kulturen (im Erscheinen)*, S. 1-2.
- Spaccapietra, S., Parent, C., Devogele, T. [1996], Analysis of Discrepancies in Spatial Data Representations, *in: Proceedings of the Int. Symposium on Cooperative Database Systems for Advanced Applications, Kyoto*, pp. 308-315.
- Spaccapietra, S. Parent, C. and Vangenot, C. [2000], GIS Databases: From Multiscale to MultiRepresentation, *in: Proceedings of the International Workshop on Emerging Technologies for Geo-based Applications, Ascona, Switzerland*, pp. 119-132.
- Timm, C., Riedemann, C., Brox, C., Kuhn, W. [1998], Möglichkeiten und Grenzen von GIS-Komponententechnologie in der Geodatenproduktion, *in: Strobl, J., Dollinger, F. (Hrsg.), Angewandte Geographische Informationsverarbeitung, Beiträge zum AGIT-Symposium in Salzburg 1998, Heidelberg*, S. 380-386.
- Tøssebro, E., Güting, R. H. [2001], Creating Representations for Continuously Moving Regions from Observations, *in: Jensen, Christian S., Schneider, M., Seeger, B., Tsotras, V. J. (eds.), Proceedings of the 7th International Symposium on Spatial and Temporal Databases (SSTD) 2001, Redondo Beach, CA, USA, July 12-15*, pp. 321-344.
- Uitermark, H. [1996], The integration of geographic databases. Realising Geodata Interoperability through the Hypermap Metaphor and a Mediator Architecture, *in: Rumor, M., McMillan, R., Otens, H. F. (eds.), Proceedings of the 2nd Joint European Conference & Exhibition on Geographical Information (JEC-GI) '96, Vol. I, Barcelona*, pp. 92-95.

- Uitermark, H., van Oosterom, P., Mars, N., Molenaar, M. [1998], Propagating Updates: Finding Corresponding Objects in a Multi-source Environment, *in*: Poiker, T. K., Chrisman, N. (eds.), Proceedings of the 8th International Symposium on Spatial Data Handling (SDH) '98, Vancouver, pp. 580-591.
- Uitermark, H., Vogels, A., van Oosterom, P. [1999a], Semantic and Geometric Aspects of Integrating Road Networks, *in*: Proceedings of the 2nd International Conference on Interoperating Geographic Information Systems, Zurich, Lecture Notes in Computer Science, Springer-Verlag, Heidelberg, Berlin, pp. 177-188.
- Uitermark, H., van Oosterom, P., Nicolaas, J., Mars, I., Molenaar, M. [1999b], Ontology-based Geographic Data Set Integration, *in*: Proceedings of the International Workshop on Spatio-Temporal Database Management STDBM '99, Lecture Notes in Computer Science, Vol. 1678, pp. 60-78.
- Van Wijngarden, F., van Putten, J., van Oosterom, P., Uitermark, H. [1997], Map Integration – Update Propagation in a Multi-source Environment, *in*: Proceedings of the 5th ACM International Workshop on Advances in GIS 97, Las Vegas, USA, pp. 71-76.
- Vazirgiannis, M., Wolfson, O. [2001], A Spatiotemporal Model and Language for Moving Objects on Road Networks, *in*: Jensen, Christian S., Schneider, M., Seeger, B., Tsotras, V. J. (eds), Proceedings of the 7th International Symposium on Spatial and Temporal Databases (SSTD) 2001, Redondo Beach, CA, USA, July 12-15, pp. 20-35.
- Volz, S., Fritsch, D., Klinec, D., Leonhardi, A., Schützner, J. [1999], NEXUS - Spatial Model Servers for Location Aware Applications on the Basis of ArcView, *in*: Proceedings of the 14th ESRI European User Conference, Munich, on CD-ROM.
- Volz, S., Sester, M. [2000], Nexus – Distributed Data Management Concepts For Location Aware Applications, *in*: Proceedings of the International Workshop on Emerging Technologies for Geo-based Applications, Ascona, Switzerland, pp. 21-35.
- Volz, S., Großmann, M., Höhle, N., Nicklas, D., Schwarz, T. [2002], Integration mehrfach repräsentierter Straßendaten für eine föderierte Navigation, *it+ti, Zeitschrift für Informationstechnik und Technische Informatik*, Bd. 5/02, S. 260-267.
- Volz, S., Bofinger, J. M. [2002], Integration of Spatial Data within a Platform for Location-based Applications, *in*: Proceedings of the Joint International Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, on CD-ROM.
- Volz, S., Walter, V. [2004], Linking Different Geospatial Databases by Explicit Relations, *in*: Proceedings of the XXth ISPRS Congress, Comm. IV, Istanbul, Turkey, pp. 152-157.
- Volz, S. [2005a], Data-driven Matching of Geospatial Schemas, *in*: Cohn, A.G., Mark, D.M. (eds.): Spatial Information Theory, *in*: Proceedings of the International Conference on Spatial Information Theory (COSIT '05), Ellicottville, NY. Lecture Notes in Computer Science 3693, pp. 115-132.
- Volz, S. [2005b], Shortest Path Search in Multi-Representation Street Databases, *in*: Proceedings of the 3rd Symposium on Location Based Services and TeleCartography, Vienna, Austria, pp. 125-130.
- Vosselman, G. [1992], Relational Matching, Springer-Verlag, Berlin, Heidelberg, New York.
- W3C [1999], Extensible Stylesheet Language Transformation Version 1.0.
<http://www.w3.org/TR/xslt>
Zugriff: 25.01.2002

