



Diplomarbeit

**FUSION VON GEODATEN UNTERSCHIEDLICHER
QUELLEN IN GEODATENINFRASTRUKTUREN AM
BEISPIEL VON ATKIS UND OPENSTREETMAP**

Stefan Wiemann
Mat.-Nr.: 3011367

Betreut durch:
Prof. Dr. Lars Bernard, TU Dresden
Prof. Dr. Dirk Burghardt, TU Dresden
Eingereicht am 05.11.2009

ZUSAMMENFASSUNG

Die Zusammenführung von Geodaten auf Basis homologer Objekte ist ein wichtiger Teilprozess zur Wissensgenerierung aus verfügbaren Geoinformationen. Forschungen im Bereich der digitalen Geodatenfusion gibt es bereits seit Anfang der 80er Jahre. Das Aufgabenspektrum umfasst dabei die Aktualisierung, Veränderungsdetektion, Informationsanreicherung und Integration verfügbarer Datensätze. Gleichzeitig vollzieht sich seit Ende der 90er Jahre ein Paradigmenwechsel hin zum Aufbau dienstebasierter Geoinformationslandschaften auf Basis serviceorientierter Architekturen (SOA). Dieser wird insbesondere durch die Entwicklung einer Geodateninfrastruktur (GDI) im öffentlichen Sektor forciert und bildet einen Schwerpunkt der aktuellen Forschung im Bereich Geoinformatik. Innerhalb dieser interoperablen Strukturen kann ein entscheidender Informationsmehrwert durch die Kombination thematisch verwandter Ressourcen geschaffen werden. Die Fusion von Daten wird daher einen zentralen Bestandteil zukünftiger Entwicklungen im Bereich Web-basierter Anwendungen darstellen.

Zur Bereitstellung von Geodaten in einer GDI hat das *Open Geospatial Consortium* (OGC) bereits zahlreiche Standards veröffentlicht. Darüber hinaus eröffnet die Entwicklung des Web 2.0 weitere, oftmals Community-gestützte, Möglichkeiten zur Bereitstellung von Geodaten außerhalb standardisierter GDI. Die Verarbeitung dieser Geodaten kann durch die Einführung des OGC *Web Processing Service* (WPS) realisiert werden. Diese Schnittstellenspezifikation ermöglicht die Verlagerung von Geoprozessierungsfunktionalitäten in eine GDI und trägt somit zur Ablösung monolithischer Geoinformationssysteme (GIS) durch verteilte dienstebasierte Strukturen bei. Für die Umsetzung komplexer Prozesse wie einer Geodatenfusion ist die Verfügbarkeit, Interoperabilität und Verkettung beteiligter Dienste von entscheidender Bedeutung.

Nach der Einführung in Grundlagen von GDI und Geodatenfusion werden in dieser Arbeit Systemarchitektur und Bestandteile einer dienstebasierten Geodatenfusion konzipiert. Im Anschluss erfolgt die Beschreibung einer proof-of-concept Implementierung wesentlicher Bestandteile unter Nutzung des *52°North* WPS-Framework. Gegenstand der Implementierung ist die Fusion von Straßendaten der Modelle ATKIS (Amtliches Topographisch-Kartographisches Informationssystem) und OSM (OpenStreetMap) durch einen Feature- und Attributtransfer. Die Metadatenverarbeitung, Generalisierung und Evaluierung im Kontext einer dienstebasierten Geodatenfusion stellen weitere Teilaspekte dieser Arbeit dar.

ABSTRACT

The conflation of spatial data is one important task concerning the generation of knowledge from available geo-information. Research in this domain has been carried out since the early 80s and incorporates updating, change detection, enhancement and integration of spatial data. At the same time a paradigm shift leads towards service-oriented Architectures (SOA) in the field of geoinformation science. In the public sector this change is promoted by the development of spatial data infrastructures (SDI). Especially within these interoperable structures, the combination of thematically comparable resources can be used to enhance available spatial information. The conflation of data in general represents a core component of future research on web-based applications.

The *Open Geospatial Consortium* (OGC) has already published various standards for spatial data dissemination. In addition, the Web 2.0 development offers the possibility of user-generated spatial data beyond standardized SDI. The conflation of institutional- and community-provided datasets can be realized by the introduction of the OGC *Web Processing Service* (WPS). The WPS interface offers geoprocessing capabilities within SDI and thus helps distributed service-oriented environments to replace monolithic Geographic Information Systems. Availability, interoperability and chaining of services are crucial for implementing complex processes, such as conflation.

After an introduction to the fundamentals of SDI and conflation, a service-based architecture for geodata conflation will be developed within this thesis. The proof-of-concept implementation is realized using the *52°North* WPS and exercises the conflation of street data. For this purpose, the data models ATKIS (Authoritative Topographic Cartographic Information System) and OSM (Open-StreetMap) were applied to perform a transfer of attributes and features. Other important aspects of this thesis related to service-based conflation include the processing of metadata, generalization and evaluation.

INHALTSVERZEICHNIS

Zusammenfassung	ii
Abstract	iii
Abbildungsverzeichnis	vi
Verzeichnis der Tabellen und Listings	vii
Abkürzungsverzeichnis	viii
1 Einleitung	1
2 Geodateninfrastrukturen	4
2.1 Geodaten	4
2.1.1 Attribute	4
2.1.2 Metadaten	6
2.2 Serviceorientierte Architektur	6
2.3 Aufbau einer Geodateninfrastruktur	8
2.4 OGC Web Services	10
2.4.1 OGC Web Processing Service	11
2.5 Metadaten in Geodateninfrastrukturen	14
2.6 Aktuelle Defizite und Lösungsansätze	16
3 Geodatenfusion	19
3.1 Klassifikation und Anwendungsgebiete	19
3.2 Vorverarbeitung der Daten	22
3.3 Matching	24
3.3.1 Differenzierung vorhandener Ansätze	25
3.3.2 Kriterien für die Suche homologer Objekte	28
3.3.3 Matching von Straßennetzwerken	30
3.4 Zusammenführung von Datensätzen	31
3.5 Geodatenfusion im Kontext einer GDI	31
3.5.1 Monolithische und dienstebasierte Methoden	32
3.5.2 Herausforderungen und Probleme	33

4	Konzeption einer dienstebasierten Geodatenfusion	35
4.1	Betrachtung möglicher Eingangsdaten	35
4.2	Architektur	37
4.2.1	Granularität der Services	37
4.2.2	Prozessablauf	38
4.3	Prozessierungsschritte	40
4.3.1	Vergleich der Eingangsdaten	40
4.3.2	Anpassung	42
4.3.3	Matching	43
4.3.4	Fusion	45
4.3.5	Nachbearbeitung und Ergebnis	47
4.4	Wichtige Teilaspekte	47
4.4.1	Generalisierung	47
4.4.2	Metadaten	49
4.4.3	Service Chaining	50
4.4.4	Automatisierung und Performanz	53
4.4.5	Evaluierung	54
5	Implementierung	56
5.1	Ausgangsdaten	57
5.1.1	ATKIS	57
5.1.2	OpenStreetMap	57
5.2	Verwendete Software	59
5.3	Aufbau	59
5.3.1	Bestandteile	60
5.3.2	Durchführung der Fusion	65
5.4	Evaluierung der Ergebnisse	67
5.4.1	Ergebnisbetrachtung	67
5.4.2	Problemanalyse und Erweiterungsmöglichkeiten	70
6	Abschließende Betrachtung und Ausblick	73
6.1	Diskussion der Ergebnisse	73
6.2	Weiterer Forschungsbedarf	77
6.3	Fazit	79
	Literatur	81
	Anhang	96
	A: UML-Sequenzdiagramm	96
	B: CD mit Quellcode	97

ABBILDUNGSVERZEICHNIS

1	Geometrie und Topologie linearer Vektordaten	5
2	Klassifikation der Geodatenfusion	20
3	Allgemeiner Ablauf einer Geodatenfusion	22
4	Ausgewählte Problemfälle im Matchingprozess	25
5	Differenzierungsmöglichkeiten des Matchingprozesses	26
6	Architekturunterschied zwischen GIS und GDI	32
7	Zentrale Elemente des OGC-Feature-Modells	36
8	Aufbau feingranularer und grobgranularer Services	37
9	Ablauf einer dienstebasierten Fusion von Geodaten	39
10	Ablauf des Vergleiches von Eingangsdaten	40
11	Ablauf der Datenvorverarbeitung	42
12	Ablauf des Matching	44
13	Sequenzdiagramm der Fusion	46
14	Manuelles und Workflow-managed Service-Chaining	51
15	Implementierung des Ausgangsdatenvergleiches	61
16	Implementierung des Matchingprozesses	62
17	Ablauf des implementierten Fusionsprozesses	64
18	Verwendete Testgebiete in der Implementierung	66
19	Ergebnis der Fusion im Innenstadtbereich	68
20	Ergebnis der Fusion in der Dresdner Heide	69
21	Häufig auftretende Probleme beim Featuretransfer	71

TABELLENVERZEICHNIS

1	Operatoren der Modellgeneralisierung	24
2	Input und Output eines Datenvergleiches	41
3	Input und Output der Vorverarbeitung von Datensätzen	43
4	Input und Output eines dienstebasierten Matchings	45
5	Input und Output der Fusion	46
6	Charakterisierung der verwendeten Testgebiete	66

LISTINGS

1	Ausschnitt aus dem Ergebnis eines Datensatzvergleiches	61
2	Ergebnis aus dem Matching von Datensätzen	63
3	Aufruf eines WPS mittels PHP und cURL	65

ABKÜRZUNGSVERZEICHNIS

ATKIS	Amtliches Topographisch-Kartographisches Informationssystem
BPEL	Business Process Execution Language
EPSG	European Petroleum Survey Group Geodesy
ESDI	European Spatial Data Infrastructure
GDI	Geodateninfrastruktur
GIS	Geoinformationssystem
GML	Geography Markup Language
GSDI	Global Spatial Data Infrastructure
HTTP	Hypertext Transfer Protocol
IMAGI	Interministerieller Ausschuss für Geoinformationswesen
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
KVP	Key Value Pairs
MRDB	Multirepräsentationsdatenbank
OGC	Open Geospatial Consortium
OSM	OpenStreetMap
RDF	Resource Description Framework
SDI	Spatial Data Infrastructure
SOA	Service Oriented Architecture
SOS	Sensor Observation Service
UID	Unique Identifier
W3C	World Wide Web Consortium
WCPS	Web Coverage Processing Service
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service
WPS	Web Processing Service
WSDL	Web Services Description Language
XML	Extensible Markup Language

1 EINLEITUNG

Die Generierung von Wissen aus der Vielzahl verfügbarer Informationen zählt zu einer der wichtigsten Aufgaben des gegenwärtigen Informationszeitalters. Daher spielt deren effiziente Aufbereitung, Anreicherung, Verarbeitung und Verwendung eine entscheidende Rolle beim Aufbau wissensorientierter Strukturen.

Im Kontext der Geowissenschaften hat sich in den letzten Jahren der Begriff Geodateninfrastruktur¹ (GDI) etabliert. Dieser steht synonym für den Wechsel von datenorientierten hin zu dienstorientierten Strukturen und umfasst dabei technische, rechtliche, organisatorische, fachliche und soziale Aspekte einer Geoinformationslandschaft [Wytzisk und Sliwinski (2004)]. GDIs sollen nach Bernard et al. (2005b) „[...] über System- und Verwaltungsgrenzen hinweg die kooperative Nutzung verteilter GI-Dienste ermöglichen.“ und beschreiben somit einen Paradigmenwechsel von bisher dominierenden monolithischen Geoinformationssystemen (GIS) hin zu offenen interoperablen Systemen.

Technisch gesehen basiert eine Geodateninfrastruktur zumeist auf den Prinzipien einer serviceorientierten Architektur (SOA), welche auf der Organisation und Nutzung verteilter Funktionalitäten, bereitgestellt durch unabhängige Services, beruhen [OASIS (2006)]. Die Vorteile einer SOA im Vergleich zu geschlossenen Systemen liegen in der Autonomie, losen Kopplung und Orchestrbarkeit von Services, bedingt durch die Verwendung standardisierter Schnittstellen [Heutschi (2007)]. Durch die Verwendung des Internets als Kommunikationsinfrastruktur erfolgt die Umsetzung von GDI-Komponenten in der Regel über Web Services, in dieser Arbeit auch als Dienste bezeichnet.

Kennzeichnende Probleme der konventionellen Geodatenhaltung sind eine hohe Anzahl verschiedener, teils proprietärer Datenformate und Schnittstellen sowie heterogene Metadatenstrukturen [McKee (2004)]. Für den Aufbau einer GDI werden daher einheitliche Standards benötigt, bei deren Erarbeitung, neben den jeweiligen Gesetzgebern, das 1994 gegründete *Open Geospatial Consortium* (OGC) eine tragende Rolle spielt. Mit der Schaffung offener Schnittstellen wird die Veröffentlichung zahlreicher Geofunktionalitäten im Internet standardisiert und die technische Grundlage für den Aufbau einer GDI gelegt. Ausge-

¹abgeleitet aus dem englischen Begriff der „Spatial Data Infrastructure“ (SDI)

hend von den Spezifikationen des OGC können derzeit Visualisierungs-, Daten-, Katalog- und Prozessierungsdienste unterschieden werden. Dienstetypen zur Prozessierung von Geodaten bilden dabei die Grundlage der vorliegenden Arbeit und werden bereits in Bernard et al. (2005a) als zentraler Bestandteil bestehender GDI-Referenzarchitekturen identifiziert. Die OGC Schnittstellenspezifikation des *Web Processing Service* (WPS) dient genau diesem Zweck und macht weite Teile vorhandener Funktionalitäten zur Geoprocessing in einer GDI nutzbar. Der WPS-Standard ermöglicht somit die nutzerspezifische Analyse und Verarbeitung verfügbarer Geodaten und, mit der Veredelung von Daten zu Informationen, die Umsetzung einer echten serviceorientierten Struktur [Stollberg et al. (2007)].

Die Akquise und Verwertung von Geodaten wird oftmals durch die Existenz zahlreicher Datensätze unterschiedlicher Herkunft erschwert, da sich diese zu meist in der Erfassungsmethodik, Datenstruktur oder attributiven Ausstattung unterscheiden. Eine wichtige Aufgabe zur Schaffung anwendungsspezifischer Datensätze stellt die Zusammenführung bzw. Fusion von Geodaten dar. Dabei werden Informationen mehrerer Quellen genutzt, um einen Informationsmehrwert durch Anreicherung von Datensätzen zu ermöglichen. Ein Ziel der aktuellen Entwicklungen ist die Schaffung automatisierter und interoperabler Verfahren zur Fusion in serviceorientierten Systemen [OGC (2009b)]. Die vorliegende Arbeit untersucht daher Möglichkeiten zur Umsetzung einer GDI-konformen Geodatenfusion unter Verwendung der WPS-Spezifikation des OGC. Den Schwerpunkt bildet die Konzeption einer dienstebasierten Geodatenfusion, bei der folgende Forschungsfragen im Vordergrund stehen:

Welche Vorteile bietet eine dienstebasierte Fusion von Geodaten im Kontext einer GDI?

Die Fusion von Geodaten ist im Bereich GIS ein etabliertes Verfahren zur Informationsveredelung. Daher ist zu prüfen, welche Vorteile bzw. Möglichkeiten sich aus der Verlagerung entsprechender Methoden in eine GDI ergeben.

Wie können Metadaten dazu beitragen, Teilprozesse und Ergebnisse einer Fusion zu optimieren?

Ein wichtiger Standard für Metadatenelemente innerhalb einer GDI ist unter Federführung der Initiative *Infrastructure for Spatial Information in Europe* (INSPIRE) bereits verabschiedet und stellt eine Kernkomponente der europäischen GDI-Entwicklung dar. Einflussmöglichkeiten von Metadaten auf die Optimierung, Automatisierung und Interoperabilität einzelner Fusionsprozesse müssen daher analysiert werden.

Welche Teilprozesse und Schnittstellen sind erforderlich, um eine Geodatenfusion innerhalb einer GDI zu ermöglichen?

Eine Fusion setzt sich in der Regel aus mehreren Komponenten zusammen. Aus diesem Grund stellt sich die Frage nach der Granularität verwendeter Dienste. Die konzipierte Systemstruktur darf den Grundsätzen einer SOA nicht widersprechen und muss gleichzeitig allen Anforderungen und Aspekten einer Geodatenfusion genügen.

Welche Anforderungen stellt eine ad-hoc Zusammenführung von Geodaten an eine GDI?

Ein bereitgestellter Dienst zur Fusion von Geodaten muss in der Lage sein, direkt auf Nutzereingaben zu reagieren und Ergebnisse dynamisch zu generieren. Zu betrachtende Schwerpunkte liegen daher auf der Performanz und Orchestrierung verwendeter Dienste. In diesem Zusammenhang müssen auch notwendige Voraussetzungen für eine dienstbasierte Fusion betrachtet werden.

Als Nachweis der Machbarkeit erfolgt neben der Konzeption eine prototypische Implementierung zentraler Teilprozesse. Darin wird eine Informationsanreicherung von Geodaten, basierend auf Straßendaten der Modelle ATKIS (*Amtliches Topographisch-Kartographisches Informationssystem*) und OSM (*OpenStreetMap*), durchgeführt.

Aufbau der Arbeit

Das folgende Kapitel beschäftigt sich mit dem Begriff der Geodateninfrastruktur und gibt einen Überblick zu Bestandteilen und Anwendungsmöglichkeiten einer GDI. Anschließend werden in Kapitel 3 Ablauf und Teilprozesse einer Geodatenfusion beschrieben. In Kapitel 4 folgt die Konzeption einer dienstbasierten Geodatenfusion sowie die Betrachtung zentraler Teilaspekte. Darauf aufbauend wird in Kapitel 5 die proof-of-concept-Implementierung und deren Evaluierung beschrieben. Abschließend werden die Ergebnisse in Bezug auf die Forschungsfragen diskutiert und bewertet (Kapitel 6).

2 GEODATENINFRASTRUKTUREN

Die zunehmende Verbreitung des Internets und digital erfasster Geodaten begründen den Trend innerhalb der Geowissenschaften zum Aufbau von Geodateninfrastrukturen. Diese soll sich neben bereits existierenden Infrastrukturen im öffentlichen Sektor etablieren und einem breiten Nutzerkreis Geoinformationen zur Verfügung stellen. Im folgenden Kapitel werden die für diese Arbeit relevanten Konzepte und Begriffe einer GDI erläutert.

2.1 GEODATEN

„Daten, die einen Raumbezug aufweisen, über den ein Lagebezug zur Erdoberfläche hergestellt werden kann“, können nach Bollmann und Koch (2001) (Band 1, S. 289) als Geodaten bezeichnet werden. Die Indizierung von Daten über geographische Positionen ist eine effektive und weit verbreitete Methode, digitale Daten zu organisieren [OGC (1998)]. Unbelegten Schätzungen nach fallen etwa 80 Prozent aller erfassten Daten in diese Kategorie.

Grundlegend lassen sich Geodaten in Primär- und Sekundärdaten unterscheiden. Primärdaten, oftmals auch als Geobasisdaten bezeichnet, sind alle direkt, beispielsweise durch GPS-Messung oder Satellitenbilder, erfasste und nicht wesentlich veränderte Daten. Dahingegen werden daraus abgeleitete Produkte, wie etwa kartographische Darstellungen, als Sekundärdaten bezeichnet.

2.1.1 Attribute

Die Eigenschaften von Objekten innerhalb eines Geodatensatzes werden in beschreibenden Attributen festgehalten. Diese umfassen sowohl Geometrie und Topologie als auch thematische Informationen der Objekte.

Geometrie und Topologie

Geometrien ermöglichen in Kombination mit Positionsangaben die Verortung von Geobjekten innerhalb eines geographischen Referenzsystems und somit einen

direkten Raumbezug. Dabei können fünf Dimensionen unterschieden werden [vgl. Longley et al. (2005)]:

- der Punkt als diskretes, auf einer Koordinate beruhendes Geoobjekt, bezeichnet als 0. Dimension,
- die Linie als Verbindung zweier Punkte im Raum als 1. Dimension,
- die Fläche als 2-dimensionales Objekt im Raum, definiert durch eine umschließende Grenzlinie,
- das Volumen als 3-dimensionales Objekt im Raum, definiert durch eine umschließende Grenzfläche,
- die Zeit mit der Erfassung temporaler Änderungen von Objekten als 4. Dimension.

Nicht verwechselt werden darf die Geometrie mit der kartographischen Darstellung, welche auf der Kombination von Geometrie und festgelegten Zeichenvorschriften beruht und so eine Visualisierung ermöglicht. Je nach Beschaffenheit und Verwendung der Datensätze können Geometrien, basierend auf Punktkoordinaten, als Vektordaten oder, basierend auf Pixeln, als Rasterdaten verwaltet werden.

Durch die Topologie werden räumliche Beziehungen der Objekte, wie beispielsweise Nachbarschaften oder Überschneidungen, festgehalten. Sie ist dabei invariant gegenüber räumlichen Transformationen und zählt zu den wichtigsten grundrissbezogenen Informationen von Geodaten. Die Topologie kann zumeist aus der Geometrie abgeleitet und zur Validierung oder Verarbeitung von Geodaten genutzt werden. Das Zusammenspiel von Geometrie und Topologie ist in Abbildung 1, exemplarisch für ein Straßennetzwerk, dargestellt.

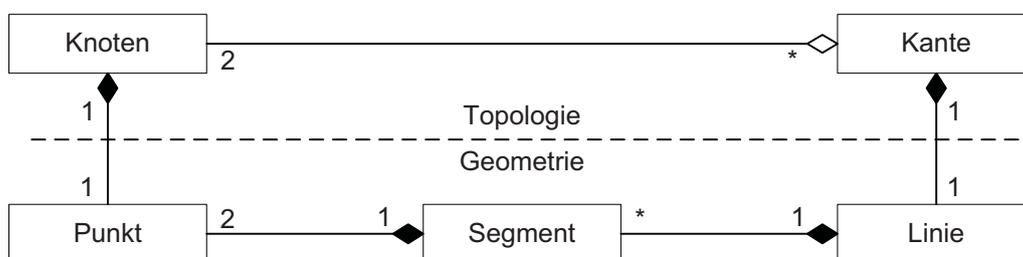


Abbildung 1 – Zusammenhang zwischen Geometrie und Topologie in linearen Vektordatensätzen, beispielsweise einem Straßennetzwerk. Abbildung nach Foerster (2004), Abb. 3, S. 19

Thematische Information

Einen Mehrwert erhalten Geodaten über näher beschreibende Fachdaten, die thematischen Informationen. Diese können entweder einzelnen Geoobjekten zugeordnet oder als Kontinuum dargestellt werden und unter anderem Beschreibungen, Messwerte oder Klassifikationen enthalten. Entsprechende Attribute basieren im Regelfall auf fachspezifischen Termini und können über Ontologien¹ und Datenmodelle abgebildet werden. Thematische Informationen sind ein essentieller Bestandteil von Geodaten und spielen bei der Analyse und Verarbeitung von Geodaten eine wesentliche Rolle.

2.1.2 Metadaten

Metadaten sind strukturierte Informationen über Geodaten und können unter anderem Informationen zur Thematik, Struktur oder geometrischen Genauigkeit eines Datensatzes beinhalten. Das Konzept der Metadaten wird in vielen Bereichen, beispielsweise dem Bibliothekswesen, angewandt und kann zur Katalogisierung, Suche und Verarbeitung von Geodaten beitragen. Auch weitergehende Analysen, wie die Suche nach Redundanzen oder Lücken in großen Datenbeständen, sind mit Hilfe von Metainformationen möglich [vgl. Longley et al. (2005)]. Nähere Informationen zu Metadaten, insbesondere hinsichtlich der Verwendung in einer Geodateninfrastruktur, finden sich im Abschnitt 2.5 dieser Arbeit.

2.2 SERVICEORIENTIERTE ARCHITEKTUR

In vielen Bereichen der IT-Branche spielt das Konzept der serviceorientierten Architektur eine wichtige Rolle. Die meist zitierte Definition einer SOA findet sich in OASIS (2006):

„Service Oriented Architecture (SOA) is a paradigm for organizing and utilizing distributed capabilities that may be under the control of different ownership domains.“

Eine allgemein gültige Definition einer SOA gibt es nicht, da viele Autoren, je nach Sichtweise, unterschiedlich interpretieren und ergänzen. Vielen gemeinsam sind folgende Merkmale [vgl. Heutschi (2007), Richter et al. (2005)]:

¹ Konzeptuelle formale Beschreibung der Objekte und Beziehungen [Dostal et al. (2005)]

- die verteilte Architektur, zusammengesetzt aus fachlichen, funktional fest definierten Anwendungsbausteinen,
- eine lose Kopplung der Bausteine über die Verwendung von Services,
- der Zugriff auf Services über einheitliche Schnittstellen.

Services, als wesentliche Bestandteile einer SOA, sind nach Dostal et al. (2005) für bestimmte Aufgaben konzipierte Softwarekomponenten, die über ein Netzwerk veröffentlicht und genutzt werden können. Die lose gekoppelte Kommunikation zwischen Services führt zu vielseitigen Nutzungsmöglichkeiten und Wiederverwendbarkeit und kann nach He (2003) durch die Minimierung künstlicher Abhängigkeiten erreicht werden. Der Austausch von Services ist ohne Beeinflussung benachbarter Komponenten möglich, wodurch insbesondere die Fehlerbehandlung und Wartung von Systemen vereinfacht werden kann. Des Weiteren kann die Entwicklung neuer Anwendungen von der Wiederverwendung bereits vorhandener Bausteine profitieren.

Web Services gelten als eine der aussichtsreichsten Möglichkeiten zur Umsetzung einer SOA, in diesem Fall in eine *Web services-oriented architecture* [Kreger (2003)]. Web Services müssen nicht von Grund auf neu entwickelt werden, sondern können auf bereits etablierten Standards, wie dem *Hypertext Transfer Protocol* (HTTP), SOAP² oder der *Extensible Markup Language* (XML), aufbauen [Papazoglou und van den Heuvel (2007)].

Die Nutzung von Diensten basiert auf der Kommunikation zwischen Nutzer, Dienstverzeichnis und Anbieter, dem sogenannten Publish-Find-Bind Paradigma [vgl. Kreger (2001)]:

1. Der Anbieter registriert einen von ihm bereitgestellten Dienst in einem Dienstverzeichnis (*Publish*).
2. Der Nutzer, oftmals eine weitere Softwarekomponente, sucht im Dienstverzeichnis nach einem für ihn geeigneten Dienst. Als Ergebnis erhält er einen Verweis auf den registrierten Dienst des Anbieters und Informationen zu dessen Schnittstelle (*Find*).
3. Der Nutzer kontaktiert den Dienst über die angebotene Schnittstelle und kann auf bereitgestellte Funktionalitäten zugreifen (*Bind*).

²SOAP: nicht als Akronym verwendet (früher: *Simple Object Access Protocol*)

Die in SOAs meist verbreiteten Dienststandards stellen nach Papazoglou und van den Heuvel (2007) die *Web Services Description Language* (WSDL) für die Schnittstellenbeschreibung, *Universal Description, Discovery and Integration* (UDDI) für Katalogdienste und SOAP für die Kommunikation dar.

Kennzeichnend für eine serviceorientierte Architektur und damit verbundene Dienstelandschaft sind [vgl. Richter et al. (2005), Papazoglou und van den Heuvel (2007), He (2003)]:

- Autonomie und lose Kopplung von Diensten,
- Kapselung technischer Details hinter der Schnittstellenbeschreibung,
- standardisierte maschinenlesbare Schnittstellen,
- Orchestrierbarkeit von Diensten.

Eine hohe Anzahl von Services und Servicevarianten kann nach Aier und Schelp (2008) allerdings zu Problemen führen, da die Komplexität der geschaffenen Netzwerke größer werden kann, als die der abgelösten Monolithen. Dies kann sich langfristig negativ auf Qualität, Performanz oder Flexibilität auswirken. Zudem fehlt in vielen Bereichen eine Einigung auf einheitliche Modelle und Schnittstellen, um das gegebene Potential ausschöpfen zu können.

2.3 AUFBAU EINER GEODATENINFRASTRUKTUR

Die Umsetzung der SOA-Prinzipien im Bereich der Geowissenschaften bildet die Grundlage für den Aufbau einer Geodateninfrastruktur. Zahlreiche Probleme bei der Verbreitung von Geodaten über das Internet sollen dabei mit dem Wechsel von daten- zu serviceorientierten Strukturen behoben werden. Zu den Zielen entstehender Geodateninfrastrukturen zählen Zeit- und Kostenreduktion sowie Steigerung von Flexibilität, Interoperabilität und Effizienz durch die dienstebasierte Bereitstellung von Geoinformationen [Rajabifard et al. (2002)].

Je nach Definition werden die technischen, prozessorientierten, organisatorischen, ökonomischen und anwenderbezogenen Aspekte einer GDI unterschiedlich gewichtet. Zahlreiche Anwendungsfelder und -hierarchien begründen dabei eine große Anzahl verschiedener Definitionen. Einige davon finden sich in Wytzisk und Sliwinski (2004), welche die Rolle des Nutzers in den Vordergrund stellen und wie folgt definieren:

„[...] a SDI can be understood as a multi-levelled, scalable, and adaptable collection of technical and human services, which are interconnected across system, organisational, and administrative boundaries via standardized interfaces.“

Hierarchisch können bestehende GDI-Initiativen relativ eindeutig voneinander abgegrenzt werden:

Die Schaffung einer *Global Spatial Data Infrastructure* (GSDI) und damit zusammenhängende Aufgaben, Herausforderungen, Probleme und weitere Aspekte sind im GSDI-Cookbook [GSDI (2004)] dargelegt.

Auf europäischer Ebene zielt die INSPIRE-Initiative der EU-Kommission auf eine einheitliche GDI innerhalb ihrer Mitgliedsstaaten zur Schaffung einer *European Spatial Data Infrastructure* (ESDI).

Die durch INSPIRE geschaffenen gesetzlichen Richtlinien fließen in Deutschland sowohl in die Bundes- als auch Landesgesetzgebung ein. Auf Bundesebene ist der *Interministerielle Ausschuss für Geoinformationswesen* (IMAGI) als Koordinationsgremium mit der Entwicklung einer deutschen GDI-DE beauftragt. Weitere Informationen zur Entwicklung einer GDI-DE und regionaler GDI in Deutschland finden sich unter anderem in BKG (2005a), Lenk (2005), Reindl (2005) und Riecken (2005).

Die Entwicklung von Geodateninfrastrukturen folgt dabei den folgenden festgelegten Prinzipien [vgl. Bilo und Bernard (2005)]:

- *Subsidiarität* (Erhebung und Pflege der Geodaten an der Stelle, wo die entsprechende Expertise vorhanden ist),
- *Interoperabilität* (Möglichkeit der direkten Kombination von Informationen unterschiedlicher Quellen),
- *Skalierbarkeit* (Austausch von Daten auf mehreren Verwaltungsebenen),
- *Datenpolitik* (ausreichende Verfügbarkeit von Geodaten in allen Bereichen),
- *Transparenz* (einfache Suche nach verfügbaren Geoinformationen und deren Nutzungsbestimmungen).

Der Aufbau einer GDI ist überall dort von Bedeutung, wo dezentral gehaltene Geoinformationen in einem Netzwerk verfügbar gemacht werden sollen. Dadurch bedingt sich eine große vertikale Streuung entsprechender Initiativen sowohl im öffentlichen als auch im privatwirtschaftlichen Sektor. Ziel kann dabei

allerdings nicht nur die Bereitstellung von Geodaten, sondern muss nach Bernard und Streit (2002) ein realer Informationsgewinn für den Nutzer sein. Wichtig ist, das öffentliche Interesse an der Entwicklung einer Geodateninfrastruktur zu wecken, indem deren realer Nutzen aufgezeigt wird. Dies kann nach Williamson (2004) am ehesten durch Anwendung in aktuellen politischen Fragestellungen, beispielsweise dem Umweltschutz, Katastrophenmanagement oder der Terrorismusbekämpfung, erreicht werden. Des Weiteren stellt sich die Frage nach der ökonomischen Effizienz einer GDI. Longhorn und Blakemore (2008) stellen dafür Kosten-Nutzen-Analysen bestehender Initiativen auf. Sie stellen fest, dass zwar der Nutzen die Kosten überwiegen wird, genaue Zahlen allerdings schwierig zu berechnen sind, da die meisten GDI-Initiativen diesbezüglich noch nicht abschließend bewertet werden können.

Den Kern der technischen Umsetzung einer GDI bilden die Schnittstellenspezifikationen für Geoinformationsdienste. Mit deren Entwicklung beschäftigt sich das 1994 gegründete *Open Geospatial Consortium*³, ein internationales Industriekonsortium mit Mitgliedern aus Wirtschaft, Wissenschaft und öffentlicher Verwaltung. In enger Kooperation mit anderen Organisationen, unter anderem der *International Organization for Standardization*⁴ (ISO), wird somit die technische Grundlage für den Aufbau einer Geodateninfrastruktur gelegt.

