
Geschäftsstelle Geodateninfrastruktur Bayern, Landesamt für Vermessung und Geoinformation

SEMANTISCHE HETEROGENITÄT UND SEMANTISCHE TRANSFORMATION IM KONTEXT VON INSPIRE

Astrid Feichtner

Zusammenfassung: Im Zuge des Aufbaus der Europäischen Geodateninfrastruktur (INSPIRE) werden europaweit gültige Datenspezifikationen definiert, um die Kombination heterogener Daten zu erleichtern. Datenanbieter stehen vor der Herausforderung, Daten aus ihren originären Schemata über semantische Transformationen in die INSPIRE-Schemata zu überführen. Für diese komplexe Aufgabe gibt es eine Reihe von unterschiedlichen Lösungsansätzen. Der vorliegende Beitrag beleuchtet die grundlegenden Konzepte der semantischen Heterogenität und semantischen Transformation. Zwei Transformationsansätze, bei denen die Schemaabbildung auf unterschiedlichen Ebenen (konzeptuelle Schemata vs. Transferformatschemata) definiert wird, werden vorgestellt, verglichen sowie deren Möglichkeiten und Grenzen eingeschätzt.

Schlüsselwörter: Semantische Heterogenität, semantische Transformation, Schemaabbildung, INSPIRE

SEMANTIC HETEROGENEITY AND SEMANTIC TRANSFORMATION IN THE CONTEXT OF INSPIRE

Abstract: With the establishment of the European Spatial Data Infrastructure (INSPIRE), community-wide data specifications are being defined in order to facilitate the combination of heterogeneous data. Data providers face the challenge of transforming data from their legacy schemas into the INSPIRE schemas using semantic transformations. There are a number of different approaches for this complex task. This paper examines the basic concepts of semantic heterogeneity and semantic transformation. Two transformation approaches, in which the schema mappings are defined at different levels (conceptual schemas vs. transfer format schemas), are presented and compared, and their possibilities and limitations are assessed.

Keywords: Semantic heterogeneity, semantic transformation, schema mapping, INSPIRE

Autorin

Dr. Astrid Feichtner
Geschäftsstelle Geodateninfrastruktur Bayern
beim Landesamt für Vermessung und Geoinformation
Alexandrastr. 4
D-80538 München
E: astrid.feichtner@lvg.bayern.de

1 EINFÜHRUNG

Viele räumliche Fragestellungen – beispielsweise im Umweltbereich – können erst durch die Kombination von verteilten, heterogenen Geodaten über die Grenzen von Geoinformationssystemen und Ländern hinweg effizient bearbeitet werden. Auf europäischer Ebene soll die gemeinsame Nutzung von Daten durch den Aufbau einer Geodateninfrastruktur erleichtert werden. Die zu diesem Zweck erlassene INSPIRE-Richtlinie (Infrastructure for Spatial Information in the European Community) verpflichtet betroffene Datenanbieter, Daten aus ihren originären Schemata (Quellschemata) über semantische Transformationen in europaweit einheitliche INSPIRE-Schemata (Zielschemata) zu überführen.

Semantische Transformation ist Gegenstand einer Reihe von Forschungsvorhaben im Geoinformationsbereich. Dazu gehören die beiden Projekte „Modellbasierter Ansatz für den Web-Zugriff auf verteilte Geodaten am Beispiel grenzübergreifender GIS-Anwendungen (mdWFS)“ (Donaubauer et al. 2008) sowie „HUMBOLDT – Development of a Framework for Data Harmonisation and Service Integration“ (Fichtinger et al. 2011). Das mdWFS-Projekt wird seit 2006 im Auftrag des Bundesamtes für Kartographie und Geodäsie und des Bundesamtes für Landestopografie swisstopo von der Technischen Universität München (TUM) in Kooperation mit der Eidgenössischen Technischen Hochschule Zürich (ETHZ) durchgeführt. Das HUMBOLDT-Projekt wurde von 2006 bis 2011 durch 26 Partner aus 13 Ländern (darunter die TUM) realisiert und von der Europäischen Kommission im 6. Forschungsrahmenprogramm gefördert. In den Projekten wurde die semantische Transformation am Beispiel topographischer Geobasisdaten aus der grenzüberschreitenden Bodenseeregion (Deutschland – Schweiz – Österreich) erprobt. Als Zielschema dienten dabei Schemata aus den INSPIRE-Datenspezifikationen (z. B. für Gewässernetz).

Dieser Beitrag basiert auf der Dissertation der Autorin (Fichtinger 2011), in die Erkenntnisse aus den oben genannten Forschungsprojekten eingeflossen sind. Zunächst wird in das Konzept der semantischen Heterogenität eingeführt und eine entsprechende Klassifikation vorgestellt. Anschließend werden Konzepte der semantischen Transformation erläutert.

Schließlich werden die Transformationsansätze aus den beiden Projekten dargestellt, verglichen sowie deren Möglichkeiten und Grenzen eingeschätzt.

2 SEMANTISCHE HETEROGENITÄT

Der erste Schritt bei der Definition einer Abbildung zwischen Quell- und Zielschema ist in der Regel die Analyse der semantischen Heterogenität der Schemata. Es gibt eine Vielzahl von Ansätzen zur Beschreibung und Klassifikation semantischer Heterogenität, beispielsweise aus dem Bereich der Datenbankintegration in der allgemeinen IT sowie aus den Ontologie- und Geoinformationsbereichen. In stark theoretischen Ansätzen werden semantische Relationen, Heterogenitäten und Abbildungen in Form von komplexen, streng logischen bzw. al-

gebraischen Ausdrücken formalisiert. Hierbei werden teilweise recht einfache bzw. für die Bearbeitung vereinfachte Schemata verwendet. In eher pragmatischen Ansätzen werden Heterogenitäten zwischen teilweise umfangreichen und komplexen Datenmodellen detailliert analysiert und Korrespondenzen definiert, die Heterogenitäten aber oft nicht systematisch beschrieben.

