

DAS RULE INTERCHANGE FORMAT (RIF) FÜR INTEROPERABLE SCHEMAMAPPING-ANWENDUNGEN IN INSPIRE

Susanne Reinwarth, Lars Bernard

Zusammenfassung: Beim Aufbau von Geodateninfrastrukturen (GDI) spielt die Harmonisierung von Geodatenätzen eine wesentliche Rolle. In diesem Zusammenhang werden Schematransformationen angewendet, die die Originaldaten anhand festgelegter Transformationsregeln – dem sogenannten Schemamapping – in ein einheitliches Datenschema überführen. INSPIRE (Infrastructure for Spatial Information in Europe) sieht hierfür eine generische Beschreibung des Schemamappings im Rule Interchange Format (RIF) vor, um für unterschiedliche Schemamapping-Systeme Interoperabilität zu erreichen. RIF dient in erster Linie dem Austausch von Regeln zwischen verschiedenen Regelsystemen. Dieser Beitrag zeigt zum einen, wie RIF für das Schemamapping von Geodaten eingesetzt werden kann, und zum anderen, dass RIF alle Anforderungen erfüllt, die INSPIRE an die Funktionalität einer solchen Mappingsprache stellt. Am Beispiel des HUMBOLDT Alignment Editors (HALE) wird die Integration von RIF in bestehende Schemamapping-Software demonstriert.

Schlüsselwörter: Rule Interchange Format (RIF), Geodateninfrastrukturen, Schematransformationsdienst, semantische Interoperabilität, INSPIRE

USING THE RULE INTERCHANGE FORMAT (RIF) FOR INTEROPERABLE SCHEMA MAPPING APPLICATIONS IN INSPIRE

Abstract: The harmonization of spatial data sets is one of the major challenges for the development of spatial data infrastructures (SDI). Therefore, schema transformations are applied to convert the original data into a unified data schema using specified transformation rules, called the schema mapping. For the implementation of INSPIRE (Infrastructure for Spatial Information in Europe) a generic description of the schema mapping in the Rule Interchange Format (RIF) is proposed to achieve interoperability of different schema mapping systems. RIF was primarily designed for the exchange of rules between different rule systems. This article shows how to deploy RIF for the schema mapping of spatial data and that RIF meets all the requirements of INSPIRE on the functionality of such a mapping language. The example of the HUMBOLDT Alignment Editor (HALE) demonstrates the integration of RIF in existing schema mapping software.

Keywords: Rule Interchange Format (RIF), spatial data infrastructures, schema transformation service, semantic interoperability, INSPIRE

Autoren

Dipl.-Ing. Susanne Reinwarth
Prof. Dr. Lars Bernard
Technische Universität Dresden
Professur für Geoinformationssysteme
Helmholtzstraße 10
01069 Dresden

E: susanne.reinwarth@tu-dresden.de
lars.bernard@tu-dresden.de

1 SCHEMAMAPPING UND SCHEMATRANSFORMATION IM KONTEXT VON INSPIRE

Ein wesentliches Ziel von Geodateninfrastrukturen (GDI) ist die effiziente Integration von Geodaten unterschiedlicher Quellen. Idealerweise sollen diese in der Regel heterogenen Geodatenätze ad hoc kombiniert und gemeinsam genutzt werden können, ohne sie erst aufwendig in einen Datenbestand zusammenzuführen (Bill 2010). Auf der Grundlage der EU-Initiative INSPIRE (Infrastructure for Spatial Information in Europe) (INSPIRE 2007) findet derzeit der Aufbau einer europäischen GDI statt. Die Integrationsfähigkeit der Geodaten hinsichtlich Inhalt und Struktur wird hier durch die Entwicklung europäisch harmonisierter Geodatenpezifikationen erreicht. In diese gemeinsamen Schemas, die als Klassendiagramme in der Unified Modeling Language (UML) modelliert und als XML-Schemadateien bereitgestellt werden, sollen sich die nationalen Geodatenätze dann transformieren lassen. Abbildung 1 zeigt beispielhaft das vereinfachte UML-Klassendiagramm eines Flurstücks im Amtlichen Liegenschaftskatasterinformationssystem (ALKIS). In ALKIS werden verschiedene Attribute eines Flurstücks gespeichert wie die amtliche Fläche in Quadratmetern, die Flurstücksnummer, die Flurstücksgeometrie etc. Dem Kataster anderer EU-Mitgliedsstaaten liegen meist ebenfalls Flurstücke zugrunde, jedoch haben deren Attribute anderssprachige Bezeichnungen und die Werte möglicherweise andere Einheiten.

Um nun das gesamte Angebot europäischer Katasterdaten integriert nutzen zu können, sind die nationalen Daten INSPIRE-konform aufzubereiten und bereitzustellen. Im Falle von Flurstücken legt die entsprechende Datenspezifikation Cadastral Parcels (INSPIRE TWG Cadastral Parcels 2010) fest, dass es sich dabei um einen geschlossenen Teil der Erdoberfläche auf nationalem Territorium handeln soll, der vollständig einheitlichem Eigentum und homogenen Eigentumsrechten unterliegt. Wie Abbildung 1 rechts zeigt, definiert das zugehörige Datenschema verschiedene Pflichtattribute, z. B. die Geometrie, einen Identifikator und eine Bezeichnung sowie optionale Attribute, darunter die Fläche und einen Referenzpunkt.

Um ein INSPIRE-konformes Abbild nationaler Geodaten zu erzeugen und so

semantische Interoperabilität zu erzielen, werden Schematransformationen angewendet, die die Originaldaten anhand festgelegter Transformationsregeln – dem sogenannten Schemamapping – in das einheitliche INSPIRE-Datenschema überführen. Abbildung 2 dient einer Abgrenzung dieser beiden Begriffe. Das Schemamapping definiert die Regeln für den Übergang vom Quell- zum Zieldatenschema. Geht man von einer objektstrukturierten Modellierung von Quell- und Zieldatenschema aus, so müssen sowohl Quell- und Zielklassen als auch Quell- und Zielattribute miteinander in Beziehung gebracht werden. Existieren keine direkten Korrespondenzen, also keine äquivalenten Bestandteile, so sind erwei-

der Extensible Stylesheet Language (XSL) beschrieben wird.