2.4 OGC WEB SERVICES

Web Services im Allgemeinen beschreibt Tidwell (2000) als: „[...] self-contained, self-describing, modular applications that can be published, located, and invoked across the Web“. Die im Kontext einer Geodateninfrastruktur entwickelten OGC Web Services bilden dabei nach Doyle und Reed (2001) „[...] a web of geoprocessing services that can be connected in dynamic, open interoperable chains to create dynamic applications“. Die bisher entwickelten Diensteschnittstellen können, angelehnt an ISO 19119, in vier abstrakte Kategorien eingeteilt werden [vgl. Müller und Portele (2005)]:

- *Data-Services* dienen zum einen der Visualisierung und zum anderen der Bereitstellung von Geodaten. Bestehende Standards sind:
 - *Web Map Service* (WMS) zur Visualisierung,
 - *Web Feature Service* (WFS) zur Bereitstellung von Vektordaten,

³bis 2004: *Open GIS Consortium*

⁴ISO/OGC-Agreement: http://www.isotc211.org/Agreements/Agreement_OGC.pdf

- *Web Coverage Service* (WCS) zur Bereitstellung von Rasterdaten,
- *Sensor Observation Service* (SOS) zur Bereitstellung von Sensordaten,
- *Processing-Services* dienen der Verarbeitung bzw. Veredelung von Geodaten. Der wichtigste OGC-Standard in diesem Bereich ist der *Web Processing Service* (WPS),
- *Library-Services* sind Katalogdienste, die bei der Suche nach Geodaten und Geodiensten helfen. Der entsprechende OGC-Standard für diese Aufgabe ist der *Web Catalogue Service* (CSW),
- *Human Interface Services* beschreiben die Clients, mit deren Hilfe ein Nutzer mit anderen Diensten kommunizieren kann.

Neben der reinen Bereitstellung von Geoinformationen gibt es den Trend hin zu transaktionalen Diensten, welche das Hinzufügen, Aktualisieren oder Löschen von Ressourcen erlauben. Über entsprechende Operationen der Schnittstelle wird beispielsweise die Pflege von Geodaten durch einen *WFS-Transactional* ermöglicht.

Prozessierungsdienste sind ein wichtiger Schritt zur Veredelung vorhandener Geodaten. Methoden des OGC *Filter Encoding* bieten bereits die Möglichkeit, einfache Prozessierungsschritte durchzuführen, können nach Kiehle et al. (2007) aber nicht als Geoprozessierung, sondern lediglich als eine andere Sicht auf Daten bezeichnet werden. Auch die Schnittstelle des *Web Coverage Processing Service* (WCPS), welche durch einen WCS implementiert werden kann und die Verarbeitung von Coverages ermöglicht [vgl. OGC (2009c)], ist an bereitgestellte Daten gebunden. Als eigenständiger Prozessierungsdienst dient daher der 2007 zum OGC Standard gereifte *Web Processing Service*.

2.4.1 OGC Web Processing Service

„WPS defines a standardized interface that facilitates the publishing of geospatial processes, and the discovery of and binding to those processes by clients.“ So definiert OGC (2007b) die WPS-Spezifikation, welche zur Durchführung beliebiger Datenprozessierungen innerhalb einer GDI genutzt werden kann. Das erklärte Ziel ist, über die Visualisierung und Datenbereitstellung hinaus zu kommen und weitere GIS-Funktionalitäten in eine GDI zu verlagern. Ein WPS ist in der Lage, jede Form von Daten, eingeteilt in *ComplexData*, *LiteralData* und *BoundingBoxData*, zu verarbeiten und als Ergebnis auszugeben. Entwickelte Pro-

zessierungsdienste können dabei nach räumlichen, thematischen, zeitlichen und Metadaten-Diensten unterschieden werden [OGC (2002)].

Eine Schlüsselrolle bei der Entwicklung von GDI spielt nach Alameh (2003) die Orchestrierung von Diensten, welche nutzerspezifische Anwendungen durch die Verkettung vorhandener Dienste ermöglicht. Nach ISO 19119 werden die folgenden drei Entwurfsmuster für das sogenannte *Service-Chaining* unterschieden und nach Sichtbarkeit aus Perspektive der Nutzer eingeteilt:

- In *User-defined Chains* entscheidet ausschließlich der Nutzer über die Verkettung von Diensten. (transparent),
- *Workflow-managed Chains* werden über eine Kontrollinstanz gesteuert. Die Dienstekette ist vom Nutzer einseh- und manipulierbar (transluzent),
- Bei Nutzung eines *Aggregate service* läuft die Verkettung von Diensten im Hintergrund ohne Kenntnis des Nutzers ab (opak).

Möglichkeiten zur Umsetzung eines WPS-Workflows auf Basis einer Orchestrierung werden in Kapitel 4.4.3 beschrieben.

Anwendungen

Bei der Umsetzung der WPS-Spezifikation lassen sich generische und problem-spezifische Implementierungen unterscheiden. Erstere stellen eine abstrakte Implementierung der WPS-Schnittstelle dar, in der konkrete Anwendungsmöglichkeiten im Hintergrund stehen. Bei der Umsetzung von Prozessierungsdiensten stellen diese Frameworks die Konformität mit der WPS-Spezifikation sicher. Im Folgenden werden drei von ihnen vorgestellt:

Das WPS Framework von *52°North*⁵ ist eine Java-basierte Umsetzung der WPS-Spezifikation in Version 1.0.0. Neben einigen Beispielprozessen wurde die SEXTANTE⁶ Bibliothek für Geoprozessierung eingebunden, wodurch ein umfangreiches Spektrum an grundlegenden Funktionalitäten bereitgestellt wird. Zudem wird in Brauner (2008) die Anbindung der Open Source GIS-Software GRASS (*Geographic Resources Analysis Support System*) an das Framework angeregt. Nähere Informationen zum Aufbau des *52°North* WPS finden sich in Foerster und Stoter (2006) sowie Foerster und Schäffer (2007).

⁵52°North WPS: <http://www.52north.org/wps>

⁶SEXTANTE Project: <http://www.sextantegis.com>

*PyWPS*⁷ basiert auf der Programmiersprache *Python* und bietet die Möglichkeit, GRASS-Funktionalitäten über eine WPS-Schnittstelle zu nutzen. Auch hier liegt eine Umsetzung der Version 1.0.0 der WPS-Spezifikation vor. Nähere Informationen zum *PyWPS*-Framework finden sich in Čepický und Becchi (2007).

Eine weitere Java-basierte Entwicklung ist das WPS-Framework von *deegree*⁸. Auch hier ist eine Umsetzung der Version 1.0.0 der WPS-Spezifikation geplant, derzeit vorliegend ist Version 0.4.0. Informationen zum *deegree* WPS finden sich in Kiehle et al. (2007).

Aufbauend auf diesen Umsetzungen können wichtige Aspekte der Geoprozessierung in einer GDI untersucht werden. Die Entwicklung eines transaktionalen WPS (WPS-T) durch Schaeffer (2008) oder die Anbindung der GRID-Computing-Technologie an einen WPS (WPS-G) durch Baranski (2008), jeweils unter Verwendung des WPS-Frameworks von *52°North*, sind Beispiele für derartige Analysen.

Anwendungsorientierte Umsetzungen bauen in der Regel ebenfalls auf generischen Implementierungen auf und behandeln problemspezifische Einsatzmöglichkeiten des WPS. Ein Mehrwert im Vergleich zu konventionellen, GIS-basierten Lösungen kann dabei insbesondere in zeitkritischen Anwendungen geschaffen werden. Am häufigsten finden sich Umsetzungen im Bereich des Katastrophen- und Risikomanagements, da hier ein ad-hoc Zugriff auf Geodaten und Geoprozessierung von großer Bedeutung sein kann. Konkrete Umsetzungen in diesem Rahmen werden in Weiser et al. (2006), Stollberg et al. (2007), Stollberg und Zipf (2007), Friis-Christensen et al. (2007) und Müller (2009) vorgestellt. Auch im Bereich der Online-Generalisierung gibt es bereits zahlreiche Forschungsansätze zur Verwendung des WPS. Nachdem bereits Burghardt et al. (2005b) die Bedeutung und mögliche Ansätze einer Web-basierten Generalisierung beschrieben haben, finden sich WPS-basierte Umsetzungen in Foerster und Stoter (2006), Foerster et al. (2009) und Bergenheim et al. (2009).

Schnittstellenbeschreibung

Die technische Umsetzung eines WPS ist unabhängig von Programmiersprache oder Betriebssystem, da Implementierungsdetails hinter der Schnittstelle gekapselt werden. Diese definiert drei mandatorische Operationen, die jeder WPS

⁷PyWPS: <http://pywps.wald.intevation.org>

⁸deegree: <http://www.deegree.org>

anbieten muss [vgl. OGC (2007b)]:

- *GetCapabilities* ist mandatorisch für alle OGC Web Services und gibt allgemeine Informationen zum Service, beim WPS im Wesentlichen eine Auflistung der angebotenen Prozesse, an den Client zurück.
- *DescribeProcess* stellt detaillierte Informationen zu angefragten Prozessen, darunter eine ausführliche Beschreibung und Angaben zu Ein- und Ausgabedaten, zur Verfügung.
- *Execute* führt einen Prozess auf Grundlage der im DescribeProcess-Dokument aufgelisteten Parameter aus. Während der Durchführung sind, sofern vom Service unterstützt, auch Informationen zum aktuellen Status der Prozessierung möglich.

Die Anfragen durch den Client können über HTTP/GET in KVP (*Key Value Pairs*) oder über HTTP/POST in XML codiert erfolgen. Mandatorisch sind HTTP/GET bei den Operationen *GetCapabilities* und *DescribeProcess* sowie HTTP/POST bei der *Execute*-Operation. Auch die im Web bereits weit verbreiteten Standards SOAP für die Datenübertragung und WSDL für die Beschreibung von Web Services sind mit dem WPS-Standard vereinbar. Eine Methode zur automatisierten Ableitung von SOAP-Interfaces und WSDL-Metadaten aus den mandatorischen Operationen des WPS diskutieren Sancho-Jiménez et al. (2008).

Zur Registrierung und eindeutigen Spezifizierung von Prozessen unterstützt der WPS Standard sogenannte *Application Profiles*. Diese bestehen mindestens aus einem eindeutigen Identifikator und einer Referenz zum DescribeProcess-Dokument des jeweiligen Prozesses. Sie sollen die Interoperabilität verbessern und die Suche nach geeigneten Prozessierungsdiensten vereinfachen [OGC (2007b)].

2.5 METADATEN IN GEODATENINFRASTRUKTUREN

Im GIS-Kontext bereits etabliert, gewinnen Metadaten auch innerhalb von Geodateninfrastrukturen zunehmend an Bedeutung, wobei zwischen geodaten- und servicebeschreibenden Metadaten unterschieden werden muss.

Beschreibung von Geodaten und Diensten

Ein im Internet weithin anerkannter und als ISO 15836 bereits standardisierter Satz von Metadatenelementen für digitale Datensätze wurde bereits durch die

1995 gegründete *Dublin Core Metadata Initiative*⁹ definiert. Dem Anspruch einer Geodateninfrastruktur können diese allerdings nur zu Teilen gerecht werden, da Geodaten im Regelfall deutlich komplexer sind und eine GDI weit mehr darstellt als eine einfache Geodatenbank [Nogueras-Iso et al. (2005)]. Für Geoinformationen wurde daher die ISO 19115 Norm „Geographic information – Metadata“ entwickelt, welche als Teil der vom ISO/TC 211 definierten 191xx-Reihe zur Normung von Geoinformationen eine wichtige Rolle beim Aufbau von Geodateninfrastrukturen spielt. Eine entsprechende Umsetzung im Zuge der INSPIRE-Initiative wird über die „Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich Metadaten“ [EU (2008)] geregelt.

Metadaten bilden die Grundlage für Daten- und Dienstekataloge und müssen daher sowohl durch Nutzer als auch Maschine interpretierbar sein. Neben der Suche spielen servicebeschreibende Metadaten vor allem bei der Orchestrierung von Geodaten- und Geoprocessingdiensten eine wichtige Rolle. Ein OGC Web Service muss daher immer ausreichend formelle Metadaten zur Verfügung stellen, um auf dieser Grundlage aufgerufen werden zu können [OGC (2002)]. Bei einem WPS sind somit mindestens Angaben zur URI (*Uniform Resource Identifier*) des Dienstes und der angebotenen Operationen sowie eine Beschreibung möglicher Ein- und Ausgabedaten notwendig. Servicebeschreibende Metadaten können oftmals aus den zu Grunde liegenden Ressourcen und deren Metadaten abgeleitet werden. Möglichkeiten dazu und weitere Informationen zur Verwendung von Metadaten in Geodateninfrastrukturen finden sich unter anderem in Nogueras-Iso et al. (2005) und Najar (2006).

Der Weg zum Semantic Web

Derzeit beziehen sich Metadaten vor allem auf syntaktische Informationen zu Daten und Diensten. Aus diesem Grund muss deren Integration in eine GDI oftmals durch Fachkräfte erfolgen [Fensel (2007)]. Bei der Erstellung automatisierter und autonom handelnder Netze gewinnt die Semantik von Daten und Diensten an Bedeutung. Diese kann in Metadaten, unter anderem in Form von Ontologien, festgehalten werden. Die Bedeutung von Daten und Diensten, unter Verwendung eines gemeinsamen Vokabulars und einer formalen Beschreibung der Struktur, kann somit maschinenlesbar beschrieben werden [Dostal et al. (2005)].

Zur Darstellung von Ontologien wurden durch das *World Wide Web Consortium* (W3C) die Metadatenstandards des *Resource Description Framework* (RDF)

⁹Dublin Core Metadata Initiative: <http://dublincore.org/documents/dcmi-terms/>

und der *Web Ontology Language* (OWL) entwickelt. Zusammen mit den bereits genannten Standards für Web Services ist die Entwicklung zu einem "service-oriented Semantic Web" möglich [Fensel (2007)]. Zuständig für die Erstellung einer entsprechenden Architektur zur Unterstützung semantischer Web Services ist das *Architecture Committee der Semantic Web Services Initiative*¹⁰ (SWSI) [Burstein et al. (2005)].

2.6 AKTUELLE DEFIZITE UND LÖSUNGSANSÄTZE

Trotz der Erfolge bei der Umsetzung von Geodateninfrastrukturen gibt es noch zahlreiche Defizite, Gegensätze und Probleme, die in den folgenden Jahren analysiert und beseitigt werden müssen. Erst dann wird sich zeigen, inwieweit die bereits gesteckten Ziele erreicht werden können.

Probleme im Aufbau einer GDI

Probleme einer traditionell GIS-dominierten Anwendungslandschaft sind zumeist heterogene Datenstrukturen, Datenformate, Schnittstellen oder Taxonomien. Dadurch isolierte Geodaten führen zu einer teuren redundanten Datenhaltung und zu Schwierigkeiten beim Austausch von Informationen [McKee (2004)].

Viele der technischen Interoperabilitätsprobleme wurden im Zuge der Entwicklung von Standards durch OGC und ISO bereits angegangen und zum Teil überwunden. Größere Probleme bereitet allerdings die Handhabung semantischer Heterogenität von Geodaten und Geoinformationsdiensten. Deren Ursache sind Unterschiede zwischen natürlichen Sprachen, verschiedenen Anwendungsbereichen, Geoinformationsgemeinschaften, Objektdefinitionen und grundlegenden Notationen [Pundt (2005)]. Die Schaffung semantischer Interoperabilität ist derzeit Gegenstand zahlreicher Forschungsansätze mit dem Ziel der Verknüpfung von Daten auf der Ebene ihrer Bedeutung. Das Ziel definiert Egenhofer (2002) wie folgt: „[...] geospatial information retrieval will become precise to the level that the results of user queries will be immediately useful, without weeding out irrelevant hits“

Die Interoperabilität von Geodiensten wird wesentlich zum Erfolg einer GDI beitragen, in der mangelnde Kooperationen, projektspezifische Umsetzungen und kurzfristige Planungen bisher oftmals zu Insellösungen führten. Um eine

¹⁰Semantic Web Services Initiative: <http://www.swsi.org>

global vernetzte, mannigfaltige GDI zu schaffen, müssen daher redundante Entwicklungen vermieden und Möglichkeiten zur Harmonisierung von Diensten geschaffen werden.

Ferner werden die fehlende Beachtung von Nutzerkenntnissen, welche in divergierenden Anforderungen an eine GDI resultieren können, fehlende Systeme zur Entscheidungsunterstützung, hohe Kosten für Geodaten und grundlegende Unterschiede in den angebotenen Geobasisdaten beteiligter Länder kritisiert [vgl. Bernard und Streit (2002), Craglia (2003)].

Defizite in der Web-basierten Prozessierung

Um eine Prozessierung von Geodaten durchführen zu können, sind eine gewisse Anzahl atomarer Prozessierungsdienste mit einem entsprechenden Spektrum an Funktionalitäten sowie Möglichkeiten zur Modellierung abstrakter Prozesse notwendig. Ob die aktuell verfügbaren Standards dafür ausreichen, ist umstritten. Fehlende formale Beschreibungen der WPS-Schnittstelle und die unzureichende Unterstützung etablierter Standards wie SOAP oder WSDL werden dabei häufig als Hemmnis der Flexibilität und Orchestrrierbarkeit von Diensten gesehen [Kiehle et al. (2007), Friis-Christensen et al. (2007), Stollberg et al. (2007)].

Eines der wesentlichen Defizite bei Betrachtung der Performanz im Vergleich zu konventionellen monolithischen Lösungen sind hohe Transferkosten zwischen Diensten, welche bei Codierung, Transport und Decodierung benötigter Daten entstehen. In diesem Zusammenhang wird auch die On-the-fly-Codierung von GML (Geography Markup Language), konzipiert als Standard für den Transfer von Vektorinformationen innerhalb einer GDI, kritisiert, da die Konvertierung von Geodatenbeständen viel Zeit kosten und in großen Datenpaketen resultiert kann [Kim et al. (2005)]. Die Reaktionszeiten einzelner Prozessierungsdienste können dadurch stark ansteigen und sich negativ auf den Gesamtprozess auswirken. Zu dieser Problematik gibt es, neben der Entwicklung leistungsfähigerer Rechner bzw. Rechnernetze und höherer Internet-Transferraten, mehrere Lösungsansätze. Zum einen kann eine Vorprozessierung von GML-Daten in Verbindung mit einem Caching-Algorithmus, wie ihn Kim et al. (2005) vorschlagen, die On-the-fly-Codierung von GML überflüssig machen. Unabhängig von GML wird zum anderen die Trennung von Geodaten in relevante und nicht relevante Teile diskutiert, um Transferkosten zu sparen, wobei eine entsprechende Selektion sowohl über räumliche als auch thematische Abfragen erfolgen kann [vgl. Friis-Christensen et al. (2007), Michaelis und Ames (2009)].

Des Weiteren wird, insbesondere im Kontext der Prozessierung großer Daten-

bestände, die Übertragung des Algorithmus an den Client, auch als *moving-code*-Paradigma bezeichnet, diskutiert [Bernard et al. (2005a), Friis-Christensen et al. (2007), Granell et al. (2007)]. Dies erfordert eine *client-side*-Strategie und entsprechende Kapazitäten des Client (*Thick Client*) [Worboys und Duckham (2004)]. Durch derzeitige OGC Standards wird dies nicht unterstützt.

Defizite im Metadatenmanagement

Das größte Problem der Metadaten in Geodateninfrastrukturen ist der Mangel an eben diesen, welcher die Katalogisierung und fachliche Nutzung von Geodaten und Geoinformationsdiensten erschwert. Köhler (2004) sieht die Ursache in bisherigen, meist projektorientierten und teils proprietären, Einzellösungen, in denen das Metadatenmanagement oft nur zweitrangig behandelt wird. Auch Najar (2006) stellt fest, dass es eine Lücke zwischen dem Status-Quo und den tatsächlichen Anforderungen an Metadaten gibt, wobei die Komplexität und unterschiedlichen Arten der Metadatenakquise sowie die separate Handhabung von Metadaten als Hauptgründe genannt werden. Die Komplexität der Metadaten bzw. Metadatenstandards führt demnach dazu, dass deren Aufbereitung und Aktualisierung nur mit entsprechender Expertise erfolgen kann, an der es jedoch zum Teil mangelt. Des Weiteren können separat gespeicherte Metadaten Zuordnungsfehler oder einen Verlust bei Datentransfers zur Folge haben, weshalb die Möglichkeit der Integration von Metadaten als direkter Bestandteil von Geodaten bzw. Geodatenformaten diskutiert wird. Dieser auch in Peng und Tsou (2003) beschriebene Ansatz würde jedoch erneut notwendige Transferleistungen erhöhen.

Die Metadateninitiative von INSPIRE ist ein erster Schritt zur Harmonisierung der Dokumentation von Daten und Diensten innerhalb einer europäischen GDI [Craglia (2003)]. Die Umsetzung der in EU (2008) festgeschriebenen Metadaten-elemente wird zukünftig sowohl die Suche als auch die Nutzung von Geodaten erleichtern und löst damit das Problem der inhaltlichen Interoperabilität.

3 GEODATENFUSION

Aufgrund der exponentiell wachsenden Zahl verfügbarer Geodaten steigt auch der Bedarf an qualifizierten Methoden zur Informationsextraktion. Ein wichtiges Element darin ist die Fusion von Geodaten, im Englischen auch als „Conflation“ bezeichnet. Diese soll redundante Datenhaltung minimieren und zu einer Aufwertung von Geoinformationen beitragen. Yuan und Tao (1999) definieren wie folgt:

„Conflation is the process of combining the information from two (or more) geodata sets to make a master data set that is superior to either source data set in either spatial or attribute aspect.“

Zentrale Bestandteile sind die Zusammenführung von Geodaten auf Basis attributiver oder semantischer Eigenschaften und eine Wertsteigerung durch den Transfer von Informationen. Theoretisch können dabei beliebig viele Datensätze fusioniert werden, wobei zusätzliche Datensätze unter anderem zur Verifizierung von Objektzuordnungen genutzt werden können. Zumeist wird jedoch, wie auch im Folgenden, von zwei Datensätzen ausgegangen.

3.1 KLASSIFIKATION UND ANWENDUNGSGEBIETE

Eine allgemeine Klassifikation der Geodatenfusion findet sich in Abbildung 2, wonach sich generell zwei Vorgehensweisen unterscheiden lassen. Die horizontale Fusion beschreibt eine Zusammenführung aneinandergrenzender, leicht überlappender Datensätze. Diese Randanpassung kann über das in Beard und Chrisman (1986) beschriebene Zipping durchgeführt werden und unter anderem zur Erweiterung von Routing-Datensätzen dienen [vgl. Foerster (2004)]. Der Schwerpunkt der vorliegenden Arbeit liegt dahingegen auf der vertikalen Geodatenfusion. Dabei werden Datensätze mit gleicher Gebietsabdeckung aufeinander registriert und fusioniert. Folgende Arten lassen sich dabei unterscheiden:

- Image–Image: Registrierung von Rasterdatensätzen über zumeist flächenbasierte Verfahren der Bildverarbeitung, beispielsweise durch Kreuzkorrela-

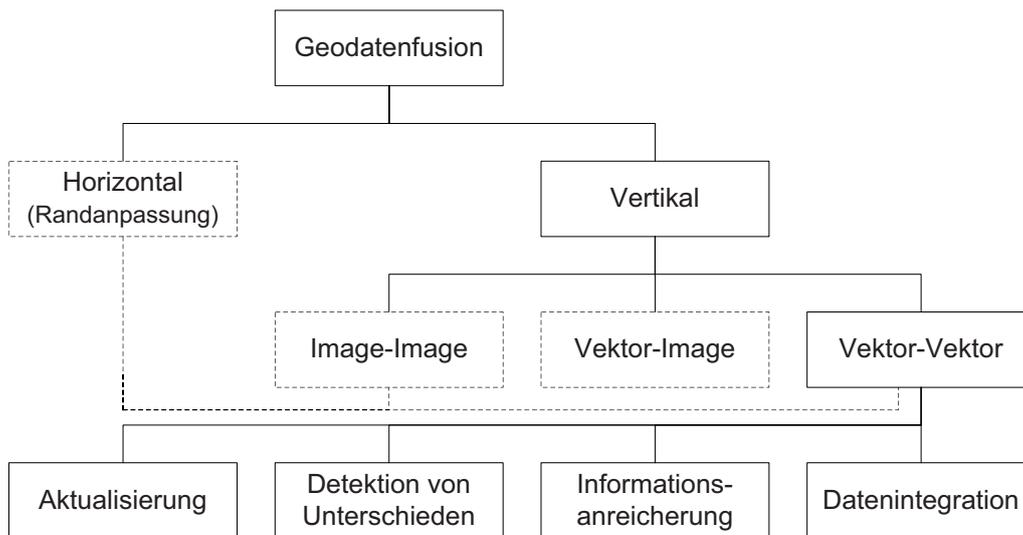


Abbildung 2 – Klassifikation der Geodatenfusion mit Fokus auf der vertikalen Vektor-Vektor Fusion (angelehnt an Yuan und Tao (1999)).

tion oder Least-Squares-Matching. Die Fusion erfolgt auf Basis homologer¹ Rasterelemente,

- Image-Vektor: Registrierung von Vektoren auf korrespondierende Elemente in einem Rasterbild. Dabei können beispielsweise radiometrische oder textuelle Informationen zu Objekten gewonnen oder Datensätze mit Hilfe von Fernerkundungsdaten aktualisiert werden,
- Vektor-Vektor: zumeist objektbasierte Zusammenführung von Vektordatensätzen über ein featurebasiertes Matching und den Transfer von Informationen.

Im weiteren Verlauf befasst sich diese Arbeit mit der Fusion von Vektordatensätzen, mit der primär folgende Anwendungen verbunden sind:

- Die Aktualisierung als qualitative Verbesserung von Datensätzen über das Ersetzen von alten Informationen durch Neuere. Dies kann entweder durch einen Attributtransfer, die Aktualisierung thematischer Informationen, oder einen Featuretransfer, die Aktualisierung geometrischer Informationen, erfolgen.
- Die Detektion von Unterschieden im Sinne einer klassischen Veränderungsdetektion oder als Möglichkeit zur Durchführung einer Fehleranalyse. Die Veränderungsdetektion (*Change detection*) bildet die Grundlage für einen

¹homolog: aus dem griechischen *homologeō* = übereinstimmen

temporalen Vergleich von Datensätzen. Dabei werden geometrische oder thematische Informationen und deren zeitliche Veränderungen erfasst und analysiert. Die Feststellung attributiver Abweichungen der Datensätze kann allerdings auch für eine Fehleranalyse verwendet werden. Der Vergleich von Aufnahmetechniken hinsichtlich geometrischer und thematischer Genauigkeit ist ein mögliches Beispiel.

- Eine Informationsanreicherung, die Steigerung der quantitativen Ausstattung von Datensätzen, welche, wie die Aktualisierung, durch einen Attribut- oder Featuretransfer erfolgen kann. Im Unterschied dazu werden allerdings ausschließlich Informationen ergänzt, nicht aktualisiert. Eine Informationsanreicherung umfasst zudem die Ableitung neuer Informationen aus bereits fusionierten Daten.
- Die Datenintegration, bei der Zuordnungen gespeichert und Ausgangsdaten nach ihrer Fusion bestehen bleiben. Dies ist insbesondere bei Datensätzen unterschiedlicher Maßstäbe, Thematiken, Aufnahmezeitpunkte oder Datenmodelle sinnvoll und kann zum Aufbau von Multirepräsentationsdatenbanken² (MRDB) genutzt werden [Medeiros et al. (1996)].

Die erste Anwendung und Implementierung einer computergestützten Geodatenfusion fand bereits Mitte der 80er Jahre in einem Verbundprojekt des *United States Geological Survey* und *Bureau of the Census* statt [Saalfeld (1988)]. Die Entwicklung einer auf Geoobjekten basierenden Fusion brachte bei der Zusammenführung von Karten städtischer Gebiete zahlreiche Vorteile gegenüber den bis dahin verwendeten, zumeist abbildungsbasierten Verfahren. Auf Grundlage der in Rosen und Saalfeld (1985) beschriebenen Matching-Kriterien konnten die Datensätze teilautomatisiert und mit, für damalige Verhältnisse, akzeptablem Zeit- und Kostenaufwand fusioniert werden.

Der grundlegende Ablauf einer Geodatenfusion, wie sie in dieser Arbeit verstanden wird, ist in Abbildung 3 skizziert. Die darin wichtigsten Bestandteile Vorverarbeitung, Matching und Fusion der Datensätze werden im nächsten Abschnitt erläutert.

²MRDB: Darstellung mehrerer Repräsentationen eines Gebietes in einer Datenbank

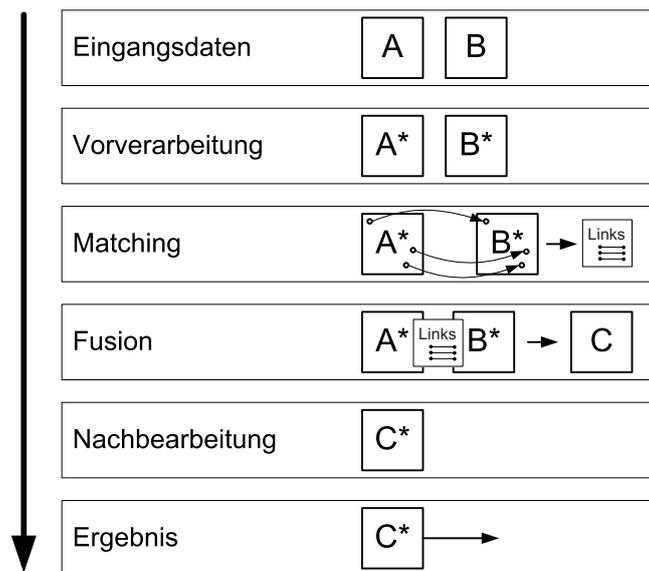


Abbildung 3 – Allgemeiner Ablauf einer Geodatenfusion. Zentrale Elemente sind die Vorverarbeitung, das Matching und die Fusion der Eingangsdaten. Das Ergebnis stellt ein anwendungsspezifisch fusionierter Datensatz dar.

3.2 VORVERARBEITUNG DER DATEN

Ausgangsdaten einer vertikalen Geodatenfusion sollten die gleiche räumliche Abdeckung besitzen, auf dem selben Koordinaten- und Referenzsystem beruhen, logisch miteinander vergleichbar sein und ausreichend Informationen für ein Matching beinhalten. Bei Letzterem spielt nicht nur die Quantität vorhandener Informationen, sondern auch die Qualität eine wichtige Rolle. Zu erfüllende Qualitätsparameter sind in der ISO-Norm 19113 "Geographic information – Quality principles" festgehalten [vgl. Kresse und Fadaie (2004)]:

- Vollständigkeit (weder Überschuss noch Mangel an Informationen),
- logische Konsistenz (keine Fehler in Datenstruktur, Geometrie, Topologie oder thematischer Information),
- Positionsgenauigkeit (absolute und relative Abweichungen geometrischer Lagebeziehungen),
- zeitliche Genauigkeit (Abweichungen von Zeitmessungen und anderen zeitbezogenen Informationen),
- thematische Genauigkeit (Genauigkeit quantitativer Attribute, Korrektheit nicht-quantitativer Attribute und der thematischen Klassifikation).

Qualitativ mangelhafte Datensätze können zu Zuordnungsproblemen, ungewünschten Fehlerfortpflanzungen, Informationsverlusten oder dem Abbruch

einer Prozessierung führen. Entsprechende Angaben sollten daher in Metadaten festgehalten sein und im weiteren Prozessverlauf berücksichtigt werden. Um eine Vergleichbarkeit herzustellen, muss oftmals eine Vorverarbeitung der Datensätze erfolgen, wobei notwendige Schritte in der Regel erst im Anwendungskontext festgelegt werden können.

Map Alignment

Unterschiedliche Aufnahmetechniken, Digitalisierungs- oder Transformationsfehler resultieren bei einer Überlagerung der Datensätze oftmals in geometrischen Abweichungen. Diese Passungenauigkeiten können Zuordnungsfehler im Matching verursachen und müssen beseitigt werden. Bereits Lupien und Moreland (1987) sehen aus diesem Grund das Feature- bzw. Map-Alignment, basierend auf einer topologieerhaltenden geometrischen Transformation, als Voraussetzung einer computergestützten Geodatenfusion. Als Referenz für die Anpassung dient im Regelfall der geometrisch genauere Datensatz.

Globale Abweichungen, wie sie beispielsweise durch systematische Fehler einer Koordinatentransformation entstehen, können zumeist durch eine einfache Affintransformation eines Datensatzes eliminiert werden. Lokale Abweichungen, etwa durch zufällige Fehler bei der Digitalisierung entstanden, verlangen dahingegen komplexere Methoden. Das am häufigsten eingesetzte Verfahren stellt dabei das *Rubber-Sheeting* dar, welches auf einer Triangulierung bekannter Punkte und anschließenden Netztransformation beruht. Weitere Informationen zur Anwendung und Durchführung einer *Rubber-Sheet-Transformation* finden sich unter anderem in Gillman (1985), Lupien und Moreland (1987) und Saalfeld (1988). Einen alternativen Ansatz zum Map Alignment beschreiben Doytsher et al. (2001), welche die Anpassung nicht über Punkte, sondern lineare Elemente des Datensatzes verfolgen.