Die hier vorgeschlagene Klassifikation (siehe Abbildung 1, Tabellen 1 und 2) wurde unter Einbeziehung bestehender Ansätze (Annoni et al. 2008; Batini et al. 1986; Kashyap & Sheth 1996; Nyerges 1989; Reitz 2010a; Scharffe 2009; Stoter et al. 2009) sowie der empirischen Erfahrung aus den eingangs beschriebenen Projekten entworfen. Auf der ersten Ebene werden drei grundlegende Fälle unterschieden:

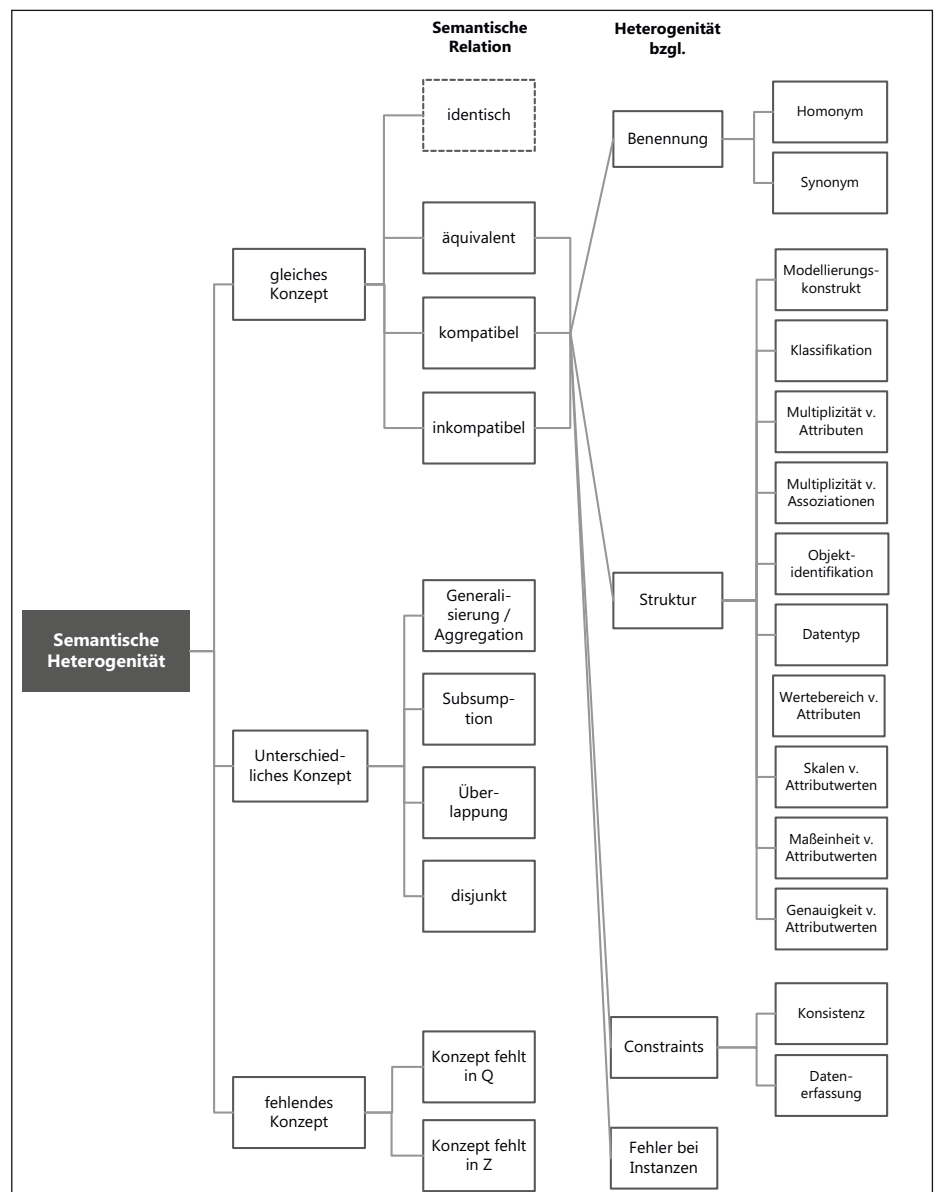


Abbildung 1: Klassifikation der semantischen Relationen und Heterogenitäten

1. dem Quellschema (Q) und dem Zielschema (Z) liegen gleiche Konzepte zugrunde, die aber auf unterschiedliche Weise modelliert wurden,
2. Q und Z liegen unterschiedliche Konzepte zugrunde,
3. ein in Q modelliertes Konzept fehlt in Z oder umgekehrt.

Auf den folgenden Ebenen werden zunächst die möglichen semantischen Relationen zwischen den Konzepten und anschließend die Heterogenitäten zwischen den Elementen (E) in Q und Z, die die Konzepte abbilden, klassifiziert. Die in der Klassifikation verwendeten Begriffe werden in den Tabellen 1 und 2 definiert. Detailliertere Beschreibungen sowie Beispiele hierzu können Fichtinger (2011) entnommen werden.

3 SEMANTISCHE TRANSFORMATION

3.1 BEGRIFFSBESTIMMUNG

Semantische Transformation wird hier im Kontext von Geodaten als Prozess mit zwei Phasen verstanden, in Anlehnung an Lehto (2007a), Eisenhut & Kutzner (2010) sowie Legler & Naumann (2007). In der *Definitionsphase* wird eine Abbildung zwischen einem Quell- und einem Zielschema (= Schemaabbildung, engl. „schema mapping“) definiert, indem Korrespondenzen zwischen Elementen im Quell- und Zielschema ggf. unter Verwendung von Transformationsfunktionen hergestellt und in Form von Abbildungsregeln beschrieben werden. In der anschließenden *Ausführungsphase* werden die Geodaten unter Verwendung der in der ersten Phase erstellten Abbildungsregeln aus dem Quell- in das Zielschema transformiert (= Datentransformation).

Während die Ausführungsphase immer auf der Ebene der Daten stattfindet, kann die Abbildung auf verschiedenen Ebenen definiert werden (siehe Abbildung 2), beispielsweise zwischen zwei Transferformaten wie Shapefiles. Dies wird als *formatbasierte* Abbildung bezeichnet und ist heute state-of-the-art in Software-Werkzeugen auf dem Markt. Die nächsthöhere Ebene ist die der Transferformatschemata (z. B. GML-Anwendungsschemata), zwischen denen eine *formatschemabasierte* Abbildung definiert werden kann. Auf der höchsten Ebene wird eine Abbildung zwischen konzeptuellen