- W3C [2002], Scalable Vector Graphics Version 1.1 Specification.
<http://www.w3.org/TR/2002/CR-SVG11-20020430/>
Zugriff: 25.01.2002
- W3C [2004], XML Schema.
<http://www.w3.org/XML/Schema>
Zugriff: 11.08.2003
- Walter, V. [1997], Zuordnung von raumbezogenen Daten – am Beispiel der Datenmodelle ATKIS und GDF, Dissertation, Fakultät für Bauingenieur- und Vermessungswesen, Universität Stuttgart, Deutsche Geodätische Kommission (DGK), Reihe C, Heft Nr. 480, 130 S.
- Walter, V., Fritsch, D. [1999], Matching Spatial Data Sets: a Statistical Approach, *Int. J. of Geographical Information Science* 13 (5), pp. 445–473.
- Weis, M., Naumann, F. [2004], Detecting Duplicate Objects in XML Documents, *in: Proceedings of the SIGMOD International Workshop on Information Quality in Information Systems (IQIS) '04*, Paris, pp. 10-19.
- Wiederhold, G. [1999], Mediation to Deal with Heterogeneous Data Sources, *in: Proceedings of the 2nd International Conference on Interoperating Geographic Information Systems, Zurich, Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg, Berlin, pp. 1-16.
- Xiong, D., Sperling, J. [2003], Semi-Automated Matching for Network Database Integration, *in: Proceedings of the „GIS for Transportation” Symposium '03*, Colorado Springs, USA.
- Yuan, S., Tao, C. [1999], Development of Conflation Components, *in: Proceedings of Geoinformatics '99 Conference*, Ann Arbor, USA, pp. 1-13.

I Darstellung der Testgebiete

Testgebiet 1: ATKIS



Testgebiet 1: GDF



Paralleldarstellung der Daten aus Testgebiet 1:

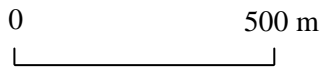
ATKIS:



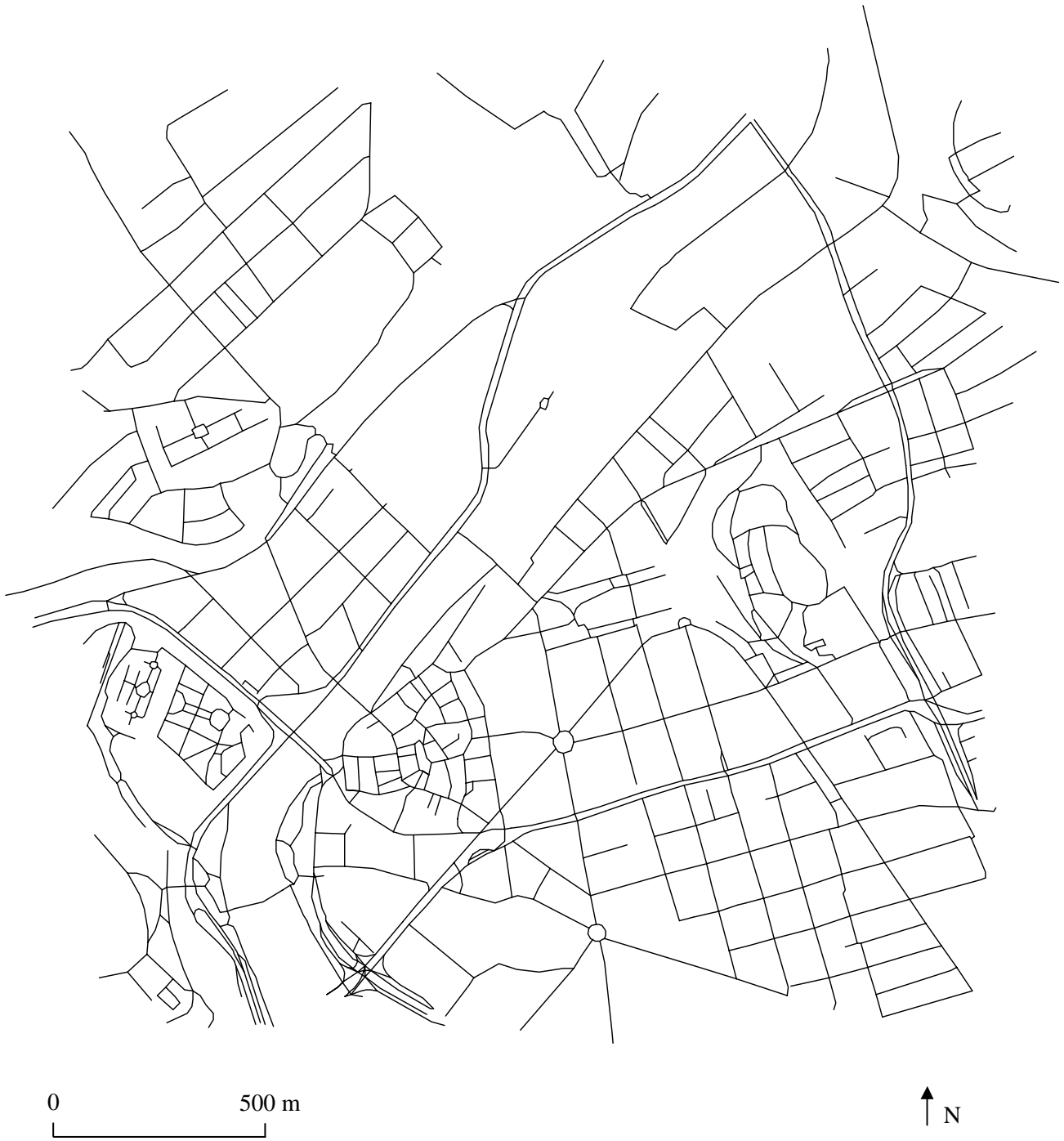
GDF:



Testgebiet 2: ATKIS



Testgebiet 2: GDF



Paralleldarstellung der Daten aus Testgebiet 2:

ATKIS:



GDF:



II Objektklassen-Korrespondenzen von GDF und ATKIS

II.A Korrespondenzen der GDF-Klassen

Alle MRep-Relationen berücksichtigt:

Korrespondenzen Road	Testgebiet 1	Testgebiet 2	Gesamt
Road→Straße	6 (7,41%)	10 (12,5%)	16 (9,94%)
Road→Fahrbahn	68 (83,95%)	49 (61,25%)	117 (72,67%)
Road→Fahrbahn/Straße	1 (1,23%)	2 (2,47%)	3 (1,86%)
Road/Road Element→Straße	-	4 (5%)	4 (2,48%)
Road/Road Element →Fahrbahn	3 (3,7%)	4 (5%)	7 (4,35%)
Road/Road Element→Fahrbahn/Straße	1(1,23%)	4 (5%)	5 (3,11%)
Road/Intersection→Fahrbahn	2 (2,47%)	7 (8,64%)	9 (5,59%)
Gesamtzahl	81	80	161