Generalisierung

Methoden der Generalisierung bieten die Möglichkeit, Referenzmaßstäbe der Datensätze dahingehend anzupassen, dass eine direkte Vergleichbarkeit bezüglich Geometrie und thematischer Attribute gewährleistet werden kann. Im Kontext der Geodatenfusion ist vor allem die Modellgeneralisierung von Bedeutung, da die Visualisierung und damit verbundene kartographische Generalisierung der Daten im Hintergrund steht.

Unterschiede von Auflösung bzw. Feinheitsgrad und Repräsentationsform enthaltener Features sind bedingt durch geometrische, meist maßstabsbedingte

Tabelle 1 – Operatoren der Modellgeneralisierung. Ein Generalisierungsprozess besteht in der Regel aus einer Verknüpfung dieser Operatoren. Die Einteilung entspricht der von Foerster et al. (2007).

Operator	Beschreibung und Beispiel
Objektauswahl	Selektion oder Filterung von Elementen auf Grundlage geometrischer oder thematischer Attribute z.B. Auswahl von Straßenelementen mit Länge ≥ 1 km
Reklassifizierung	Bildung neuer Objektklassen durch Aggregation bestehender Klassen, zumeist auf Basis thematischer Attribute z.B. Laubwald + Mischwald = Wald (neue Objektklasse)
Collapse	Änderung der Repräsentation zu einer stärker abstrahierenden Darstellungsform (Polygon \rightarrow Polylinie \rightarrow Punkt) z.B. Skelettierung [Blum (1967)] oder Mittelachsenbildung [Thom (2005)]
Zusammenfassung	Bildung neuer Geometrien durch Vereinigung benachbarter Objekte z.B. Clustering [Burghardt et al. (2005a)]
Vereinfachung	Reduzierung der Datenmenge durch Stützpunktreduktion unter Beibehaltung von Grundform und Topologie z.B. Douglas–Peucker Algorithmus [Douglas und Peucker (1976)]
Vereinigung	Gruppierung von Objekten mit gleichen Eigenschaften zu neuen Objekten, ohne die Geometrie direkt zu beeinflussen z.B. Gruppierung von Objekten in der reklassifizierten Wald-Objektklasse

Differenzen der Datensätze. Attributive Differenzen sind dagegen Folgen unterschiedlicher thematischer Auflösungen und müssen dementsprechend separat gehandhabt werden. Eine Auflistung vorhandener Operatoren zur Modellgeneralisierung findet sich in Tabelle 1. Da zahlreiche Abhängigkeiten zwischen den Operatoren existieren, bestehen Generalisierungsprozesse zumeist aus deren Kombinationen.

Je nach Beschaffenheit und Anwendung ist es möglich, Differenzen zwischen Datensätzen auch direkt im Matching zu behandeln. Dabei können Kenntnisse über Attribute und die Semantik der Daten genutzt werden, um entsprechende Assoziationen zwischen Features herzustellen, ohne einen Informationsverlust durch Generalisierung in Kauf nehmen zu müssen.

3.3 MATCHING

Der Erfolg einer Fusion von Geodaten hängt wesentlich von der Identifizierung homologer Objekte in den Datensätzen ab. Dies ist aufgrund der teilweise sehr heterogen beschaffenen Geodaten allerdings eine der größten Herausforderungen. Das Problem ist, dass gleiche Objekte in der realen Welt im Datensatz nicht automatisch gleich abgebildet werden und daher unterschiedliche Lagebeziehungen, Repräsentationsformen, Auflösungen oder Klassifikationen aufweisen kön-

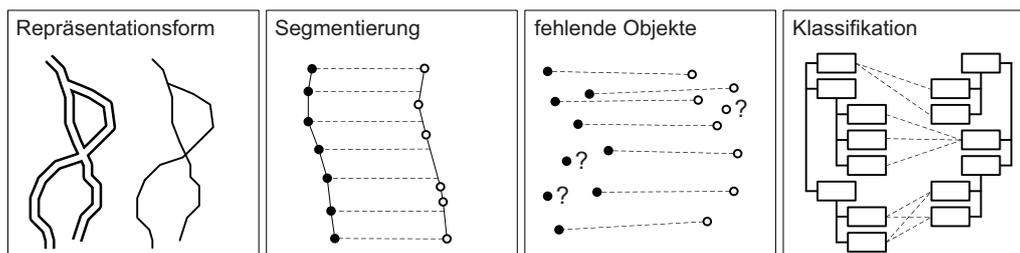


Abbildung 4 – Ausgewählte Problemfälle im Matchingprozess. Unterschiede in der Repräsentationsform, Segmentierung, Objektanzahl und Klassifikation sind oftmals Hindernisse bei der Suche nach homologen Objekten.

nen. Zudem ist es möglich, dass Objekte gänzlich fehlen oder sich in Geometrie, Topologie oder Thematik grundlegend voneinander unterscheiden. Mögliche Konflikte sind in Abbildung 4 dargestellt. Ursachen auftretender Differenzen sind:

- unterschiedliche Datenmodelle und damit unterschiedliche Definitionen, Bedeutungen und Handhabungen von Objekten,
- abweichende Auflösungen, Segmentierungen oder Formatstrukturen, resultierend aus verschiedenen Aufnahmetechniken,
- unterschiedliche Zeitpunkte der Aufnahmen.

Auch die geometrische Informationsdichte in Datensätzen kann Einfluss auf den Matchingprozess nehmen. Mit der Informationsdichte oder diesbezüglichen Unterschieden in den Datensätzen steigt oftmals die Wahrscheinlichkeit von Mismatchings [vgl. Walter (1996), Lüscher (2006)].

In der Literatur sind bereits zahlreiche Ansätze zum Matching von Geodaten vorhanden, weisen allerdings bislang wenig Verknüpfungspunkte auf. Der Bezug auf spezielle Einzelfälle lässt die Erfolgsquote der meisten Ansätze bei Anwendung auf andere Testgebiete erheblich sinken [Zhang et al. (2005)]. Die Entwicklung generischer Algorithmen bzw. deren Parametrisierung erweist sich aufgrund der bereits erwähnten Heterogenität von Geodaten als schwierig.

3.3.1 Differenzierung vorhandener Ansätze

Der Matchingprozess lässt sich über verwendete Eigenschaften, Aggregationshierarchie und Geometrie der Eingangsdaten differenzieren. Die entsprechenden Kategorisierungen sind in Abbildung 5 dargestellt.

Die erste Einteilung basiert auf den Methoden, die im Matchingprozess angewendet werden. Diese können auf geometrischen, topologischen und semantischen Eigenschaften oder deren Kombination beruhen.

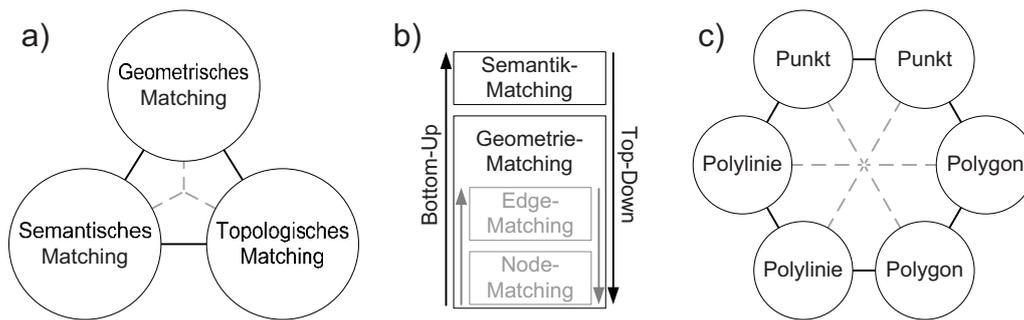


Abbildung 5 – Differenzierungsmöglichkeiten des Matchingprozesses. Unterschieden wird dabei nach verwendeten Eigenschaften (a), Aggregationshierarchie (b) und Geometrie der Eingangsdaten (c).

Geometrisches Matching nutzt den direkten Raumbezug von Geodaten aus und stellt Assoziationen zwischen Objekten auf Basis absoluter Lagebeziehungen her. Die Grundlage bilden im Regelfall Einzelobjekte und deren geometrische Parameter.

Topologisches Matching wird häufig in Kombination mit geometrischem Matching durchgeführt, wobei neben den Einzelobjekten auch deren direkten oder indirekten Nachbarschaftsbeziehungen analysiert werden. Diese Informationen gehen als zusätzliche Parameter in den Matchingprozess ein und werden oftmals zur Verifizierung geometrischer Zuordnungen eingesetzt.

Semantisches Matching hat das Ziel, Objekte auf Grundlage ihrer Bedeutung zuzuordnen. Im Falle der Nutzung einer gemeinsamen Ontologie läuft dies zumeist auf ein Schema-Matching hinaus. Sobald die genutzten Ontologien jedoch verschieden sind, muss zuvor ein Ontologie-Matching durchgeführt werden [Hakimpour und Timpf (2001)]. Dabei könnten sowohl die von Kuhn (2003) vorgeschlagenen semantischen Referenzsysteme mit entsprechenden Transformationsvorschriften als auch die in Bishr et al. (1999) beschriebene Struktur aus Semantic-Wrapper und Semantic-Mapper zum Einsatz kommen. Da ein Großteil bisher vorhandener Geodaten nur bedingt formale semantische Informationen bereitstellt, erfolgen Zuordnungen auf semantischer Ebene oftmals noch interaktiv.

Als zweite Möglichkeit können Matchingprozesse über deren Richtung, entweder als Top-Down- oder Bottom-Up-Ansatz, klassifiziert werden. Diese Differenzierung kann einerseits auf geometrischer und andererseits auf methodischer Ebene erfolgen. Bei Ersterem wird die Reihenfolge von Knoten- und Kantenmatching betrachtet. Der Bottom-Up-Ansatz beschreibt dabei die Zuordnung der Kan-

ten auf Basis zuvor gematchter Knoten. Beim Top-Down-Ansatz werden dahingegen zunächst die Kanten und anschließend die Knoten zugeordnet. In zahlreichen Arbeiten wird eine Verbindung der beiden Ansätze verfolgt. Als Zwischenschritt beschreibt Xiong (2000) ein Segment-Matching, um die Zuordnung linearer Elemente zu verfeinern. Auf methodischer Ebene führt die Reihenfolge von geometrischem und semantischem Matching zu einer entsprechenden Differenzierung. Beim Bottom-Up Ansatz werden Objekte zunächst auf Grundlage ihrer Geometrie gematcht und darauf basierend semantisch zugeordnet. Der Top-Down-Ansatz beschreibt dahingegen die Verknüpfung von Objekten auf semantischer und anschließende Zuordnung auf geometrischer Ebene. Welcher der Ansätze am besten geeignet ist, hängt von den jeweils verwendeten Datenmodellen und verfügbaren Informationen ab [Walter (1996)]. Alle Ansätze können iterativ durchgeführt werden, um die Anzahl von Verknüpfungen zu erhöhen, diese zu verifizieren und gegebenenfalls falsche Zuordnungen zu eliminieren.

Eine dritte Differenzierung des Matchingprozesses erfolgt aus der Tatsache heraus, dass viele der bereits entwickelten Methoden auf der Verarbeitung bestimmter Geometrietypen beruhen:

Das Matching von Punktgeometrien kann oftmals über weniger komplexe Ansätze gelöst werden. Dabei verwendete Methoden können zumeist auch auf Polylinien und Polygone übertragen werden. Beispiele für ein Punktmatching sind die von Samal et al. (2004) verwendeten Nachbarschaftsgraphen oder die von Beerli et al. (2005) beschriebene lagebasierte Verknüpfung mehrerer Datensätze.

Der Vergleich von Polylinien kann über Punkte, Segmente oder Eigenschaften des gesamten Objektes erfolgen. Filin und Doytsher (1999) beschreiben beispielsweise ein Verfahren zum linienbasierten Matching über kumulative Distanzen und Detektion von Eckpunkten. Des Weiteren untersuchen Cobb et al. (1998) am Beispiel von Vektordaten den Einfluss der Eingangsdaten auf das Ergebnis und stellen dabei auch die Datenqualität in den Fokus der Betrachtung.

Das Matching von Polygonen ähnelt, abgesehen von flächenspezifischen Parametern, dem Matching von Polylinien. Ein entsprechendes Beispiel, die Integration von Wasserflächen aus ATKIS-Daten und der Geologischen Karte 1:25 000, beschreiben v. Gösseln und Sester (2004).

Kombinationen verschiedener Geometrien und einige Spezialfälle adressieren dabei auch die bereits genannte Problematik der Generalisierung und müssen in

der Vorverarbeitung oder direkt im Matchingprozess berücksichtigt werden. Die Anpassung unterschiedlicher Repräsentationsformen kann beispielsweise durch Generalisierung (z.B. Polygon \rightarrow Polylinie durch Skelettierung) oder Erweiterung (z.B. Polylinie \rightarrow Polygon durch Bufferbildung) erfolgen. Eine weitere Möglichkeit ist ein Matching über die, in Egenhofer und Herring (1992) beschriebenen, topologischen Beziehungen von Objekten (z.B. Punkt in Polygon). Welche der Varianten Verwendung findet, muss im jeweiligen Anwendungskontext entschieden werden.

3.3.2 Kriterien für die Suche homologer Objekte

Weil Darstellungen von Objekten in unterschiedlichen Datensätzen im Regelfall voneinander abweichen, müssen Ähnlichkeits- bzw. Vergleichsmaße ermittelt werden, um Verknüpfungen zu ermöglichen. Je höher diese sind, desto höher ist die Wahrscheinlichkeit, dass es sich um dasselbe Objekt in der Realität handelt. Auch wenn im Folgenden nach geometrischen, topologischen und semantischen Kriterien differenziert wird, ist deren richtige Kombination und Gewichtung der Schlüsselfaktor für ein erfolgreiches Matching.

Geometrische Vergleiche basieren zum Großteil auf Distanzmessungen und Formparametern. Einige grundlegende Vergleichsmaße diesbezüglich sind in McMaster (1986), Walter (1996) und Veltkamp und Hagedoorn (2001) aufgeführt:

- Distanzmessungen:
 - die *Nearest Neighbor*-Methode, welche die Suche nach nächstgelegenen Punkten, z.B. über paarweisen Vergleich oder *Voronoi*-Diagramme, beschreibt,
 - die *Hausdorff*-Distanz als maximale Minimaldistanz zwischen zwei Punktmengen. Zu jedem Punkt der Menge A wird die Minimaldistanz zur Punktmenge B bestimmt. Das Maximum dieser Distanzen wird als *Hausdorff*-Distanz bezeichnet und kann wie folgt dargestellt werden:

$$d_H(A, B) = \max \left\{ \max_{x \in X} \left\{ \min_{y \in Y} d(x, y) \right\}, \max_{y \in Y} \left\{ \min_{x \in X} d(x, y) \right\} \right\}$$

- die *Fréchet*-Distanz als parametrisierte Distanzmessung zwischen zwei Linien. Dabei werden die Objekte abgetastet und die Distanz an den jeweiligen Positionen bestimmt. Diese Methode ist schwieriger zu implementieren, liefert aber im Regelfall genauere Aussagen zur Similarität von Kurven als die *Hausdorff*-Distanz. Mathematisch

beschrieben wird die *Fréchet*-Distanz wie folgt, wobei A und B stetige lineare Objekte darstellen:

$$d_F(A, B) = \inf_{\alpha, \beta} \max_{t \in [0,1]} \left\{ d\left(A(\alpha(t)), B(\beta(t))\right) \right\}$$

- die *Turning-Function* mit der Darstellung von Linienelementen als Funktion von Segmentlängen und -winkeln. Anhand des Abstandes, z.B. der mittleren Distanz oder Fläche zwischen den Funktionen, kann ein Vergleichsmaß errechnet werden,
- die Ermittlung der prozentualen Überlappung oder Differenz zweier Flächenelemente,
- mathematische Vergleichsmessungen auf Basis geometrischer Parameter, wie beispielsweise Länge, Umfang, Winkel, Stützpunktzahl, Linearität oder Vektordifferenz.

Durch topologische Kriterien werden Ähnlichkeitsmaße von Objekten aus Nachbarschaftsbeziehungen abgeleitet. Beispiele dafür sind:

- die Anzahl von einem Knoten abgehender Kanten, auch als Grad eines Knotens bezeichnet,
- die *Spider-Function* als Kombination der Anzahl von einem Knoten abgehender Kanten und deren Winkel. Durch die Einteilung in acht Sektoren ergibt sich dabei ein charakteristisches 8-Bit-Muster für jeden Knoten [vgl. Rosen und Saalfeld (1985)],
- der *Round-Trip-Walk*, bei dem eine Verifizierung eines gematchten Knotens über seine direkten Nachbarn erfolgt [vgl. Filin und Doytsher (2000)].

Semantische Kriterien beziehen sich auf die thematische Attributierung, Struktur oder Ontologie der Geodaten. Entsprechende Ähnlichkeitsmaße können abgeleitet werden aus:

- der Struktur- und Inhaltsanalyse von XML-Bäumen, in der ein automatisierter Vergleich von XML-Dokumentstrukturen beispielsweise durch die in Weis und Naumann (2004) beschriebene *threshold-similarity-function* oder den in Giunchiglia et al. (2007) verwendeten *S-Match*-Algorithmus erfolgen kann,
- einem Attributvergleich:

- Vergleich skalerer Zahlwerte über eine einfache oder normierte Differenzbildung. Eine Ausnahme bilden beschreibende Zahlwerte in Nominal- oder Ordinalskalen, bei denen Cobb et al. (1998) die Bildung von Vergleichskoeffizienten vorschlagen,
- Vergleich beliebiger Zeichenketten über die Zerlegung in Einzelbestandteile. Entstandene Fragmente können anschließend direkt oder mit Hilfe der *Damerau-Levenshtein*-Distanz verglichen werden. Nähere Informationen zu Letzterem finden sich in Samal et al. (2004).

3.3.3 Matching von Straßennetzwerken

Straßendaten werden in der Literatur häufig als Grundlage für einen Matchingprozess verwendet. Einerseits ist sicher die Verfügbarkeit und Abdeckung von Straßendaten ein Grund dafür, andererseits ist die Verkehrsinfrastruktur aber auch ein wichtiger Wirtschaftsfaktor, der insbesondere im Bereich Transport und Logistik zahlreiche Anwendungsmöglichkeiten aufweist.

Ein Matching von Straßendaten aus den Datenmodellen ATKIS und GDF (*Geographic Data File*) beschreiben Walter und Fritsch (1999) und nutzen als Grundlage statistische Erhebungen aus manuell bearbeiteten Testgebieten. Dabei kommt das bereits in Walter (1996) vorgestellte Buffer-growing zum Einsatz, welches genutzt werden kann, um lage- und formähnliche Linienobjekte auch bei unterschiedlicher Segmentierung zu matchen. Im Rahmen des *Nexus*³-Projektes wird dieser Ansatz aufgegriffen und erweitert. Volz und Walter (2004) führen dabei das Matching auf Instanz- und Schemaebene über einen Bottom-Up-Ansatz durch. Darauf aufbauend werden in Volz (2005) die Korrelationen verschiedener Straßenklassifikationen zwischen den Datensätzen untersucht. Die Speicherung von Links zwischen ATKIS und GDF-Datensätzen erfolgt schließlich in einer XML-basierten Struktur, der sogenannten *MultiRepresentational Relation Language* (MRRL) [Volz (2006)]. Implementierungen innerhalb dieses Projektes erfolgten zumeist auf Basis des Java-basierten Open Source JUMP⁴-Framework.

Matchingansätze zur Integration von ATKIS- und TeleAtlas-Datensätzen und deren Umsetzung in ArcGIS-Tools wurden in mehreren Projekten an der TU München untersucht. Zhang et al. (2005) beschreiben dabei die Entwicklung eines generischen Algorithmus, bei dem notwendige Parameter zum Matching dynamisch ermittelt und verfeinert werden. In Zhang et al. (2006) wird der Ver-

³<http://www.nexus.uni-stuttgart.de/index.en.html>

⁴<http://www.jump-project.org>

such unternommen, bekannte Geometrien in der Vorverarbeitung zu trennen und separat zu matchen. Als Beispiele dafür wurden ein Matching paralleler Straßen und kreisförmiger Geometrien implementiert. Im weiteren Verlauf beschreiben Zhang und Meng (2007), basierend auf dem Ansatz von Walter (1996), die Entwicklung eines asymmetrischen Buffer-growing, welches in der Lage ist, systematische Verschiebungen der Datensätze auszugleichen. Eine Alternative zum Buffer-growing, der *Delimited-Stroke-Oriented-Algorithm*, wird in Zhang und Meng (2008) entwickelt. Dieser zeichnet sich vor allem durch die Nutzung komplexer topologischer Informationen und seine Performanz aus.

Weitere Arbeiten befassen sich unter anderem mit dem Problem unterschiedlicher Repräsentationsformen. So widmen sich beispielsweise Uitermark et al. (1999) der polygonalen Darstellung von Straßen in großen Maßstäben und stellen formale Definitionen von Straßensegmenten und Kreuzungen auf. Mit dem Matching von Straßendaten unterschiedlicher Maßstäbe befassen sich Lüscher (2006) mit dem Fokus auf der Erstellung einer MRDB, und Xiaomeng et al. (2007) mit dem Schwerpunkt auf routingfähigen Datensätzen.

3.4 ZUSAMMENFÜHRUNG VON DATENSÄTZEN

Nachdem die Zuordnung der einzelnen Objekte abgeschlossen ist, folgt die eigentliche Fusion der Datensätze. Attribute können während der Fusion erhalten, neu berechnet, ersetzt, hinzugefügt oder entfernt werden. Dies trifft sowohl auf die Geometrie als auch die thematischen Informationen zu und kann von Auflösung, Genauigkeit oder Aktualität der verglichenen Attribute abhängig gemacht werden. Auch die Struktur der Datensätze kann in diesem Zusammenhang vereinheitlicht werden und dabei entweder auf einem Referenzdatensatz oder einer Neubildung beruhen.

Bei der Ermittlung geeigneter Prozessierungsschritte sollten Eingangsdaten und anschließende Verwendung des Ergebnisses Berücksichtigung finden. Der Anwendungskontext beeinflusst damit maßgeblich den Ablauf einer Fusion.

3.5 GEODATENFUSION IM KONTEXT EINER GDI

Die Möglichkeiten datensatzüberschreitender Analysen innerhalb einer Geodateninfrastruktur sind derzeit noch sehr beschränkt. Eine steigende Anzahl über Diensteschnittstellen bereitgestellter Daten begründet, zusammen mit den Zielen der Vermeidung redundanter Informationen und Schaffung von Informations-

mehrwert, die Notwendigkeit einer Geodatenfusion. Die Abgrenzung von der GIS-basierten Geodatenfusion und vorhandene technische und konzeptionelle Einschränkungen definieren dabei das mögliche Aufgabenspektrum.

3.5.1 Monolithische und dienstebasierte Methoden

Die allgemeinen strukturellen Unterschiede zwischen monolithischem GIS und dienstorientierter GDI in Bezug auf die Bereitstellung und Prozessierung von Geodaten sind in Abbildung 6 dargestellt.

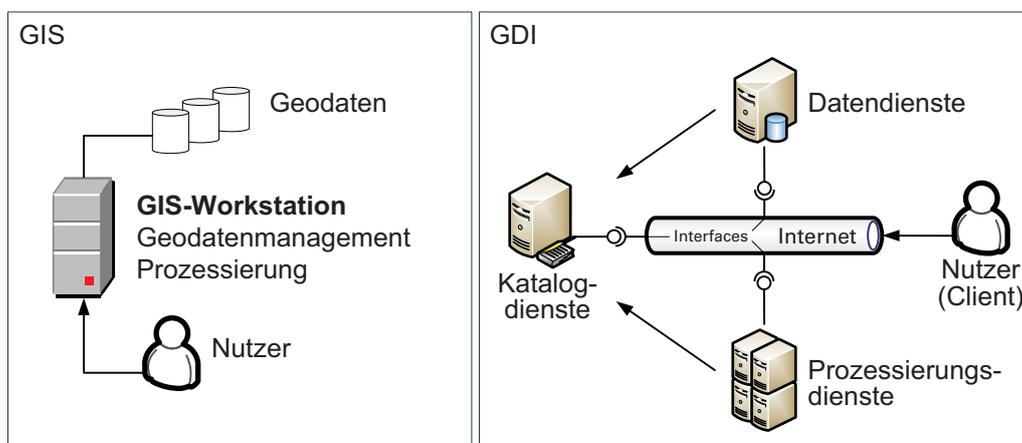


Abbildung 6 – Architekturunterschied zwischen GIS und GDI. Während GIS vorwiegend monolithisch geprägt sind, ist eine GDI als dienstebasierte Anwendungslandschaft konzipiert. Dargestellt sind die, für eine Geodatenfusion wesentlichen Elemente, Datenbereitstellung und Datenprozessierung.

In konventionellen, monolithisch geprägten GIS, sind grundlegende Funktionalitäten der Geodatenverarbeitung bereits implementiert. Vorhandene Tools im Bereich Matching und Fusion von Geodaten bieten allerdings nur Teilfunktionalitäten an. Beispiele aus dem proprietären Bereich sind das *Spatial Adjustment Tool* in *ArcGIS Desktop* und der *MRF Cleaner*⁵ in der *Feature Manipulation Engine* (FME). Ein Vorteil dieser Softwarepakete ist eine zuverlässige und zumeist einfach bedienbare Arbeitsumgebung, in der nutzerspezifische Workflows entwickelt werden können. Aufgrund der zumeist einheitlichen Systemstruktur stellt die Verkettung von Prozessen und der Einbau von Interaktionsmöglichkeiten in der Regel keine großen Hindernisse dar. Zur Automatisierung von Prozessen sind darüber hinaus Schnittstellen zu Skript- oder Programmiersprachen implementiert. Die Anwendung von Prozessierungstools auf lokal vorgehaltenen Daten und

⁵Plug-In der Firma MRF Geosystems Corporation

der Wegfall von Transferleistungen sind Ursachen einer hohen Performanz gegenüber Web-basierten Anwendungen. Abhängig von verwendeten Methoden lag die Dauer der Prozessierung der in Kapitel 3.3.3 beschriebenen Matchings bei durchschnittlich zwei bis drei Minuten, teilweise auch im Bereich weniger Sekunden. Des Weiteren trägt eine zumeist umfassende Dokumentation etablierter GIS-Software zur Nutzerfreundlichkeit bei.

Die Geodatenfusion innerhalb einer GDI kann mit der Stabilität und Performanz monolithischer Desktop-GIS Systeme nicht konkurrieren. Das Ziel kann daher nicht die Zusammenführung großer Datenbestände, sondern muss vielmehr die ad-hoc Fusion kleinerer Datenbestände und deren Nutzung durch Entscheidungsträger sein. Die Stärke einer GDI liegt dabei vor allem in der Verarbeitung hochaktueller, teilweise in-situ aufgenommener Datenbestände. Der Zugriff auf Daten, Services und Ergebnisse ist dezentral möglich, weshalb beliebige Endgeräte, vom *Personal Computer* (PC) bis zum *Personal Digital Assistant* (PDA), als Client fungieren können. Algorithmen für die Fusion von Geodaten innerhalb einer GDI können unabhängig von Programmiersprache oder Betriebssystem entwickelt und über die WPS-Schnittstelle veröffentlicht und genutzt werden. Durch die fehlende Bindung an bestehende Datenstrukturen oder vorhandene Prozesse steigt somit die Flexibilität der Implementierung im Vergleich zu vielen monolithischen Lösungen.

Da GIS und GDI nicht konträr arbeiten, sondern zahlreiche Verknüpfungspunkte aufweisen, können sie sich gegenseitig ergänzen und voneinander profitieren. Beispiele dafür sind die Einbindung von GDI-Funktionalitäten in einem GIS oder die bereits erwähnte Kapselung von GIS-Funktionalitäten zur Geoprozessierung durch einen WPS.

3.5.2 Herausforderungen und Probleme

Wichtige Informationen zu Vorverarbeitung oder Matching von Geodatenätzen können aus Metadaten abgeleitet werden. Diese geben Aufschluss über das Datenmodell, das geographische Koordinaten- und Referenzsystem, die Qualität der Daten oder die dargestellte Thematik. Problematisch ist jedoch, dass Metadaten oftmals unvollständig oder zu wenig formalisiert sind. Der Nutzer muss diese Defizite durch eigene Expertise ausgleichen, was die Nutzung der Daten einschränkt und die Automatisierung von Prozessen erschwert.

Ein weiterer kritischer Punkt ist die Suche nach nutzbaren Ressourcen. Insbesondere der Mangel an semantischen Informationen kann die Suche nach Daten und Diensten schwierig gestalten und bremst die Entwicklung entscheidungsun-

terstützender Systeme. Des Weiteren ist der Pool verfügbarer Geoinformationen noch relativ klein, da viele nicht öffentlich zugänglich sind oder nicht über eine standardisierte Diensteschnittstelle verfügen.

Auch das Defizit der WPS-Schnittstelle hinsichtlich fehlender Interaktionsmöglichkeiten kann ein Hindernis für die Geodatenfusion darstellen. Viele in der Literatur vorgestellte Matching-Algorithmen besitzen eine maximale Zuordnungsrate von 90 bis 95 Prozent. Um einen zusätzlichen Datentransfer zu vermeiden, werden daher entsprechende Interaktionsmechanismen benötigt, die eine interaktive Nachbearbeitung verbliebener Objekte ermöglichen. Ein dazu notwendiges *stateful processing* muss die Speicherung temporärer Daten und entsprechende Möglichkeiten zur Weiterverarbeitung anbieten. Ein Ansatz dafür wurde bereits in OGC (2007c) diskutiert und sieht die Adaption des *Web Services Resource Framework* (WSRF) der *Organization for the Advancement of Structured Information Standards* (OASIS) an den WPS vor.

4 KONZEPTION EINER DIENSTEBASIERTEN GEODATENFUSION

In diesem Kapitel wird auf Basis der beschriebenen Grundlagen eine Konzeption für die vertikale Fusion von Geodaten innerhalb einer Geodateninfrastruktur entwickelt. Betrachtet werden mögliche Eingangsdaten, die Systemarchitektur, einzelne Prozessierungsschritte sowie Besonderheiten der dienstebasierten Fusion. Eine prototypische Implementierung der Konzeption folgt in Kapitel 5.

4.1 BETRACHTUNG MÖGLICHER EINGANGSDATEN

Zentrale Elemente des in Abbildung 7 skizzierten OGC-Feature-Modells und damit die Basis für eine dienstebasierte Fusion sind das *Feature* und die *FeatureCollection*¹. Features sind nach OGC (1998) reale Entitäten oder eine Abstraktion der realen Welt und werden aus einem *FeatureType* abgeleitet. Dieser wird wiederum über eine Menge von *FeatureAttributeTypes* definiert, welche positionelle, geometrische, thematische oder temporale Attribute beschreiben können. Die Strukturen und Inhalte dieser Attribute bilden die Grundlage für eine Geodatenfusion und beeinflussen maßgeblich die verwendeten Matching- und Fusionsstrategien. Um eine Verarbeitung zu ermöglichen, gelten folgende Kriterien für die verwendeten Ausgangsdaten:

- die bereits erwähnten Qualitätsnormen der ISO-Norm 19113 müssen eingehalten werden. Dies betrifft im Wesentlichen die Vollständigkeit, Konsistenz und Genauigkeit vorhandener Attribute.
- bei einer vertikal durchgeführten Fusion muss eine entsprechende räumliche Überlappung der Datensätze gewährleistet sein, da nur in mindestens zwei Datensätzen enthaltene Features fusioniert werden können.
- Attribute müssen miteinander vergleichbar sein, da strukturelle oder inhaltliche Unterschiede zu Konflikten bei der Zuordnung und Fusion führen können.