Neben den typischen Diensten einer GDI zum Suchen, zur Visualisierung und zum Download von Geodaten sind in der INSPIRE-Direktive auch Transformationsdienste gefordert, die eine interoperable Überführung nationaler Geodaten in die für INSPIRE festgelegten Referenzsysteme (Koordinatensysteme, Schemas etc.) gewährleisten sollen. In den aktuellen Empfehlungen zur technischen Umsetzung von INSPIRE-Schematransformationsdiensten (Howard et al. 2010) wird der Aufbau und die Verwendung von Registern vorgeschlagen, die formalisierte und entsprechend generische Beschreibungen zu den

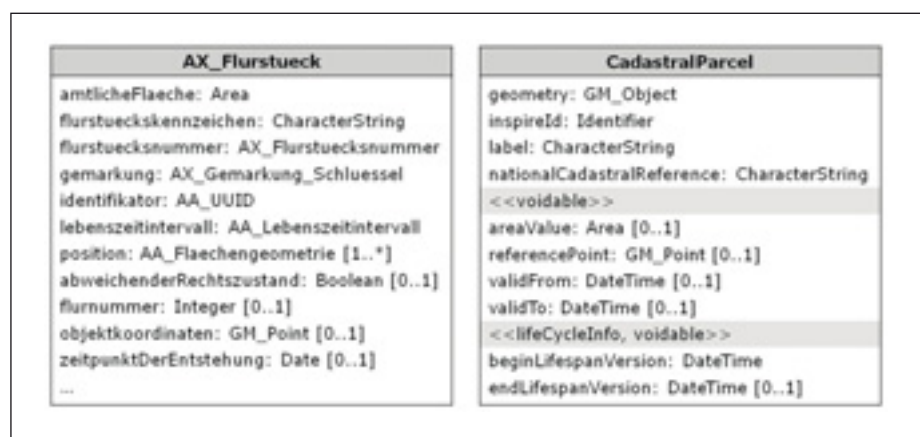


Abbildung 1: Flurstücke in ALKIS (links) und INSPIRE (rechts) nach Adv (2008) und INSPIRE TWG Cadastral Parcels (2010)

terte Transformationsregeln zu definieren, die es ermöglichen, aus der Struktur der Quelldaten die Zieldatenstruktur zu erstellen. Dies kann beispielsweise durch Kombination mehrerer Attribute oder Berechnung der geforderten Werte erfolgen. Informationen können gänzlich fehlen und Zielattribute somit leer bleiben. Meist sind auch nicht alle Informationen des Quelldatenschemas für das Zieldatenschema relevant.

Während das Schemamapping nur die Ebene der Datenschemas betrifft, ist die Schematransformation die eigentliche Datentransformation auf Grundlage des jeweiligen Schemamappings und überführt den Quelldatenatz aus der Struktur des Quelldatenschemas in einen Zieldatenatz mit der Struktur des Zieldatenschemas. Dazu gilt es, aus dem Schemamapping eine geeignete Transformationsvorschrift abzuleiten, die beispielsweise für in der Extensible Markup Language (XML) codierte Daten in

Schemamappings sowie den Quell- und Zieldatenschemas vorhalten. So kann das Schemamapping losgelöst von einem spezifischen Dienst verfügbar gemacht werden und verschiedene Implementierungen der Schematransformationsdienste ermöglichen eine erhöhte Flexibilität.

Die Definition des Schemamappings erfolgt idealerweise in einer grafischen Nutzerumgebung, die dann die Mappingregeln in eine generische Mappingsprache überführt und verwaltet. Derzeit verfügbare Softwareprodukte bieten zwar grafische Nutzeroberflächen, arbeiten allerdings meist mit proprietären Formaten, was die Wiederverwendbarkeit des formulierten Schemamappings verhindert. Eine für INSPIRE durchgeführte State-of-the-Art-Analyse (Beare et al. 2010a) hat ergeben, dass noch kein allgemein gebräuchlicher Standard für Mappingsprachen existiert. In diesem Zusammenhang wurden verschiedene

infrage kommende Formate hinsichtlich ihrer Ausdrucksfähigkeit und verschiedener Implementierungskriterien untersucht. Daraufhin wurde RIF für den INSPIRE-Schematransformationsdienst als Austauschformat für Schemamappings empfohlen (Howard et al. 2010).

2 RIF – EINE REGELSPRACHE

Der Begriff „Regel“ ist nicht klar definiert, sondern ein Ausdruck für verschiedene Konzepte (Boley et al. 2007). Im Bereich der Wissensformen verknüpft eine Regel eine oder mehrere Bedingungen (Wenn-Teil) mit einer oder mehreren Schlussfolgerungen (Dann-Teil) (Bill 1999). Bezogen auf das Schemamapping wird die Struktur

mit den Daten festzulegen, sodass daraus Informationen abgeleitet werden können (W3C 2012). Hierbei dient RIF dem Austausch von Regeln zwischen verschiedenen Regelsystemen (Kifer & Boley 2010) und fungiert somit als nicht-proprietäres, intermediäres Regelformat. Damit ist RIF zwar nicht in erster Linie eine Mappingsprache, aber die Möglichkeit der Nutzung für die Integration unterschiedlich strukturierter Daten wurde bereits bei der Konzeption von RIF in einem Anwendungsfall betrachtet (Paschke et al. 2008).

Wie von Boley et al. (2007) ausgeführt und von Nash et al. (2011) zusammengefasst sind im Semantic-Web-Kontext drei Arten von Regeln von Bedeutung:

Die Klassen und Attribute des Zieldatenschemas werden im Aktionsteil der Regel neu erzeugt und belegt.

Jede Regelsprache, oder auch Regelformat, besitzt eine Syntax mit festgelegter Semantik zum Ausdruck der Wenn-Dann-Struktur. Einige Regelsprachen beinhalten zusätzlich einen Teil für die Definition von alternativen Schlussfolgerungen für den Fall, dass die Bedingungen nicht zutreffen. Einen solchen Else-Teil besitzt RIF nicht.

2.1 KONZEPT UND AUFBAU VON RIF

RIF ist nicht als universale Regelsprache konzipiert, sondern als Familie mehrerer Sprachdialekte (Kifer & Boley 2010). Diese nutzen eine gemeinsame Syntax und Semantik und werden formal im XML-Format ausgetauscht. Neben der normativen XML-Syntax existiert zu Anschauungszwecken die stark verkürzte und nicht-normative Präsentationssyntax, die auch im Folgenden für Beispiele genutzt wird.

Die RIF-Sprachfamilie ist erweiterbar, sodass unterschiedliche Nutzergruppen weitere RIF-Dialekte gemäß deren Anforderungen definieren können (Kifer & Boley 2010). Die Grundlage für die Spezifizierung neuer Dialekte bildet das RIF Framework for Logic Dialects (RIF-FLD). Es fungiert als Katalog von syntaktischen und semantischen Charakteristika für den Aufbau neuer logischer Dialekte. Derzeit bestehen drei RIF-Dialekte als offizielle W3C-Vorschläge:

- ▶ RIF Basic Logic Dialect (RIF-BLD) enthält die logischen Grundlagen für die Repräsentation deduktiver Regeln mit logischen Aussagen im Schlussfolgerungsteil.
 - ▶ RIF Production Rule Dialect (RIF-PRD) basiert auf RIF-BLD und erweitert den Schlussfolgerungsteil der Regel um Aktionen. Somit werden mit RIF-PRD Produktionsregeln formuliert. Außerdem erlaubt RIF-PRD bisher als einziger Dialekt die Formulierung von Negationen.
 - ▶ RIF Core umfasst Teilmengen der beiden Hauptdialekte RIF-BLD und RIF-PRD.
- Die RIF-Dialekte greifen auf die Datentypen, Funktionen und Prädikate zu, die in RIF Datatypes and Built-Ins (RIF-DTB) definiert sind und auf XML Schema und XPath (XML Path Language) basieren.