	Art der semantischen Relation	Definition
Gleiches Konzept	Identisch	EQ und EZ sind exakt gleich, d. h. gleiche Konzepte, denen die gleichen Vorstellungen zugrunde liegen, wurden auf gleiche Art und Weise modelliert.
	Äquivalent	EQ und EZ sind nicht exakt gleich, d. h. gleiche Konzepte, denen die gleichen Vorstellungen zugrunde liegen, wurden auf unterschiedliche, aber äquivalente Art und Weise modelliert. Zwischen EQ und EZ kann eine 1:1 Korrespondenz hergestellt werden oder EZ kann aus EQ durch eine Äquivalenz erhaltende Transformation abgeleitet werden.
	Kompatibel	EQ und EZ sind weder identisch noch äquivalent, aber die für die gleichen Konzepte verwendeten Modellierungsstrukturen, Vorstellungen sowie Bedingungen und Einschränkungen sind kompatibel und widersprechen sich nicht.
	Inkompatibel	EQ und EZ sind inkompatibel, obwohl sie das gleiche Konzept abbilden.
Unterschiedliches Konzept	Generalisierung/Aggregation	Ein Konzept wird in Q und Z als Objektklasse auf unterschiedlicher Generalisierungsebene abgebildet, d. h. EQ ist Subtyp/Supertyp von EZ oder umgekehrt. Ein Konzept wird in einem der beiden Schemata (Q oder Z) als Aggregation mehrerer Elemente abgebildet und im anderen nicht.
	Subsumption	Ein Konzept wird in einem Schema als Element abgebildet, das von einem Element im anderen Schema subsummiert wird.
	Überlappung	Ein Konzept wird in den Schemata durch sich überlappende Elemente abgebildet.
	Disjunkte Konzepte	Ein Konzept wird in den Schemata durch komplett disjunkte Elemente abgebildet.
Fehlendes Konzept		Ein bestimmtes Konzept fehlt entweder in Q oder Z ganz.

Tabelle 1: Definition der semantischen Relationen

Schemata (z. B. UML-Schemata) definiert, was als *schemabasiert* bezeichnet wird. Im vorliegenden Beitrag werden die beiden oberen Ebenen näher betrachtet (siehe Abschnitt 4).

Um die Transformation automatisiert ausführen zu können, müssen die Abbildungsregeln auf formale, maschineninterpretierbare Weise mithilfe einer Abbildungssprache beschrieben werden.

	Art der semantischen Heterogenität	Definition
Benennung	Homonym	Gleiche Begriffe werden für unterschiedliche Konzepte in Q und Z verwendet.
	Synonym	Unterschiedliche Begriffe werden für gleiche Konzepte in Q und Z verwendet.
Struktur	Modellierungskonstrukt	Gleiche Konzepte wurden in Q und Z unter Verwendung unterschiedlicher Typen von Modellierungskonstrukten abgebildet.
	Klassifikation	In Q und Z werden unterschiedliche Klassifikationssysteme bei der Einteilung in Objektklassen oder in Attributwerte verwendet.
	Multiplizität von Attributen	Attribute haben in Q und Z eine unterschiedliche Multiplizität (z. B. 0..1, 1, 1..*).
	Multiplizität von Assoziationen	Mehrere Objekte sind in Q über Assoziationen mit anderen Multiplizitäten verbunden als in Z.
	Objektidentifikation	Gleichen Konzepten in Q und Z sind unterschiedliche Arten von Schlüsseln zugeordnet.
	Datentyp von Attributen	Attribute in Q und Z haben einen unterschiedlichen Datentyp.
	Wertebereich von Attributwerten	Für Attribute in Q und Z ist ein unterschiedlicher Wertebereich definiert.
	Skalierung von Attributwerten	Attributwerte in Q und Z haben ein unterschiedliches Skalenniveau.
	Maßeinheit von Attributwerten	Attributwerte in Q und Z werden in unterschiedlichen Maßeinheiten angegeben.
	Genauigkeit von Attributwerten	Die Genauigkeit oder Auflösung von Attributwerten unterscheidet sich in Q und Z.
Constraint	Konsistenzbedingungen	Ein Attribut hat nur in einem von beiden Schemata (Q oder Z) Konsistenzbedingungen oder die Attribute in Q und Z haben unterschiedliche, nicht kompatible Konsistenzbedingungen.
	Datenerfassungskriterien	Die Bedingungen für die Erfassung und Unterteilung von Objekten der realen Welt unterscheiden sich in Q und Z.
Fehler bei Instanzen		Die Instanzen sind fehlerhaft z. B. durch verletzte Integritätsbedingungen, fehlende oder falsche Attributwerte.

Tabelle 2: Definition der semantischen Heterogenitäten

Im Gegensatz zu einer *syntaktischen Transformation*, bei der nur die Syntax der Geodaten umgebaut wird – d. h. nur eine Formatkonvertierung durchgeführt wird – werden die Daten bei einer *semantischen Transformation* so umstrukturiert, dass sie inhaltlich einem anderen Schema entsprechen (Eisenhut & Kutzner 2010). Die Transformation eines Shapefiles in eine GML-Datei ist ein Beispiel für eine syntaktische Transformation, wohingegen die Transformation von Instanzen einer Objektklasse „Fluss“ im Quellschema in Instanzen der Objektklasse „Watercourse“ im Zielschema beispielsweise einer semantischen Transformation entspricht.

Der Begriff der *semantischen Transformation* wird hier dem beispielsweise bei INSPIRE verwendeten Begriff der *Schema-transformation* (engl. *schema transformation*) vorgezogen, da gemäß der eingangs gegebenen Definition als Prozess mit zwei Phasen streng genommen nicht die Schemata, sondern die Daten transformiert werden, basierend auf einer semantischen Abbildung zwischen den Schemata.

3.2 ANFORDERUNGEN AN LÖSUNGEN ZUR SEMANTISCHEN TRANSFORMATION

Die Definition und die Ausführung semantischer Transformationen sind komplexe Aufgaben. Basierend auf den Erfahrungen aus den in Abschnitt 1 angesprochenen Projekten sowie weiteren, in der Literatur beschriebenen Ansätzen und Projekten (Annoni et al. 2008; Beare et al. 2010a; Eisenhut & Kutzner 2010; Fichtinger et al. 2010; Legler & Naumann 2007; Lehto 2007; Mens et al. 2005) können eine Reihe von Anforderungen an Lösungen zur semantischen Transformation im Kontext von Geodaten und Geodateninfrastrukturen ermittelt werden. Die Wichtigsten werden im Folgenden kurz zusammengefasst.