¹FeatureCollection: Aggregation von Features, welche wiederum selbst als Feature definiert werden kann

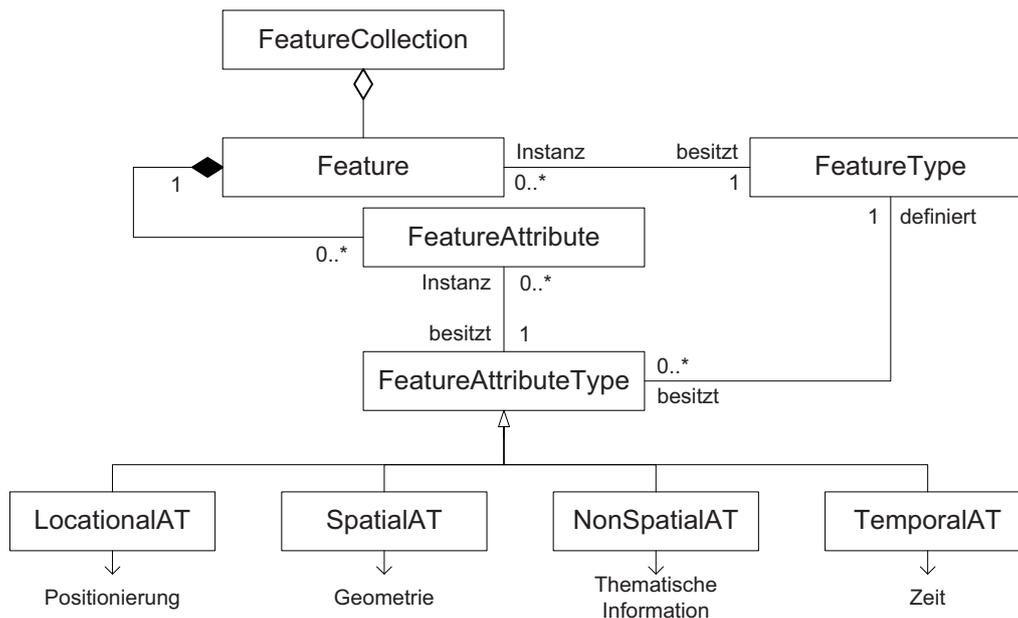


Abbildung 7 – Vereinfachte Darstellung des OGC-Feature-Modells nach OGC (2009a). Für eine dienstebasierte Fusion von Geodaten spielen vor allem *FeatureAttributes*, beschrieben durch positionelle, räumliche, thematische oder temporale *AttributeTypes* (AT), eine wichtige Rolle.

nen. Je nach Prozesskonfiguration können im Vorfeld bereits bekannte oder detektierte Unterschiede durch eine angepasste Fusionsstrategie behandelt werden.

- Grundlegende Metadaten, beispielsweise mit Informationen zur Thematik, Genauigkeit und räumlichen Referenz der Daten, sollten verfügbar sein. Insbesondere beim Vergleich und Matching von Geodaten können diese Informationen eine wichtige Rolle spielen.

Features können nach OGC (2007b) als *ComplexData* eingestuft werden und müssen daher Angaben zu *MimeType*, Codierung und Schema enthalten. Als Ausgangsdaten unterstützte Datentypen hängen von der jeweiligen WPS-Implementierung ab und sind im jeweiligen *DescribeProcess*-Dokument beschrieben. Möglich sind sowohl eine direkte Übertragung von Geodaten, beispielsweise im Shapefile-Format, als auch die Nutzung von Datendiensten, beispielsweise einem WFS.

Die Bereitstellung von Features über die WFS-Schnittstelle erfolgt als GML-Stream im Ergebnis eines *GetFeature*-Request. Für Angaben zum entsprechenden *FeatureType* kann die *DescribeFeatureType*-Operation genutzt werden. Meta-informationen können unter anderem im optionalen Element `<MetadataURL>` eines *GetCapabilities*-Response referenziert sein oder direkt aus bereitgestellten Features ermittelt werden.

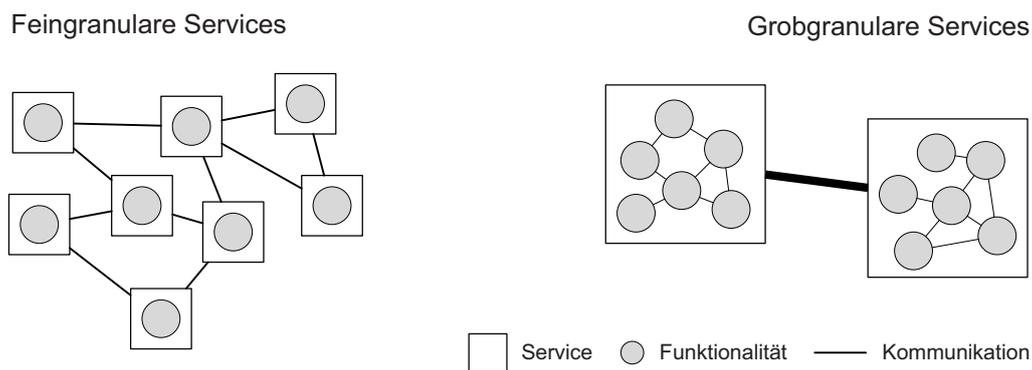


Abbildung 8 – Aufbau von fein- und grobgranularen Services in einer serviceorientierten Architektur, angelehnt an McGovern et al. (2003). Während feingranulare Services ein komplexes Netz atomarer Services bilden, werden Funktionalitäten in grobgranularen Services zu logischen Teilprozessen aggregiert.

4.2 ARCHITEKTUR

Im Folgenden wird die konzeptionelle Systemstruktur für eine dienstebasierte Fusion von Geodaten vorgestellt. Dazu werden zunächst Aspekte der Granularität von Teilprozessen betrachtet und anschließend die Gesamtarchitektur für die Fusion entwickelt.

4.2.1 Granularität der Services

Die Granularität beschreibt den Umfang der implementierten Anwendungslogik eines Services [McGovern et al. (2003)]. Dabei spielen sowohl der Funktionsumfang bzw. fachliche Granularität des Services als auch der Umfang der zwischen Client und Service ausgetauschten Daten, die Granularität der Schnittstelle, eine Rolle. Bei einem WPS entspricht dies dem Angebot an verfügbarer Geoprozessierungsfunktionalität und der Parametrisierung des Dienstes durch notwendige Input-Parameter.

In Abbildung 8 wird der strukturelle Unterschied zwischen feingranularen und grobgranularen Services dargestellt. Feingranulare Services sind durch atomare Funktionalitäten und geringe Datentransferleistungen gekennzeichnet, wodurch schnelle Reaktionszeiten einzelner Services erreicht werden. Die hohe Anzahl benötigter Services in einem Gesamtsystem macht allerdings eine häufige Codierung und Decodierung transferierter Daten notwendig und kann überdies zu komplexen Netzstrukturen führen. Grobgranulare Services bieten dahingegen ein größeres Maß an Funktionalität und stellen im Prinzip eine "Aggregation" atomarer Services dar. Die Kombination mit einer zumeist höheren Parametrisierung bedingt jedoch entsprechend längere Reaktionszeiten, komplexere Prozessstruk-

turen und höhere Spezialisierungsgrade. Dies kann zu Problemen bei der Anpassung bzw. Änderung sowie zu Einschränkungen in der Wiederverwendbarkeit von Services führen [vgl. Wilkes und Veryard (2004), Schwinn und Winter (2005)].

Für verteilte Dienste sollte nach McGovern et al. (2003) in der Regel eine eher grobgranulare Struktur angestrebt werden, in der Dienste als fachliche Teilprozesse umgesetzt werden. Dementsprechend sollte auch bei der dienstebasierten Fusion von Geodaten eine Einteilung in logische, fachlich voneinander abgegrenzte Teilprozesse erfolgen. Dabei müssen insbesondere Aspekte der Wiederverwendbarkeit und Interoperabilität von Diensten Beachtung finden.

4.2.2 Prozessablauf

Teilprozesse der Fusion müssen in sich geschlossene funktionale Einheiten bilden und möglichst flexibel angelegt werden, um auch in einem anderen Anwendungskontext verwendbar zu sein. Zwischenergebnisse sollten eine eigenständige, in sich logische Aussage beinhalten und direkt als Input darauf folgender Dienste geeignet sein. Aus den Vorbetrachtungen und angelehnt an Yuan und Tao (1999) ergeben sich folgende, in Abbildung 9 dargestellte, Teilprozesse:

1. Im ersten Schritt erfolgt ein Vergleich der Ausgangsdaten hinsichtlich geometrischer, thematischer und struktureller Eigenschaften. Auf dieser Grundlage können notwendige Schritte zur Anpassung der Datensätze festgelegt und im Preprocessing durchgeführt werden. Zudem können daraus abgeleitete Informationen die Prozessbildung beeinflussen. Der Vergleich von Ausgangsdaten ist in der Regel mandatorisch, kann allerdings entfallen, wenn die Vergleichbarkeit durch den Nutzer gewährleistet wird.
2. Eine optionale Anpassung der Datensätze erfolgt im Preprocessing. Dabei können bereits vorhandene Dienste zur Geoprocessing genutzt werden, um die Ausgangsdaten einander anzugleichen. Dieser Schritt kann auch iterativ durchgeführt und durch einen erneuten Vergleich der Datensätze verifiziert werden.
3. Das Matching der Ausgangsdaten beinhaltet die Suche und Zuordnung homologer Objekte in den Datensätzen und ist der wichtigste Faktor einer erfolgreichen Fusion. Potentielle Matches können in den Ausgangsdaten, einer externen Datei oder einer MRDB festgehalten und Parameter, auf denen ein Match beruht, für eine anschließende Evaluierung verwendet werden.

4. Die eigentliche Fusion basiert auf dem vorangegangenen Matching und hat einen Informationsgewinn unter Nutzung der Ausgangsdaten zum Ziel. In der Regel erfolgt in diesem Schritt die Anreicherung des Referenzdatensatzes durch einen Attribut- oder Featuretransfer. Basierend auf der Fusion kann zudem eine Analyse und Auswertung im Kontext einer Gesamtevaluierung des Prozesses durchgeführt werden.
5. In einem optionalen Postprocessing erfolgt die Anpassung des Fusionsergebnisses an Nutzervorgaben. Wie beim Preprocessing können auch hier bereits vorhandene Dienste zur Anpassung von Datensätzen genutzt werden.

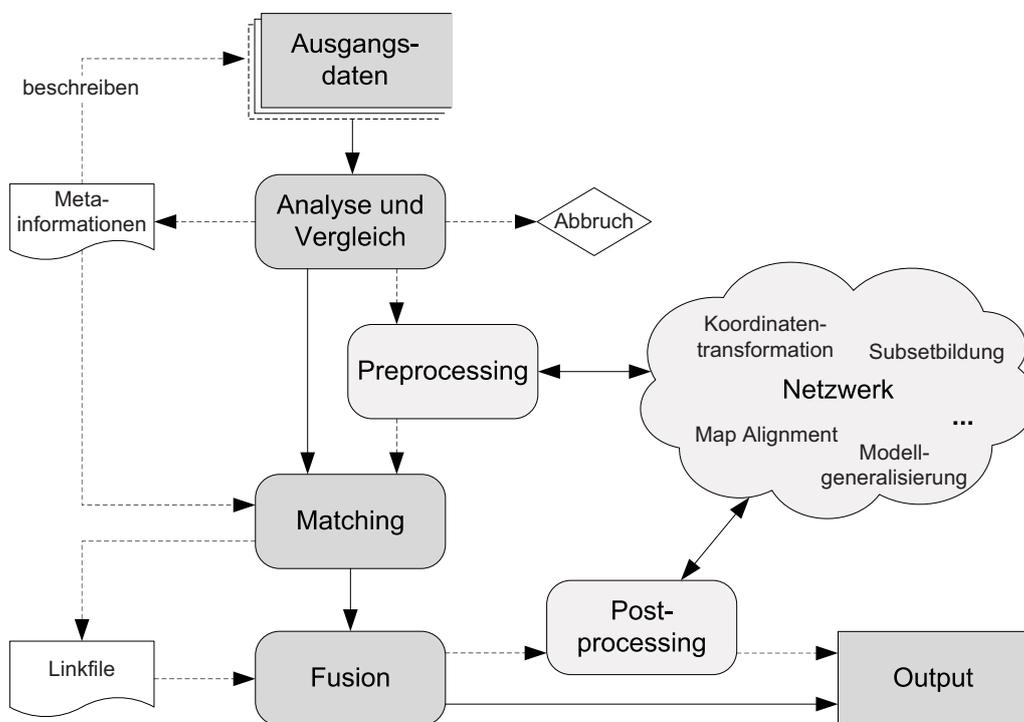


Abbildung 9 – Ablauf einer dienstebasierten Fusion mit den mandatorischen Bestandteilen Vergleich, Matching und Fusion der Datensätze. Optional ist die Anpassung von Ausgangsdaten und des Ergebnisses durch externe Dienste möglich.

Das Ergebnis der Prozessierung ist der fusionierte Datensatz. Optional können auch Metadaten zum Prozess, Zwischenergebnisse oder eine statistische Auswertung der Fusion bereitgestellt werden. Im nächsten Abschnitt erfolgt eine detaillierte Beschreibung der Teilprozesse und notwendiger Schnittstellen. Ein UML-Sequenzdiagramm des Gesamtprozesses befindet sich im Anhang A dieser Arbeit.

4.3 PROZESSIERUNGSSCHRITTE

Im Folgenden werden die festgelegten Teilprozesse einer dienstebasierten Geodatenfusion näher beschrieben. Betrachtet werden insbesondere deren Verwendung, Ablauf und Schnittstellenbeschreibung.

4.3.1 Vergleich der Eingangsdaten

Die Prüfung auf Vergleichbarkeit der Ausgangsdaten ist in der Regel erforderlich, um die in Kapitel 3.2 beschriebenen Vorgaben für eine Fusion sicherzustellen. Der allgemeine Ablauf des Vergleiches ist in Abbildung 10 dargestellt.

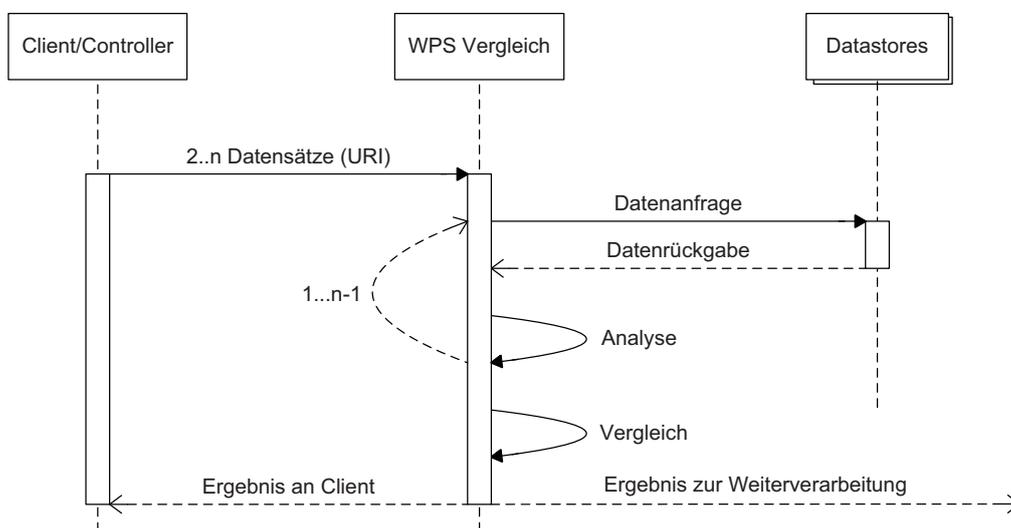


Abbildung 10 – Sequenzdiagramm des Vergleiches von Ausgangsdaten einer Geodatenfusion unter Nutzung externer Datastores. Das Ergebnis kann direkt vom Client oder zur Weiterverarbeitung genutzt werden.

In einem ersten Schritt erfolgt die Analyse von Ausgangsdaten auf Grundlage vorhandener Informationen zu Geometrie, Thematik, Struktur und Metadaten. Auch die Qualität von Attributen, beispielsweise die logische Konsistenz oder Genauigkeit, kann dabei erfasst bzw. ermittelt werden. Zur Definition entsprechender Parameter können unter anderem die von Werder (2009) beschriebenen Constraints der Datenintegration, welche wesentliche Parameter für einen Vergleich beinhalten, verwendet werden. Ein wichtiger Faktor der Analyse ist das Vorhandensein von Metadaten, welche notwendige Informationen über Qualität, strukturelle Beschaffenheit und Thematik der Datensätze geben können. Die Detektion von Unterschieden und Gemeinsamkeiten in den Ausgangsdaten kann unter anderem auf folgenden Elementen beruhen:

- dem Koordinatenreferenzsystem (z.B. auf Grundlage von EPSG-Codes²),
- der räumlichen Lage von Ausgangsdaten (z.B. Überlappungsbereich der BoundingBox),
- der Gesamtanzahl und geometrischen Parameter (z.B. Geometrietypen) von Features in den Datensätzen,
- dem Generalisierungsmaßstab,
- der Genauigkeit von geometrischen und thematischen Attributen,
- der durchschnittlichen räumlichen Abweichung zwischen Features.

Das Ziel des Vergleiches ist eine Aussage zur Wahrscheinlichkeit einer erfolgreichen Fusion und die Festlegung notwendiger Schritte zu Anpassung der Datensätze. Letzteres kann sowohl eine Auflistung benötigter Anpassungsschritte, als auch bereits die Suche nach entsprechenden Diensten zu deren Durchführung beinhalten. Sollte eine Anpassung nicht erwünscht oder nicht möglich sein, ist als Alternative die Suche nach geeigneten Diensten zur direkten Weiterverarbeitung der Ausgangsdaten ohne Anpassung oder die Auswahl besser geeigneter Ausgangsdaten möglich. Der Vergleich nimmt damit direkten Einfluss auf die Erstellung der Prozesskette. Sind vorhandene Ausgangsdaten unter gegebenen Umständen nicht fusionierbar, wird die Prozessierung abgebrochen.

Tabelle 2 – Mandatorische und optionale Input- und Outputelemente eines Datenvergleiches mittels WPS im Kontext einer Geodatenfusion.

Name	Beschreibung	Anzahl
Input		
Datensatz	Referenz (URI) oder direkter Transfer der zu vergleichenden Eingangsdaten	2..n
Metadaten	Referenz (URI) oder direkter Transfer von Metadaten zu bereitgestellten Datensätzen	0..n
Parameter	Eigenschaften bzw. Parameter, auf denen der Vergleich beruhen soll	0..n
Output		
Vergleichbarkeit	Wahrscheinlichkeit einer erfolgreichen Fusion	0..1
Anpassung	Notwendige Schritte zur Angleichung der Daten	0..n

²EPSG-Code: Durch die EPSG (European Petroleum Survey Group Geodesy) festgelegte Schlüsselnummer für Koordinatenreferenzsysteme

Im DescribeProcess-Dokument einer entsprechenden WPS-Implementierung müssen eine Beschreibung möglicher Eingangsdaten und die Parameter, auf denen der Vergleich beruht, enthalten sein. Des Weiteren sind Angaben zur Codierung und zum Schema des Ergebnisses notwendig, um eine direkte Weiterverarbeitung zu gewährleisten. Mögliche Input- und Output-Elemente des Datenvergleiches in einem WPS sind in Tabelle 2 aufgelistet.

4.3.2 Anpassung

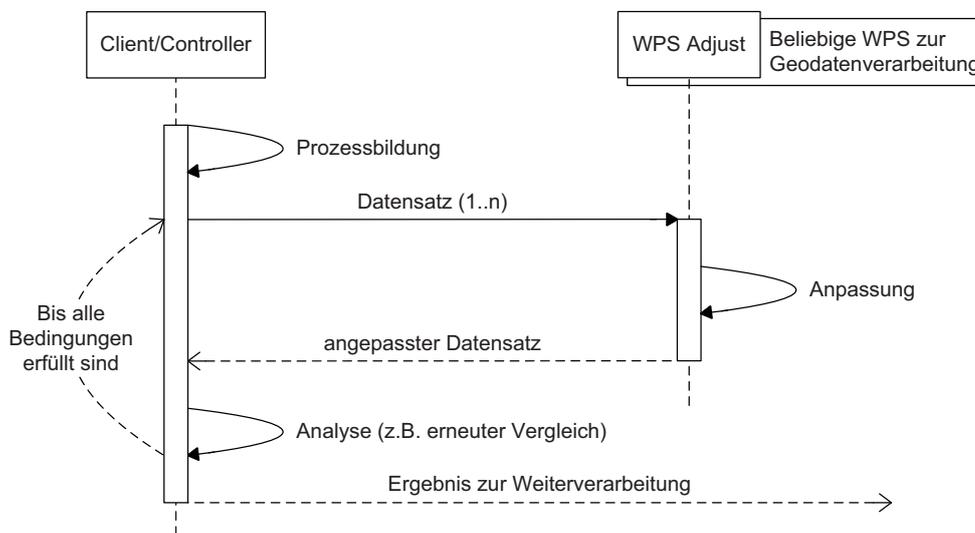


Abbildung 11 – Sequenzdiagramm der Datenvorverarbeitung mit Anpassung des Datensatzes über einen beliebigen WPS sowie abschließender Analyse und Bewertung des Ergebnisses.

Notwendige Schritte für eine Anpassung der Datensätze resultieren aus den im Vergleich festgestellten Unterschieden der Ausgangsdaten und können unter Verwendung externer Ressourcen durchgeführt werden. Die Steuerung des Prozesses kann manuell oder über eine Kontrollinstanz erfolgen, wobei benötigte Prozessierungsdienste entweder aus dem vorangegangenen Vergleich oder aus einer Suche in entsprechenden Katalogdiensten entnommen werden können. Nach einer erfolgten Anpassung kann ein erneuter Vergleich folgen, um sicherzustellen, dass alle notwendigen Voraussetzungen für eine Fusion der Daten erfüllt sind. Der allgemeine Ablauf der Anpassung von Daten ist in Abbildung 11 dargestellt und entsprechende Input- und Outputdaten einer WPS-basierten Kontrollinstanz in Tabelle 3 aufgelistet. Beispiele für eine gegebenenfalls notwendige Angleichung von Datensätzen sind:

- eine Koordinatentransformation, beispielsweise mit Hilfe eines *Web Coordinate Transformation Service* (WCTS) [OGC (2007d)], um einen direkten

Vergleich von Positionsangaben zu ermöglichen.

- ein Map Alignment durch Rubbersheet-Transformation zur Minimierung geometrischer Abweichungen der Features und Unterstützung einer anschließenden Zuordnung homologer Objekte.
- eine geometrische oder thematische Modellgeneralisierung zur Angleichung der Referenzmaßstäbe, beispielsweise durch Stützpunktreduktion der Geometrie oder Reklassifikation thematischer Attribute.
- eine Schnittmengenbildung der Datensätze, um nur in einem Datensatz vorhandene und damit nicht fusionierbare Objekte von der weiteren Prozessierung auszuschließen.

Tabelle 3 – Mögliche Input- und Outputdaten eines Workflow enactment service für die Datenvorverarbeitung mittels WPS im Kontext einer Geodatenfusion.

Name	Beschreibung	Anzahl
Input		
Datensatz	Referenz (URI) oder direkter Transfer der Eingangsdaten	1..n
Parameter	notwendige Anpassungsschritte	1..n
Referenz	Referenzdatensatz für die Anpassung	0..1
Registry	Katalogdienst für die Suche nach geeigneten Diensten	0..n
Output		
Response	den Anforderungen angepasster Datensatz	1..n

Für die Erstellung einer entsprechenden Prozesskette muss eine Möglichkeit zur formalen Beschreibung vorhandener Differenzen geschaffen werden. Diese muss eine Ableitung notwendiger Anpassungsschritte ermöglichen und die Grundlage für die Umsetzung einer Kontrollinstanz bilden. Auf syntaktischer Ebene können diese Prozessierungsschritte über standardisierte WPS-Profile definiert werden. Eine Erweiterung durch die semantische Beschreibung von Geoprozessen ist ebenfalls möglich, nach Brauner et al. (2009) allerdings noch Gegenstand der aktuellen Forschung.

4.3.3 Matching

Den wichtigsten Faktor einer erfolgreichen Fusion stellt das Matching dar, dessen allgemeiner Ablauf innerhalb einer GDI in Abbildung 12 dargestellt ist. Die Aufgabe eines entsprechenden Dienstes ist die Zuordnung homologer Objekte in den Datensätzen und die Speicherung gefundener Matches. Ermittelte

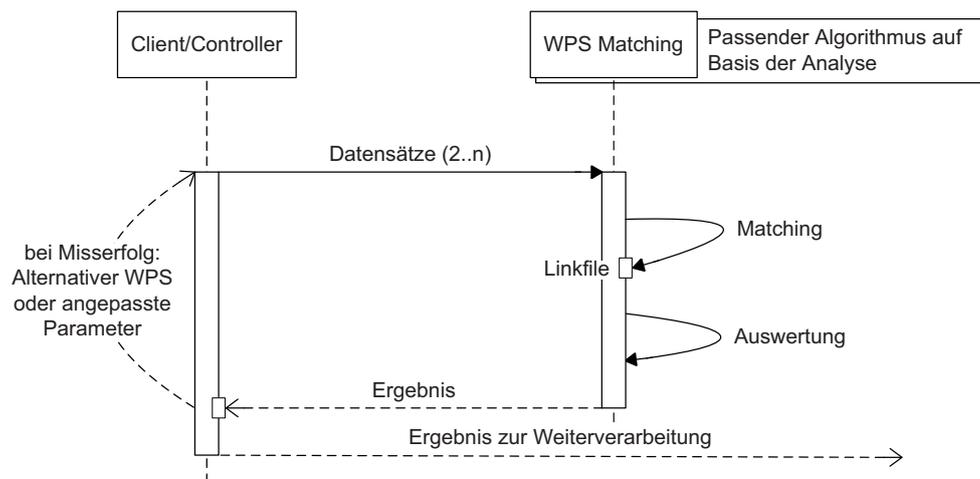


Abbildung 12 – Sequenzdiagramm mit einem Linkfile als Ergebnis des Matchingprozesses. Auf Grundlage einer Auswertung kann die Qualität des Ergebnisses analysiert werden.

Vergleichs- bzw. Ähnlichkeitsmaße zwischen Features bilden dabei die Grundlage für eine Suche, Analyse und Filterung potentieller Matches.

Im Matchingalgorithmus genutzte und somit für eine Prozessierung erforderliche Parameter müssen im DescribeProcess-Dokument des WPS enthalten sein und können unter anderem folgende Angaben enthalten:

- geometrische Toleranzen und Maximalabstände für die Zuordnung von Features über distanzbasierte Verfahren,
- ein Schema-Mapping, welches die Zuordnung der Attribute in den Datensätzen beschreibt und zur Verifizierung von Matches genutzt werden kann,
- Ausschlusskriterien für einen Match, wie beispielsweise unvereinbare Attributkombinationen oder Geometrieformen,
- die Mindestwahrscheinlichkeit eines Matches, um in eine Liste potentieller Matches übernommen zu werden,
- eine Liste manuell zugeordneter Referenzpunkte.

Des Weiteren müssen Typ, Codierung und Schema des Ergebnisses festgehalten sein, um eine anschließende Fusion auf Basis des Matchings zu ermöglichen. Notwendige Eingangsdaten und Ergebnisse eines WPS-basierten Matchings sind in Tabelle 4 aufgelistet.

Für die Speicherung des Matching-Ergebnisses gibt es im Wesentlichen drei Möglichkeiten. Die erste Möglichkeit ist die Speicherung der Links als zusätzliches Attribut im Referenzdatensatz. Dies reduziert das Datenaufkommen, kann

Tabelle 4 – Benötigte Input- und Outputdaten bei einem dienstebasierten Matching mittels WPS. Matches müssen im Ergebnis gekennzeichnet und können anschließend zur Weiterverarbeitung genutzt werden.

Name	Beschreibung	Anzahl
Input		
Datensatz	Referenz (URI) oder direkter Transfer der Eingangsdaten	2..n
Parameter	notwendige Parameter für den Matchingprozess	0..n
Output		
Ergebnis	Kennzeichnung homologer Objekte in den Datensätzen	1
Statistik	Auswertung des Matchings	0..n

allerdings die Flexibilität in der Weiterverarbeitung einschränken. Bei der Speicherung von Zuordnungen in einem Linkfile können dahingegen zahlreiche Zusatzinformationen zum Matchingprozess aufgenommen werden. Dieses Linkfile muss jedoch separat gespeichert und von Folgeprozessen interpretierbar sein. Daher müssen mindestens Codierung und Schema sowie optional eine semantische Beschreibung des Ergebnisses zur Verfügung gestellt werden. Die dritte Möglichkeit stellt die Speicherung von Links in einer MRDB dar, in deren Anschluss eine Integration von Datensätzen erfolgen kann.

Sowohl nicht gematchte als auch multiple Matches (1:n, m:1, m:n) müssen im jeweiligen Ergebnis gekennzeichnet werden. Diese können in der Fusion ausgewertet und Nutzervorgaben entsprechend verarbeitet werden. Nicht gematchte Features bilden in der Regel die Grundlage für einen Featuretransfer, multiple Matches können unter anderem zur Aktualisierung einer Geometrie beitragen. Ein weiterer Aspekt ist die Filterung von Matches. Je nach Anwendungszweck kann entschieden werden, ob die Liste der Matches im Ergebnis des Teilprozesses eindeutig sein muss oder lediglich potentielle Matches mit zugehörigen Vergleichsparametern beinhalten kann. Letzere können in einer interaktiven Nachbearbeitung des Gesamtprozesses dazu beitragen, Ergebnisse zu evaluieren, Informationen zu alternativen Matches zu bewahren und gegebenenfalls Parameter des WPS anzupassen. Sollte das Ergebnis des Matchings nicht eindeutig sein, muss eine entsprechende Filterung zur Eliminierung fehlerhafter Zuordnungen im Fusionsprozess erfolgen.

4.3.4 Fusion

Auf Grundlage des Matchings erfolgt die Zusammenführung der Ausgangsdaten und Schaffung eines nutzerspezifischen Mehrwertes. Die Auswahl eines ge-

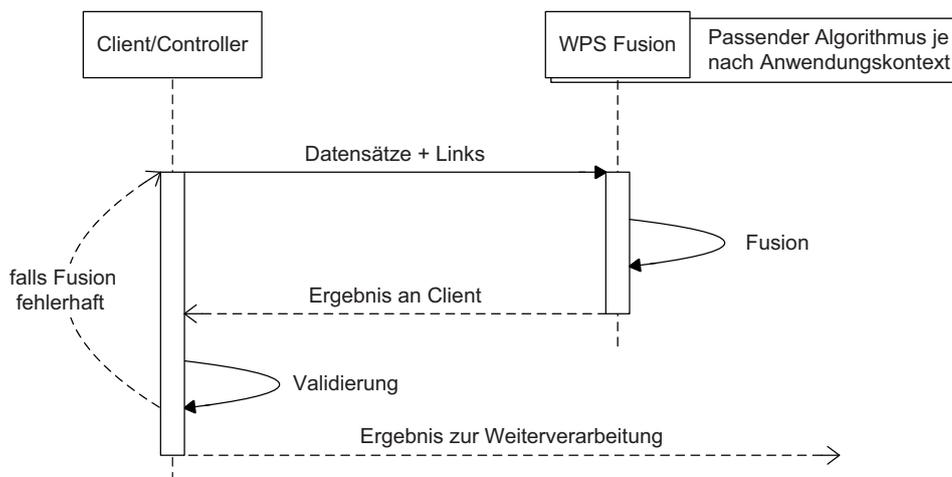


Abbildung 13 – Sequenzdiagramm der Geodatenfusion mit anschließender Validierung des Ergebnisses durch den Client.

eigneten Fusionsdienstes muss dabei abhängig vom Anwendungsfall erfolgen. Wesentliches Merkmal und Ziel der Fusion ist ein Informationsgewinn durch geometrische oder thematische Anreicherung, eine strukturelle Integration oder die Detektion attributiver Abweichungen.

Falls nicht im Matchingprozess erfolgt, müssen vorhandene Matches vor einer Fusion gefiltert und validiert werden, um eindeutige Zuordnungen zu erreichen. Auch die Behandlung multipler Matches muss je nach Fusionszweck entschieden werden. Entsprechende Möglichkeiten sind die Anpassung der Geometrien, attributive Mehrfachzuordnung oder die Reduzierung auf den Match mit der höchsten Wahrscheinlichkeit.

Das Ergebnis der Fusion stellt in der Regel der fusionierte Datensatz dar, kann aber auch in anderer Form, beispielsweise einer separaten Analyse, erfolgen. Der allgemeine Ablauf des Fusionsprozesses mittels WPS ist in Abbildung 13 dargestellt. Notwendige Input- und Outputdaten sind in Tabelle 5 aufgelistet.

Tabelle 5 – mögliche Input- und Outputdaten einer Fusion von Geodaten mittels WPS. Eine Auswertung des Ergebnisses kann als Grundlage einer anschließenden Evaluation genutzt werden.

Name	Beschreibung	Anzahl
Input		
Datensatz	Referenz (URI) oder direkter Transfer der Eingangsdaten	2..n
Links	Angaben zur Verlinkung von Objekten (z.B. Linkfile)	1
Output		
Ergebnis	Ergebnis der Fusion	1
Statistik	Auswertung der Fusion (Grundlage einer Evaluation)	0..n

4.3.5 Nachbearbeitung und Ergebnis

Je nach Anwendung können Nutzervorgaben eine Nachbearbeitung des Fusionsergebnisses erforderlich machen. Diese kann, wie die Anpassung der Ausgangsdaten, über externe Ressourcen erfolgen und beispielsweise eine erneute Koordinatentransformation oder Generalisierung beinhalten. Auch in diesem Fall können entsprechende Constraints verwendet werden, um eine Anpassung an Zielvorgaben oder eine Validierung der Fusion durchzuführen. Sollten bei der Validierung bestimmte Constraints nicht erfüllt sein, so kann eine erneute Fusion der Ausgangsdaten unter Verwendung angepasster Parameter oder alternativer Dienste initiiert werden.