Mappingsprachen sind Regelsprachen, die für die Definition von Schemamappings genutzt werden und gehören damit zu den regelbasierten Formen der Wissensreprä-

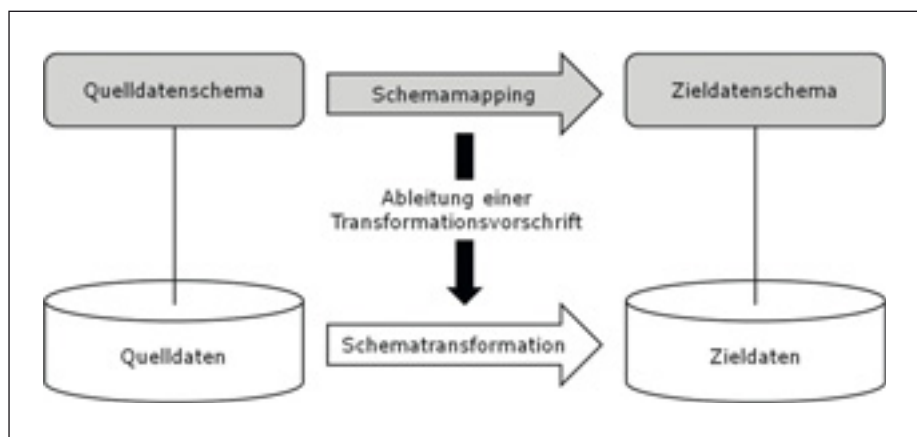


Abbildung 2: Schemamapping und Schematransformation

des Zieldatenschemas als Schlussfolgerung generiert, wenn die formulierten Bedingungen auf das Quelldatenschema bzw. den Quelldatensatz zutreffen. Eine weitere Interpretation von Regeln liefert die Sicht auf den Systemzustand. Demnach existieren verschiedene Zustände – Situationen, die in einer Anwendungsumgebung vorkommen können – und mit einer Regel wird die Logik für den Übergang von einem Startzustand zu einem Zielzustand festgelegt (Gašević et al. 2006). Im Falle des Schemamappings stellt das Quelldatenschema den Startzustand dar und das Zieldatenschema den Zielzustand.

RIF ist ein Standard, der im Rahmen der Semantic-Web-Aktivitäten des World Wide Web Consortiums (W3C) entwickelt wurde. Semantic-Web-Technologien sollen es künftig ermöglichen, das Internet als Datenspeicher zu nutzen, Vokabulare aufzubauen und Regeln für den Umgang

1. Deduktive Regeln beschreiben statische Abhängigkeiten, um implizites Wissen von explizitem Wissen abzuleiten.
2. Normative Regeln dienen nicht der Ableitung neuer Erkenntnisse, sondern werden für die Einschränkung von Daten oder Anwendungslogik genutzt. Normative Regeln werden häufig als deduktive oder reaktive Regeln implementiert.
3. Reaktive Regeln führen Aktionen aus, bei denen es sich um alle denkbaren Operationen handeln kann, z. B. das Ändern von Daten. Handelt es sich um ereignisbasierte Regeln, spricht man von Event-Condition-Action-(ECA-)Regeln. Produktionsregeln hingegen sind nicht ereignisbasiert und die definierten Aktionen werden ausgeführt wann immer die Daten die Bedingungen erfüllen.

Ein Schemamapping lässt sich in Form einer oder mehrerer Produktionsregeln formu-

Nr.	INSPIRE-Zielattribut	Belegung aus ALKIS-Attribut(en)	Bedingung
1	areaValue	amtlicheFlaeche	
2	beginLifespanVersion	AA_Lebenszeitintervall.beginnt	
3	endLifespanVersion	AA_Lebenszeitintervall.endet	endet EXISTS
4	endLifespanVersion	void, „unpopulated“	endet NOT EXISTS
5	geometry	position	
6	inspireId.localId	identifikator	
7	inspireId.namespace	„DEAAA“	
8	label	AX_Flurstuecksnummer.zaehler + „/“ + AX_Flurstuecksnummer.nenner	nenner EXISTS
9	label	AX_Flurstuecksnummer.zaehler	nenner NOT EXISTS
10	nationalCadastralReference	flurstueckskennzeichen	
11	referencePoint	objektkoordinaten	objektkoordinaten EXISTS
12	referencePoint	PointOnSurface(position)	objektkoordinaten NOT EXISTS
13	validFrom	zeitpunktDerEntstehung	zeitpunktDerEntstehung EXISTS
14	validFrom	AA_Lebenszeitintervall.beginnt	zeitpunktDerEntstehung NOT EXISTS
15	validTo	AA_Lebenszeitintervall.endet	endet EXISTS
16	validTo	void, „unpopulated“	endet NOT EXISTS

Tabelle 1: Schemamapping von AX_Flurstueck nach CadastralParcel

sensation. Sie lassen sich wie Wissensformen in prozedurale und deklarative Mappingsprachen unterteilen (Bill 1999). RIF-PRD ist prozedural, oder auch imperativ, aufgebaut und ähnelt in seiner Struktur einer prozeduralen Programmiersprache, die explizit beschreibt, wie Variablen deklariert, belegt und verarbeitet werden. In deklarativen Sprachen hingegen wird ein Problem oder einen Sachverhalt formuliert, ohne auf die konkrete Vorgehensweise zur Lösung einzugehen. Daher gehört zu deklarativem Wissen auch immer ein Verfahren, das die Auswertung der Beschreibung übernimmt. Ein Beispiel für eine deklarative Mappingsprache ist die geospatial Ontology Mapping Language (gOML), die die Schemamapping-Software HALE (HUMBOLDT Alignment Editor) nutzt. Soll beispielsweise in dem Schemamapping von ALKIS nach INSPIRE Cadastral Parcels der Wert des Quellattributs `amtlicheFlaeche` für das Zielattribut `area` übernommen werden, so gibt die deklarative Mappingsprache gOML lediglich die Transformationsfunktion an, in diesem Fall `RenameAttributeFunction(amtlicheFlaeche, area)`, und die konkrete Implementierung dieser Funktion bleibt verborgen. Im Gegensatz dazu

wird in der prozeduralen Sprache RIF zunächst eine Variable deklariert, diese mit dem Wert des Quellattributs belegt und dieser dann dem Zielattribut zugewiesen. Dies ist für alle Zielklassen und -attribute erforderlich, sodass ein prozedurales Schemamapping zwar sehr detailliert alle Arbeitsschritte beschreibt, jedoch deutlich weniger kompakt ist als ein in einer deklarativen Mappingsprache formuliertes Schemamapping, welches auf die konzeptuelle Ebene beschränkt ist.