► *Expressivität*: Die Abbildungssprache muss die Beschreibung der Abbildungsregeln auf formale, eindeutige und maschineninterpretierbare Art und Weise ermöglichen. Die Abbildungssprache muss Konstrukte für alle benötigten Transformationsfunktionen bereitstellen und sollte komplexe, aus mehreren Schritten zusammengesetzte Transformationsfunktionen unterstützen. Abbildung 3 zeigt die Obergruppen der in den Transformationsfällen in den Projek-

ten HUMBOLDT und mdVFS benötigten Transformationsfunktionen.

- **Erweiterbarkeit:** Die Abbildungssprache sollte um nutzerdefinierte Funktionen erweitert werden können.
- **Benutzerfreundlichkeit:** Die Syntax der Abbildungssprache sollte weder zu komplex noch zu umfangreich sein, sodass Abbildungen übersichtlich und effizient ausgedrückt werden können. Idealerweise sollte die Abbildungssprache in einem Werkzeug mit einer grafischen Benutzeroberfläche, das den Anwender beim Erstellen der Abbildungsregeln unterstützen kann, verwendbar sein.
- **Systemunabhängigkeit/Offenheit:** Die Abbildungssprache sollte eine generische Beschreibung der Abbildungsregeln ermöglichen, aus der Codes für die Ausführung der Transformation in unterschiedlichen Systemen abgeleitet werden kann. Die Abbildungssprache sollte nicht proprietär sein und soweit wie möglich auf offenen, internationalen Standards und Normen basieren.
- **Automatisierbarkeit:** Die Ausführung der Transformation basierend auf den Abbildungsregeln sollte automatisierbar sein.
- **Skalierbarkeit:** Die performante Verarbeitung großer Datenmengen und komplexer Datenmodelle sollte möglich sein.
- **Qualitätsprüfung und Fehlerbehandlung:** Die Vollständigkeit und Richtigkeit der erstellten Abbildungsregeln sowie die Vollständigkeit und Richtigkeit des Transformationsergebnisses (d. h. der transformierten Daten) sollte überprüft werden können.

Gemäß der INSPIRE-Richtlinie können „Geodatensätze (...) entweder durch Anpassung der bestehenden Geodatensätze oder durch (...) Transformationsdienste verfügbar gemacht [werden]“ (Europäisches Parlament und Rat 2007). Grundlegende Anforderungen an INSPIRE-Transformationsdienste – beispielsweise hinsichtlich der zu unterstützenden Operationen sowie der Kapazität und Verfügbarkeit – werden in der Verordnung zu Download- und Transformationsdiensten (Europäische Kommission 2010) festgelegt. Das im Auftrag des Joint Research Centres der Europäischen Kommission durch ein Firmen-Konsortium erstellte „Technical Guidance“ Dokument

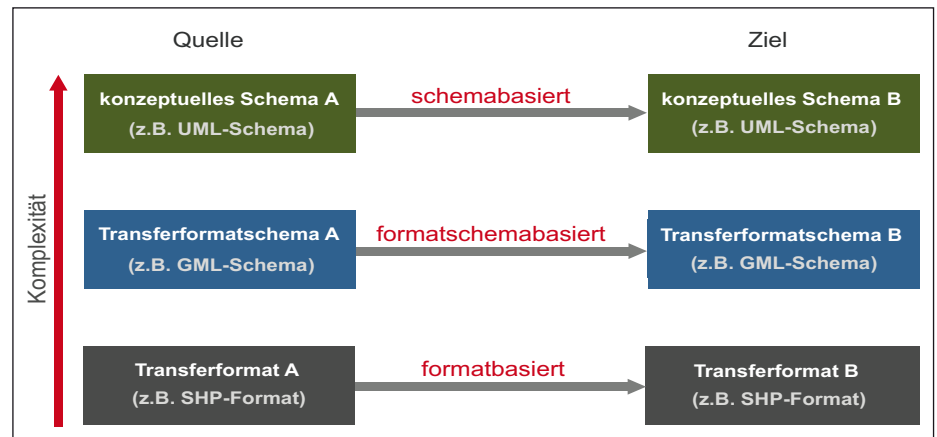


Abbildung 2: Ebenen semantischer Transformation

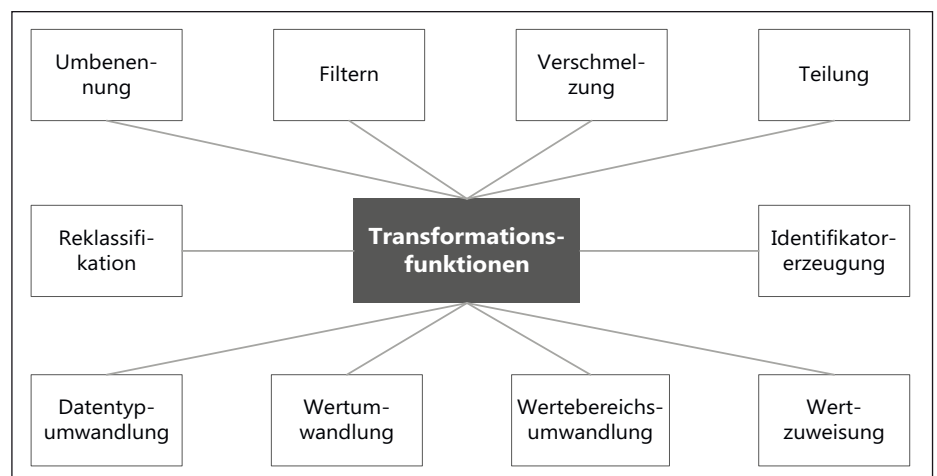


Abbildung 3: Transformationsfunktionen

für INSPIRE-Schematransformationsdienste (Howard et al. 2010) spezifiziert technische Details und empfiehlt die Verwendung eines Austauschformats für Abbildungsregeln (siehe Abschnitt 3).

3.3 TRANSFORMATIONSANSÄTZE

In der allgemeinen IT und der Geoinformatik existiert mittlerweile eine ganze Reihe von Ansätzen zur semantischen Transformation. Diese lassen sich anhand der in Abbildung 4 vorgestellten Unterscheidungsmerkmale charakterisieren und vergleichen. Im Folgenden wird näher auf die Transformationsansätze aus den in Abschnitt 1 eingeführten Projekten „mdVFS“ und „HUMBOLDT“ eingegangen.

Bei beiden Ansätzen findet eine horizontale Transformation zwischen Schemata auf derselben Abstraktionsebene statt. Im Projekt mdVFS wird die Abbildung zwischen konzeptuellen Schemata (UML-Schemata) definiert, während dies im Projekt HUMBOLDT auf der Ebene der Transferfor-

matschemata (GML-Anwendungsschemata) geschieht.