Tabelle II-1: Korrespondenzen der Klasse Road

Korrelationen Road	Werte gesamt
Road→Straße	20,5 (13,85%)
Road→Weg	0 (0%)
Road→Fahrbahn	127,5 (86,15%)
Gesamtzahl	148

Tabelle II-2: Korrelationswerte der Klasse Road

Korrespondenzen Intersection	Testgebiet 1	Testgebiet 2	Gesamt
Intersection→Straße	11 (33,33%)	17 (56,67%)	28 (44,44%)
Intersection→Fahrbahn	15 (45,45%)	1 (3,33%)	16 (25,4%)
Intersection/Road Element→Straße	4 (12,12%)	5 (16,67%)	9 (14,29%)
Intersection/Road Element→Fahrbahn	1 (3,03%)	-	1 (1,59%)
Intersection/Road→Fahrbahn	2 (6,06)	7 (23,33%)	9 (14,29%)
Gesamtzahl	33	30	63

Tabelle II-3: Korrespondenzen der Klasse Intersection

Korrelationen Intersection	Werte gesamt
Intersection→Straße	32,5 (60,75%)
Intersection→Weg	0 (0%)
Intersection→Fahrbahn	21 (39,25%)
Gesamtzahl	53,5

Tabelle II-4: Korrelationswerte der Klasse Intersection

Nur MRep-Relationen mit GM > 80 berücksichtigt:

Korrespondenzen Road Element	Testgebiet 1	Testgebiet 2	Gesamt
Road Element→Straße	430 (84,98%)	362 (82,09%)	792 (83,63%)
Road Element→Weg	45 (8,89%)	55 (12,47%)	100 (10,56%)
Road Element→Fahrbahn	9 (1,78%)	2 (0,45%)	11 (1,16%)
Road Element→Straße/Weg	17 (3,36%)	11 (2,49%)	28 (2,96%)
Road Element/Intersection→Straße	2 (0,39%)	3 (0,68%)	5 (0,53%)
Road Element/Intersection→Fahrbahn	1 (0,20%)	-	1 (0,11%)
Road Element/Road→Straße	-	4 (0,91%)	4 (0,42%)
Road Element/Road→Fahrbahn	2 (0,39%)	2 (0,45%)	4 (0,42%)
Road Element/Road→Fahrbahn/Straße	-	2 (0,45%)	2 (0,21%)
Gesamtzahl	506	441	947

Tabelle II-5: Korrespondenzen der Klasse Road Element (GM > 80)

Korrelationen Road Element	Werte gesamt
Road Element→Straße	811 (86,37%)
Road Element→Weg	114 (12,14%)
Road Element→Fahrbahn	14 (1,49%)
Gesamtzahl	939

Tabelle II-6: Korrelationswerte der Klasse RoadElement (GM > 80)

Korrespondenzen Road	Testgebiet 1	Testgebiet 2	Gesamt
Road→Straße	1 (2,04%)	6 (10,91%)	7 (6,73%)
Road→Fahrbahn	44 (89,80%)	35 (63,64%)	79 (75,96%)
Road→Fahrbahn/Straße	-	2 (3,64%)	2 (1,92%)
Road/Road Element→Straße	-	4 (7,27%)	4 (3,85%)
Road/Road Element →Fahrbahn	2 (4,08%)	2 (3,64%)	4 (3,85%)
Road/Road Element→Fahrbahn/Straße	-	2 (3,64%)	2 (1,92%)
Road/Intersection→Fahrbahn	2 (4,08%)	4 (7,27%)	6 (5,77%)
Gesamtzahl	49	55	104

Tabelle II-7: Korrespondenzen der Klasse Road (GM > 80)

Korrelationen Road	Werte gesamt
Road→Straße	10,5 (10,94%)
Road→Weg	0 (0%)
Road→Fahrbahn	85,5 (89,06%)
Gesamtzahl	96

Tabelle II-8: Korrelationswerte der Klasse Road (GM > 80)