2.2 GEO-SCHEMAMAPPING MIT RIF-PRD

RIF-PRD eignet sich für die Definition von Schemamappingregeln, da die Aktionen die Generierung von Klassen und Attributen des Zielschemas erlauben. Konkret definiert RIF-PRD drei Aktionen: `Assert`, `Modify` und `Retract` zum Erzeugen, Verändern und Löschen von Klasseninstanzen oder Attributen. Ein komplettes Schemamapping besteht aus mehreren Produktionsregeln, die nach OMG (2009) in folgenden Schritten abgearbeitet werden:

1. Übereinstimmung: Alle Regeln, deren Bedingungen aufgrund des Quelldatensatzes erfüllt sind, werden instanziiert.
2. Konfliktlösung: Ein Konflikt entsteht,

wenn zu einem Zeitpunkt die Bedingungen mehrerer Regeln erfüllt sind, und somit mehrere Regeln instanziiert wurden. In diesem Fall greift ein Entscheidungsalgorithmus, der die auszuführende Regel ermittelt.

3. Aktion: Die Aktionen der ausgewählten Regel werden ausgeführt. Ist kein finaler Stand erreicht, springt das System zum ersten Schritt (Übereinstimmung) zurück. So entsteht nach und nach der Zieldatensatz.

Ein Schemamapping kann prinzipiell mit den Produktionsregeln, die RIF-PRD bereits abdeckt, beschrieben werden. Um speziell das Schemamapping von Geodaten vollständig beschreiben zu können, sind lediglich die RIF-Datentypen und Built-In-Funktionen um passende räumliche Datentypen und Funktionen zu erweitern. Hier können bereits existierende Geostandards für die Bereitstellung räumlicher Konzepte genutzt werden, sodass eine Bibliothek mit räumlichen Datentypen und Funktionen entsteht (Nash et al. 2011). Dabei gilt es nicht, komplexe Funktionsalgorithmen zu definieren, sondern nur die Funktionssignaturen mit den entsprechenden Ein- und Ausgabeparametern festzulegen. Da RIF nur als Zwischenformat fungiert, obliegt die

korrekte Implementierung der räumlichen Datentypen und Funktionen dann dem jeweiligen Schematransformationswerkzeug. Diesen Weg schlägt auch die technische Anleitung für den INSPIRE-Schematransformationsdienst (Howard et al. 2010) vor und empfiehlt die Spezifikation „Simple Feature Access“ des Open Geospatial Consortiums (OGC) (Herring 2011) als Basis für die Benennung von Platzhaltern für räumliche Datentypen, Funktionen und Prädikate.

Diesem Ansatz folgend konzipieren Nash et al. (2011) mit „GeoRIF“ eine Teilmenge der Funktionen der OGC-Spezifikation „Simple Features for SQL“ und zeigen, wie drei verschiedene Anwendungsfälle als Regeln in GeoRIF-Syntax formuliert werden können. Während die von Nash et al. (2011) gewählten Anwendungsfälle die räumliche Funktionalität im Bedingungs- und Regelteil benötigen, braucht das Schemamapping die räumlichen Funktionen häufig im Schlussfolgerungsteil, beispielsweise „Existiert im Quelldatensatz kein Referenzpunkt für das Flurstück, so berechne einen Punkt aus der Flurstücksgeometrie“.

Tabelle 1 zeigt das Schemamapping zwischen den in Abbildung 1 skizzierten Klassen `AX_Flurstueck` aus ALKIS und

`CadastralParcel` für INSPIRE. Dieses Schemamapping besteht überwiegend aus direkten Korrespondenzen, d. h. das Quellattribut ist lediglich in das Zielattribut umzubenennen und der Wert kann direkt übernommen werden. Einige der Quellattribute sind optional, erkennbar an der Kardinalität `[0..1]`. Bei der Verwendung optionaler Attribute muss sowohl der Fall abgedeckt werden, dass das Attribut im Quelldatensatz existiert (Bedingung `EXISTS`) als auch, dass es nicht existiert (Bedingung `NOT EXISTS`). So kann z. B. die Flurstücksnummer nur aus dem Zähler bestehen. In diesem Fall existiert das Attribut `nummer` nicht (Zeile 9) und `zaehler` muss nur in `label` umbenannt werden. Oder die Flurstücksnummer setzt sich aus Zähler und Nenner zusammen. In diesem Fall existiert das Attribut `nummer` (Zeile 8) und das Zielattribut `label` ist durch die Verkettung von `zaehler` und `nenner` mit dem Separator `„/“` zu bilden. Neben der Umbenennung und Verkettung von Attributen kommt als nicht-räumliche Funktion noch das Setzen fester Werte vor, z. B. im Subattribut `namespace` des Zielattributs `inspireId` (Zeile 7). Als einzige räumliche Funktion wird für die Belegung von `referencePoint` die Simple-Feature-Access-Funktion `pointOnSurface()`

angewendet, die aus der Flurstücksgeometrie einen Referenzpunkt des Flurstücks berechnet, wenn der Quelldatensatz das Attribut `objektkoordinaten` nicht führt (Zeile 12).

Den Aufbau des in RIF-PRD formulierten Schemamappings zeigt Listing 1. Quell- und Zieldatenschema werden durch die Festlegung der Prefix-Elemente `source` und `target` eindeutig angegeben. Der Bezug auf den OGC-Standard „Simple Feature Access“ für die Nutzung räumlicher Funktionen erfolgt ebenso unkompliziert über das Prefix-Element `spatial`. Die erste Regel enthält das Mapping zwischen der Quell- und der Zielklasse sowie den Attributen, für die keine Bedingung formuliert werden muss, beispielsweise `amtlicheFlaeche`. Für jedes Attribut-Mapping mit Bedingung würde sich nun eine weitere Regel anschließen. Da mit RIF kein Else-Teil formuliert werden kann, müsste die Beschreibung der Alternative jeweils in zusätzlichen Regeln erfolgen. Somit hätte das komplette Schemamapping noch weitere zehn Regeln.