Mögliche Fehler bei der Prozessierung oder besonderer Objektkonstellationen in Datensätzen machen im Regelfall eine interaktive Kontrolle des Ergebnisses notwendig. Um diesen oftmals mit hohen Kosten verbundenen Aufwand zu minimieren, können Metadaten oder Zwischenergebnisse beteiligter Prozesse verwendet werden. Das Ergebnis der dienstebasierten Fusion kann über eine beliebige Netzwerkressource oder eine Diensteschnittstelle bereitgestellt werden. Letzteres kann sowohl die Visualisierung über einen WMS, die Bereitstellung der Daten über einen WFS als auch die Weiterverarbeitung durch einen WPS beinhalten. Folgeprozesse können unter anderem die Analyse des Ergebnisses im Kontext einer Veränderungsdetektion, die Nutzung einer durch Featuretransfers angereicherten Geometrie zur Routenplanung oder den Aufbau einer Multirepräsentationsdatenbank zum Ziel haben.

4.4 WICHTIGE TEILASPEKTE

Die im Folgenden beschriebenen Randbedingungen tragen grundlegend zum Erfolg einer dienstebasierten Geodatenfusion bei. Betrachtet werden zunächst Möglichkeiten zur Generalisierung von Ausgangsdaten sowie die Nutzung verfügbarer Metainformationen. Anschließend werden Aspekte der Verkettung, Automatisierung und Performanz von Diensten sowie Möglichkeiten zur Evaluierung des Gesamtprozesses analysiert.

4.4.1 Generalisierung

Eine korrekte Zuordnung von realweltlichen Objekten in den Datensätzen ist unter anderem von deren Abstraktionsebenen abhängig. Eine Anpassung geometrischer und struktureller Unterschiede kann daher notwendig sein und ist durch

eine Modellgeneralisierung in der Vorverarbeitung der Datensätze möglich.

Bei einem geometrischen oder topologischen Matching können Unterschiede im Referenzmaßstab beteiligter Datensätze, welcher unter anderem den Auflösungsabstand beeinflusst, zu Konflikten führen. Des Weiteren besteht die Möglichkeit, dass sich Repräsentationsformen von Objekten durch unterschiedliche thematische Gewichtung oder Aufnahmeverfahren, und damit unabhängig vom Maßstab, voneinander unterscheiden. Diese Unterschiede müssen entweder durch eine Generalisierung minimiert oder im Matchingprozess berücksichtigt werden. Die Generalisierung kann über beliebige Prozessierungsdienste erfolgen und passt Ausgangsdaten in der Regel an den kleinsten vorhandenen Referenzmaßstab an. Dabei können unter anderem die in Beard (1991) beschriebenen Constraints der Generalisierung angewandt werden. Diese Constraints werden durch einen Referenzdatensatz definiert und müssen nach einer Generalisierung von allen Ausgangsdaten erfüllt werden. Verwendbare Operatoren der geometrischen Modellgeneralisierung umfassen die Objektauswahl, Anpassung der Repräsentationsform, Zusammenfassung von Objekten und Vereinfachung der Geometrie auf Basis eines definierten Auflösungsabstandes.

Differenzen in der thematischen Struktur von Datensätzen basieren auf unterschiedlichen Datenmodellen, in denen Attribute oftmals anders klassifiziert oder kategorisiert werden. Diese resultieren in unterschiedlichen Wertebereichen, Detaillierungsgraden oder Klassengrenzen korrespondierender Attribute und sind Gegenstand der thematischen Generalisierung. Wesentlich für den Erfolg der Generalisierung ist ein Schema-Mapping zwischen den Datensätzen, in dem Attribute mit gleicher thematischer Aussage aufeinander abgebildet werden. Dieses kann interaktiv oder durch ein Schema-Matching erreicht werden und bildet die Grundlage für eine Reklassifikation oder thematische Auswahl von Objekten. Ohne Schema-Mapping ist keine kontrollierte thematische Generalisierung möglich, da notwendige Referenzattribute nicht bestimmt werden können.

Die Entscheidung über zu treffende Generalisierungsmaßnahmen kann aus einem Vergleich der Datensätze oder manuellen Vorgaben heraus erfolgen. Je nach Anwendung und Ausgangskonfiguration sind unterschiedliche Anforderungen an eine Generalisierung geknüpft. Bei der Möglichkeit einer On-the-fly Generalisierung von Ausgangsdaten, wie sie unter anderem in van Oosterom et al. (2006) beschrieben wird, ist keine geometrische Generalisierung im Fusionsprozess notwendig, da erforderliche Referenzmaßstäbe dynamisch aus einem vorgehaltenen Datensatz abgeleitet werden können. Eine thematische Generalisierung kann in diesem Fall weiterhin von Bedeutung sein, da die thematische Attributierung eines Datensatzes nicht direkt von der räumlichen Auflösung abhängig ist.

4.4.2 Metadaten

Metadaten können in vielen Bereichen der Geoprozessierung notwendige Informationen bereitstellen. Notwendige Metadatenelemente in einer GDI werden in EU (2008) sowohl für Geodaten als auch Geodaten- und Geoprozessierungsdienste festgelegt. Im Folgenden werden daraus zentrale Elemente und ihre Bedeutung bzw. mögliche Verwendung in einer dienstebasierten Fusion aufgelistet:

- Ein Ressourcenbezeichner ist notwendig für die eindeutige Identifizierung einer bereitgestellten Geoinformation. Bei mehrmaligem Zugriff auf eine Ressource stellt dieser sicher, dass es sich um den selben Datensatz oder Dienst handelt.
- Die Angabe zur Thematik ist ein wesentlicher Faktor bei der themengestützten Suche nach geeigneten Daten und Diensten.
- Ein Schlüsselwort ist insbesondere bei Prozessierungsdiensten einer Fusion von Bedeutung. Es definiert die Art des Dienstes und ermöglicht sowohl die Suche nach geeigneten Diensten als auch die Suche nach entsprechenden Alternativen. Von Bedeutung sind unter anderem:
 - *metadataStatisticalCalculationServices*: Dienste für statistische Berechnungen (z.B. Analyse von Ausgangsdaten und Ergebnis),
 - *spatialFeatureManipulationServices*: Objektbearbeitungsdienste (z.B. Map Alignment oder Konsistenzprüfung),
 - *spatialFeatureMatchingServices*: Vergleichsdienste für die Suche homologer Objekte,
 - *thematicChangeDetectionServices*: Erkennungsdienst für Veränderungen als mögliches Anwendungsbeispiel einer Fusion
- Die geographische BoundingBox kann zum Vergleich der räumlichen Lage der Daten verwendet werden. Eine Ableitung der BoundingBox ist, neben den Metadaten, auch über die Features möglich.
- Angaben zum Zeitbezug geben Informationen zur Aktualität der Daten und können für temporale Veränderungsmessungen genutzt werden.
- Die Qualität oder Genauigkeit der Datensätze kann bei der Festlegung von Toleranzen im Matchingprozess verwendet werden.
- Die räumliche Auflösung ist maßgeblich für eine geometrische Modellgeneralisierung.

- Zugangs- und Nutzungsbeschränkungen der Ressourcen können die Verwendung von Daten und Diensten in einer Fusion rechtlich einschränken und müssen daher beachtet werden.
- Metadaten zu Metadaten geben unter anderem Auskunft zur verwendeten Sprache der Metadaten und ermöglichen somit deren Auswertung.

Durch ein großes Spektrum möglicher Ausgangsdaten variieren auch mögliche Quellen datenbezogener Metainformationen. In vielen Rasterdatenformaten sind diese beispielsweise im Header eines Datensatzes gespeichert. Auch eine Trennung der Metadaten von weiteren Geoinformationen ist möglich. Ein Beispiel dafür ist das Shapefile-Format, bei dem Geometrie, Attribute, Projektion und Metadaten separat vorgehalten werden. Die dritte, und im Kontext einer GDI zu favorisierende, Möglichkeit ist die Bereitstellung von Metadaten über die Schnittstelle vorhandener Geodatendienste. Diese können direkt abgerufen und verarbeitet oder durch einen Katalogdienst verfügbar gemacht werden.

Auch Dienstemetadaten können über die jeweiligen Schnittstellen abgefragt werden. Zu diesem Zweck müssen alle OGC Web Services eine *GetCapabilities*-Operation implementieren, deren Response syntaktische und semantische Beschreibungen der Dienste beinhalten kann [OGC (2007a)]. Dies ermöglicht die Katalogisierung von Diensten und bildet die Grundlage für die Automatisierung von Prozessabläufen.

4.4.3 Service Chaining

Die Aggregation bestehender Services zu Serviceketten wird oft als wesentlicher Faktor für die Wertschöpfung einer serviceorientierten Architektur gesehen [vgl. Einspanier et al. (2003)]. Durch die Orchestrierung vorhandener Dienste lassen sich anwendungsspezifische Prozesse mit geringen Implementierungs-, Integrations- und Wartungskosten erstellen. Diese abstrakten Dienste beinhalten beliebig viele atomare Prozesse, von denen jeder einen Diensteaufruf beinhaltet [McGovern et al. (2003)].

Die Durchführung eines Service-Chaining kann manuell, teilautomatisiert oder vollautomatisiert erfolgen [vgl. Alameh (2003), Kiehle et al. (2007)]. Die manuelle Verkettung von Diensten wird ausschließlich vom Nutzer koordiniert. Sie beinhaltet einen sequenziellen, manuellen Aufruf benötigter Dienste und setzt dabei eine entsprechende fachliche Expertise des Nutzers voraus. Als halbautomatisch wird die zentrale Koordination eines Chains durch einen Mediation-Dienst beschrieben. Die Grundlage für diesen bildet ein Skript- oder Konfigurationsfile

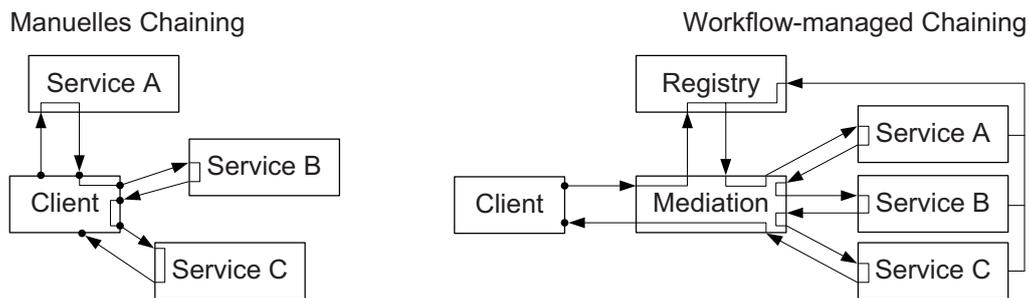


Abbildung 14 – Vergleich von manuellem und Workflow-managed Chaining von Web Services. Während das manuelle Chaining einen sequentiellen, nutzerkoordinierten Aufruf von Services beinhaltet, wird ein Workflow-managed Chaining mit Hilfe eines Mediation-Dienstes durchgeführt.

zur Beschreibung einer entsprechenden Servicesequenz. Der strukturelle Unterschied zwischen den beiden genannten Möglichkeiten ist in Abbildung 14 vereinfacht dargestellt. Die vollautomatisierte Zusammenstellung einer Dienstekette basiert dahingegen auf selbstorganisierenden Netzen und ist eng mit der Entwicklung des Semantic Web verbunden.

Im Folgenden werden drei Möglichkeiten zur Modellierung eines Service-Chains vorgestellt: die *Business Process Execution Language* (BPEL), die Nutzung der WPS-Schnittstelle und die skriptbasierte Aggregation von Diensten durch den Client.

BPEL ist XML-basiert und nach van der Aalst et al. (2005) der de-facto Standard für die Orchestrierung von Web Services. Mit BPEL wird das Verhalten von Web Services in einem Geschäftsprozess modelliert und aufbauend auf WSDL, dem Standard des W3C für die Beschreibung von Netzwerkdiensten, die Erstellung von Prozessketten über einen zentralen Orchestrierungs-Dienst ermöglicht. Entsprechende WSDL-Dokumente für Web Processing Services müssen manuell erstellt oder aus der vorhandenen Schnittstelle abgeleitet werden. Die Nutzung von BPEL zur Orchestrierung von OGC Web Services wird unter anderem in Kiehle et al. (2007), Stollberg und Zipf (2007) und Schaeffer (2008) beschrieben.

Die WPS-Schnittstelle bietet zwei Möglichkeiten zur Erstellung von Service-Chains [vgl. Stollberg und Zipf (2007)]. Zunächst ist eine Verwendung des WPS als Mediation- bzw. Workflow enactment service möglich. Dieser "Composite WPS" führt nach dem Aufruf eine dynamische oder im Vorfeld definierte Verkettung vorhandener Dienste aus. Eine zweite Möglichkeit stellt die Verschachtelung eines WPS-Request dar, in dem Ausgangsdaten eines Request wiederum einen WPS-Aufruf beinhalten können. Durch die

Verwendung eines kaskadierenden Aufrufes können Datentransferleistungen durch die direkte Kommunikation beteiligter Dienste verringert werden. Dies ist jedoch, aufgrund der zunehmenden Komplexität der Anfrage, nur bei der Erstellung einfacher Prozessketten praktikabel.

Die Aggregation von Daten- und Prozessierungsdiensten kann auch über ein clientseitiges Skript erfolgen, wobei der Aufruf vorhandener Dienste durch den Nutzer parametrisiert werden kann. Entsprechende Skriptsprachen müssen Nutzereingaben verwalten, einen WPS auf dieser Grundlage aufrufen und einen entsprechenden WPS-Response verarbeiten können. Dieser Ansatz ist weniger flexibel als BPEL, kann in einigen Anwendungsfällen jedoch ausreichend sein.

Bei der Umsetzung von Service-Chains müssen zahlreiche Aspekte berücksichtigt werden. Wesentliche Merkmale werden in Alameh (2003) und Kiehle et al. (2007) beschrieben und beinhalten:

- die Transparenz von Dienstketten, welche je nach äußerer Sichtbarkeit einzelner Prozessabläufe in transparent, transluzent und opak unterschieden wird. Eine hohe Transparenz eines Prozesses kann zu einem besseren Verständnis der Prozessierung und des Ergebnisses beitragen, setzt für deren Interpretation allerdings eine entsprechende fachliche Expertise des Nutzers voraus [Schaeffer (2008)].
- die Mitführung von Prozessinformationen, festgehalten in Metadaten des Ergebnisses, welche dem Nutzer wichtige Informationen zur Verarbeitung und Fusion von Geodaten liefern können.
- die Fehlerberichterstattung, die vom Nutzer verwendet werden kann, um gegebenenfalls Probleme zu beheben. Dabei sind sowohl Art als auch Ort des Fehlers anzugeben. Nach McGovern et al. (2003) sollte nach Auftreten eines Fehlers der Ausgangszustand vor dem ersten Serviceaufruf wieder hergestellt werden.
- die Anzahl verwendeter Dienste und deren Granularität, welche direkten Einfluss auf Reaktionszeiten und Transferleistungen innerhalb des Gesamtprozesses haben.
- die Registrierung verwendeter WPS in einem Katalogdienst, welche deren Beschreibung und Parametrisierung umfasst. Diese Informationen können unter anderem in WPS-Profilen festgehalten und für eine Verkettung der Dienste verwendet werden.

4.4.4 Automatisierung und Performanz

Die Automatisierung trägt entscheidend zur Effizienz einer dienstebasierten Fusion bei und betrifft die Suche, Verkettung und Parametrisierung geeigneter Teilprozesse unter Minimierung notwendiger Nutzerinteraktionen. Die Suche nach Diensten erfolgt im Regelfall über deren Veröffentlichung in Katalogdiensten, in denen sowohl Angaben zu den Diensten als auch zu angebotenen Prozessen, beispielsweise über WPS-Profile, festgehalten werden. Um einen automatisierten Abgleich von Dienstebeschreibungen und Nutzeransprüchen zu ermöglichen, sind jedoch semantische Beschreibungen notwendig. Eine Umsetzung unter Verwendung von Ontologien zur Beschreibung von Geoprozessen wird in Lutz (2005) ausführlich diskutiert. Auf dieser Grundlage ist auch eine dynamische Verkettung von Diensten möglich, bei der ein Ergebnis ohne Interaktion direkt generiert und dem Nutzer zur Verfügung gestellt werden kann [Kiehle et al. (2007)]. Eine automatische Parametrisierung von Prozessen erfordert dahingegen eine syntaktische und semantische Beschreibung der Ausgangsdaten. Je nach Anwendungszweck und Prozesskonfiguration können dies Informationen zur Geometrie, Thematik, Struktur oder Metadaten der Datensätze sein. Diese können über die entsprechende Schnittstelle eines Geodatendienstes oder direkt in den Ausgangsdaten bereitgestellt oder referenziert werden. Um Parameter ohne Nutzerinteraktion ermitteln zu können, sind die semantischen Beschreibungen von zentraler Bedeutung.

Auf der einen Seite Kennzeichen einer verteilten Architektur, auf der anderen Seite wesentlicher Faktor bei Betrachtung der Performanz einer Prozesskette ist der Datentransfer zwischen Diensten. Dieser kann daher nicht verhindert, sollte aber auf das Wesentliche reduziert werden. Ein erster Schritt zur Minimierung dieser Transferleistungen ist die Beschränkung der Ausgangsdaten auf die zur Fusion notwendigen Features. Diese Selektion kann sowohl durch thematische als auch räumliche Einschränkungen erfolgen. Zur Filterung von Daten in einem WFS *GetFeature*-Request kann zu diesem Zweck das optionale Query-Element genutzt werden. Wenn Daten durch einen Dienst nicht verändert oder auf einem Server zwischengespeichert werden, stellt auch die Verwendung von Datenreferenzen eine Möglichkeit zur Reduktion des transferierten Datenvolumens dar. Da die Kommunikation zwischen Diensten zumeist in leistungsstarken Netzwerkkombinationen stattfindet, gilt dies vor allem für die Kommunikation zwischen Client und Server [Foerster und Schäffer (2007)].

Bei der Verarbeitung großer Datenbestände rückt die Leistung des eigentlichen Prozesses in den Fokus der Betrachtung, weshalb Möglichkeiten zur Op-

timierung von Algorithmen oder Prozessstrukturen an Bedeutung gewinnen. Ein erster Ansatz ist hierbei die Subsetbildung der Ausgangsdaten über ein Raster oder definierte Features im Datensatz. Diese Fragmente können über eine asynchrone Client-Server Kommunikation parallel verarbeitet, beispielsweise gematcht, und anschließend zu einem Ergebnis zusammengesetzt werden [Scholten et al. (2006)]. Ein Workflow enactment service kann dabei zur Steuerung der Sequenz eingesetzt werden. Als Erweiterung dieses Ansatzes kann die Verwendung der GRID-Computing-Technologie [Foster und Kesselman (1998)] gesehen werden, welche die Verteilung von Prozessen innerhalb eines Netzwerkes vorsieht. Umsetzungen der parallelen Verarbeitung von Geodaten im Kontext eines WPS finden sich in Baranski (2008) und Lanig et al. (2008). Des Weiteren kann, je nach Anwendung und Ausgangskonfiguration, ein Caching zur Performanzsteigerung eines Prozessierungsdienstes beitragen. Diese Möglichkeit wird unter anderem in Scholten et al. (2006) und Foerster und Schäffer (2007) diskutiert und kann bei mehrmaliger Nutzung eines Dienstes mit gleichen Ausgangsdaten angewandt werden.

4.4.5 Evaluierung

Aus Informationen zum Prozessablauf und einem Vergleich des Ergebnisses mit gestellten Qualitätsvorgaben kann eine Aussage zur *Quality of Service* (QoS) abgeleitet werden. Diese spielt eine entscheidende Rolle bei der Bewertung und Qualitätssicherung verwendeter Dienste und Dienstketten hinsichtlich ihrer Eignung für eine dienstebasierte Fusion von Geodaten. Eine entsprechende Analyse kann sowohl durch einen darauf spezialisierten Dienst als auch manuell erfolgen.

Prozessinformationen sollten verwendete Dienste und deren Beschreibung, die Parametrisierung der Teilprozesse sowie Angaben zum Ablauf der gesamten Prozessierung beinhalten. Ein Großteil dieser Informationen kann direkt in den Metadaten des Ergebnisses gespeichert werden. Zusätzlich können Zwischenergebnisse verfügbar gemacht und gegebenenfalls zur Analyse herangezogen werden. Die Speicherung von Zwischenergebnissen kann sowohl am Server als auch am Client erfolgen. Im Folgenden werden wichtige prozessbeschreibende Elemente aufgeführt, die zu einer Evaluierung des Fusionsprozesses beitragen können:

- die Laufzeit der Prozessierung gibt Auskunft zur Performanz des Prozesses und setzt sich aus Transfer- und Prozessleistungen zusammen.
- die Beschreibung verwendeter Daten und Dienste ermöglicht eine Folge-

prozessierung ähnlich strukturierter Daten über die selben Dienste. Zudem kann anhand dieser Informationen eine Bewertung der Dienstekette, beispielsweise hinsichtlich ihrer Eignung für bestimmte Ausgangskonfigurationen, erfolgen.

- verwendete Parameter in den Teilprozessen, beispielsweise Toleranzen oder transferierte Attribute, können die Grundlage für deren iterative Anpassung während einer Prozessierung bilden. Des Weiteren ermöglichen sie, in Kombination mit Angaben zu verwendeten Daten und Diensten, eine exakte Reproduktion des Prozesses.
- Zwischenergebnisse, wie die angepassten Datensätze oder ein Linkfile, dienen der Evaluierung von Teilprozessen und können in einer Fehleranalyse verwendet werden. Nach einem Prozessabbruch oder einer fehlerhaften Fusion kann über die Analyse von Zwischenergebnissen der Einstiegspunkt für eine erneute Prozessierung unter Verwendung alternativer Dienste oder angepasster Parameter festgelegt werden.
- statistische Angaben zum Matching können wichtige Informationen zur Qualität des Matchingprozesses geben. Zudem ist die Verwendung dieser Angaben in einer interaktiven oder automatisierten Nachbearbeitung möglich, da potentielle Fehlzuordnungen schneller erkannt und gegebenenfalls korrigiert werden können.
- die Kennzeichnung nicht gematchter Features bildet die Grundlage für eine Kontrolle bzw. Verifizierung eines Feature-Transfers oder ein interaktives Matching von Features in der Nachbearbeitung einer Fusion. Des Weiteren können Angaben zu transferierten Features bei deren Integration in den Referenzdatensatz oder zur Konsistenzprüfung des Ergebnisses verwendet werden.

Weitere Möglichkeiten der Evaluierung und des Qualitätsmanagements basieren in der Regel auf Prozessmetadaten, dem fusionierten Datensatz oder bereitgestellten Zwischenergebnissen. Im Abschluss eines Fusionsprozesses sollte möglichst eine zusätzliche manuelle Kontrolle erfolgen, um die Qualität des Fusionsergebnisses sicherzustellen. Eine automatisierte Auswertung, unter Verwendung statistischer Analysen und bereitgestellter Prozessinformationen, kann den dabei erforderlichen manuellen Aufwand allerdings stark reduzieren.

5 IMPLEMENTIERUNG

Im Folgenden wird die prototypische Umsetzung zentraler Bausteine der Konzeption beschrieben und somit die Machbarkeit einer Fusion von Geodaten im Kontext einer GDI demonstriert. Die implementierten Teilprozesse basieren auf dem OGC WPS-Standard und können daher direkt in einer GDI integriert werden. Für die Umsetzung werden Straßendaten der Modelle ATKIS und OSM verwendet. Folgende Anwendungsmöglichkeiten stehen dabei im Vordergrund:

1. die Aktualisierung bzw. Erweiterung von ATKIS-Daten durch einen Transfer von OSM-Features, welche zu einer Kostenersparnis bei der Fortführung von ATKIS-Datensätzen führen kann. Vorhandene ATKIS-Daten werden dabei als Referenz genutzt, da sie in der Regel geometrisch genauer und thematisch umfangreicher attributiert sind. Die Eignung von OSM-Daten für die Erweiterung von ATKIS-Daten hängt stark von der anschließenden Verwendung ab und muss je nach Anwendungsfall geprüft werden.
2. die Verifizierung der OSM-Straßenklassifikation über einen Attributtransfer. Durch den Vergleich der Klassifikation korrespondierender Objekte in den ATKIS- und OSM-Datensätzen kann eine Aussage zur Genauigkeit der OSM-Klassifikation getroffen werden. Grundsätzlich wird dabei von einer fehlerfreien thematischen Attributierung der ATKIS-Daten ausgegangen.

Zur Durchführung der Fusion wurden drei aufeinander aufbauende, aber voneinander unabhängige WPS implementiert. Diese ermöglichen den Vergleich von Ausgangsdaten auf Basis definierter Merkmale, das Matching enthaltener Objekte auf Grundlage von Ähnlichkeitsmaßen sowie die Fusion der Daten durch Feature- und Attributtransfers. Die Verkettung der Dienste erfolgt durch eine skriptbasierte Aggregation und kann über ein HTML-Formular parametrisiert werden. Eine dienstebasierte Generalisierung von Daten wurde in der Implementierung nicht durchgeführt, da die verwendeten Datensätze geometrisch vergleichbar aufgelöst sind.

In den folgenden Abschnitten werden zunächst die verwendeten Ausgangsdaten und Softwarekomponenten beschrieben. Anschließend erfolgt die Beschreibung der implementierten Prozesse und eine Evaluierung der erreichten Ergebnisse.

5.1 AUSGANGSDATEN

Als Grundlage für die Implementierung wurden Straßendaten des Stadtgebietes Dresden aus den Datenmodellen ATKIS¹ und OSM² verwendet. Diese enthalten einfache lineare Elemente mit entsprechender Attributierung und werden über eine WFS-Schnittstelle bereitgestellt.

5.1.1 ATKIS

ATKIS ist Bestandteil des AAA-Modells der *Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland*³ (AdV). Dieses besteht aus dem *Amtlichen Festpunktinformationssystem* (AFIS), dem *Amtlichen Liegenschaftskatasterinformationssystem* (ALKIS) sowie dem *Amtlichen Topographisch-Kartographischen Informationssystem* (ATKIS) und stellt die Geobasisdaten der Bundesländer in digitaler Form dar.

Die Entwicklung von ATKIS, einer amtlichen Beschreibung der Erdoberfläche mit Hilfe digitaler Landschafts- und Geländemodelle, wurde 1989 durch die AdV initiiert [AdV (2008)]. Die in der vorliegenden Arbeit verwendeten Daten wurden dem ATKIS Basis-DLM mit einem Referenzmaßstab von 1:10 000 bis 1:25 000 entnommen. Gefiltert wurden die Objektarten 3001 (Straße), 3002 (Weg) und 3005 (Straßenkörper) des für die verfügbaren Daten gültigen Objektartenkataloges in BKG (2005b). Weitere Modifikationen der Daten wurden nicht vorgenommen. Für die Bereitstellung von ATKIS-Daten über einen Geodatendienst stellt die AdV entsprechende GML-Schemata zur Verfügung. Aus pragmatischen Gründen wurde auf deren Verwendung allerdings verzichtet und ein dynamisch erstelltes GML-Schema verwendet.

Da es sich um amtlich erhobene Daten handelt, wird der ATKIS-Datensatz als Referenz für die Fusion genutzt. Im Vergleich zu den OSM-Daten wird eine höhere geometrische und thematische Genauigkeit sowie ein entsprechendes Qualitätsmanagement vorausgesetzt.

5.1.2 OpenStreetMap

Das Prinzip von OpenStreetMap entspricht einer *Commons-based peer production*, dem Aufbau von großen Datenbeständen durch eine Internet-Community

¹ATKIS-Daten: Stand: April 2008; Landesvermessungsamt Sachsen (Erlaubnis-Nr. 296/2008)

²OSM-Daten: Stand: März 2008; verfügbar unter <http://www.openstreetmap.org>

³AdV: <http://www.adv-online.de>

[vgl. Benkler (2002)]. Jeder in dem Projekt registrierte Nutzer kann Geodaten erstellen oder editieren und somit am Aufbau eines globalen, frei verfügbaren Geodatenbestandes mitwirken.

Das OSM-Projekt wurde 2004 gegründet, um einen freien Zugang zu Geoinformationen ohne Lizenzgebühren, wie sie bei amtlich geführten Geodaten zumeist erhoben werden, zu ermöglichen. Seitdem wird ein stetiges exponentielles Wachstum an Nutzern und Daten verzeichnet. Im September 2009 zählte OSM nach eigenen Angaben bereits über 160 000 registrierte Benutzer und mehr als 33 Millionen gekennzeichnete Wege weltweit. Des Weiteren existieren zahlreiche Unternehmen und Softwaretools, welche auf die Verarbeitung von OSM-Daten spezialisiert sind und Sekundärprodukte ableiten können.

Aufgenommen werden die Geodaten zumeist über GPS (*Global Positioning System*)-fähige mobile Endgeräte. Anschließend werden die Features vom Erfasser attribuiert und in OSM integriert. Für den Datenzugriff und -austausch stehen eine eigens entwickelte Schnittstelle sowie ein auf XML-basierendes Austauschformat zur Verfügung. Diese sind nicht OGC-konform, was die Nutzung von OSM-Daten in einer GDI nur durch Integration der Datenbestände, beispielsweise über Sekundärprodukte, möglich macht. Des Weiteren führt die Erfassung durch eine Nutzergemeinschaft zu einer ungleichmäßigen Abdeckung der Erdoberfläche und einem mangelhaften Qualitätsmanagement [Haklay und Weber (2008)]. Diese Unsicherheitsfaktoren müssen bei der Verarbeitung von OSM-Daten stets in Betracht gezogen werden.

Der vorliegende OSM-Datensatz enthält zahlreiche Objekte, die nicht in den ATKIS-Daten erfasst wurden, sowie ein großes Spektrum vorhandener Straßenklassifikationen. Daraus resultieren die bereits genannten Ziele der Implementierung: der Transfer von OSM-Features und die Verifizierung der OSM-Klassifikation.

Im Vorfeld der Fusion wurde der OSM-Datensatz an den ATKIS-Datensatz angepasst. Zunächst wurden die Daten in das *Gauß-Krüger*-Koordinatensystem (5. Meridianstreifen) transformiert und anschließend die topologische Konsistenz durch Verschneidung von Straßenelementen hergestellt. Diese Schritte hätten auch innerhalb einer GDI unter Verwendung entsprechender WPS erfolgen können, wurden aus pragmatischen Gründen jedoch manuell durchgeführt.

5.2 VERWENDETE SOFTWARE

Die Kernkomponente der Implementierung stellt das *Java*-basierte *52°North* WPS-Framework dar, welches zur Umsetzung OGC-konformer WPS genutzt werden kann. Durch den generischen Charakter dieses Open Source Frameworks können entwickelte Algorithmen ohne großen Aufwand eingebunden und mit einer WPS-Schnittstelle versehen werden. Bereits vorhandene Parser und Generatoren für Geodatenformate sowie Beispielprozesse unterstützen die Entwicklung eigener Prozessierungsdienste. Notwendig sind die Entwicklung des Algorithmus, die Erstellung eines entsprechenden DescribeProcess-Dokumentes sowie die Veröffentlichung des Prozesses im Framework.

Für die Umsetzung wurde die Entwicklungsumgebung *Eclipse*⁴ gewählt. Ein Tutorial für die Einbindung des *52°North* WPS-Frameworks in *Eclipse* unter Verwendung von *Subversion* (SVN) und *Apache Maven*⁵, findet sich in Schaefer (2009). Durch die Kompilierung des Frameworks wird ein *Web Application Archive* (WAR) erstellt, welches anschließend über eine *Apache Tomcat*⁶-Serverinstanz veröffentlicht werden kann.

Die Bereitstellung der Ausgangsdaten wurde über das ebenfalls *Java*-basierte Open Source-Projekt *Geoserver*⁷ realisiert. Dieses unterstützt mit dem WMS, WCS und WFS zentrale OGC-Standards für die Visualisierung und Bereitstellung von Geodaten. Unter Nutzung von *Geoserver* konnten daher alle in dieser Arbeit verwendeten ATKIS- und OSM-Daten mit einer WFS-Schnittstelle ausgestattet und veröffentlicht werden.

Für die Verkettung von Prozessierungsdiensten wurde die Skriptsprache PHP (*Hypertext Preprocessor*) inklusive der darin enthaltenen *libcurl*-Bibliothek gewählt. Das erstellte Skript wird ebenfalls über eine Serverinstanz bereitgestellt und fungiert in Kombination mit einem HTML (*Hyper Text Markup Language*)-Formular als Client-Interface der Implementierung.

5.3 AUFBAU

Die Umsetzung der prototypischen dienstebasierten Fusion von Geodaten erfolgte in drei Schritten:

⁴Eclipse: <http://www.eclipse.org>; verwendete Version: 3.3.2

⁵Apache Maven: <http://maven.apache.org>; verwendete Version: 2.0.9

⁶Apache Tomcat: <http://tomcat.apache.org>; verwendete Version: 6.0

⁷Geoserver: <http://geoserver.org>; verwendete Version: 1.7.6

1. Selektion und Aufbereitung der Ausgangsdaten sowie Einbindung in den *Geoserver* zur Bereitstellung über eine WFS-Schnittstelle,
2. Entwicklung der Algorithmen für den Vergleich, das Matching und die Fusion der Ausgangsdaten, Einbindung in das *52°North* WPS-Framework sowie Veröffentlichung als *Java*-Servlet über eine *Apache Tomcat*-Serverinstanz,
3. Erstellung des PHP-Skriptes für das Chaining der entwickelten WPS und Bereitstellung der Funktionalität über ein HTML-Formular.