Die Abbildungsregeln werden in beiden Ansätzen auf formale, maschineninterpretierbare und systemunabhängige Art und Weise unter Verwendung einer nicht proprietären Abbildungssprache definiert. Während die im HUMBOLDT-Projekt verwendete geo Ontology Mapping Language (gOML) (Reitz et al. 2010) eine rein deklarative Sprache ist, verbindet die im mdVFS-Projekt entwickelte UML Transformations (UMLT) (Staub 2009) deklarative und operationale Ansätze. Bei deklarativen Sprachen liegt der Schwerpunkt auf dem „Was“, d. h. welche Elemente im Quellschema beispielsweise über Funktionen oder logische Ausdrücke auf welche Elemente im Zielschema abgebildet werden. Bei operationalen (auch „imperativ“ genannten) Sprachen hingegen liegt der Fokus auf dem „Wie“, d. h. der Art und Weise, wie die Transformation ausgeführt wird – beispielsweise als eine Abfolge von

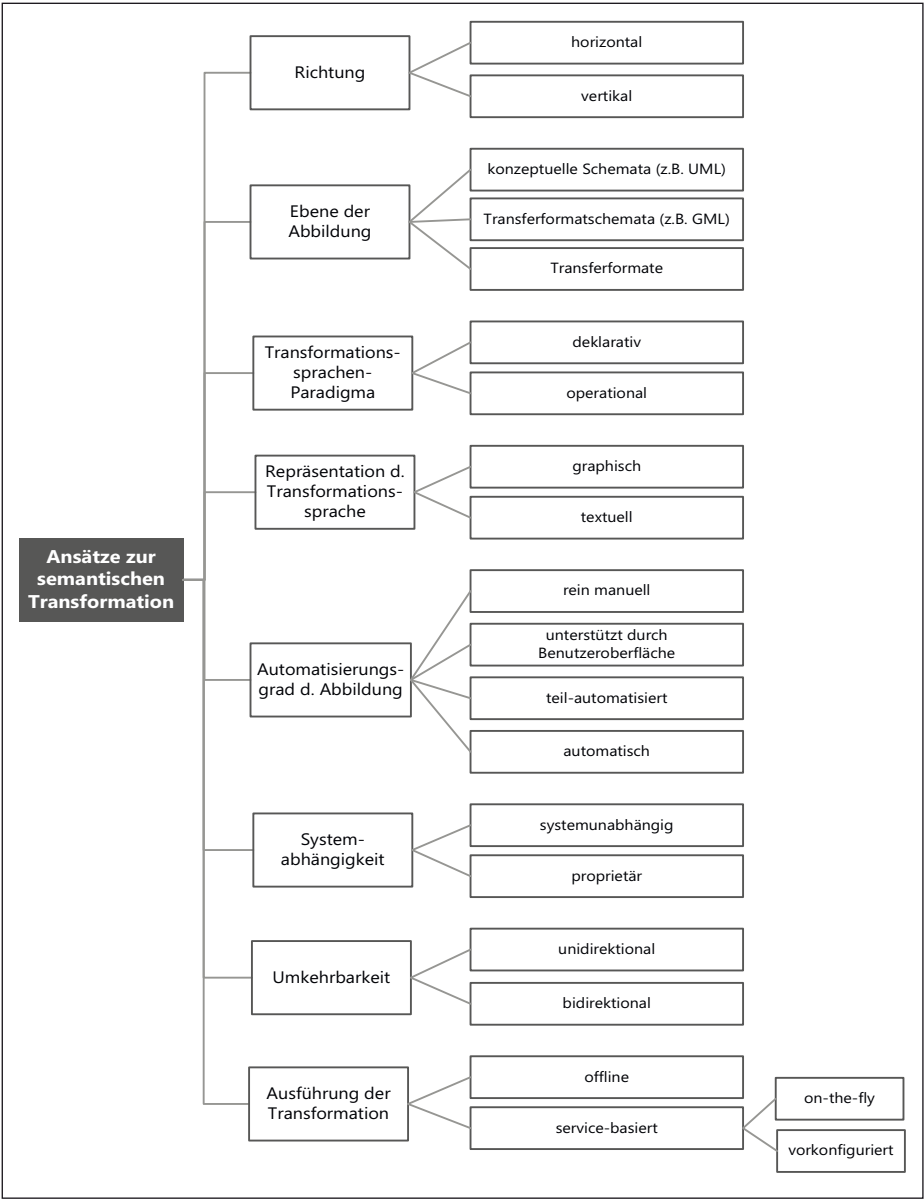


Abbildung 4: Klassifikation von Transformationsansätzen

Prozessschritten. Dies ermöglicht eine gezieltere Steuerung der Reihenfolge, in der die einzelnen Transformationsschritte durchgeführt werden sollen, bereits bei der Definition der Abbildungsregeln. Bei einer deklarativen Abbildungssprache muss diese Logik in dem Werkzeug implementiert werden, mit dem die Transformation ausgeführt wird. Dadurch, dass prozedurale Details bei deklarativen Sprachen nicht explizit formuliert sind, sind die Abbildungsregeln hier oft kompakter und leichter verständlich (Mens et al. 2005). Der deklarative Ansatz ähnelt darüber hinaus stärker dem in der Praxis oft angewendeten Verfahren der Abbildungsbeschreibung in Tabellen. Über eine graphische Repräsentation zusätzlich zur textuellen Notation verfügt nur die UMLT.

Da es bisher in der Geoinformationswelt keinen allgemein akzeptierten Standard für eine Abbildungssprache gibt, wird die Sprachenvielfalt in den verschiedenen Transformationsansätzen bis auf Weiteres bestehen bleiben. Um dennoch eine interoperable Nutzung von Abbildungsregeln mit standardisierten INSPIRE-Transformationsdiensten zu ermöglichen, empfiehlt das „Technical Guidance“ Dokument zu den INSPIRE-Transformationsdiensten die Verwendung des Rule Interchange Formats (RIF) für den systemunabhängigen Austausch von Abbildungsregeln. Der W3C Standard RIF verwendet XML-Syntax und umfasst sowohl deklarative als auch operationale Bestandteile (Beare et al. 2010a). In der Geoinformationswelt ist das RIF bisher nicht sehr weit verbreitet. Im Rahmen

der Erarbeitung des „Technical Guidance“ Dokuments wurde ein Plug-in für das aus dem HUMBOLDT-Projekt hervorgegangene Werkzeug HUMBOLDT Alignment Editor (HALE) prototypisch entwickelt, das den Export der Abbildungsregeln in das RIF grundsätzlich ermöglicht (Beare et al. 2010a). Für die Definition von Abbildungsregeln zwischen UML-Schemata ist RIF nicht geeignet (Eisenhut & Kutzner 2010).