3 DIE AUSDRUCKSFÄHIGKEIT VON RIF-PRD

Am Beispiel des konkreten Schemamappings von `AX_Flurstueck` nach `Ca-`

```
Document (
  Prefix (source <http://www.adv-online.de/namespaces/adv/gid/6.0>)
  Prefix (target <urn:x-inspire:specification:gmlas:CadastralParcels:3.0>)
  Prefix (spatial <http://www.opengeospatial.org/standards/sfa/>)

  Group (
    (***) Klassen- und Attribut-Mapping ohne Bedingungen (***)
    If (Exists (
      ?ax_flurstueck-instance # source:AX_Flurstueck
      ?ax_flurstueck-instance [
        source:amtlicheFlaeche -> ?ax_flurstueck-amtlicheflaeche
        ...]
    ) )
    Then Do (Assert (
      ?cadastralparcel-instance New()
      ?cadastralparcel-instance # target:CadastralParcel
      ?cadastralparcel-instance [
        target:areaValue -> ?ax_flurstueck-amtlicheflaeche
        ...]
    ) )
    ...
  ) )
```

Listing 1: Ausschnitt des Schemamappings von `AX_Flurstueck` nach `CadastralParcel` in RIF-PRD-Präsentationssyntax

`dastralParcel` wird im Folgenden die Ausdrucksfähigkeit von RIF-PRD anhand von sechs Schematransformationsleveln (Beare et al. 2010a) diskutiert. Diese beinhalten unterschiedlich komplexe, insbesondere auch räumliche Transformationsarten, die für Schematransformationen im INSPIRE-Kontext relevant sind. Dabei nimmt die Komplexität der Funktionen mit steigendem Level zu und jedes Level beinhaltet die Funktionalität der darunter liegenden Level.

3.1 LEVEL 1 – UMBENENNUNG VON KLASSEN UND ATTRIBUTEN

Das niedrigste Schematransformationslevel ermöglicht die Umbenennung von Klassen und Attributen, bei denen direkte Korrespondenzen zwischen Quell- und Zielelement vorliegen und die Werte beibehalten werden.

Die Umbenennung von Klassen wird mit RIF realisiert, indem für jede Instanz der Quellklasse eine entsprechende Instanz der Zielklasse angelegt wird. Hierfür kann `Exists` die Existenz der Quellklasseninstanz prüfen, wie es im Beispiel `AX_Flurstueck/CadastralParcel` in Listing 1 der Fall ist. Nach den Regeln für die Abarbeitung von Produktionsregeln wird die formulierte Regel so oft angewendet, bis alle Flurstücke des Quelldatensatzes abgearbeitet sind und damit der finale Zustand erreicht ist. Ebenso kann aber mit dem Universalquantor `ForAll` die Regel von vornherein für alle vorhandenen Flurstücke aufgestellt werden. Hiermit wäre der finale Zustand bereits nach der ersten Anwendung der Regel erreicht.

Die Umbenennung von Attributen erfolgt durch die Zuweisung von Quellattributwerten an Zielattribute. Dazu ist zunächst der Wert des Quellattributs in einer Variablen zwischenspeichern. Bei der Belegung des Zielattributs kann auf den Wert dieser Variablen zugegriffen werden.

3.2 LEVEL 2 – EINFACHE ABLEITUNG VON ATTRIBUTEN

Die Schematransformation auf diesem Level findet immer nur innerhalb einer einzelnen Klasse statt, d. h. eine Zielklasse und deren Attribute können durch Untersuchung einer einzigen Quellklasse abgeleitet werden. Dabei kommen einfache nicht-räumliche und räumliche Funktionen zur Anwendung. RIF-DTB stellt elementare nicht-räumliche

Funktionen und Prädikate bereit wie Konvertierung von einfachen Datentypen, Vergleichsoperatoren, einfache numerische Operatoren und Funktionen zur Bearbeitung von Zeichenketten und Listen. Grundlegende räumliche Funktionalität wird durch den OGC-Standard „Simple Feature Access“ zur Verfügung gestellt. Dazu gehören u. a. `Envelope()` zur Berechnung des umgebenden Rechtecks einer Geometrie, `GetArea()` zur Berechnung der Fläche von einfachen oder komplexen Oberflächen oder `GetX()` und `GetY()` für den Zugriff auf Punktkoordinaten.

Schematransformationslevel 2 sieht außerdem das Ersetzen von Werten basierend auf Lookup-Tabellen vor, was mit RIF in Form von Listen mit übereinstimmenden Listenindizes realisiert werden kann.

3.3 LEVEL 3 – AGGREGATION

Diese Transformationsstufe bezieht sich auf die Datenhaltung in relationalen Datenbanken, wo Daten in Form von Tabellen gespeichert werden. Im Zuge der Normalisierung werden oft Daten einer einzelnen Objektklasse auf verschiedene Tabellen aufgeteilt. Mit Schematransformationslevel 3 sollen derart verteilte Daten wieder zusammengeführt werden können. Zum einen müssen also geeignete Funktionen für die Untersuchung von Schlüsselattributen vorhanden sein und zum anderen Funktionen für die Aggregation numerischer, temporal, nominaler und räumlicher Attributwerte. Wie für Schematransformationslevel 2 enthält RIF-DTB die benötigten nicht-räumlichen Funktionen. Die Funktionen zur räumlichen Aggregation, beispielsweise `Union()`, werden durch den OGC-Standard „Simple Feature Access“ abgedeckt.

3.4 LEVEL 4 – KOMPLEXE ABLEITUNG UND DYNAMISCHE TYPSELEKTION

Schematransformationslevel 4 ist die Erweiterung des Level 2 um komplexere Funktionalität und ermöglicht die Auswertung logischer Operationen. Dazu gehören Bedingungstests, Tests von skalaren Beziehungen und die logische Kombination von Tests. Da RIF eine Regelsprache ist, sind Bedingungstests ein grundsätzlicher Bestandteil und kommen schon ab Schematransformationslevel 1 zur Anwendung. RIF beinhaltet kein Else-Konstrukt, dafür aber die Möglichkeit, Negationen auszudrücken.

Somit ist für die Formulierung des Alternativfalls jeweils die ursprüngliche Bedingung zu negieren. Durch die Verknüpfungsooperatoren `Or` und `And` können Bedingungen kombiniert werden.

Für die Auswertung skalarer Beziehungen stellt RIF-DTB eine Reihe numerischer Prädikate bereit wie `pred:numeric-equal(?arg1 ?arg2)` oder `pred:numeric-greater-than(?arg1 ?arg2)`.

Level 4 sieht des Weiteren die Iteration über Attribute oder abgeleitete Werte vor, z. B. über jedes Element eines Arrays oder für alle Teile einer komplexen Geometrie. Hierfür kann der Universalquantor `ForAll` eingesetzt werden (siehe RIF Working Group (2009)).