Korrespondenzen Intersection	Testgebiet 1	Testgebiet 2	Gesamt
Intersection→Straße	3 (23,08%)	8 (53,33%)	11 (39,29%)
Intersection→Fahrbahn	5 (38,46%)	-	5 (17,86%)
Intersection/Road Element→Straße	2 (15,38%)	3 (20,0%)	5 (17,86%)
Intersection/Road Element→Fahrbahn	1 (7,69%)	-	1 (3,57%)
Intersection/Road→Fahrbahn	2 (15,38%)	4 (26,67%)	6 (21,43%)
Gesamtzahl	13	15	28

Tabelle II-9: Korrespondenzen der Klasse Intersection (GM > 80)

Korrelationen Intersection	Werte gesamt
Intersection→Straße	13,5 (61,36%)
Intersection→Weg	0 (0%)
Intersection→Fahrbahn	8,5 (38,64%)
Gesamtzahl	22

Tabelle II-10: Korrelationswerte der Klasse Intersection (GM > 80)

II.B Korrespondenzen der ATKIS-Klassen:

Alle MRep-Relationen berücksichtigt:

Korrespondenzen Straße	Testgebiet 1	Testgebiet 2	Gesamt
Straße→Road Element	526 (92,61%)	429 (88,82%)	955 (90,87%)
Straße→Road	6 (1,06%)	10 (2,07%)	16 (1,52%)
Straße→Intersection	11 (1,94%)	17 (3,52%)	28 (2,66%)
Straße→Road/Road Element	-	4 (0,83%)	4 (0,38%)
Straße→Road Element/Intersection	4 (0,70%)	5 (1,03%)	9 (0,86%)
Straße/Weg→Road Element	18 (3,17%)	13 (2,69%)	31 (2,95%)
Straße/Fahrbahn→Road Element	1 (0,18%)	-	1 (0,09%)
Straße/Fahrbahn→Road	1 (0,18%)	2 (0,41%)	3 (0,28%)
Straße/Fahrbahn→Road Element/Road	1 (0,18%)	3 (0,62%)	4 (0,38%)
Gesamtzahl	568	483	1051

Tabelle II-11: Korrespondenzen der Klasse Straße

Korrelationen Straße	Werte gesamt
Straße→Road Element	978,5 (94,86%)
Straße→Road	20,5 (1,99%)
Straße→Intersection	32,5 (3,15%)
Gesamtzahl	1031,5

Tabelle II-12: Korrelationswerte der Klasse Straße

Korrespondenzen Weg	Testgebiet 1	Testgebiet 2	Gesamt
Weg→Road Element	67 (78,82%)	107 (89,17%)	174 (84,88%)
Weg/Straße→Road Element	18 (21,18%)	13 (10,83%)	31 (15,12%)
Gesamtzahl	85	120	205

Tabelle II-13: Korrespondenzen der Klasse Weg

Korrelationen Weg	Werte gesamt
Weg→Road Element	189,5 (100%)
Weg→Road	0 (0%)
Weg→Intersection	0 (0%)
Gesamtzahl	189,5

Tabelle II-14: Korrelationswerte der Klasse Weg

Korrespondenzen Fahrbahn	Testgebiet 1	Testgebiet 2	Gesamt
Fahrbahn→Road Element	15 (14,02%)	5 (7,04%)	20 (11,24%)
Fahrbahn→Road	68 (63,55%)	49 (69,01%)	117 (65,73%)
Fahrbahn→Intersection	15 (14,02%)	1 (1,41%)	16 (8,99%)
Fahrbahn→Road Element/Road	3 (2,8%)	4 (5,63%)	7 (3,93%)
Fahrbahn→Road Element/Intersection	1 (0,93%)	-	1 (0,56%)
Fahrbahn→Road/Intersection	2 (1,87%)	7 (9,86%)	9 (5,06%)
Fahrbahn/Straße→Road Element	1 (0,93%)	-	1 (0,56%)
Fahrbahn/Straße→Road	1 (0,93%)	2 (2,82%)	3 (1,69%)
Fahrbahn/Straße→Road Element/Road	1 (0,93%)	3 (4,22%)	4 (2,25%)
Gesamtzahl	107	71	178

Tabelle II-15: Korrespondenzen der Klasse Fahrbahn

Korrelationen Fahrbahn	Werte gesamt
Fahrbahn→Road Element	25,5 (14,65%)
Fahrbahn→Road	127,5 (73,28%)
Fahrbahn→Intersection	21 (12,07%)
Gesamtzahl	174