Die wichtigsten Bestandteile und gleichzeitig wesentlichen Eigenleistungen werden im folgenden Abschnitt beschrieben. Dies betrifft den Aufbau der entwickelten Algorithmen und das Skript zur Verkettung der Teilprozesse.

5.3.1 Bestandteile

Vergleich der Ausgangsdaten

Der Aufbau des implementierten WPS zum Vergleich der Ausgangsdaten ist vereinfacht in Abbildung 15 dargestellt. Der Aufruf erfolgt, wie bei jedem der entwickelten Dienste, über HTTP/POST. Als Eingangsparameter werden Verbindungsinformationen zu den Geodatendiensten, die URL der WFS, die Bezeichnung der *Featurelayer* sowie das Attribut mit der *Unique Identifier* (UID) der Features übergeben. Diese werden genutzt, um anschließend eine Verbindung zu den Ausgangsdaten aufzubauen und notwendige Informationen zu den Schemata und Features zu erlangen. Der für diese Aufgabe konzipierte *WFSClnt* bildet damit die Grundlage für den darauf folgenden Vergleich der Datensätze. Erfasste Eigenschaften der Ausgangsdaten, auf denen der durchgeführte Vergleich beruht, sind:

- das Koordinatenreferenzsystem,
- der Geometriotyp,
- die gegenseitige Überlappungsfläche der Bounding Boxes,
- die Anzahl jeweils enthaltener Features,
- die durchschnittliche Anzahl von Stützpunkten pro Feature,
- die durchschnittliche Länge von Features.

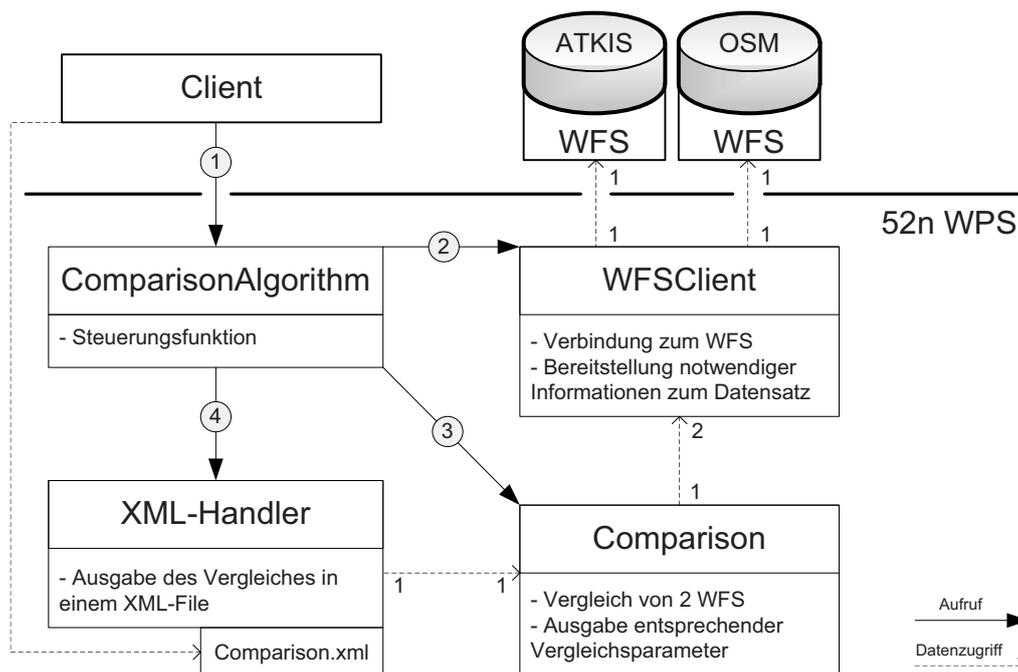


Abbildung 15 – Implementierung des Ausgangsdatenvergleiches. Notwendige Informationen werden über die WFS-Schnittstellen bereitgestellt. Das Ergebnis des Vergleiches wird XML-codiert gespeichert und abschließend im WPS-Response referenziert.

Ergebnisse des Vergleiches werden, neben den anfangs erfassten Verbindungs- und Informationen zu den Ausgangsdaten, in XML codiert und gespeichert. Ein Ausschnitt aus einer entsprechenden Ausgabe ist in Listing 1 aufgeführt. Das an den Client zurückgegebene Ergebnis des Teilprozesses enthält eine Referenz auf die erstellte XML-Datei.

Listing 1 – Ausschnitt aus dem Ergebnis eines Datensatzvergleiches. Dargestellt sind ermittelte Werte für den Vergleich der durchschnittlichen Länge, durchschnittlichen Anzahl von Stützpunkten und Gesamtanzahl enthaltener Features sowie die jeweilige Flächenüberlagerung der Bounding Boxes.

```

<comparison>
  <result>
    <avgLengthOfFeatFactor>1.46</avgLengthOfFeatFactor>
    <avgNumOfVertFactor>1.13</avgNumOfVertFactor>
    <numOfFeatFactor>1.72</numOfFeatFactor>
    <overlap1>79.65</overlap1>
    <overlap2>98.84</overlap2>
  </result>
  ...
</comparison>

```

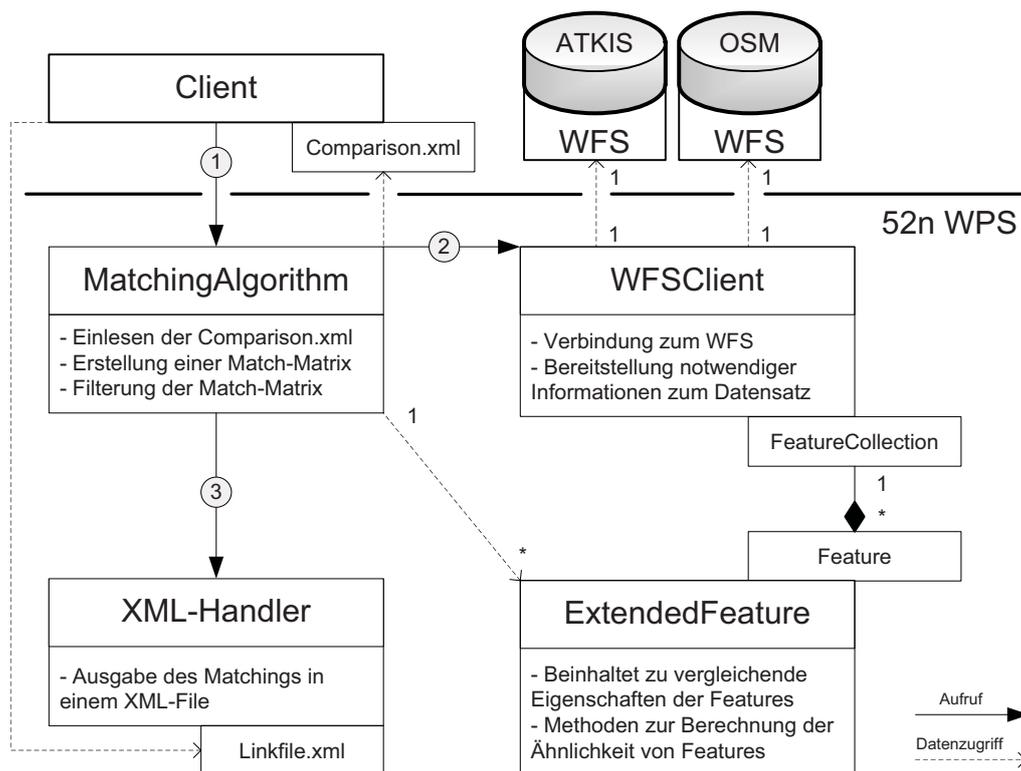


Abbildung 16 – Implementierung des Matchingprozesses. Nach dem Aufruf wird zunächst die Verbindung zu den Ausgangsdienstleistungen hergestellt. Features werden um zusätzliche Parameter ergänzt und auf dieser Basis aufeinander gematcht. Abschließend wird ein Linkfile erzeugt und im Response des WPS referenziert.

Matching

Die zuvor im Vergleich der Datensätze erzeugte XML-Datei wird als Eingangsparameter des Matchingprozesses verwendet, da sie neben den Vergleichsergebnissen auch notwendige Verbindungsinformationen zu den Ausgangsdaten enthält. Zudem werden im Algorithmus Toleranzen, zu vergleichende Attributfelder, welche in diesem Fall die Straßennamen beinhalten, sowie weitere prozessrelevante Parameter definiert. Der Ablauf des implementierten Matchingprozesses ist vereinfacht in Abbildung 16 dargestellt.

Das Einlesen der Ausgangsdaten erfolgt erneut über den implementierten *WFSCient*, dessen wesentliche Aufgabe in der Bereitstellung der *FeatureCollection* besteht. Daraufhin erfolgt eine Erweiterung aller enthaltenen Features um zusätzliche, im Matchingprozess relevante, Attribute. Aufbauend auf diesen Attributen wurden Methoden entwickelt, die einen Vergleich bzw. eine Aussage zur Ähnlichkeit von Features ermöglichen. Die darin ermittelte Wahrscheinlichkeit eines Matches wird durch folgende Parameter beeinflusst [vgl. Kapitel 3.3]:

- den Vergleich von Sinusität, Basiswinkel und Gesamtlänge der Features,

- die Berechnung der *Damerau-Levenshtein*-Distanz zwischen vorhandenen Straßennamen,
- den prozentualen Anteil eines Features in einem definierten Buffer des Vergleichsfeatures (buffer intersection),
- die *Hausdorff*-Distanz zwischen Features.

Durch den sequentiellen Vergleich aller in den Ausgangsdaten enthaltenen Features wird eine Matrix aufgebaut, in der potentielle Matches und Angaben zu deren Wahrscheinlichkeit enthalten sind. Diese wird anschließend, auf Basis eines definierten Schwellenwertes für die Wahrscheinlichkeit eines Matches, gefiltert und bildet die Grundlage für die Erstellung des XML-basierten Linkfile. In diesem werden alle potentiellen Matches und ihre zugehörigen Vergleichsparameter gespeichert. Zudem werden nicht gematchte Features der Ausgangsdatensätze sowie Verbindungsinformationen zu den Ausgangsdaten festgehalten. Durch dieses Verfahren werden auch multiple Matches gespeichert, welche im anschließenden Fusionsprozess gefiltert werden. Listing 2 stellt die Codierung eines potentiellen Matches im Linkfile dar. Im WPS-Response des Matchingprozesses ist erneut eine Referenz zum erstellten Ergebnis enthalten.

Listing 2 – Ergebnis aus dem Matching von Datensätzen. Bei einem potentiellen Match werden die entsprechende UID der Features, die Matchingwahrscheinlichkeit und ermittelte Vergleichsparameter festgehalten. Diese Informationen können für eine spätere Auswertung des Matchings verwendet werden.

```
⋮  
<match>  
  <id1>47433</id1>  
  <id2>X014POT005</id2>  
  <possibility>100</possibility>  
  <hausdorffDistancel1>5.69</hausdorffDistancel1>  
  <hausdorffDistancel2>9.19</hausdorffDistancel2>  
  <relativeLengthDifference>0.0291</relativeLengthDifference>  
  <angleDifference>2.8436</angleDifference>  
  <relativeSinuosityDifference>0.0314</relativeSinuosityDifference>  
  <damerauLevenshteinDistance>0</damerauLevenshteinDistance>  
</match>  
⋮
```

Fusion

Als Ausgangsdaten für den Fusionsprozess werden die Referenz auf das erstellte Linkfile und der Bezeichner des zu transferierenden Attributes verwendet. Nach

dem Aufruf wird zunächst das Linkfile eingelesen und verarbeitet. Multiple Matches werden in diesem Schritt auf Basis der ermittelten Wahrscheinlichkeiten gefiltert, woraufhin jedes Feature im Referenzdatensatz eine eindeutige Zuordnung im Vergleichsdatensatz erhält. Eine Ausnahme bilden die im Matchingprozess nicht zugeordneten Objekte, welche im Linkfile bereits als "not matched" gekennzeichnet wurden.

Im darauf folgenden Schritt wird ein neuer *FeatureType* zur Beschreibung der Features im Fusionsergebnis erzeugt. Dabei wird der *FeatureType* des Referenzdatensatzes um drei Attribute erweitert. Dies sind die UID der zugeordneten Features im Vergleichsdatensatz, das für den Transfer vorgesehene Attribut und die Wahrscheinlichkeit von Matches zur Unterstützung einer interaktiven Nachbearbeitung.

Auf Grundlage der vorangegangenen Schritte erfolgt die eigentliche Fusion der Datensätze. Grundsätzlich bleibt dabei die Geometrie des Referenzdatensatzes bei allen Matches erhalten. In diesem Fall erfolgt die Anreicherung der Daten durch den Attributtransfer. Mit den zuvor nicht gematchten Features des Vergleichsdatensatzes wird anschließend der Featuretransfer durchgeführt, bei denen ursprüngliche Attribute des Referenzdatensatzes auf Null und die Matchingwahrscheinlichkeit auf den Wert 8888 gesetzt werden. Nicht gematchte Features des Referenzdatensatzes bleiben dahingegen unverändert bestehen und werden mit dem Wert 9999 in der Matchingwahrscheinlichkeit gekennzeichnet.

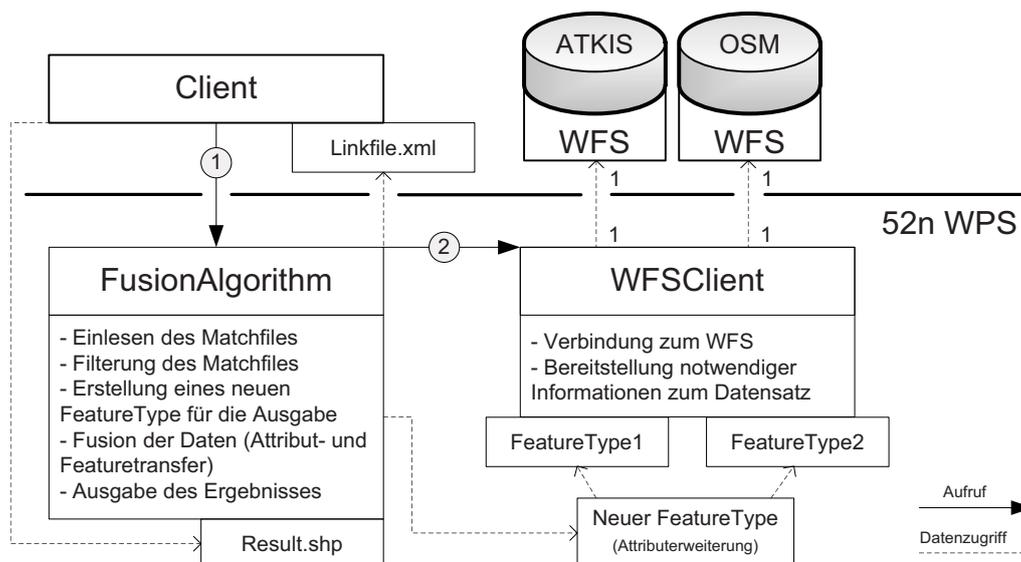


Abbildung 17 – Ablauf des implementierten Fusionsprozesses. Dieser setzt sich im Wesentlichen aus der Filterung des Linkfiles, der Bildung eines neuen FeatureTypes sowie der Fusion der Daten über einen Attribut- und Featuretransfer zusammen. Das Ergebnis wird als Shapefile bereitgestellt und im WPS Response referenziert.

Der vereinfachte Ablauf des Fusionsprozesses ist in Abbildung 17 dargestellt. Aufgrund der Nachbearbeitung und besseren Unterstützung durch verwendete Softwaretools wird im Ergebnis des Fusionsprozesses kein GML-Stream, sondern ein Shapefile erstellt. Im Response des WPS wird dieses referenziert und kann anschließend extern visualisiert, evaluiert oder weiterverarbeitet werden.

Client

Als Client für die dienstebasierte Fusion wurde ein einfaches Web-Interface auf Basis von HTML und PHP realisiert. Über ein HTML-Formular werden alle notwendigen Parameter für die Prozessierung eingelesen. Dies sind die *WFS-FeatureLayer* und UID's der Ausgangsdaten sowie der Bezeichner des zu transferierenden Attributes im Vergleichsdatensatz. Der sequentielle Aufruf der implementierten WPS wird über die *libcurl*-Bibliothek in PHP gesteuert. Ergebnisse der Teilprozesse werden nach dem WPS-Aufruf eingelesen und darin enthaltene Ergebnisreferenzen im Client dargestellt sowie gegebenenfalls als Input-Parameter der Folgeprozessierung verwendet. Ein WPS-Beispielaufruf mittels PHP und cURL ist in Listing 3 dargestellt.

Listing 3 – Aufruf eines WPS mittels PHP und cURL. Der in der Variable `$requestXML` enthaltene WPS-Request wird via HTTP/POST an eine zuvor definierte WPS-Instanz gesendet. Der WPS-Response wird anschließend in der Variable `$responseXML` gespeichert und ausgewertet.

```
:
:
$requestURL = 'http://localhost:8080/wps/WebProcessingService';
$chApi = curl_init($requestURL);
curl_setopt($chApi, CURLOPT_RETURNTRANSFER, 1);
curl_setopt($chApi, CURLOPT_POST, 1);
curl_setopt($chApi, CURLOPT_POSTFIELDS, $requestXML);
curl_setopt($chApi, CURLOPT_HTTPHEADER, $requestHeader);
$responseXML = simplexml_load_string(curl_exec($chApi));
:
:
```

5.3.2 Durchführung der Fusion

Zur Bewertung der implementierten Prozesse wurden drei Testgebiete aus den vorhandenen ATKIS- und OSM-Datensätzen gebildet. Dies sind ein ca. 4 km² großes Areal im Innenstadtbereich Dresdens sowie jeweils ca. 16 km² große Areale aus der Dresdner Heide und dem Stadtrandbereich. Eine Charakterisierung dieser, in Abbildung 18 in Ausschnitten dargestellten, Testgebiete befindet sich in Tabelle 6.

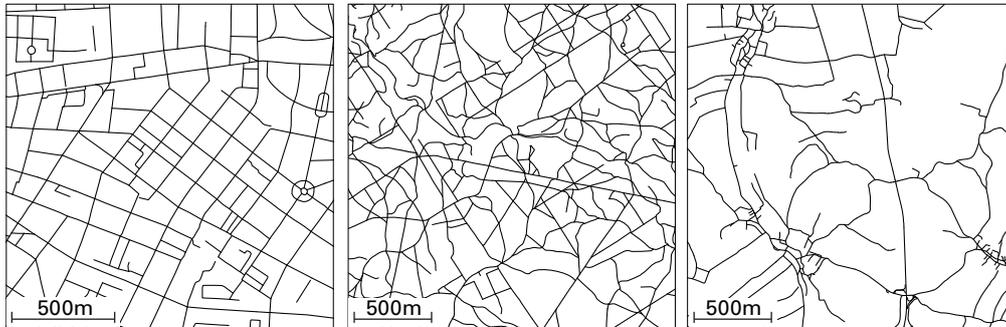


Abbildung 18 – Verwendete Testgebiete in der Implementierung. Dargestellt sind Ausschnitte der ATKIS-Testdatensätze aus dem Innenstadtbereich (links), der Dresdner Heide (mitte) sowie dem Stadtrandbereich (rechts).

Tabelle 6 – Charakterisierung der verwendeten Testgebiete hinsichtlich Objektanzahl, Struktur, Linienbeschaffenheit, Homogenität und Dichte der Ausgangsdatensätze.

Datensatz	Features (ATKIS:OSM)	Charakterisierung
Innenstadt	700:1078	Klare Strukturen erkennbar, viele kurze und gerade Straßenabschnitte, ATKIS- und OSM-Daten relativ homogen, hohe Dichte an Features pro Flächeneinheit.
Dresdner Heide	1009:991	Keine Strukturen erkennbar, Mischung verschiedener Straßenlängen und Formen, große Abweichungen zwischen ATKIS- und OSM-Daten, hohe Dichte an Features pro Flächeneinheit.
Stadtrand	564:467	Kaum Strukturen erkennbar, viele lange und kurvige Straßenelemente, ATKIS- und OSM-Daten weichen teilweise voneinander ab, relativ geringe Dichte der Features.

Die verwendeten UID's der Ausgangsdaten sind die Attribute "ID" in den OSM- und "OBJ_ID" in den ATKIS-Datensätzen. Um einen anschließenden Vergleich der Klassifikationen beider Datensätze zu ermöglichen, wurde das Attribut "Typ", welches die Straßenklassifikation der OSM-Features enthält, für den Attributtransfer gewählt. Die im Matchingprozess verglichenen Straßennamen werden den Attributen "Streetname" in OSM und "GN" in ATKIS entnommen.

5.4 EVALUIERUNG DER ERGEBNISSE

Die Bewertung der implementierten Prozesskette erfolgt auf Grundlage der drei verwendeten Testgebiete. In der Analyse erreichter Ergebnisse werden Probleme aufgezeigt und gegebenenfalls Lösungsansätze vorgestellt.

5.4.1 Ergebnisbetrachtung

Innenstadtbereich

Das farblich codierte Fusionsergebnis im Innenstadtbereich ist in Abbildung 19 abgebildet. Die Farbcodierung vereinfacht eine visuelle Bewertung des Ergebnisses durch die Darstellung der Matchingwahrscheinlichkeit sowie Kennzeichnung transferierter und nicht gematchter Features.

Insgesamt wurden im gewählten Areal 637 von 700 Features aus dem ATKIS-Datensatz gematcht. Von den übrigen 63 wurden 19 nicht gematcht, obwohl sie ein jeweils korrespondierendes Element im OSM-Datensatz besaßen. Die durchschnittliche Matchingwahrscheinlichkeit liegt bei rund 70 Prozent, wobei über die Hälfte aller Matches eine Wahrscheinlichkeit von mehr als 80 Prozent besitzen. Manuelle Stichproben haben allerdings gezeigt, dass bereits bei einer Wahrscheinlichkeit von 20 bis 40 Prozent in mehr als der Hälfte der Fälle eine korrekte Zuordnung getroffen wurde. Sobald die Wahrscheinlichkeit über 40 Prozent betrug, lag die Quote der korrekten Zuordnungen bei über 90 Prozent. Die reale Wahrscheinlichkeit eines Matches liegt demnach deutlich höher als die durch das Matching ermittelte.

Durch den Featuretransfer wurden insgesamt 370 Features aus den OSM-Daten in das Fusionsergebnis übernommen. Von diesen wurden 16 Prozent transferiert, obwohl jeweils entsprechende Elemente im ATKIS-Datensatz vorhanden waren. Der Großteil der transferierten Features wurde in OSM als "footway" (69 Prozent) oder "service" (22 Prozent) klassifiziert.

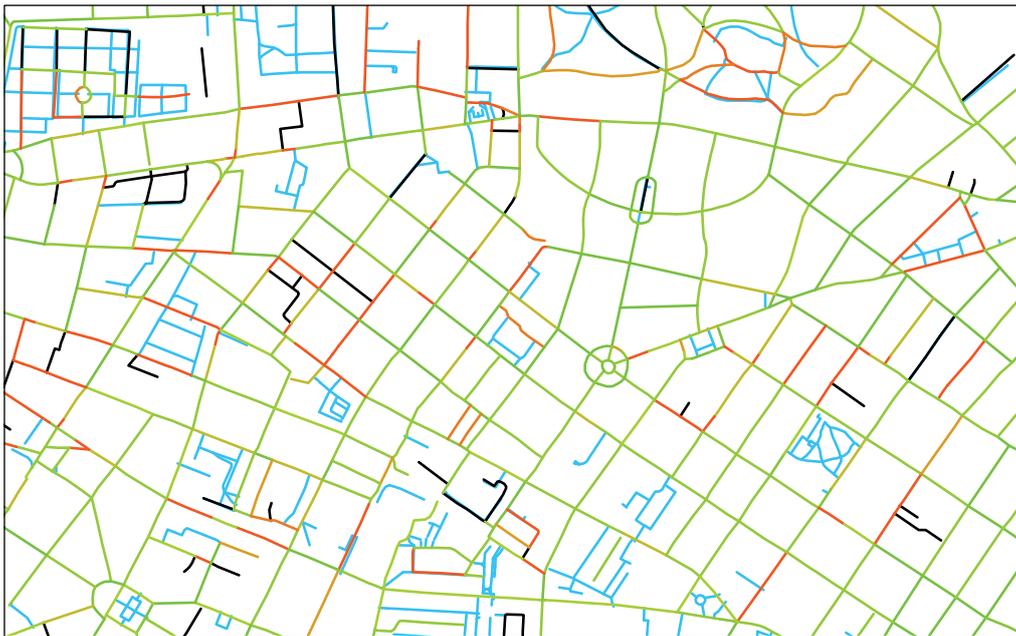


Abbildung 19 – Ergebnis der Fusion im Innenstadtbereich. Die Matchingwahrscheinlichkeit ist von rot (niedrig) bis grün (hoch) farbcodiert dargestellt. Nicht gematchte Features der Ausgangsdaten sind schwarz (ATKIS) bzw. blau (OSM) abgebildet.

Dresdner Heide und Stadtrandbereich

Die Analyse der Fusionsergebnisse in den Gebieten Dresdner Heide und Stadtrand werden hier zusammengefasst, da sich die Ergebnisse nicht signifikant voneinander unterscheiden. In Abbildung 20 wird das Fusionsergebnis in der Dresdner Heide unter Verwendung der bereits in Abbildung 19 verwendeten Symbolisierung dargestellt.

Insgesamt wurden in beiden Beispielen rund 80 Prozent der ATKIS-Objekte gematcht. Von den übrigen 20 Prozent wurde allerdings nahezu die Hälfte nicht gematcht, obwohl korrespondierende Elemente im OSM-Datensatz existierten, womit die Fehlerquote gefundener Matches höher liegt als im Innenstadtbereich. Die durchschnittliche Matchingwahrscheinlichkeit fällt mit rund 60 Prozent niedriger aus, was insbesondere auf eine Häufung von Zuordnungen mit einer Wahrscheinlichkeit zwischen 20 und 40 Prozent zurückzuführen ist. Die manuelle Stichprobe zeigt, dass diese vor allem durch die hohe Anzahl kurzer Straßenabschnitte in den Datensätzen hervorgerufen wird. Die durch den Algorithmus ermittelte Wahrscheinlichkeit entspricht daher im Durchschnitt der realen Wahrscheinlichkeit von Matches.

Im Featuretransfer wurden aus den OSM-Daten 222 Objekte in der Dresdner Heide sowie 53 Objekte im Stadtrandgebiet in des Fusionsergebnis übernommen. Diese sind größtenteils als "track", "footway" oder "secondary" klassifi-



Abbildung 20 – Ergebnis der Fusion in der Dresdner Heide. Die Matchingwahrscheinlichkeit ist von rot (niedrig) bis grün (hoch) farbcodiert dargestellt. Nicht gematchte Features der Ausgangsdaten sind schwarz (ATKIS) bzw. blau (OSM) abgebildet.

ziert. Die Anzahl durch Mismatchings verursachter und damit fehlerhafter Featuretransfers ist auch an dieser Stelle höher als im Innenstadtbereich. Während die Fehlerquote bei Featuretransfers in der Dresdner Heide bei rund 30 Prozent liegt, fällt diese im Stadtrandgebiet mit 50 Prozent deutlich schlechter aus. Dies kann größtenteils mit der bereits erwähnten hohen Anzahl kurzer Straßensegmente in den Ausgangsdaten begründet werden.

Auswertung der Ergebnisse

Trotz eines rein featurebasierten und ohne topologische Kontrollen durchgeführten Matchingprozesses konnte ein akzeptables Ergebnis in der Zuordnung homologer Objekte erreicht werden. Ohne Nutzerinteraktionen wurden in unstrukturierten Testgebieten knapp 75 Prozent und in Innenstadtbereichen bis zu 90 Prozent der ATKIS-Features richtig zugeordnet. Auf dieser Basis konnte sowohl der Attribut- als auch der Featuretransfer erfolgreich durchgeführt werden. Des Weiteren konnte die interaktive, visuelle Nachbearbeitung und Evaluierung durch Ermittlung der Matchingwahrscheinlichkeit und eine entsprechend farbcodierte Darstellung der Fusionsergebnisse wesentlich erleichtert werden.

Durch den Featuretransfer wurden die drei ATKIS-Testdatensätze um insgesamt 645 OSM-Features erweitert. Ein Großteil der fehlerhaften Transfers, zu meist hervorgerufen durch die unterschiedliche Segmentierung der Ausgangs-

daten, konnte durch die Kennzeichnung der transferierten Daten visuell leicht identifiziert und behoben werden. Korrekte, aber dennoch im Fusionsprozess eliminierte Matches konnten durch eine Überlagerung des Ergebnisses mit den Ausgangsdaten ergänzt werden. Dies war notwendig, da gefilterte Features im Fusionsprozess nicht erneut gematcht werden.

Mit Hilfe des durchgeführten Attributtransfers konnte im Anschluss ein Vergleich der Straßenklassifikationen erfolgen. Dabei wurden die Features im ATKIS-Datensatz um ein entsprechendes Attribut erweitert und die Klassifikation des jeweils korrespondierenden OSM-Features zugeordnet. In den ATKIS-Daten wurden die Objektarten "Straße", "Weg" und "Straßenkörper" für den Vergleich verwendet. Weitere Attribute zur Differenzierung dieser Klassifikation wurden nicht in Betracht gezogen. Im OSM-Datensatz sind dahingegen 12 Straßentypen klassifiziert: "track", "footway", "residential", "service", "living_street", "secondary", "tertiary", "path", "cycleway", "motorway", "motorway_link" und "unclassified". Die häufigsten Zuordnungen zur ATKIS-Objektart "Straße" sind "residential" (54%), "secondary" (14%) und "track" (10%). Der Objektart "Weg" sind dahingegen primär "footway" (54%) und "track" (43%) zugewiesen worden. Da lediglich 17 Objekte der ATKIS-Datensätze als "Straßenkörper" klassifiziert waren, konnte in diesem Fall keine sichere Aussage zu möglichen Zuordnungen getroffen werden. Im Idealfall sollte, unter Annahme eines thematisch korrekt attribuierten ATKIS-Datensatzes, eine OSM-Klasse genau einer ATKIS-Objektart zugeordnet sein. Vorhandene Überschneidungen, wie etwa bei der Klassifizierung "track", sind daher Matchingfehlern oder, mit höherer Wahrscheinlichkeit, Mängeln im Qualitätsmanagement der OSM-Datensätze zuzuschreiben.

5.4.2 Problemanalyse und Erweiterungsmöglichkeiten

Eines der grundlegenden Probleme der implementierten Fusion ist die teils heterogene Beschaffenheit der ATKIS- und OSM-Datensätze. Insbesondere die fehlende fachliche Expertise bei der Erfassung der OSM-Daten und deren Kombination mit einem mangelhaften Qualitätsmanagement machen eine Vorverarbeitung notwendig. Ein Schwachpunkt ist beispielsweise die fehlende topologische Konsistenz der OSM-Daten, welche Zuordnungsprobleme im Matchingprozess verursachen kann. Des Weiteren führen unterschiedliche Schreibweisen oder vereinzelt unterschiedliche Benennungen von Straßen zu Konflikten beim Matching auf thematischer Ebene.

Aus den genannten Unterschieden resultiert ein Großteil der fehlerhaften Zuordnungen von Objekten. Ursache dafür ist das rein featurebasierte Matching

ohne Verarbeitung kontextbezogener oder semantischer Informationen. In einigen Fällen kann das Matching bereits durch Anpassung verwendeter Parameter verbessert werden. Oftmals ist dies allerdings nicht ausreichend. Daher müssten verwendete Matchingkriterien erweitert oder neue Kriterien implementiert werden. Die Einführung bzw. Erweiterung topologischer und semantischer Vergleichsmaße könnte das bisherige Matching und darin insbesondere die Filterung von Matches unterstützen. Zur Verarbeitung unterschiedlicher Segmentierungen und Verifizierung multipler Matches kann zudem das im Vorfeld genannte Buffergrowing-Verfahren eingesetzt werden. Als Erweiterung der *Hausdorff*-Distanz ist auch die Berechnung der *Fréchet*-Distanz möglich, um die Fehlerquote bei der Zuordnung von Objekten zu verringern.

Der Attributtransfer ist direkt abhängig vom Matchingprozess, da falsche oder nicht getroffene Zuordnungen die Auswertung transferierter Attribute direkt beeinflussen. Daher müssen Matchingfehler beseitigt und gegebenenfalls Fehler-toleranzen in die Bewertung des Attributtransfers einbezogen werden.