3.5 LEVEL 5 – ABLEITUNG AUS UNTERSCHIEDLICHEN KLASSEN

Die bisherigen Schematransformationslevel erlauben die Bildung von Zielklassen und Zielattributen durch die Untersuchung einer einzigen Quellklasse. Ab Level 5 können Informationen für die Ableitung von Werten der Zielklasse auch aus mehreren Quellklassen gewonnen werden. Als Beispiel dient hier die räumliche Aggregation von Flurstücken zur Ableitung der Geometrie einer Katasterzone (INSPIRE-Zielklasse `CadastralZoning`, Abbildung 3 rechts). In der korrespondierenden ALKIS-Quellklasse `AX_Gemarkung` (Abbildung 3 links) werden keine Geometrien, sondern nur Sachdaten gespeichert. Im INSPIRE-Datenschema `Cadastral Parcels` ist aber für Katasterzonen das Attribut `geometry` verpflichtend. Somit ist die Geometrie der Gemarkung durch die räumliche Aggregation aller Flurstücke mit identischem Gemarkungsschlüssel mit der Funktion `Union()` zu bilden.

Neben der Untersuchung mehrerer Quellklassen beinhaltet Schematransformationslevel 5 auch komplexere Funktionen für die Erstellung der Zielklassen und -attribute. Hierzu gehören z. B. Universal- und Existenzquantor, die RIF mit `ForAll` und `Exists` abdeckt. Bei der Formulierung von Schemamappings mit RIF-PRD ist für die Instanziierung einer Zielklasse immer einer dieser beiden Quantoren nötig. Komplexere räumliche Funktionen wie `Buffer()` oder `Intersection()` stellt auch für dieses Level der OGC-Standard „Simple Feature Access“ bereit.

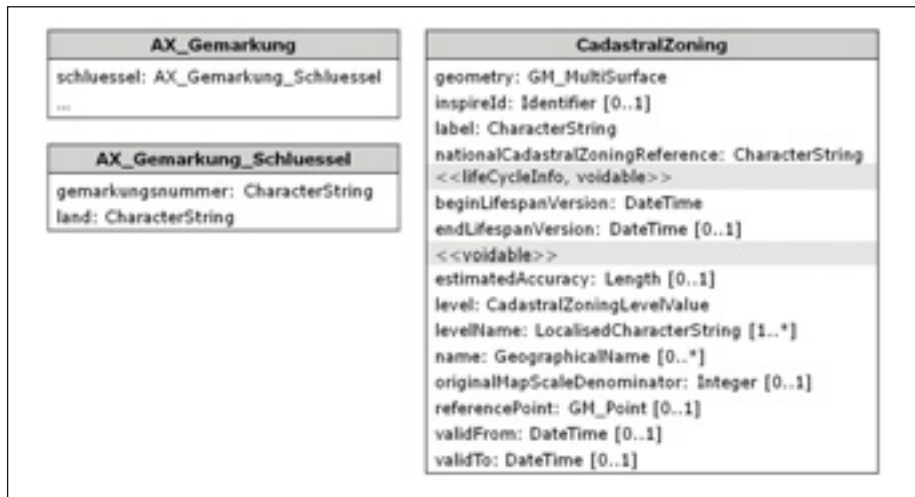


Abbildung 3: Gemarkungen in ALKIS (links) und INSPIRE (rechts) nach Adv (2008) und INSPIRE TVWG Cadastral Parcels (2010)

3.6 LEVEL 6 – AUFTEILUNG UND ZUSAMMENFÜHRUNG

Mit dem obersten Schematransformationslevel kann erstens eine Zielklasse aus verschiedenen Quellklassen gebildet werden, die entweder aus dem gleichen oder aus unterschiedlichen Datenschemas stammen. Zweitens ist es möglich, verschiedene Zielklassen aus einer einzelnen Quellklasse zu erzeugen.

Den ersten Fall deckt das Beispiel der Aggregation von Flurstücksgeometrien für die Ableitung der Gemarkungsgeometrie bereits mit ab. Hier stammen die Sachdaten aus der Klasse `AX_Gemarkung`, während die Geometrie aus den zugehörigen Flurstücksgeometrien resultiert. Auf weitere Quelldatenschemas könnte problemlos über weitere Prefix-Elemente zugegriffen werden, die die Namensräume zusätzlicher Quelldatenschemas festlegen.

Für den zweiten Fall ist die Bildung separater Zielklassen für unterschiedliche Geometrietyphen, wie Polygone oder Punkte, aus einer einzelnen Quellklasse denkbar. Hier wären mit RIF-PRD im Schlussfolgerungsteil der Regel entsprechend mehrere Zielklassen zu instanzieren.

4 PROTOTYPISCHE UMSETZUNG EINES RIF-BASIERTEN SCHEMAMAPPINGS

Grundlage der prototypischen Implementierung eines in RIF-PRD beschriebenen ALKIS-INSPIRE-Schemamappings ist der HUMBOLDT Alignment Editor (HALE). HALE ist eine im Rahmen des EU-Projektes HUMBOLDT entwickelte Open-Source-Software mit einer grafischen Benutzeroberfläche

zur Definition von Schemamappings auf Basis der XML-Schemadateien von Quell- und Zieldatenschema. Standardmäßig wird das Schemamapping vom HALE in der deklarativen Mappingsprache gOML (geospatial Ontology Mapping Language) gespeichert. Die Firma „1Spatial“ hat für HALE einen RIF-Exporter entwickelt, der das gOML-Schemamapping in den prozeduralen RIF-PRD übersetzt und in der XML-Syntax ausgibt (Beare et al. 2010b).

Für das hier diskutierte Schemamapping von ALKIS nach INSPIRE wurde der HALE-RIF-Exporter um folgende Funktionalität erweitert:

- ▶ Verkettung von Strings, um beispielsweise die Attribute `zaehler` und `nenner` für das Zielattribut `label` zu verketteten,
- ▶ Zuweisung des Schwerpunkts einer Geometrie an ein Zielattribut, um einen Punkt innerhalb des Flurstücks angeben zu können,
- ▶ Generieren eines INSPIRE-Identifikators,
- ▶ Formulierung von Bedingungen für das Attribut-Mapping, insbesondere Existenz-Bedingungen, um die Existenz der optionalen Attributwerte im Quelldatensatz zu prüfen.