Tabelle II-16: Korrelationswerte der Klasse Fahrbahn

Nur MRep-Relationen mit GM > 80 berücksichtigt:

Korrespondenzen Straße	Testgebiet 1	Testgebiet 2	Gesamt
Straße→Road Element	430 (94,92%)	362 (90,95%)	792 (93,07%)
Straße→Road	1 (0,22%)	6 (1,51%)	7 (0,82%)
Straße→Intersection	3 (0,66%)	8 (2,01%)	11 (1,29%)
Straße→Road/Road Element	-	4 (1,0%)	4 (0,47%)
Straße→Road Element/Intersection	2 (0,44%)	3 (0,75%)	5 (0,59%)
Straße/Weg→Road Element	17 (3,75%)	11 (2,76%)	28 (3,29%)
Straße/Fahrbahn→Road	-	2 (0,5%)	2 (0,23%)
Straße/Fahrbahn→Road Element/Road	-	2 (0,5%)	2 (0,23%)
Gesamtzahl	453	398	851

Tabelle II-17: Korrespondenzen der Klasse Fahrbahn (GM > 80)

Korrelationen Straße	Werte gesamt
Straße→Road Element	811 (97,13%)
Straße→Road	10,5 (1,26%)
Straße→Intersection	13,5 (1,61%)
Gesamtzahl	835

Tabelle II-18: Korrelationswerte der Klasse Fahrbahn (GM > 80)

Korrespondenzen Weg	Testgebiet 1	Testgebiet 2	Gesamt
Weg→Road Element	45 (72,58%)	55 (83,33%)	100 (78,12%)
Weg/Straße→Road Element	17 (27,42%)	11 (16,67%)	28 (21,87%)
Gesamtzahl	62	66	128

Tabelle II-19: Korrespondenzen der Klasse Weg (GM > 80)

Korrelationen Weg	Werte gesamt
Weg→Road Element	114 (100%)
Weg→Road	0 (0%)
Weg→Intersection	0 (0%)
Gesamtzahl	114

Tabelle II-20: Korrelationswerte der Klasse Weg (GM > 80)

Korrespondenzen Fahrbahn	Testgebiet 1	Testgebiet 2	Gesamt
Fahrbahn→Road Element	9 (14,29%)	2 (5,25%)	11 (10,0%)
Fahrbahn→Road	44 (69,84%)	35 (74,47%)	79 (71,81%)
Fahrbahn→Intersection	5 (7,94%)	-	5 (4,54%)
Fahrbahn→Road Element/Road	2 (3,17%)	2 (4,25%)	4 (3,64%)
Fahrbahn→Road Element/Intersection	1 (1,59%)	-	1 (0,91%)
Fahrbahn→Road/Intersection	2 (3,17%)	4 (8,51%)	6 (5,45%)
Fahrbahn/Straße→Road	-	2 (4,25%)	2 (1,82%)
Fahrbahn/Straße→Road Element/Road	-	2 (4,25%)	2 (1,82%)
Gesamtzahl	63	47	110

Tabelle II-21: Korrespondenzen der Klasse Fahrbahn (GM > 80)

Korrelationswerte der Klasse Fahrbahn:

Korrelationen Fahrbahn	Werte gesamt
Fahrbahn→Road Element	14 (12,96%)
Fahrbahn→Road	85,5 (79,17%)
Fahrbahn→Intersection	8,5 (7,87%)
Gesamtzahl	108

Tabelle II-20: Korrelationswerte der Klasse Fahrbahn (GM > 80)

Lebenslauf

von

Steffen Volz

11. August 1969	Geboren in Bietigheim-Bissingen
1977 – 1981	Grundschule in Sachsenheim/Württ.
1981 – 1990	Gymnasium im Ellental, Bietigheim-Bissingen
1990 – 1991	Zivildienst
1991 – 1998	Studium der Geographie an der Universität Stuttgart
1998 – 2006	Wissenschaftlicher Mitarbeiter in der Forschungsgruppe Geo-Informationssysteme am Institut für Photogrammetrie der Universität Stuttgart