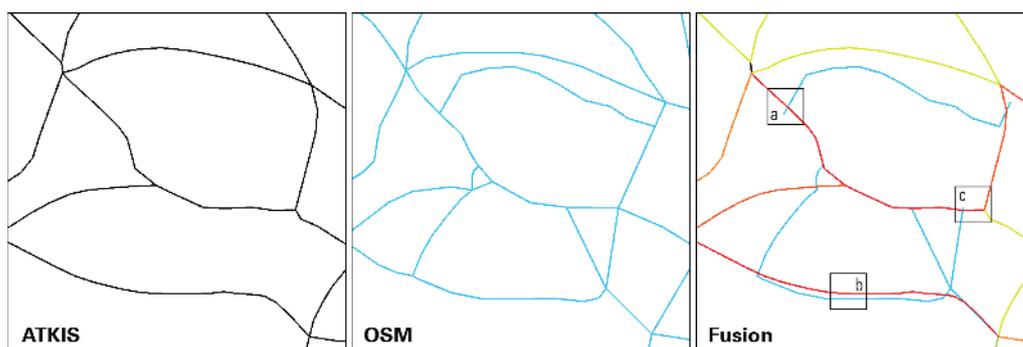


Abbildung 21 – Häufig auftretende Probleme beim Featuretransfer. Dargestellt sind die fehlerhafte Einpassung eines Objektes (a), die Entstehung redundanter Objekte (b) sowie die Bildung eines fehlerhaften Kreuzungselementes (c).

Fehler beim durchgeführten Featuretransfer basieren ebenfalls auf der Heterogenität der ATKIS- und OSM-Datensätze. Häufig auftretende Probleme beim Featuretransfer in der vorgestellten Implementierung sind exemplarisch in Abbildung 21 dargestellt. Dies sind:

- a) geometrische Passungenauigkeiten, welche zu Überlappungen (overshoots) oder Lücken (undershoots) im Ergebnis führen. Diese können in einem erweiterten Fusionsprozess oder einer Nachbearbeitung durch Festlegung einer Snapping-Toleranz eliminiert werden, wodurch ein Element automatisch auf das nächstliegende Element innerhalb der definierten Toleranz gezogen wird. Dieser Schritt ist notwendig, um die topologische Konsistenz des Ergebnisses sicherzustellen und eine Weiterverarbeitung, z.B. ein Routing auf dem Fusionsergebnis, zu ermöglichen.

- b) die Bildung redundanter Objekte durch den Transfer von Features trotz vorhandenem Matchingpartner im Referenzdatensatz. Diese kann entweder durch die bereits genannten Erweiterungen des Matchingalgorithmus' oder eine Konsistenzprüfung unter Verwendung der Ausgangsdaten vermieden werden.
- c) das Zustandekommen fehlerhafter Kreuzungselemente durch unterschiedliche Geometrien. Auch hier kann eine Snapping-Toleranz angewendet werden, welche im dargestellten Beispiel jedoch unzureichend wäre. Daher müsste eine topologische Verifizierung von Matches oder eine Topologieprüfung in der Nachbearbeitung des Ergebnisses erfolgen.

Neben der Verfeinerung verwendeter Algorithmen ist eine Weiterentwicklung der Systemstruktur notwendig, um die implementierten Dienste vollständig in eine GDI zu integrieren. Die Nutzung einer BPEL-Engine zur Verkettung der Dienste kann dazu beitragen die Interoperabilität und Flexibilität der Anwendung zu erhöhen. Zudem besteht die Möglichkeit, Fusionsalgorithmen in bestehende Datendienste einzubinden. Dabei könnte der in Foerster et al. (2008) diskutierte Ansatz zur Nutzung von Prozessierungsfunktionalitäten im *Geoserver* weiterentwickelt werden und eine ad-hoc Anreicherung über WFS-Schnittstellen bereitgestellter Daten ermöglichen.

6 ABSCHLIESSENDE BETRACHTUNG UND AUSBLICK

Das Ziel dieses Kapitels ist, die im Laufe der Arbeit gewonnenen Erkenntnisse darzustellen und zu bewerten. Zunächst erfolgt eine Diskussion der Ergebnisse auf Basis der gestellten Forschungsfragen. Aus dieser werden bestehende Probleme und ein damit verbundener Forschungsbedarf abgeleitet. Abschließend erfolgt ein Fazit der vorliegenden Arbeit.

6.1 DISKUSSION DER ERGEBNISSE

In dieser Arbeit werden zunächst Grundlagen zu Geodateninfrastrukturen und der Fusion von Geodaten gelegt. Darauf aufbauend wird eine dienstebasierte Fusion unter Verwendung des OGC *Web Processing Service* konzipiert. Mit der Schnittstellenbeschreibung, Granularität, Orchestrierung und Automatisierung von Diensten bzw. Dienstketten wird der Schwerpunkt auf die technische Umsetzung gelegt. Als proof-of-concept Implementierung dient die Zusammenführung von Straßendaten der Modelle ATKIS und OSM, welche im Laufe dieser Arbeit erfolgreich umgesetzt wurde. Im Folgenden werden die in der Einleitung der Arbeit gestellten Forschungsfragen aufgegriffen und diesbezügliche Ergebnisse diskutiert.

Welche Vorteile bietet eine dienstebasierte Fusion von Geodaten im Kontext einer GDI?

Die zunehmende Verlagerung verfügbarer Geodaten von monolithischen auf dienstebasierte Strukturen führt zu einem stetig wachsenden Angebot an Geoinformationen im Internet. Da viele geowissenschaftliche Analysen und Prozesse auf der Kombination von Geodaten mehrerer Ressourcen basieren, müssen entsprechende Methoden zur Zusammenführung dieser Daten entwickelt werden. Zudem kann eine Fusion zur Minimierung redundanter Geoinformationen, einem der grundlegenden Ziele einer Geodateninfrastruktur, beitragen. Die Abgrenzung von monolithischen GIS-basierten Methoden ist für die Definition des

Anwendungsspektrums von besonderer Bedeutung, da mit aktuell verfügbaren Standards nicht alle Funktionalitäten im Bereich der Geodatenfusion in eine GDI übertragen werden können. Dies betrifft insbesondere die Verarbeitung umfangreicher Datenbestände und den Interaktionsbedarf zahlreicher Anwendungen.

Wesentlicher Vorteil einer dienstebasierten Geoinformationslandschaft ist die Möglichkeit auf Primärquellen und somit auf hochaktuelle Daten zugreifen zu können. Die ad-hoc Fusion dieser Datenbestände kann insbesondere in zeitkritischen Anwendungen zur Informationsgenerierung genutzt werden und als Grundlage für Entscheidungsträger dienen. Durch derzeitige Einschränkungen in der Performanz vieler Dienste trifft dies insbesondere auf die Fusion kleinerer Datenbestände zu. Alle Ergebnisse einer dienstebasierten Prozessierung sind über das Internet verfügbar und können softwareunabhängig verwertet werden. Da die Prozessierung durch externe Dienste erfolgt, kann der Zugriff über beliebige Endgeräte, beispielsweise Laptops oder PDAs, erfolgen (*Thin Client* Struktur).

Wichtig für die Durchführung einer dienstebasierten Fusion ist die Existenz entsprechender Prozessierungsdienste innerhalb einer GDI. Je mehr Dienste und Dienstalternativen verfügbar sind, desto flexibler kann eine Anwendung auf gegebene Ausgangskonfigurationen reagieren. Für die Suche nach fusionierbaren Datensätzen zur Anreicherung gegebener Geodaten, ist neben der Verfügbarkeit von Prozessierungsdiensten auch ein entsprechendes Spektrum verfügbarer Daten von Bedeutung. Um die Zusammenführung von Geodaten mehrerer Quellen zu automatisieren, müssen thematische und strukturelle Verknüpfungspunkte zwischen den Datensätzen vorhanden und beschrieben sein. Sobald die genannten Kriterien erfüllt sind, kann eine Dienststruktur zur spontanen, interoperablen und autonomen Fusion von Geodaten in einer GDI geschaffen werden.

Inwieweit können Metadaten dazu beitragen, Teilprozesse und Ergebnisse einer Fusion zu optimieren?

Viele essentielle Informationen zu Geodaten und Diensten können aus deren Metadaten entnommen werden. Diese können im Fusionsprozess verwendet werden und zur Verringerung notwendiger Nutzerinteraktionen und erforderlicher Nutzerexpertise beitragen. Vorhandene Standards im Bereich Metadaten und deren Anwendung in einer Geodatenfusion werden in den Kapiteln 2.5 und 4.4.2 beschrieben.

Bereits in der Analyse von Ausgangsdaten können Metadaten verwendet werden, um Informationen zur Geometrie, Thematik oder Struktur der Datensätze zu erlangen. Die Festlegung notwendiger Anpassungsschritte nach dem Vergleich

von Ausgangsdaten beruht zum großen Teil auf diesen Informationen. Um Interoperabilität zu gewährleisten, sind die in EU (2008) standardisierten Metadaten-elemente von zentraler Bedeutung. Die Suche nach geeigneten Diensten und deren Verkettung erfolgt dabei über dienstebeschreibende Metadaten. Formelle Schnittstellenbeschreibungen, beispielsweise über WPS-Profile, können zur Automatisierung dieser Schritte beitragen. Eine weitere Möglichkeit zur Verwendung von Metadaten ist die Prozessanalyse, festgehalten im Ergebnis einer Fusion. Diese kann verwendet werden, um das erzielte Ergebnis zu analysieren und gegebenenfalls eine erneute Prozessierung unter Verwendung angepasster Ausgangsparameter durchzuführen. Neben dieser iterativen Optimierung kann eine Evaluierung vorgenommen und somit das Nutzerverständnis für Teilprozesse und Ergebnis erhöht werden.

Die Bereitstellung von Metainformationen erfolgt in der Regel über die Schnittstellen der Geodaten- und Prozessierungsdienste. Die Interpretierbarkeit durch Mensch und Maschine muss dabei gewährleistet sein, um einerseits das Verständnis für die Prozessierung zu erhalten und andererseits die automatisierte Erzeugung und Ausführung eines Workflows zu ermöglichen. Letzteres hängt wesentlich von der Bereitstellung semantischer Metainformationen zu Daten und Diensten ab, da erst diese eine autonome, dynamische Verkettung von Diensten innerhalb einer GDI ermöglichen. Die maschinenlesbare Darstellung der Semantik von Daten und Diensten bildet daher einen Schwerpunkt aktueller Forschung im Kontext von Geodateninfrastrukturen [vgl. Egenhofer (2002), Lutz (2005), Kiehle et al. (2007), Brauner et al. (2009)].

Welche Teilprozesse und Schnittstellen sind erforderlich, um eine Geodatenfusion innerhalb einer GDI zu ermöglichen?

Die wichtigsten Schnittstellen bei der Durchführung einer Geodatenfusion sind Datendienste zur Bereitstellung erforderlicher Ausgangsdaten sowie Prozessierungsdienste zur Durchführung der Fusion. Zur Bereitstellung der Ausgangsdaten können unter anderem die OGC-Standards zu WFS, WCS und SOS genutzt werden. Für die Prozessierung wurde der WPS-Standard entwickelt.

Zentraler Punkt bei der Konzeption einer dienstebasierten Geodatenfusion ist die Frage nach der Granularität einzelner Teilprozesse. In der vorliegenden Arbeit wurde dafür der Ansatz von McGovern et al. (2003) verfolgt, welche die grobgranulare Zerlegung des Gesamtprozesses in logische, fachlich voneinander abgegrenzte Bausteine vorsehen. Gekennzeichnet wird dies durch eine geringe Komplexität der Netzstruktur sowie die Minimierung notwendiger Datentransfer-

leistungen. Im Ergebnis wurden folgende Prozessierungsschritte konzipiert:

1. Evaluierung und Vergleich der Ausgangsdaten bezüglich ihrer Eignung für eine dienstebasierte Fusion. Die Auswertung geometrischer, thematischer und struktureller Eigenschaften führt im Ergebnis zur Definition möglicher und notwendiger Anpassungsschritte.
2. (optional) Angleichung der Ausgangsdaten unter Nutzung verfügbarer externer Dienste. Diese Vorprozessierung muss erfolgen, sobald eine Weiterverarbeitung der Ausgangsdaten unter gegebenen Umständen nicht erfolgen kann.
3. Matching der Datensätze über die Suche homologer Objekte. Die korrekte Zuordnung von Objekten bildet die Basis einer erfolgreichen Geodatenfusion.
4. Informationsgewinn durch Fusion der Datensätze. Dies kann eine Aktualisierung, Detektion von Unterschieden, Informationsanreicherung oder Datenintegration beinhalten.
5. (optional) Nachbearbeitung des Fusionsergebnisses. Dieses kann unter Nutzung vorhandener externer Dienste an Nutzervorgaben angepasst und anschließend als Endergebnis der Fusion bereitgestellt werden.

Durch diese fünf abstrakten Teilprozesse sind alle Aspekte einer Geodatenfusion abgedeckt. Die entsprechenden Dienste müssen unabhängig voneinander konzipiert werden, um auch in einem anderen Anwendungskontext einsetzbar zu sein. Die Verwendung optionaler, externer Dienste ist anwendungsabhängig und sollte bei der Erstellung des Gesamtprozesses berücksichtigt werden.

Welche Anforderungen stellt eine ad-hoc Zusammenführung von Geodaten an eine GDI?

Grundlegende Voraussetzungen für eine Zusammenführung von Geodaten innerhalb einer GDI sind die Verfügbarkeit fusionierbarer Daten sowie entsprechender Dienste zur Prozessierung. Metainformationen ermöglichen dabei die katalogbasierte Registrierung und Suche von Daten und Diensten. Die dort enthaltenen Schnittstellenbeschreibungen machen eine Orchestrierung vorhandener Dienste möglich, in der abstrakte Prozesse zur Lösung fachspezifischer Probleme erstellt werden können. Möglichkeiten zur Verkettung mehrerer WPS werden in Kapitel

4.4.3 beschrieben, wobei die Nutzung von BPEL den am häufigsten verwendeten und diskutierten Ansatz darstellt [vgl. van der Aalst et al. (2005), Kiehle et al. (2007), Schaeffer (2008)]. Für die Erstellung von BPEL-Dokumenten muss allerdings die Unterstützung von SOAP und WSDL durch die jeweiligen OGC Dienste gewährleistet sein [Kiehle et al. (2007)].

Interoperabilität zwischen Diensten ist die Grundvoraussetzung einer dienstebasierten Geoinformationslandschaft. Die Beschreibung vorhandener Dienste und Schnittstellen ist daher von besonderer Bedeutung. Der WPS Standard sieht daher die Verwendung von WPS-Anwendungsprofilen zur Umsetzung einer "fully-automated interoperability" vor [OGC (2007b)]. Anwendungen in diesem Bereich gibt es jedoch kaum, da Regeln zur konkreten Umsetzung fehlen. Ansätze zur Strukturierung von WPS-Profilen diskutiert Nash (2008). Weitere Möglichkeiten zur Charakterisierung von Diensten basieren unter anderem auf der Kombination syntaktischer und semantischer Informationen [Lemmens et al. (2007)] sowie der Entwicklung komplexer beschreibender Ontologien [vgl. Lutz (2005)].

Insbesondere bei der Entwicklung zeitkritischer Anwendungen und der Verarbeitung großer Geodatenbestände spielt die Performanz verwendeter Prozessierungsdiensten eine zentrale Rolle. Dies betrifft sowohl die Prozessleistung als auch den notwendigen Datentransfer zwischen Diensten. Brauner et al. (2009) identifizieren das *moving-code*-Paradigma, Möglichkeiten des GRID-Computing und *stateful processing* als vielversprechende Möglichkeiten zur Steigerung der Performanz von Prozessierungsdiensten. Weitere im Kontext einer GDI diskutierte Ansätze sind die asynchrone Kommunikation zwischen Client und Server, die Implementierung von Caching-Verfahren sowie die Filterung relevanter Daten [vgl. Scholten et al. (2006), Friis-Christensen et al. (2007)].

6.2 WEITERER FORSCHUNGSBEDARF

Zahlreiche Algorithmen zur Fusion von Geodaten sind bereits Teil aktueller Geoinformationssysteme. Für eine Umsetzung in einer GDI müssen diese als Prozessierungsdienst veröffentlicht und nutzbar gemacht werden. Bisherige Entwicklungen stellen jedoch zumeist Insellösungen dar und sind damit zu spezialisiert, um interoperabel in einer GDI verwendet werden zu können. Die Forschung an generischen Algorithmen ist daher essentiell, um die Geodatenfusion in heterogenen Geoinformationslandschaften zu ermöglichen. Die Flexibilität der Anwendung ist neben der Qualität allerdings auch von der Quantität angebotener Dienste abhän-

gig. Neben der Verarbeitung von Vektordaten bietet eine GDI ein weites Spektrum an Möglichkeiten zur Geodatenfusion. Diese umfassen unter anderem die Fusion in Echtzeit aufgenommener Daten im Bereich des Sensor-Web oder über die WCS-Schnittstelle bereitgestellter Coverages. Zudem müssen Möglichkeiten zur Anpassung von Geodaten, beispielsweise über die geometrische oder thematische Generalisierung, im Vorfeld einer Fusion analysiert und entwickelt werden. Wichtige Randbedingungen bilden dabei die Erreichbarkeit, Zuverlässigkeit, Interoperabilität, Skalierbarkeit und Sicherheit entwickelter Dienste [OGC (2009b)].

Die Performanz ist entscheidend für den praktikablen Einsatz einer dienstebasierten Fusion von Geodaten. In diesem Kontext vorhandene Ansätze müssen daher analysiert und weiterentwickelt werden. Die Nutzung verteilter Systemstrukturen für die parallele Prozessierung von Daten ist dabei einer der vielversprechendsten Möglichkeiten. Anwendungen des GRID-Computing stehen dabei im Vordergrund und müssen bezüglich ihrer Eignung für die Geoprozessierung geprüft und bewertet werden. Damit verbunden steht die Frage wie und wann Prozesse oder Geodaten optimal geteilt werden können [Brauner et al. (2009)]. Für ein Matching scheint dabei die räumliche Teilung von Datensätzen, ein sogenanntes *Tiling*, am besten geeignet. Da viele bisher entwickelte Algorithmen der Geodatenfusion eine interaktiver Nachbearbeitung notwendig machen, sollte der WPS-Standard in Zukunft *stateful processing* unterstützen, um eine direkte Weiterverarbeitung der Ergebnisse zu ermöglichen. Zur Verringerung des Transfervolumens bei der Verarbeitung großer Datenbestände müssen zudem Möglichkeiten zum Transfer der Algorithmen anstelle der Daten entwickelt werden. In diesem Zusammenhang ist auch die Verwendung eines Caching im Kontext von Prozessierungsdiensten zu analysieren.

Neben der Performanz einzelner Dienste ist die Möglichkeit zur Verkettung von Diensten in einer GDI zu beachten. Die Entwicklung von Workflow enactment services zur Generierung und Steuerung von Service-Chains muss vorangetrieben werden, um abstrakte Dienste zur Lösung problemspezifischer Anwendungen zu erstellen. Dafür ist insbesondere eine formale Beschreibung von Geodaten und -diensten notwendig. Eine Kategorisierung, beispielsweise über WPS-Anwendungsprofile, kann die Suche und dynamische Verkettung von Diensten vereinfachen. Neben syntaktischen sind jedoch vor allem semantische Beschreibungen relevant, um die Bildung autonom handelnder Netze und vollständige Automatisierung von Prozessen zu ermöglichen. Auch in einer Fusion können entsprechende Hintergrundinformationen zur Verarbeitung von Ausgangsdaten hilfreich sein. Die Semantik von Daten und Diensten ist daher bereits Gegenstand zahlreicher Forschungen und stellt nach Kiehle et al. (2007) „one of the

next frontiers in Geographical Information Science“ dar. Möglichkeiten zur standardisierten maschinenlesbaren Beschreibung von semantischen Informationen müssen in diesem Kontext weiterentwickelt und evaluiert werden.

Bisherige Betrachtungen zur Geodatenfusion sind fast ausschließlich auf die Entwicklung von Geodateninfrastrukturen bezogen. Die Verbreitung von Geoinformationen lässt sich aber auch zunehmend außerhalb einer, insbesondere auf institutioneller Ebene verbreiteten, GDI beobachten. Der Informationsfluss im Web 2.0 bewegt sich dabei in zwei Richtungen (Nutzer ↔ Nutzer), während die Kommunikation in einer GDI primär in eine Richtung läuft (Anbieter → Nutzer) [vgl. Mac Gillavry (2003)]. Dies resultiert im Aufbau von *Commons-based peer productions* und der Nutzung von Bevölkerungsteilen als “Voluntary Sensors“ [Goodchild (2007)]. Etablierte interaktive Projekte sind beispielsweise OpenStreetMap, Geocommons¹ oder Up2Maps². Zudem gibt es eine Vielzahl von Anbietern für frei zugängliche, jedoch nicht frei verfügbare, Geodaten. Allen gemein ist die fehlende Unterstützung vorhandener OGC-Standards. Die Kopplung an eine GDI ist zumeist nur über Sekundärprodukte möglich. Um diese Geodaten in einer Fusion nutzbar zu machen, müssen vorhandene Schnittstellen und Datenstrukturen analysiert und aufeinander abgebildet werden. Eine wichtige Aufgabe besteht daher in der Bewertung verschiedener Möglichkeiten zu deren Integration in eine GDI.

6.3 FAZIT

Die Geodatenfusion stellt eines der wichtigsten Verfahren zur Generierung eines Informationsmehrwertes aus bestehenden Ressourcen dar. Die vorliegende Arbeit thematisiert dabei deren Integration in eine Geodateninfrastruktur. In der Einleitung (Kapitel 1) werden in diesem Zusammenhang relevante Forschungsfragen definiert und im Laufe der Arbeit aufgegriffen. Basierend auf den Grundlagen zur GDI (Kapitel 2) und Geodatenfusion (Kapitel 3) wird eine dienstebasierte Geodatenfusion konzipiert (Kapitel 4), in der Teilprozesse und notwendige Schnittstellen beschrieben werden. Betrachtungen zur Interoperabilität, Performanz, Automatisierung und Evaluierung der Dienste sowie Metadatenanalyse und Generalisierung von Ausgangsdaten spielen dabei eine zentrale Rolle. Die Machbarkeit einer Geodatenfusion innerhalb einer GDI wird in einer prototypischen Implementierung (Kapitel 5) zur Fusion von Straßendaten der Modelle ATKIS und OSM bewie-

¹Geocommons: (<http://www.geocommons.com>)

²Up2Maps: <http://www.up2maps.net>

sen, in der drei Prozessierungsdienste mit Hilfe des *52°North* WPS-Framework umgesetzt und zu einer Dienstekette aggregiert wurden.

Dass die Fusion von Geodaten in einer GDI zunehmend an Bedeutung gewinnt, zeigt der Aufruf des OGC zu einer "Fusion Standards Study" [OGC (2009b)]. Diese soll zur Evaluierung existierender Standards bezüglich ihrer möglichen Verwendung in einer Geodatenfusion beitragen und unterscheidet dabei nach Sensor Fusion, Feature Fusion und Decision Fusion. Die Studie bildet zugleich eine der Grundlagen für die OGC *Web Services 7 Interoperability Initiative* (OWS-7).

In einer GDI ist sowohl die ad-hoc Fusion kleinerer als auch die Zusammenführung größerer Datenbestände möglich. Die Entwicklung kann dabei auch von klassischen GIS profitieren. Dies betrifft sowohl die Nutzung von GIS-Funktionalitäten im Backend eines WPS als auch die Einbindung von OGC Diensten in bestehende Desktop GIS. Zudem sind GDI-spezifische Anwendungen, wie die Fusion von Geoinformationen im Bereich des Sensor-Web, Gegenstand aktueller Forschung.

Parallel zur GDI etablieren sich Community-basierte interaktive Projekte zur Erstellung und Veröffentlichung von Geodaten im Internet. Durch die Einbindung dieser Daten in eine GDI ist ein fusionsgestützter Informationsgewinn auf thematischer oder geometrischer Ebene möglich. Die Fusion von ATKIS und OSM in dieser Arbeit hat gezeigt, wie dies technisch umgesetzt und ein Datensatz anhand von Nutzervorgaben attributiv angereichert werden kann. Des Weiteren wird in Auer et al. (2009) bereits die semantischen Aufwertung von OSM-Daten durch Transformation in das RDF-Datenmodell diskutiert. Eine Vielzahl solcher Eigenentwicklungen wird jedoch unweigerlich zu Problemen in der Interoperabilität und Flexibilität Web-basierter Anwendungen führen. Die Schaffung und Einhaltung entsprechender Standards wird daher entscheidend zur Evolution des *Semantic Spatial Web* [Egenhofer (2002)] beitragen.

Zur abschließenden Betrachtung wird erneut das, in der Einleitung aufgeführte, Zitat von John Naisbitt aufgegriffen: „We are drowning in information, but we are starved for knowledge“. Dieses stammt aus dem Jahr 1982 und besitzt auch heute noch Gültigkeit. Umgeben von einer Flut an Informationen, insbesondere über das Internet, fällt es schwer, konkretes Wissen daraus abzuleiten. Aktuell steht in vielen Bereichen die Sammlung von Daten im Vordergrund, da Mechanismen zur effektiven Auswertung fehlen. Die Fusion verfügbarer Informationen ist daher ein wichtiger Schritt zu einer möglichen Wissensbildung und wird ein Kernelement der zukünftigen Forschung im Kontext Web-basierter Anwendungen darstellen.

LITERATUR

- [AdV 2008] ARBEITSGEMEINSCHAFT DER VERMESSUNGSVERWALTUNGEN DER LÄNDER DER BUNDESREPUBLIK DEUTSCHLAND (ADV) (Hrsg.): *Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok)*. April 2008. – Version 6.0
- [Aier und Schelp 2008] AIER, Stephan ; SCHELP, Joachim: EAI und SOA – Was bleibt nach dem Hype? In: *Multikonferenz Wirtschaftsinformatik 2008*. GITO-Verlag, Berlin, 2008, S. 1469–1480
- [Alameh 2003] ALAMEH, Nadine: Chaining Geographic Information Web Services. In: *IEEE Internet Computing* 7 (2003), September, Nr. 5, S. 22–29. – URL <http://dx.doi.org/10.1109/MIC.2003.1232514>
- [Auer et al. 2009] AUER, Sören ; LEHMANN, Jens ; HELLMANN, Sebastian: *LinkedGeoData – Adding a Spatial Dimension to the Web of Data*. The 8th International Semantic Web Conference. Oktober 2009. – URL <http://linkedgeo.org/files/paper.pdf>
- [Baranski 2008] BARANSKI, Bastian: *Grid Computing Enabled Web Processing Service*. GI-Days, Münster. Juni 2008
- [Beard 1991] BEARD, M. K.: *Constraints on rule formation*. S. 121–135. In: BUTTENFIELD, Barbara P. (Hrsg.) ; MCMASTER, Robert B. (Hrsg.): *Map Generalization: Making Rules for Knowledge Representation*, Longman, London, 1991
- [Beard und Chrisman 1986] BEARD, M. K. ; CHRISMAN, Nicholas R.: Zipping: New Software for Merging Map Sheets. In: *Proceeding of American Congress on Surveying and Mapping 46th Annual Meeting*, 1986, S. 153–161
- [Beeri et al. 2005] BEERI, Catriel ; DOYTSHER, Yerach ; KANZA, Yaron ; SAFRA, Eliyahu ; SAGIV, Yehoshua: Finding Corresponding Objects when Integrating Several GeoSpatial Datasets. In: *Proceedings of the 13th annual ACM international workshop on Geographic information systems*, 2005, S. 87–96

- [Benkler 2002] BENKLER, Yochai: Coase's Penguin, or, Linux and The Nature of the Firm. In: *The Yale Law Journal* 12 (2002), Nr. 3, S. 369–446. – URL <http://www.yalelawjournal.org/pdf/112-3/BenklerFINAL.pdf>
- [Bergenheim et al. 2009] BERGENHEIM, Wolf ; SARJAKOSKI, L. T. ; SARJAKOSKI, Tapani: *A Web Processing Service for GRASS GIS to Provide on-line Generalisation*. 12th AGILE International Conference on Geographic Information Science, Hannover. Juni 2009
- [Bernard et al. 2005a] BERNARD, Lars ; CRAGLIA, Max ; GOULD, Michael ; KUHN, Werner: Towards an SDI Research Agenda. In: *11th EC GI & GIS Workshop – ESDI: Setting the Framework – Abstracts Handbook*, 2005, S. 147–151
- [Bernard et al. 2005b] BERNARD, Lars ; CROMPVOETS, Joep ; FITZKE, Jens: *Geodateninfrastrukturen – ein Überblick*. S. 3–8. In: BERNARD, Lars (Hrsg.) ; FITZKE, Jens (Hrsg.) ; WAGNER, Roland M. (Hrsg.): *Geodateninfrastruktur: Grundlagen und Anwendung*, Herbert Wichmann Verlag, Heidelberg, 2005
- [Bernard und Streit 2002] BERNARD, Lars ; STREIT, Ulrich: Geodateninfrastrukturen und Geoinformationsdienste: Aktueller Stand und Forschungsprobleme. In: *Publikationen der Deutschen Gesellschaft für Photogrammetrie und Fernerkundung, Band 11*, 2002, S. 11–20
- [Bilo und Bernard 2005] BILO, Michael ; BERNARD, Lars: *INSPIRE – Aufbau einer Infrastruktur für raumbezogene Informationen in Europa*. S. 18–28. In: BERNARD, Lars (Hrsg.) ; FITZKE, Jens (Hrsg.) ; WAGNER, Roland M. (Hrsg.): *Geodateninfrastruktur: Grundlagen und Anwendung*, Herbert Wichmann Verlag, Heidelberg, 2005
- [Bishr et al. 1999] BISHR, Yaser A. ; PUNDT, Hardy ; RÜTHER, Christoph: *Proceeding on the Road of Semantic Interoperability – Design of a Semantic Mapper Based on a Case Study from Transportation*. S. 203–215. In: VČKOVSKI, Andrej (Hrsg.) ; BRASSEL, Kurt E. (Hrsg.) ; SCHEK, Hans-Jörg (Hrsg.): *Interoperating Geographic Information Systems*, Springer-Verlag, Berlin - Heidelberg, 1999
- [BKG 2005a] BUNDESAMT FÜR KARTOGRAPHIE UND GEODÄSIE (Hrsg.): *Geoinformation und moderner Staat*. Informationsschrift des Interministeriellen Ausschusses für Geoinformationswesen (IMAGI). 2005. – 4. Auflage
- [BKG 2005b] BUNDESAMT FÜR KARTOGRAPHIE UND GEODÄSIE (Hrsg.): *Vektordaten Bundesrepublik Deutschland: Digitales Basis-Landschaftsmodell Basis-*

- DLM*. Januar 2005. – URL http://www.atkis.de/dstinfo/dstinfo2.dst_gliederung2?dst_ver=dst
- [Blum 1967] BLUM, Harry: A transformation for extracting new descriptors of shape. In: *Proceedings of the Symposium on Models for the Perception of Speech and Visual Form*, 1967, S. 362–380
- [Bollmann und Koch 2001] BOLLMANN, Jürgen (Hrsg.) ; KOCH, Wolf G. (Hrsg.): *Lexikon der Kartographie und Geomatik*. Spektrum Akademischer Verlag, Heidelberg - Berlin, 2001
- [Brauner 2008] BRAUNER, Johannes: Anbindung von GIS-Funktionalitäten an eine Geodateninfrastruktur über eine Web Processing Service Schnittstelle. In: *GIS – Zeitschrift für Geoinformatik* (2008), Nr. 3, S. 18–25
- [Brauner et al. 2009] BRAUNER, Johannes ; FOERSTER, Theodor ; SCHAEFFER, Bastian ; BARANSKI, Bastian: *Towards a Research Agenda for Geoprocessing Services*. 12th AGILE International Conference on Geographic Information Science, Hannover. Juni 2009
- [Burghardt et al. 2005a] BURGHARDT, Dirk ; EDWARDES, Alistair ; PURVES, Ross ; WEIBEL, Robert: Automatische Generalisierung dynamisch generierter Karten für mobile Endgeräte. In: *Kartographische Nachrichten* 55 (2005), Juni, Nr. 3, S. 119–125
- [Burghardt et al. 2005b] BURGHARDT, Dirk ; NEUN, Moritz ; WEIBEL, Robert: Generalization Services on the Web – Classification and an Initial Prototype Implementation. In: *Cartography and Geographic Information Science* 32 (2005), Oktober, Nr. 4, S. 257–268
- [Burstein et al. 2005] BURSTEIN, Mark ; BUSSLER, Christoph ; ZAREMBA, Michal ; FININ, Tim ; HUHNS, Michael N. ; PAOLUCCI, Massimo ; SHETH, Amit P. ; WILLIAMS, Stuart: A Semantic Web Services Architecture. In: *IEEE Internet Computing* 9 (2005), Nr. 5, S. 72–81. – URL <http://doi.ieeecomputersociety.org/10.1109/MIC.2005.96>
- [Čepický und Becchi 2007] ČEPICKÝ, Jáchym ; BECCHI, Lorenzo: Geospatial Processing via Internet on Remote Servers – PyWPS. In: *OSGeo Journal* (2007), Mai, Nr. 1
- [Cobb et al. 1998] COBB, Maria A. ; CHUNG, Miyi J. ; FOLEY III, Harold ; PETRY, Frederick E. ; SHAW, Kevin B. ; MILLER, H. V.: A Rule-based Approach for the