Der gesamte Prozess der Übersetzung eines gOML-Schemamappings läuft in mehreren Schritten ab, bei denen vier Zwischenformate des Schemamappings erzeugt werden. Die Weiterentwicklung erfolgte durch Erweiterung und Anpassung der Klassen, die die jeweiligen Metadaten auf dem Weg von gOML nach RIF-PRD erzeugen. Als limitierend für den Ausbau

des HALE-RIF-Exporters erwies sich an einigen Stellen die Funktionalität des HALE selbst. War eine benötigte Funktion schon im HALE nicht implementiert, so konnte sie auch im RIF-Exporter nicht umgesetzt werden, da fehlende HALE-Funktionen im gOML-Schemamapping nicht auftauchen. Der RIF-Exporter ist jedoch so konzipiert, dass er das gOML-Schemamapping nach RIF-PRD übersetzt. Dies betraf z. B. die Definition von Bedingungen für tief geschachtelte Attribute wie `nenner` des Datentyps `AX_Flurstueck`. Mit der kontinuierlichen Weiterentwicklung von HALE sollten diese Probleme jedoch behoben werden.

5 FAZIT UND AUSBLICK

Für das zugrunde gelegte ALKIS-INSPIRE-Schemamapping hat sich RIF-PRD als geeignetes und ausreichend ausdrucksstarkes Schemamapping-Austauschformat gezeigt. Die angeführten sechs Schematransformationslevel konnten ausnahmslos mit RIF-PRD umgesetzt werden. Die geforderte nicht-räumliche Funktionalität deckt RIF-DTB ab. Die Einbindung räumlicher Datentypen, Funktionen und Prädikate ist unkompliziert und basiert in den gezeigten Beispielen auf dem OGC-Standard „Simple Feature Access“, der sowohl grundlegende als auch komplexe räumliche Konzepte bereithält.

In einigen Punkten ist die Komplexität von RIF sogar umfangreicher als nötig. So kommt ein Schemamapping prinzipiell ohne Konfliktlösungsstrategie aus, da alle formulierten Regeln genau einmal anzuwenden sind. Des Weiteren können manche Sachverhalte auf unterschiedliche Art und Weise ausgedrückt werden, z. B. ist die Aktion `Modify` durch `Retract` und `Assert` formulierbar.

Des Weiteren wurde die Integration von RIF-PRD in die bestehende Schemamapping-Software HALE demonstriert. Für die Ausgabe des gesamten Schemamappings von ALKIS nach INSPIRE ist der HALE-RIF-Exporter um verschiedene Funktionen erweitert worden. Der prozedurale Charakter von RIF-PRD erwies sich hier als weniger intuitiv im Vergleich zur deklarativen Mappingsprache gOML, d. h. ein Schemamapping ist besser als solches lesbar, wenn es in einer deklarativen Sprache formuliert ist. Auch der Aufbau und die Struktur von prozeduralen und deklarativen Sprachen sind sehr verschieden, sodass die programmier-

technische Überwindung dieser Unterschiede mehrere Zwischenschritte erforderte.

Ebenso wie die Integration von RIF in bestehende Schemamapping-Software ist auch die Integration in existierende Schematransformationswerkzeuge noch wenig verbreitet. Nach Beare et al. (2010b) ist die Software „Radius Studio“ der Firma „1Spatial“ in der Lage, Schematransformationen auf der Grundlage von RIF-PRD-Dokumenten auszuführen. Für die Übersetzung in das proprietäre Transformationsformat von Radius Studio sind insgesamt acht Translator-Klassen nötig. Der Programmieraufwand war damit hö-

her als beim HALE-RIF-Exporter. Allerdings konnten relevante Codeteile von beiden Komponenten genutzt werden. Beare et al. (2010b) gehen davon aus, dass auch andere Schematransformationswerkzeuge derart erweiterbar sind.

RIF ist ein neues Regelformat mit großem Potenzial, das aktuell jedoch noch wenig verbreitet ist. Derzeit ein Schemamapping in RIF-PRD zu formulieren ist aufwendig und erfordert gute RIF-Kenntnisse. Die zukünftige Akzeptanz von RIF-PRD als implementierungsneutrales Format für die Beschreibung von Schemamappings wird abhängig sein von der Weiterentwicklung bestehender

Schemamapping- bzw. Schematransformations-Software, die RIF automatisiert ausgeben bzw. interpretieren kann. Aufgrund der INSPIRE-Empfehlung kann hier vermutet werden, dass RIF-PRD breitere Verwendung als Mappingsprache findet.

Darüber hinaus ist der Einsatz von RIF in vielen weiteren Geo-Anwendungen denkbar. Neben den in Nash et al. (2011) genannten Beispielen kommen etwa die interoperable Formulierung von Generalisierungsregeln oder die Ergänzung derzeitiger Geo-Metadatenbestände um entsprechend in RIF formulierte Anwendungsregeln für die beschriebenen Geodaten infrage.

IMPRESSUM // PUBLICATION INFORMATION

gis.SCIENCE – Die Zeitschrift für Geoinformatik ISSN 1869-9391 // Redaktion: Gerold Olbrich, olbrich@vde-verlag.de, Tel.: +49(0)69-840006-1121 // Hauptschriftleiter: Prof. Dr.-Ing. Ralf Bill, ralf.bill@uni-rostock.de, Tel +49(0)381-498-3200 // Editorial Board: Prof. Dr. Lars Bernard, TU Dresden; Dr. Andreas Donaubaue, TU München; Prof. Dr. Max Egenhofer, University of Maine Orono; Prof. Dr. Manfred Ehlers, Universität Osnabrück; Prof. Dr. Klaus Greve, Universität Bonn; Dr. Stefan Lang, Universität Salzburg; Prof. Dr. Stephan Nebiker, Fachhochschule Nordwestschweiz, Prof. Dr. Matthäus Schilcher, TU München, Prof. Dr. Josef Strobl, Universität Salzburg // Anzeigen: Katja Hanel, VDE VERLAG GMBH, Telefon +49(0)69/840006-1341, hanel@vde-verlag.de // Abonnement-Service und Adressänderung: Kerstin Grzechnik, +49(0)69/840006-1380, Fax: -1398 // gis.Science erscheint im: Wichmann Verlag im VDE VERLAG GMBH, Bismarckstraße 33, 10625 Berlin, Tel. +49(0)30/34 80 01-0, Fax +49(0)30/34 80 01-9088, www.wichmann-verlag.de // Geschäftsführung: Dr.-Ing. Stefan Schlegel, Margret Schneider // Verlagsleiter Zeitschriften: Ronald Heinze // Druck: H. Heene-mann GmbH & Co. KG, Berlin // Erscheinungsweise: 12 x jährlich, davon 4 Ausgaben gis.SCIENCE, 6 Ausgaben gis.BUSINESS plus 2 Sonderthemenhefte gis.TRENDS+MARKETS // Jahresabonnement (12 Hefte): Inland 157,25 EUR inkl. Versandkosten, europäisches Ausland 166 EUR, sonstiges Ausland 178 EUR inkl. Versandkosten, Studenten/Auszubildende 90,45 EUR inkl. Versandkosten, Mitglieder des Deutschen Dachverbandes für Geoinformation e.V. (DDGI) erhalten das Abo im Rahmen ihrer Mitgliedschaft // Bezugszeitraum: Ein Abonnement gilt für mindestens ein Jahr und verlängert sich jeweils um das folgende Kalenderjahr, wenn es nicht bis zum 15. November eines Jahres gekündigt wurde. Bei Nichterscheinen aus technischen Gründen oder höherer Gewalt entsteht kein Anspruch auf Ersatz. // Alle in gis.SCIENCE erscheinenden Beiträge, Abbildungen und Fotos sind urheberrechtlich geschützt. Reproduktion, gleich welcher Art, können nur nach schriftlicher Genehmigung des Verlags erfolgen. // © 2012 VDE VERLAG, Berlin • Offenbach.