Die Abbildungen werden in beiden Ansätzen durch den Nutzer manuell definiert, wobei dieser jeweils durch ein Werkzeug mit grafischer Benutzeroberfläche – dem UMLT-Editor im mdWFS-Projekt bzw. dem HALE – unterstützt wird. Basierend auf den erstellten Abbildungsregeln kann die Transformation in beiden Ansätzen automatisiert ausgeführt werden und zwar prinzipiell sowohl dienstebasiert (vorkonfiguriert) als auch offline. Bei der dienstebasierten Variante führt ein Transformationsdienst die Transformation immer dann durch, wenn ein Nutzer eine entsprechende Anfrage an den Dienst stellt, woraufhin der Dienst transformierte Daten zurückliefert (Eisenhut & Kutzner 2010). Die Transformation wiederum kann dabei on-the-fly (d. h. quasi in Echtzeit) oder vorkonfiguriert ablaufen. Bei einer On-the-fly-Transformation werden beliebige Abbildungsregeln als Teil der Anfrage an den Dienst zur Laufzeit übergeben, während bei der vorkonfigurierten Variante Abbildungsregeln für bestimmte Transformationsfälle beim Dienst gespeichert sind und in der Anfrage nur eine Referenz auf Abbildungsregeln für einen bestimmten Transformationsfall übergeben wird (Lehto 2007). Beare et al. (2010a) beschreiben drei verschiedene Architekturalternativen für INSPIRE-Transformationsdienste (s. Abbildung 5). Sowohl im mdWFS- als auch im HUMBOLDT-Projekt wurde die Variante c) (eigenständiger Transformationsdienst) gewählt. Dabei übergibt der Nutzer die Quelldaten (ggf. abgerufen von einem Downloaddienst), das Quell- und das Zielschema sowie die Abbildungsregeln als Parameter an den Transformationsdienst und erhält die transformierten Daten zurück.

In beiden Projekt-Ansätzen findet eine gerichtete Abbildung von den Quellschemata in das INSPIRE-Zielschema statt. Die Abbildung ist also unidirektional angelegt und auch nicht vollständig umkehrbar, da aufgrund der Heterogenität der Schemata

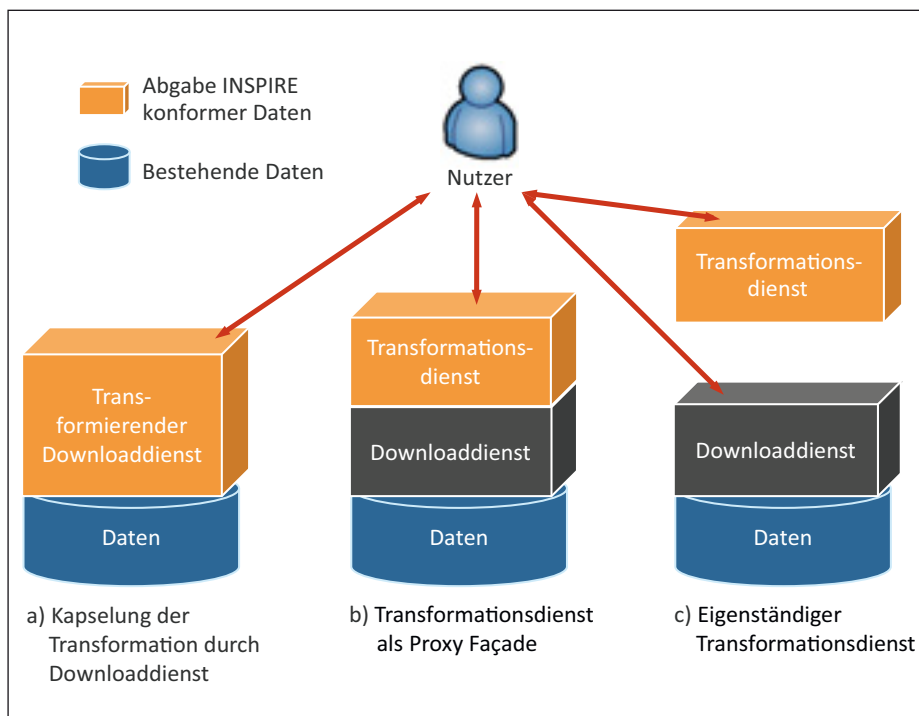


Abbildung 5: Architekturen für INSPIRE-Transformationsdienste (Fichtinger & Kutzner 2010)

keine vollständige Abbildung ohne Informationsverlust erreicht werden kann.

Die Heterogenitäten und Abbildungsregeln zwischen den Quell- und Zielschemata in den Transformationsfällen aus den beiden oben genannten Projekten sind in Fichtinger (2011) detailliert beschrieben. Bevor die Regeln unter Verwendung der beiden Abbildungssprachen UMLT und gOML formalisiert wurden, wurden jene zunächst in leicht lesbaren Abbildungstabellen abstrakt beschrieben. Hierbei wurden auch die Art der Heterogenität, die Art und Multiplizität der Korrespondenz zwischen Elementen der Schemata (z. B. 1:1-Abbildung zwischen Attributen) sowie die benötigten Transformationsfunktionen, ihre Bedingungen und Operatoren festgehalten.

4 SCHEMABASIERTE VS. FORMATSCHEMABASIERTE TRANSFORMATION – MÖGLICHKEITEN UND GRENZEN

Basierend auf den Erkenntnissen aus den Forschungsprojekten und dem Vergleich der beiden Transformationsansätze können die Möglichkeiten und Grenzen der semantischen Transformation auf Ebene der *konzeptuellen Schemata* (schemabasiert) auf der einen und auf Ebene der *Transferformatschemata* (formatschemabasiert) auf der anderen Seite eingeschätzt werden.