- Conflation of Attributed Vector Data. In: *Geoinformatica 2* (1998), März, Nr. 1, S. 7–35. – URL <http://dx.doi.org/10.1023/A:1009788905049>
- [Craglia 2003] CRAGLIA, Max (Hrsg.): *Contribution to the extended impact assessment of INSPIRE*. September 2003. – URL http://inspire.jrc.ec.europa.eu/reports/fds_report.pdf
- [Dostal et al. 2005] DOSTAL, W. ; JECKLE, M. ; MELZER, I. ; ZENGLER, B.: *Service-orientierte Architekturen mit Web Services: Konzepte–Standards–Praxis*. Spektrum Akademischer Verlag, Elsevier GmbH, München, 2005
- [Douglas und Peucker 1976] DOUGLAS, David H. ; PEUCKER, Thomas K.: Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. In: *The Canadian Cartographer* 10 (1976), Dezember, Nr. 2, S. 112–122
- [Doyle und Reed 2001] DOYLE, Allan ; REED, Carl: *Introduction to OGC Web Services*. White paper, Open GIS Consortium. Mai 2001. – URL http://portal.opengeospatial.org/files/?artifact_id=14973&version=1&format=pdf
- [Doytsher et al. 2001] DOYTSHER, Yerahmiel ; FILIN, Sagi ; EZRA, Eti: Transformation of Datasets in a Linear-based Map Conflation Framework. In: *Surveying and Land Information Systems* 61 (2001), September, Nr. 3, S. 159–169
- [Egenhofer 2002] EGENHOFER, Max J.: Toward the Semantic Geospatial Web. In: *Proceedings of the 10th ACM international symposium on Advances in geographic information systems, 2002*, S. 1–4
- [Egenhofer und Herring 1992] EGENHOFER, Max J. ; HERRING, John R.: *Categorizing Binary Topological Relations Between Regions Lines, and Points in Geographic Databases*. Technical Report, Department of Surveying Engineering, University of Maine. 1992
- [Einspanier et al. 2003] EINSPANIER, Udo ; LUTZ, Michael ; SENKLER, Kristian ; SIMONIS, Ingo ; SLIWINSKI, Adam: *Toward a Process Model for GI Service Composition*. GI-Days, Münster. Juni 2003
- [EU 2008] EUROPÄISCHEN UNION (Hrsg.): *Verordnung (EG) Nr. 1205/2008 der Kommission zur Durchführung der Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates hinsichtlich Metadaten*. Amtsblatt der Europäischen Union. Dezember 2008. – URL <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:0030:DE:PDF>

- [Fensel 2007] FENSEL, Dieter: *Enabling Semantic Web Services – The Web Service Modeling Ontology*. Springer-Verlag, Berlin - Heidelberg, 2007
- [Filin und Doytsher 1999] FILIN, Sagi ; DOYTSHER, Yerahmiel: A Linear Mapping Approach to Map Conflation: Matching of Polylines. In: *Surveying and Land Information Systems* 59 (1999), Nr. 2, S. 107–114
- [Filin und Doytsher 2000] FILIN, Sagi ; DOYTSHER, Yerahmiel: The Detection of Corresponding Objects in a Linear-Based Map Conflation. In: *Surveying and Land Information Systems* 60 (2000), Nr. 2, S. 117–128
- [Foerster 2004] FOERSTER, Theodor: *Geometrische Datenintegration für ein transnationales Routing-Szenario*, Westfälische Wilhelms-Universität Münster, Diplomarbeit, Dezember 2004
- [Foerster et al. 2009] FOERSTER, Theodor ; BURGHARDT, Dirk ; NEUN, Moritz ; REGNAULD, Nicolas ; SWAN, Jerry ; WEIBEL, Robert: *Towards an Interoperable Web Generalisation Services Framework – Current Work in Progress*. 11th ICA Workshop on Generalisation and Multiple Reslides, Montpellier. März 2009
- [Foerster und Schäffer 2007] FOERSTER, Theodor ; SCHÄFFER, Bastian: *A Client for Distributed Geo-processing on the Web*. S. 252–263. In: WARE, J. M. (Hrsg.) ; TAYLOR, George E. (Hrsg.): *Web and Wireless Geographical Information Systems*, Springer-Verlag, Berlin - Heidelberg, 2007
- [Foerster und Stoter 2006] FOERSTER, Theodor ; STOTER, Jantien: *Establishing an OGC Web Processing Service for generalization processes*. Workshop of the ICA Commission on Map Generalisation and Multiple Representation. Juni 2006
- [Foerster et al. 2007] FOERSTER, Theodor ; STOTER, Jantien ; KÖBBEN, Barend: *Towards a formal classification of generalization operators*. ICC, Moscow, Russia. 2007. – URL <http://www.durpondergronden.nl/ICC2007.pdf>
- [Foerster et al. 2008] FOERSTER, Theodor ; STOTER, Jantien ; LEMMENS, Rob: *An interoperable Web Service Architecture to Provide Base Maps Empowered by Automated Generalisation*. S. 255–275. In: RUAS, Anne (Hrsg.) ; GOLD, Christopher (Hrsg.): *Headway in Spatial Data Handling*, Springer-Verlag, Berlin - Heidelberg, 2008
- [Foster und Kesselman 1998] FOSTER, Ian (Hrsg.) ; KESSELMAN, Carl (Hrsg.): *The Grid: Blueprint for a New Computing Infrastructure: Blueprint for the New Computing Infrastructure*. Morgan Kaufmann Publishers, 1998

- [Friis-Christensen et al. 2007] FRIIS-CHRISTENSEN, Anders ; LUTZ, Michael ; OSTLÄNDER, Nicole ; BERNARD, Lars: Designing Service Architectures for Distributed Geoprocessing: Challenges and Future Directions. In: *Transactions in GIS* 11 (2007), Dezember, Nr. 6, S. 799–818
- [Gillman 1985] GILLMAN, Daniel W.: Triangulations for Rubber-Sheeting. In: *Proceedings of AUTOCARTO 7*, 1985, S. 191–199
- [Giunchiglia et al. 2007] GIUNCHIGLIA, Fausto ; YATSKEVICH, Mikalai ; SHVAIKO, Pavel: *Semantic Matching: Algorithms and Implementation*. S. 1–38. In: SPACCAPIETRA, Stefano (Hrsg.): *Journal on Data Semantics IX*, Springer-Verlag, Berlin - Heidelberg, 2007
- [Goodchild 2007] GOODCHILD, Michael F.: Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. In: *International Journal of Spatial Data Infrastructures Research* 2 (2007), S. 24–32. – URL <http://ijmdir.jrc.ec.europa.eu/index.php/ijmdir/article/view/28/22>
- [Granell et al. 2007] GRANELL, Carlos ; DÍAZ, Laura ; GOULD, Michael: Managing Earth Observation Data with Distributed Geoprocessing Services. In: *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, 2007, S. 4777–4780
- [GSDI 2004] NEBERT, Douglas D. (Hrsg.): *Developing Spatial Data Infrastructures: The SDI Cookbook*. Online. Januar 2004. – URL <http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf>. – Version 2.0
- [v. Gösseln und Sester 2004] GÖSSELN, G. v. ; SESTER, M.: Intergration of geoscientific data sets and the German digital map using a matching approach. In: *International Archives of Photogrammetry and Remote Sensing, Vol. 35, ISPRS, Istanbul*, 2004, S. 1249–1254
- [Hakimpour und Timpf 2001] HAKIMPOUR, Farshad ; TIMPF, Sabine: Using Ontologies for Resolution of Semantic Heterogeneity in GIS. In: *Proceedings 4th AGILE Conference on Geographic Information Science, Brno*, 2001, S. 385–395
- [Haklay und Weber 2008] HAKLAY, Mordechai (. ; WEBER, Patrick: OpenStreet-Map: User-Generated Street Maps. In: *IEEE Pervasive Computing* 7 (2008), Nr. 4, S. 12–18. – URL <http://dx.doi.org/10.1109/MPRV.2008.80>

- [He 2003] HE, Hao: *What is Service-Oriented Architecture*. Online. September 2003. – URL <http://www.xml.com/pub/a/ws/2003/09/30/soa.html>
- [Heutschi 2007] HEUTSCHI, Roger: *Serviceorientierte Architektur: Architekturprinzipien und Umsetzung in die Praxis*. Springer-Verlag, Berlin - Heidelberg, 2007
- [Köhler 2004] KÖHLER, Petra: Webbasierte Bereitstellung von Geodaten und Metadaten – Erfahrungen aus der Katastrophenforschung. In: *Zeitschrift für Agrarinformatik* 12 (2004), Nr. 4, S. 76–81. – URL http://www.gil.de/publications/zai/archiv/R8_20040016.pdf
- [Kiehle et al. 2007] KIEHLE, Christian ; GREVE, Klaus ; HEIER, Christian: Requirements for Next Generation Spatial Data Infrastructures-Standardized Web Based Geoprocessing and Web Service Orchestration. In: *Transactions in GIS* 11 (2007), Dezember, Nr. 6, S. 819–834. – URL <http://dx.doi.org/10.1111/j.1467-9671.2007.01076.x>
- [Kim et al. 2005] KIM, Minsoo ; KIM, Mijeong ; LEE, Eunkyuu ; JOO, Inhak: Web Services Framework for Geo-spatial Services. In: *Web and Wireless Geographical Information Systems*. Springer-Verlag, Berlin - Heidelberg, 2005, S. 1–13. – URL http://dx.doi.org/10.1007/11427865_1
- [Kreger 2001] KREGER, Heather ; GROUP, IBM S. (Hrsg.): *Web Services Conceptual Architecture (WSCA 1.0)*. Online. Mai 2001. – URL <http://www.cs.uoi.gr/~zarras/mdw-ws/WebServicesConceptualArchitectu2.pdf>
- [Kreger 2003] KREGER, Heather: Fulfilling the Web services promise. In: *Communications of the ACM* 46 (2003), Juni, Nr. 6, S. 29–34. – URL <http://doi.acm.org/10.1145/777313.777334>
- [Kresse und Fadaie 2004] KRESSE, Wolfgang ; FADAIE, Kian: *ISO standards for geographic information*. Springer-Verlag, Berlin - Heidelberg, 2004
- [Kuhn 2003] KUHN, Werner: Semantic reference systems. In: *International Journal of Geographical Information Science* 17 (2003), Nr. 5, S. 405–409. – URL <http://dx.doi.org/10.1080/1365881031000114116>
- [Lanig et al. 2008] LANIG, Sandra ; SCHILLING, Arne ; STOLLBERG, Beate ; ZIPF, Alexander: Towards Standards-Based Processing of Digital Elevation Models for Grid Computing through Web Processing Service (WPS). In: *Computational Science and Its Applications – ICCSA 2008*, 2008, S. 191–203

- [Lemmens et al. 2007] LEMMENS, Rob ; WYTZISK, Andreas ; DE BY, Rolf ; GRANELL, Carlos ; GOULD, Michael ; VAN OOSTEROM, Peter: Enhancing Geo-Service Chaining through Deep Service Descriptions. In: *Transactions in GIS* 11 (2007), Nr. 6, S. 849–871. – URL <http://dx.doi.org/10.1111/j.1467-9671.2007.01079.x>
- [Lenk 2005] LENK, Martin: IMAGI – das Gremium des Bundes zur Koordinierung des Geoinformationswesens in der Bundesverwaltung. In: *DVW Bayern e.V.* (2005), April, Nr. 4. – URL http://www.gdi-de.org/de/download/DVW_BY4-05_Lenk.pdf
- [Longhorn und Blakemore 2008] LONGHORN, Roger A. ; BLAKEMORE, Michael: *Geographic Information: Value, Pricing, Production, and Consumption*. CRC Press, Taylor & Francis Group, 2008
- [Longley et al. 2005] LONGLEY, Paul A. ; GOODCHILD, Michael F. ; MAGUIRE, David J. ; RHIND, David W.: *Geographic Information Systems and Science*. Second Edition. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, 2005
- [Lüscher 2006] LÜSCHER, Patrick: *Matching von Strassendaten stark unterschiedlicher Massstäbe und Aufbau einer Multirepräsentationsdatenbank*, Universität Zürich, Diplomarbeit, Januar 2006
- [Lupien und Moreland 1987] LUPIEN, Anthony E. ; MORELAND, William H.: A general approach to map conflation. In: *Proceedings of AUTOCARTO 8*, 1987, S. 630–639
- [Lutz 2005] LUTZ, Michael: *Ontology-based Discovery and Composition of Geographic Information Services*, Westfälische Wilhelms-Universität Münster, Dissertation, 2005
- [Mac Gillavry 2003] MAC GILLAVRY, Edward: *Collaborative mapping: grassroots LBS*. 39th Society of Cartographers Summer School, Reading University. September 2003. – URL <http://www.webmapper.net/carto2003>
- [McGovern et al. 2003] MCGOVERN, James ; TYAGI, Sameer ; STEVENS, Michael ; MATHEW, Sunil: *Java Web Services Architecture: Chapter 2 – Service-Oriented Architecture*. Morgan Kaufmann, San Francisco, 2003
- [McKee 2004] MCKEE, Lance: *The Spatial Web*. White paper, Open GIS Consortium. Juni 2004. – URL http://portal.opengeospatial.org/files/?artifact_id=3859&version=2&format=pdf

- [McMaster 1986] MCMMASTER, Robert B.: A Statistical Analysis of Mathematical Measures for Linear Simplification. In: *The American Cartographer* 13 (1986), Nr. 2, S. 103–116
- [Medeiros et al. 1996] MEDEIROS, C. B. ; BELLOSTA, M.J. ; JOMIER, G.: Managing multiple representations of georeferenced elements. In: *Proceedings of 7th International Workshop on Database and Expert Systems Applications*, 1996, S. 364–370
- [Michaelis und Ames 2009] MICHAELIS, Christopher D. ; AMES, Daniel P.: Evaluation and Implementation of the OGC Web Processing Service for Use in Client-Side GIS. In: *Geoinformatica* 13 (2009), März, Nr. 1, S. 109–120. – URL <http://dx.doi.org/10.1007/s10707-008-0048-1>
- [Müller und Portele 2005] MÜLLER, Markus ; PORTELE, Clemens: *GDI-Architekturmodelle*. S. 83–92. In: BERNARD, Lars (Hrsg.) ; FITZKE, Jens (Hrsg.) ; WAGNER, Roland M. (Hrsg.): *Geodateninfrastruktur: Grundlagen und Anwendung*, Herbert Wichmann Verlag, Heidelberg, 2005
- [Müller 2009] MÜLLER, Matthias: *Interoperable Geoinformationsdienste für die multikriterielle Entscheidungsunterstützung im Katastrophen- und Risikomanagement*, Technische Universität Dresden, Diplomarbeit, Januar 2009
- [Najar 2006] NAJAR, Christine R.: *A Model-Driven Approach to Management of Integrated Metadata - Spatial Data in the Context of Spatial Data Infrastructures*, ETH Zürich, Dissertation, 2006
- [Nash 2008] NASH, Edward: *WPS Application Profiles for Generic and Specialised Processes*. GI-Days, Münster. Juni 2008
- [Nogueras-Iso et al. 2005] NOGUERAS-ISO, Javier ; ZARAZAGA-SORIA, F. J. ; MURO-MEDRANO, Pedro R.: *Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and Information Retrieval*. Springer-Verlag, Berlin - Heidelberg, 2005
- [OASIS 2006] ORGANIZATION FOR THE ADVANCEMENT OF STRUCTURED INFORMATION STANDARDS (Hrsg.): *Reference Model for Service Oriented Architecture 1.0*. 2006. – URL <http://docs.oasis-open.org/soa-rm/v1.0/soa-rm.pdf>
- [OGC 1998] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *The OpenGIS Guide*. Juni 1998. – Third Edition

- [OGC 2002] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *The OpenGIS Abstract Specification, Topic 12: OpenGIS Service Architecture*. Januar 2002. – Version 4.3
- [OGC 2007a] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *OGC Web Services Common Specification*. Februar 2007. – Version 1.1.0
- [OGC 2007b] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *OpenGIS Web Processing Service*. Juni 2007. – Version 1.0.0
- [OGC 2007c] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *OWS-4 WPS IPR: Discussions, findings, and use of WPS in OWS-4*. Mai 2007. – URL http://portal.opengeospatial.org/files/?artifact_id=19424. – Version 0.9.1
- [OGC 2007d] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *Web Coordinate Transformation Service (WCTS) Interface Engineering Report*. Discussion Paper. Oktober 2007. – Version 0.4.0
- [OGC 2009a] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *The OpenGIS Abstract Specification, Topic 5: Features*. Januar 2009. – Version 5.0
- [OGC 2009b] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *Request for Information (RFI) for OGC Fusion Standards Study*. Juli 2009
- [OGC 2009c] OPEN GEOSPATIAL CONSORTIUM INC. (Hrsg.): *Web Coverage Processing Service (WCPS) Language Interface Standard*. März 2009. – Version 1.0.0
- [Papazoglou und van den Heuvel 2007] PAPAZOGLU, Mike P. ; VAN DEN HEUVEL, Willem-Jan: Service oriented architectures: approaches, technologies and research issues. In: *The VLDB Journal* 16 (2007), Juli, Nr. 3, S. 389–415. – URL <http://dx.doi.org/10.1007/s00778-007-0044-3>
- [Peng und Tsou 2003] PENG, Zhong-Ren ; TSOU, Ming-Hsiang: *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. John Wiley & Sons Ltd, Inc., Hoboken, New Jersey, 2003
- [Pundt 2005] PUNDT, Hardy: *Evaluating the Relevance of Spatial Data in Time Critical Situations*. S. 779–788. In: VAN OOSTEROM, Peter (Hrsg.) ; ZLATANOVA, Siyka (Hrsg.) ; FENDEL, Elfriede M. (Hrsg.): *Geo-information for Disaster Management*, Springer-Verlag, Berlin, 2005

- [Rajabifard et al. 2002] RAJABIFARD, Abbas ; FEENEY, Mary-Ellen F. ; WILLIAMSON, Ian P.: Future directions for SDI development. In: *International Journal of Applied Earth Observation and Geoinformation* 4 (2002), Nr. 1, S. 11–22. – URL [http://dx.doi.org/10.1016/S0303-2434\(02\)00002-8](http://dx.doi.org/10.1016/S0303-2434(02)00002-8)
- [Reindl 2005] REINDL, Wolfgang: Geodateninfrastruktur in Deutschland – Die Geschäfts- und Koordinierungsstelle GDI-DE in Frankfurt a. Main. In: *DVW Bayern e.V.* (2005), April, Nr. 4. – URL http://www.gdi-de.org/de/download/DVW_BY4-05_Reindl.pdf
- [Richter et al. 2005] RICHTER, Jan-Peter ; HALLER, Harald ; SCHREY, Peter: Serviceorientierte Architektur. In: *Informatik-Spektrum* 28 (2005), Oktober, Nr. 5, S. 413–416. – URL <http://dx.doi.org/10.1007/s00287-005-0021-4>
- [Riecken 2005] RIECKEN, Jens: *Geodateninfrastruktur Nordrhein-Westfalen*. S. 64–70. In: BERNARD, Lars (Hrsg.) ; FITZKE, Jens (Hrsg.) ; WAGNER, Roland M. (Hrsg.): *Geodateninfrastruktur: Grundlagen und Anwendung*, Herbert Wichmann Verlag, Heidelberg, 2005
- [Rosen und Saalfeld 1985] ROSEN, Barbara ; SAALFELD, Alan: Match criteria for automatic alignment. In: *Proceedings of AUTOCARTO 7*, 1985, S. 456–462
- [Saalfeld 1988] SAALFELD, Alan: Conflation: Automated map compilation. In: *International Journal of Geographical Information Science* 2 (1988), Nr. 3, S. 217–228
- [Samal et al. 2004] SAMAL, Ashok ; SETH, Sharad ; CUETO, Kevin: A feature-based approach to conflation of geospatial sources. In: *International Journal of Geographical Information Science* 18 (2004), Nr. 5, S. 459–489. – URL <http://dx.doi.org/10.1080/13658810410001658076>
- [Sancho-Jiménez et al. 2008] SANCHO-JIMÉNEZ, Gonzalo ; BÉJAR, Rubén ; LATRE, M.A. ; MURO-MEDRANO, Pedro R.: A Method to Derivate SOAP Interfaces and WSDL Metadata from the OGC Web Processing Service Mandatory Interfaces. In: *Advances in Conceptual Modeling – Challenges and Opportunities*. Springer-Verlag, Berlin - Heidelberg, 2008, S. 375–384. – URL http://dx.doi.org/10.1007/978-3-540-87991-6_44
- [Schaeffer 2008] SCHAEFFER, Bastian: *Towards a Transactional Web Processing Service (WPS-T)*. GI-Days, Münster. Juni 2008

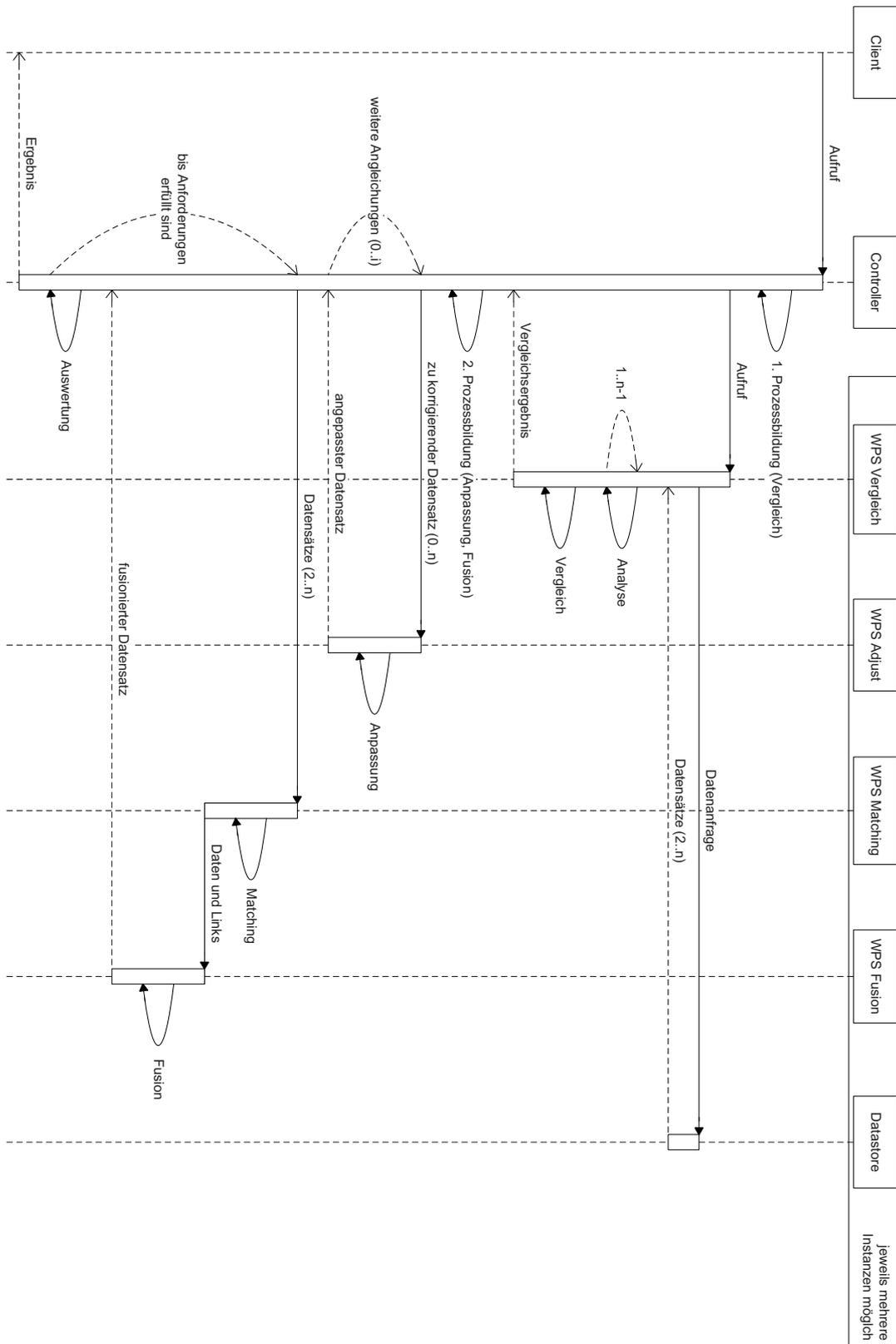
- [Schaeffer 2009] SCHAEFFER, Bastian: *Standardized Geoprocessing with 52°North Open Source Software: Web Processing Service Tutorial*. Januar 2009. – URL https://52north.org/twiki/bin/viewfile/Processing/52nWebProcessingService?rev=1;filename=WPS_Tutorial.pdf
- [Scholten et al. 2006] SCHOLTEN, Marius ; KLAMMA, Ralf ; KIEHLE, Christian: Evaluating Performance in Spatial Data Infrastructures for Geoprocessing. In: *IEEE Internet Computing* 10 (2006), Nr. 5, S. 34–41. – URL <http://doi.ieeecomputersociety.org/10.1109/MIC.2006.97>
- [Schwinn und Winter 2005] SCHWINN, Alexander ; WINTER, Robert: Entwicklung von Zielen und Messgrößen zur Steuerung der Applikationsintegration. In: *7. Internationale Tagung Wirtschaftsinformatik (WI2005), Bamberg, 2005*, S. 587–606
- [Stollberg et al. 2007] STOLLBERG, Beate ; LUTZ, Michael ; OSTLÄNDER, Nicole ; BERNARD, Lars: Geoprozessierung in Geodateninfrastrukturen: Aufgaben für die nächste Generation. In: *GIS – Zeitschrift für Geoinformatik* (2007), April, S. 24–29
- [Stollberg und Zipf 2007] STOLLBERG, Beate ; ZIPF, Alexander: OGC Web Processing Service Interface for Web Service Orchestration: Aggregating Geoprocessing Services in a Bomb Threat Scenario. In: *Web and Wireless Geographical Information Systems*. Springer-Verlag, Berlin - Heidelberg, 2007, S. 239–251
- [Thom 2005] THOM, Stuart: *A Strategy for Collapsing OS Integrated Transport Network dual carriageways*. 8th ICA WORKSHOP on Generalisation and Multiple Representation, A Coruña. Juli 2005
- [Tidwell 2000] TIDWELL, Doug: *Web Services: The Web's next Revolution*. IBM DeveloperWorks. 2000
- [Uitermark et al. 1999] UITERMARK, Harry ; VOGELS, Anton ; VAN OOSTEROM, Peter: *Semantic and Geometric Aspects of Integrating Road Networks*. S. 177–188. In: VČKOVSKI, Andrej (Hrsg.) ; BRASSEL, Kurt E. (Hrsg.) ; SCHEK, Hans-Jörg (Hrsg.): *Interoperating Geographic Information Systems*, Springer-Verlag, Berlin - Heidelberg, 1999
- [van der Aalst et al. 2005] VAN DER AALST, W.M.P. ; DUMAS, M. ; TER HOFSTEDÉ, A.H.M. ; RUSSELL, N. ; VERBEEK, H.M.W. ; WOHED, P.: *Life After BPEL?* S. 35–

50. In: *Formal Techniques for Computer Systems and Business Processes*, 2005
- [van Oosterom et al. 2006] VAN OOSTEROM, Peter ; DE VRIES, Marian ; MEIJERS, Martijn: *Vario-scale data server in a web service context*. Workshop of the ICA Commission on Map Generalisation and Multiple Representation. Juni 2006
- [Veltkamp und Hagedoorn 2001] VELTKAMP, Remco C. ; HAGEDOORN, Michiel: *State of the Art in Shape Matching*. S. 87–119. In: LEW, Michael S. (Hrsg.): *Principles of visual information retrieval*, Springer-Verlag, London, UK, 2001
- [Volz 2005] VOLZ, Steffen: *Data-Driven Matching of Geospatial Schemas*. S. 115–132. In: COHN, Anthony G. (Hrsg.) ; MARK, David M. (Hrsg.): *Spatial Information Theory*, Springer-Verlag, Berlin - Heidelberg, 2005
- [Volz 2006] VOLZ, Steffen: *An iterative approach for matching multiple representations of street data*. ISPRS Workshop: Multiple representation and interoperability of spatial data, Hannover. Februar 2006
- [Volz und Walter 2004] VOLZ, Steffen ; WALTER, Volker: Linking different geospatial databases by explicit relations. In: *Proceedings of the XXth ISPRS Congress, Comm. IV, Istanbul, 2004*, S. 152–157
- [Walter 1996] WALTER, Volker: *Zuordnung von raumbezogenen Daten – am Beispiel der Datenmodelle ATKIS und GDF*, Universität Stuttgart, Dissertation, 1996
- [Walter und Fritsch 1999] WALTER, Volker ; FRITSCH, Dieter: Matching spatial data sets: a statistical approach. In: *International Journal of Geographical Information Science* 13 (1999), Nr. 5, S. 445–473
- [Weis und Naumann 2004] WEIS, Melanie ; NAUMANN, Felix: Detecting Duplicate Objects in XML Documents. In: *Proceedings of the 2004 international workshop on Information quality in information systems, Paris, 2004*, S. 10–19
- [Weiser et al. 2006] WEISER, A. ; NEIS, P. ; ZIPF, A.: Orchestrierung von OGC Web Diensten im Katastrophenmanagement – am Beispiel eines Emergency Route Service auf Basis der OpenLS Spezifikation. In: *GIS – Zeitschrift für Geoinformatik* (2006), September, Nr. 4, S. 35–41

- [Werder 2009] WERDER, Stefan: *Formalization of Spatial Constraints*. 12th AGILE International Conference on Geographic Information Science, Hannover. Juni 2009
- [Wilkes und Veryard 2004] WILKES, Lawrence ; VERYARD, Richard: Service-Oriented Architecture: Considerations for Agile Systems. In: *Microsoft Architects Journal* (2004), April, Nr. 2, S. 11–23
- [Williamson 2004] WILLIAMSON, Ian P.: *Building SDIs – the challenges ahead*. Proceedings of the 7th International Conference: Global Spatial Data Infrastructure, Bangalore, India. Februar 2004
- [Worboys und Duckham 2004] WORBOYS, Michael (Hrsg.) ; DUCKHAM, Matt (Hrsg.): *GIS: A Computing Perspective*. Second Edition. CRC Press, 2004
- [Wytzisk und Sliwinski 2004] WYTZISK, Andreas ; SLIWINSKI, Adam: Quo Vadis SDI? In: *7th AGILE Conference on Geographic Information Science, Heraklion, Greece*, 2004, S. 43–49
- [Xiaomeng et al. 2007] XIAOMENG, Liu ; YANHUI, Wang ; HAO, Meng: *Linking Multi-Scale Representations in a Navigable Database*. ISPRS Workshop on Updating Geo-spatial Databases with Imagery & The 5th ISPRS Workshop on DMGISs. 2007
- [Xiong 2000] XIONG, Demin: *A three stage computational approach to network matching*. S. 71–89. In: THILL, Jean-Claude (Hrsg.): *Geographic Information Systems in Transportation Research*, Emerald Group Publishing, 2000
- [Yuan und Tao 1999] YUAN, Shuxin ; TAO, Dr. C.: Development of Conflation Components. In: *Proceedings of Geoinformatics, Ann Arbor*, 1999, S. 1–13
- [Zhang und Meng 2007] ZHANG, Meng ; MENG, Liqiu: An iterative road-matching approach for the integration of postal data. In: *Computers, Environment and Urban Systems* 31 (2007), September, Nr. 5, S. 597–615. – URL <http://dx.doi.org/10.1016/j.compenvurbsys.2007.08.008>
- [Zhang und Meng 2008] ZHANG, Meng ; MENG, Liqiu: Delimited Stroke Oriented Algorithm-Working Principle and Implementation for the Matching of Road Networks. In: *Journal of Geographic Information Sciences* 14 (2008), Juni, Nr. 1, S. 44–53

-
- [Zhang et al. 2005] ZHANG, Meng ; SHI, Wei ; MENG, Liqiu: *A Generic Matching Algorithm for Line Networks of Different Resolutions*. 8th ICA Workshop on Generalisation and Multiple Representation, A Coruña. Juli 2005
- [Zhang et al. 2006] ZHANG, Meng ; SHI, Wei ; MENG, Liqiu: *A matching approach focused on parallel roads and looping crosses in digital maps*. International Symposium of Theoretical Cartography and Geo-Information Science, Wuhan, China. Oktober 2006

ANHANG A: UML-SEQUENZDIAGRAMM



ANHANG B: CD MIT QUELLCODE

Die beiliegende CD enthält den gesamten Quellcode und verwendete Beispieldatensätze der Implementierung sowie eine digitale Fassung dieser Arbeit.

ERKLÄRUNG

Ich erkläre, dass ich die vorliegende Arbeit mit dem Titel "Fusion von Geodaten unterschiedlicher Quellen in Geodateninfrastrukturen am Beispiel von ATKIS und OpenStreetMap" selbständig, unter Angabe aller Zitate und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Dresden, 05.11.2009