Titelbild // Cover image: Environmental facets of Dresden in geovisualisations from the research projects ANTI-Resist and EO2HEAVEN (Professorship GIS, TU Dresden, <http://tu-dresden.de/fgh/geo/gis>)

Literatur

- Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) (2008): Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens (GeoInfoDok). Erläuterungen zu ALKIS, Version 6. <http://www.adv-online.de/icc/extdeu/broker.jsp?uCon=68470b36-de06-8a01-e1f3-351ec0023010&uBasVariantCon=11111111-1111-1111-1111-11111111>, Zugriff 03/2012.
- Beare, M.; Howard, M.; Payne, S.; Watson, P. (2010a): Development of Technical Guidance for the INSPIRE Transformation Network Service. State Of The Art Analysis, Version 2.0. http://inspire.jrc.ec.europa.eu/documents/Network_Services/JRC_INSPIRE-TransformService_SAA_v2.pdf, Zugriff 03/2012.
- Beare, M.; Payne, S.; Sunderland, R. (2010b): Prototype Report for the INSPIRE Schema Transformation Network Service. Version: 3.0. http://inspire.jrc.ec.europa.eu/documents/Network_Services/JRC_INSPIRE-TransformService_ProtoRpt_v3-0.pdf, Zugriff 03/2012.
- Bill, R. (1999): Grundlagen der Geo-Informationssysteme. Band 2 – Analysen, Anwendungen und neue Entwicklungen. 2. Aufl. Wichmann, Heidelberg.
- Bill, R. (2010): Grundlagen der Geo-Informationssysteme. 5. Aufl. Wichmann, Berlin/Offenbach.
- Boley, H.; Hallmark, G.; Kifer, M.; Paschke, A.; Polleres, A.; Reynolds, D. (2010): RIF Core Dialect. W3C Recommendation 22 June 2010. W3C (World Wide Web Consortium). <http://www.w3.org/TR/rif-core/>, Zugriff 03/2012.
- Boley, H.; Kifer, M. (2010a): RIF Basic Logic Dialect. W3C Recommendation 22 June 2010. W3C (World Wide Web Consortium). <http://www.w3.org/TR/rif-bld/>, Zugriff 03/2012.
- Boley, H.; Kifer, M. (2010b): RIF Framework for Logic Dialects. W3C Recommendation 22 June 2010. W3C (World Wide Web Consortium). <http://www.w3.org/TR/riffld/>, Zugriff 03/2012.
- Boley, H.; Kifer, M.; Patrãnjã, P.-L.; Polleres, A. (2007): Rule Interchange on the Web. In: Antoniou, G.; Aßmann, U.; Baroglio, C.; Decker, S.; Henze, N.; Patrãnjã, P.-L.; Tolksdorf, R. (Eds.): Reasoning Web, Ausg. 4636. Springer, Berlin/Heidelberg (Lecture Notes in Computer Science), S. 269-309. http://dx.doi.org/10.1007/978-3-540-74615-7_5, Zugriff 03/2012.
- De Sainte Marie, C.; Hallmark, G.; Paschke, A. (2010): RIF Production Rule Dialect. W3C Recommendation 22 June 2010. W3C (World Wide Web Consortium). <http://www.w3.org/TR/rif-prd/>, Zugriff 03/2012.
- Gašević, D.; Djurić, D.; Devedžić, V. (2006): Model Driven Architecture and Ontology Development. Springer, Berlin.
- Herring, J. R. (2011): OpenGIS® Implementation Standard for Geographic information – Simple feature access – Part 1: Common architecture. Open Geospatial Consortium Inc. OpenGIS® Implementation Standard, Version 1.2.1, OGC 06-103r4. <http://www.opengeospatial.org/standards/sfa>, Zugriff 03/2012.
- 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.
- INSPIRE (2007): Richtlinie 2007/2/EG des Europäischen Parlaments und des Rates vom 14. März 2007 zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft (INSPIRE). ABl. EU Nr. L 108 vom 25.04.2007, S. 1-14. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:01:DE:PDF>, Zugriff 03/2012.
- INSPIRE TWG Cadastral Parcels (2010): D2.8.1.6 INSPIRE Data Specification on Cadastral Parcels – Guidelines. http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_CP_v3.0.1.pdf, Zugriff 03/2012.
- ISO 19107:2003: Geographic information – Spatial schema.
- Kifer, M.; Boley, H. (2010): RIF Overview. <http://www.w3.org/2005/rules/wiki/Overview>, Zugriff 03/2012.
- Nash, E.; Nikkilä, R.; Wiebenson, J.; Walter, K.; Bill, R. (2011): Interchange of Geospatial Rules – Towards Georules Interchange Format (GeoRIF)? In: gis.SCIENCE, 03/2011, S. 82-94.
- Object Management Group (OMG) (2009): Production Rule Representation (PRR). OMG-Spezifikation. Dokumentnummer: formal/2009-12-01. Version 1.0. 2009. <http://www.omg.org/spec/PRR/1.0/PDF/>, Zugriff 03/2012.
- Paschke, A.; Hirtle, D.; Ginsberg, A.; Patrãnjã, P.-L.; McCabe, F. (2008): RIF Use Cases and Requirements. <http://www.w3.org/TR/rif-ucr/>, Zugriff 03/2012.
- Polleres, A.; Boley, H.; Kifer, M. (2010): RIF Datatypes and BuiltIns 1.0. W3C Recommendation 22 June 2010. W3C (World Wide Web Consortium). <http://www.w3.org/TR/rif-dtb/>, Zugriff 03/2012.
- Reinwarth, S. (2011): Untersuchungen zur Verwendung des Rule Interchange Formats (RIF) für die dienstbasierte Transformation von Geodaten. Diplomarbeit, Technische Universität Dresden (unveröffentlicht)
- RIF Working Group (2009): Modify loop. http://www.w3.org/2005/rules/wiki/Modify_loop, Zugriff 03/2012.
- W3C (2012): Semantic Web. <http://www.w3.org/standards/semanticweb/>, Zugriff 03/2012.