Die größten Vorteile der Abbildungsdefinition auf der Ebene der *konzeptuellen Schemata* liegen in der kompletten Plattformunabhängigkeit und hohen Wiederverwendbarkeit der erstellten Abbildungsregeln. Aus ihnen kann theoretisch beliebiger, für die Ausführung der Transformation in unterschiedlichen Software-Werkzeugen benötigter Code abgeleitet werden. Ferner ermöglicht dieser Ansatz den Umgang mit Konzepten der objektorientierten Modellierung wie z. B. Vererbung, Aggregation und Komposition. UML-Diagramme als Informationsquellen für die Abbildung sind in der Regel übersichtlicher und für Fachexperten leichter lesbar als z. B. GML-Schemata. Darüber hinaus enthalten UML-Schemata mehr Informationen als GML-Schemata (z. B. Constraints). Grenzen hat dieser Lösungsansatz beispielsweise dann, wenn für die Quelldaten keine explizit modellierten konzeptuellen Schemata vorliegen, was im Rahmen von INSPIRE in Europa voraussichtlich häufiger vorkommen wird. Ferner bereiten unterschiedliche Metamodelle und UML-Profile, die bei der Modellierung in unterschiedlichen Ländern verwendet wurden, Probleme (Eisenhut & Kutzner 2010). Gleiches gilt für das XML-Austauschformat für UML-Schemata, das von verschiedenen Werkzeugen nicht konsistent umgesetzt wird. Insgesamt ist dieser Ansatz sehr komplex, u. a. da beim Schritt von den Abbil-

dungsregeln auf konzeptueller Ebene zur Ausführung der Transformation auf Datenebene auch die Kodierungsregeln berücksichtigt werden müssen.

Auf der Ebene der *Transferformatschemata* definierte Abbildungsregeln weisen immer noch einen höheren Wiederverwendungsgrad auf als auf der Datenformatebene definierte Abbildungsregeln, da sie potenziell für alle Datensätze verwendet werden können, die dem Schema entsprechen. Es ist anzunehmen, dass der Schritt von den Abbildungsregeln zur Ausführung der Transformation auf der Datenebene etwas weniger komplex ist, da nur zwei Ebenen berücksichtigt werden müssen. Die formatschemabasierte Abbildung zwischen GML-Schemata wird zudem im „Technical Guidance“ Dokument zu den INSPIRE-Transformationsdiensten (Howard et al. 2010) empfohlen. Bei den Grenzen ist anzuführen, dass GML-Schemata – wie bereits erwähnt – weniger Informationen enthalten als UML-Schemata. Ferner können Konzepte der objektorientierten Modellierung wie Aggregationen und Kompositionen nicht problemlos behandelt werden. Schließlich können Probleme durch unterschiedliche GML-Versionen und Profile entstehen. So wird beispielsweise ab GML Version 3 ein anderes Geometrie-Modell verwendet.

Unabhängig vom gewählten Ansatz hat die semantische Transformation aber auch noch praktische Grenzen. So ist je nach dem Grad der Heterogenität zwischen Quell- und Zielschemata oft nur eine unvollständige Abbildung, eine Abbildung mit Informationsverlust oder Abbildungsungenauigkeit sowie unter Umständen sogar gar keine Abbildung möglich. Ferner können große Datensätze und komplexe Transformationen zu Performanzproblemen speziell bei der dienstebasierten Ausführung der Transformation führen. Daher tendieren Datenanbieter derzeit in der Praxis eher dazu, die Transformation in Form einer internen Prä-Prozessierung, bei der ein INSPIRE-konformer Sekundärdatenbestand erzeugt wird, und nicht als On-the-fly-Transformationsdienst zu realisieren.

5 WEITERER FORSCHUNGSBEDARF UND AUSBLICK

Im Geoinformationbereich ist semantische Transformation ein relativ junges Feld, in dem weiterer Forschungs- und Entwick-

lungsbedarf beispielsweise in den folgenden Bereichen besteht:

- Effiziente Verarbeitung von Beziehungen zwischen Objekten (z. B. Assoziationen, Aggregationen, Kompositionen) während der Transformation.
- Umgang mit unterschiedlichen Profilen der zur Datenmodellierung verwendeten Schemasprache UML (Eisenhut & Kutzer 2010).
- Funktionen zur Prüfung der Qualität der Abbildungsregeln und des Transforma-

tionsergebnisses sowie zur Beschreibung von Abbildungsungenauigkeiten.

- Lösungen für die Interoperabilität zwischen unterschiedlichen Abbildungssprachen, um den Austausch und die Wiederverwendbarkeit von Abbildungsregeln zu ermöglichen.

Der Ansatz aus dem mdWFS-Projekt wird an der TUM derzeit dahin gehend weiterentwickelt, dass die auf konzeptueller Ebene erstellten Transformationsregeln mithilfe der Software Feature Manipulation

Engine (FME) ausgeführt werden können (Donaubauer et al. 2011). Parallel dazu wird in einem vom Runder Tisch GIS e. V. koordinierten Projekt die formatbasierte semantische Transformation am Beispiel von Geobasisdaten aus der grenzüberschreitenden Bodenseeregion unter Verwendung eines Software-Werkzeuges (AED-SICAD FUSION Data Service) erprobt und anschließend mit der schemabasierten Transformation verglichen (Schönherr et al. 2011).



Wichmann

Das deutsche Vermessungs- und Geoinformationswesen 2012

Hrsg.: Kummer, K. / Frankenberger, J.

2011, XXII, 482 Seiten

ISBN 978-3-87907-511-9

73,- €

Preisänderungen und Irrtümer vorbehalten.

Es gelten die Liefer- und Zahlungsbedingungen des VDE VERLAGS.

Jetzt gleich hier bestellen: www.vde-verlag.de/120223



VDE VERLAG GMBH · Berlin · Offenbach

Bismarckstraße 33 · 10625 Berlin

Tel.: (030) 34 80 01-222 · Fax: (030) 34 80 01-9088

Mail: kundenservice@vde-verlag.de

www.wichmann-verlag.de
www.vde-verlag.de/wichmann

Literatur

- Annoni, A.; Friis-Christensen, A.; Lucchi, R.; Lutz, M. (2008): Requirements and Challenges for Building a European Spatial Information Infrastructure: INSPIRE. In: van Oosterom, P.; Zlatanova, S. (Eds.): *Creating spatial information infrastructures: Towards the spatial Semantic Web*. CRC Press, Boca Raton, S. 1-18.
- Batini, C.; Lenzerini, M.; Navathe, S. B. (1986): A comparative analysis of methodologies for database schema integration. In: *ACM Computing Surveys*, 18 (4), S. 323-364.
- Beare, M.; Howard, M.; Payne, S.; Watson, P. (2010): Development of Technical Guidance for the INSPIRE Transformation Network Service: State Of The Art Analysis. http://inspire.jrc.ec.europa.eu/documents/Network_Services/JRC_INSPIRE-TransformService_SAA_v2.pdf, Zugriff 01/2012.
- Beare, M.; Payne, S.; Sunderland, R. (2010): Prototype Report for the INSPIRE Schema Transformation Network Service. http://inspire.jrc.ec.europa.eu/documents/Network_Services/JRC_INSPIRE-TransformService_ProtoRpt_v3-0.pdf, Zugriff 03/2012.
- Donaubauer, A.; Kutzner, T.; Straub, F. (2011): Modellbasierte semantische Transformation – ein Werkstattbericht. In: Schilcher, M. (Hrsg.): *INSPIRE-GMES-Informationsbroschüre*, S. 60-62. http://www.rtg.bv.tum.de/templates/beeze_20/index.inc.filters.clickstats.php?file=images/stories/downloads/projektarbeit/projekte_topaktuell/INSPIREGMES/INSPIRE-GMES-Broschuere_V7_de.pdf, Zugriff 01/2012.
- Donaubauer, A.; Staub, P.; Straub, F.; Fichtinger, A. (2008): Web-basierte Modelltransformation – eine Lösung für INSPIRE? In: *GIS*, 2, S. 26-33.
- Eisenhut, C.; Kutzner, T. (2010): Vergleichende Untersuchungen zur Modellierung und Modelltransformation in der Region Bodensee im Kontext von INSPIRE. Technische Universität München, Fachgebiet Geoinformationssysteme, München.
- Europäische Kommission (2010): Verordnung (EU) Nr. 1088/2010 der Kommission vom 23. November 2010 zur Änderung der Verordnung (EG) Nr. 976/2009 hinsichtlich Downloaddiensten und Transformationsdiensten. In: *Amtsblatt der Europäischen Union*, L 323, 8.12.2010, S. 1-10. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0001:0010:DE:PDF>, Zugriff 01/2012.
- Europäisches Parlament und Rat (2007): Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE). In: *Amtsblatt der Europäischen Union*, L 108, 25.04.2007. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:DE:PDF>, Zugriff 01/2012.
- Fichtinger, A. (2011): Semantische Transformation im Kontext von INSPIRE – dargestellt am Beispiel der grenzüberschreitenden Bodensee-region. Dissertation, Fakultät für Bauingenieur- und Vermessungswesen, Technische Universität München.
- Fichtinger, A.; Kutzner, T. (2010): Datenharmonisierung im Kontext von INSPIRE. In: Schilcher, M. (Hrsg.): *Geoinformationssysteme: Beiträge zum 15. Münchner Fortbildungsseminar*. abcv Verlag, Heidelberg, S. 30-47.
- Fichtinger, A.; Rix, J.; Schäffler, U.; Michi, I.; Gone, M.; Reitz, T. (2011): Data Harmonisation Put into Practice by the HUMBOLDT Project. In: *International Journal of Spatial Data Infrastructures Research*, 6, S. 234-260. <http://ijmdir.jrc.ec.europa.eu/index.php/ijmdir/article/view/191/248>, Zugriff 01/2012.
- Howard, M.; Payne, S.; Sunderland, R. (2010): Technical Guidance for the INSPIRE Schema Transformation Network Service. Version 3.0. http://inspire.jrc.ec.europa.eu/documents/Network_Services/JRC_INSPIRE-TransformService_TG_v3-0.pdf, Zugriff 03/2012.
- Kashyap, V.; Sheth, A. (1996): Semantic and schematic similarities between database objects: a context-based approach. In: *The VLDB Journal*, 5 (4), S. 276-304.
- Legler, F.; Naumann, F. (2007): A Classification of Schema Mappings and Analysis of Mapping Tools. <http://subs.emis.de/LNI/Proceedings/Proceedings103/gi-proc-103-027.pdf>, Zugriff 01/2012.
- Lehto, L.: Schema Translations in a Web Service Based SDI. http://people.plan.aau.dk/~enc/AGILE2007/PDF/29_PDF.pdf, Zugriff 01/2012.
- Mens, T.; Czarnecki, K.; Van Gorp, P. (2005): A Taxonomy of Model Transformations. In: Bezzin, J.; Heckel, R. (Eds.): *Language Engineering for Model-Driven Software Development*. Internationales Begegnungs- und Forschungszentrum für Informatik, Dagstuhl (Dagstuhl Seminar Proceedings 04101).
- Nyerges, T. L. (1989): Schema integration analysis for the development of GIS databases. In: *International Journal of Geographical Information Science*, 3 (2), S. 153-183.
- Reitz, T. (2010): A Mismatch Description Language for Conceptual Schema Mapping and Its Cartographic Representation. In: Fabrikant, S. I. et al. (Hrsg.): *Geographic Information Science*. Springer, Berlin (Lecture Notes in Computer Science 6292), S. 204-218.
- Reitz, T.; Schäffler, U.; Klien, E.; Fitzner, D. (2010): Efficient Conceptual Schema Translation for Geographic Vector Data Sets. In: Painho, M. et al. (Eds.): *Geospatial Thinking*. http://agile2010.dsi.uminho.pt/pen/ShortPapers_PDF%5C112_DOC.pdf, Zugriff 01/2012.
- Scharffe, F. (2009): Correspondence Patterns Representation. Dissertation, Fakultät für Mathematik, Informatik und Physik, Universität Innsbruck.
- Schönherr, H.; Schilcher, M.; Illert, A.; Kutzner, T.; Heß, D. (2011): Grenzüberschreitende Transformation von Geodaten nach INSPIRE in der Region Bodensee – Herausforderung für Verwaltung, Wirtschaft und Wissenschaft. In: Schilcher, M. (Hrsg.): *INSPIRE-GMES-Informationsbroschüre*, S. 51-55.
- Staub, P. (2009): Über das Potenzial und die Grenzen der semantischen Interoperabilität von Geodaten: Ein operationelles Verfahren zur Nutzung verteilter Systeme in Geodaten-Infrastrukturen. Dissertation, Eidgenössische Technische Hochschule Zürich (Diss. ETH Nr. 18201).
- Stoter, J.; Quak, W.; Hofman, A. (2009): Harmonising and Integrating Two Domain Models Topography. In: van Loenen, B. et al. (Eds.): *SDI Convergence: Research, Emerging Trends, and Critical Assessment*. Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission, Delft, S. 89-106.